

**Understanding the replication biology of
Providence virus: elucidating the function
of non-structural proteins.**

Thesis

Submitted in the fulfilment of the requirements for the
degree of

Doctor of Philosophy

(Microbiology)

In the

FACULTY OF SCIENCE

RHODES UNIVERSITY

By

RITAH NAKAYINGA

December 2013

ABSTRACT

Tetraviruses are non-enveloped, small insect RNA viruses with a single stranded positive RNA genome that is either monopartite or bipartite. Providence virus (PrV) is the only member of the three tetravirus families with a viral replicase similar to the replicases of tombusviruses and umbraviruses. The principle aim of this thesis was to study PrV replication, focusing on subcellular localization and potential interactions between PrV replication proteins.

The first objective of this study was to generate an anti-p104 antibody that does not cross-react with p40. Expression of the C-terminal portion of p104 in *E. coli* resulted in no detectable protein. Further expression in an insect cell based expression system resulted in the production of an insoluble protein. Attempts to improve protein solubility with a range of solubilization treatments were unsuccessful. Bioinformatic analysis was used to detect an antigenic region at the C-terminus of p104 and the peptide was used to raise anti-p104 antibodies. These antibodies did not detect native protein by western blot detection however they were used for immunoprecipitation.

The establishment of the subcellular localization of PrV required two approaches; immunofluorescence in persistently infected *Helicoverpa zea* MG8 cells using anti-p40 and anti-dsRNA antibodies and the expression of EGFP-replicase fusion protein in *Spodoptera frugiperda* Sf9 cells. Replication of PrV was found to take place in cytosolic punctate structures. Co-immunoprecipitation experiments revealed that p40 self-interacts and interacts with p104.

Bioinformatic analysis of PrV p104 suggests that the RdRp is similar to viral RdRps of the carmo-like supergroup II. Potential RNA binding regions are present within p104. A potential p40 interaction domain that shares hydrophilic and surface exposed properties with the TBSV p33 interaction domain is present. A putative arginine-rich region and disordered C-terminal region is present in p130.

In conclusion, PrV p104 is the viral replicase. The resemblance of the expression strategy and putative functional domains with tombusviruses and umbraviruses suggest that PrV replication is related to the replication system of the tombusviruses and umbraviruses. This has led to propose that tetravirus replication strategies are diverse and raises questions on the origin and evolution of PrV.

Key words: *Tetraviridae*, *Tombusviridae*, *Umbraviridae*, *Replication*

TABLE OF CONTENTS

Abstract.....	ii
Acknowledgements.....	viii
List of abbreviations and virus names	ix
List of tables and figures.....	xiii
 Chapter 1: Literature review	 1
1.1 Introduction.....	1
1.2 The replicases of positive sense RNA viruses	1
1.2.1 Shape, structure and function of the RdRp	4
1.2.2 Classification of (+ve) ss RNA viral replicases	6
1.2.3 Translation strategies for the expression of the replicase among (+ve) ssRNA viruses	 8
1.3 The Tetraviruses	13
1.3.1 Host range and pathology.....	14
1.3.2 Taxonomic classification of tetraviruses.....	15
1.3.3 Capsid structure, assembly and maturation.....	15
1.4 The <i>Tombusviridae</i>	26
1.4.1 Genome organisation	26
1.4.2 The tombusvirus replicase and replication complexes.....	29
1.4.3 Replication of viral RNA	30
1.4.4 Chaperone activity of tombusvirus replication proteins	32
1.4.5 <i>Tombusviridae</i> replication complex-host interactions	33
1.4.6 Subcellular localization of tombusvirus replication proteins.....	35
1.4.7 Establishment of persistent infections.....	37
1.5 The <i>Umbraviridae</i>	38
1.6 Research motivation	41

1.7 Objectives	42
Chapter 2: Production of antiserum for the detection of PrV replicase.....	43
2.1 Introduction.....	43
2.2 Materials and methods	43
2.2.1 Bacterial strains	43
2.2.2 Construction and expression of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ in <i>E. coli</i>	44
2.2.3 Baculovirus-mediated expression of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ in <i>Spodoptera frugiperda</i> (Sf9) cells	46
2.3 Results.....	49
2.3.1 Expression of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ in <i>E. coli</i>	49
2.3.2 Baculovirus-mediated expression of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ in Sf9 cells.....	50
2.4 Discussion.....	59
Chapter 3: Subcellular localization of PrV replication	61
3.1 Introduction.....	61
3.2 Materials and methods	62
3.2.1. Cell culture	62
3.2.2 Immunofluorescence staining of PrV infected MG8 cells	62
3.2.3 Construction of native and mutant EGFP-tagged PrV replicase expression plasmids.....	63
3.2.4 Expression of EGFP-tagged PrV replicase plasmids pRN58, pRN59 and pCW57 in <i>S. frugiperda</i> Sf9 cells.....	67
3.3 Results.....	67
3.3.1 Immunofluorescence co-localization of PrV replication proteins and replicating vRNA in <i>H. zea</i> MG8 cells	68
3.3.2 Expression of PrV replicase in <i>S. frugiperda</i> Sf9 cells	71
3.3.3 Subcellular localization of EGFP-replicase fusion proteins	74
3.4 Discussion	75
Chapter 4: Protein-protein interactions between PrV p40 and p104	77

4.1 Introduction.....	77
4.2 Materials and methods	77
4.2.1 Conjugation of antibodies to magnetic dynabeads.....	77
4.2.2 Co-immunoprecipitation of PrV replication proteins	78
4.2.3 Co-immunoprecipitation of PrV replication proteins in infected <i>S. fugiperda</i> Sf9 cells.....	78
4.3 Results.....	79
4.3.1 Conjugation of anti-p104 ₍₈₈₉₋₉₀₂₎ antibodies to magnetic dynabeads.....	79
4.3.2 Optimisation of p104 immunoprecipitation procedure	80
4.3.3 Co-immunoprecipitation of PrV replication proteins from virus-infected cells	81
4.4 Discussion.....	85
Chapter 5: Identification of potential functional domains in the PrV replicase and p130	87
5.1 Introduction.....	87
5.2 Materials and methods	88
5.3 Results.....	89
5.3.1 Identification of the putative RdRp domain within p104.....	89
5.3.2 Other potentially functional domains of p104	93
5.3.3 Potential transmembrane domains and topology of p104.....	97
5.3.4 Potential phosphorylation sites within p104	104
5.3.5 Identification of putative motifs and domains within p130	105
5.4 Discussion.....	110
Chapter 6: General discussion and conclusions.....	113
6.1 Subcellular localization of the site of PrV replication	113
6.2 A comparison between PrV, tombusviruses and umbraviruses.....	114
6.3 A model for the replication of PrV	116
6.4 The origin of providence virus.....	121
References.....	124
Appendix A: Construction of recombinant plasmids	154

Table of contents

Appendix B: Alternative transmembrane topology models of PrV Replicase	158
Appendix C	161

ACKNOWLEDGEMENTS

Armed with fairly basic knowledge of Molecular Virology, I embarked on a long journey that has been enriching, amazing, exciting yet very challenging. I would like to thank several people whose guidance and support has enabled me get this far. First and foremost, i am grateful to my supervisor Prof Rosemary Dorrington for her tireless support, extensive guidance and advice throughout this project. I will be forever grateful to her for her patience that enabled me to develop a better understanding of the research project.

I am indebted to Lab 417 members both past and present for the support they rendered to me during my stay at Rhodes University. I would like to especially thank Cheryl for setting the foundation upon which I built my research, James for his willingness to teach and train me with valuable techniques, Petra for helping out with the basics of molecular biology, Dr Knox for always being positive even when life seemed to take the opposite direction, Mez and Gwynneth for their invaluable guidance during the final stages of my thesis. Surely, i wouldn't have made it without all of them and would therefore like to thank them all !! I would also like to extend my special thanks to Val for her patience, support and keeping the 417 lab the best functioning laboratory there could be.

I owe my deepest gratitude to my husband, Mr Martin Erone and daughter, Malaika Zoe Akiror, for being very understanding and putting up with the stress that came with the pursuit for a PhD. A special thank you to my father and mother for giving me an education, the bedrock to all my academic achievements; I credit this achievement to you. To my sisters, brothers and friends, thanks for your never ending love and support during this time.

My sincere gratitude to the National Research Foundation (NRF) for the bursary that supported this research throughout my stay at Rhodes. All the glory and honour back to GOD.

LIST OF ABBREVIATIONS AND VIRUS NAMES

GENERAL ABBREVIATIONS

6xHis	hexahistidine tag
(+ve)	positive sense
(-ve)	negative sense
3D ^{pol}	RNA-dependent RNA polymerase of <i>Polio virus</i>
AcMNPV	<i>Autographa californica</i> multiple nucleopolyhedrovirus
BLAST	basic local alignment search tool
BSA	bovine serum albumin
BV	baculovirus
CDS	coding sequence
CMV	cytomegalovirus
DAPI	4',6-diamidino-2-phenylindole
DdRp	DNA-depedent RNA polymerase
DdDp	DNA-dependent DNA polymerase
ds	double stranded
DTT	dithiothreitol
DI	defective interfering
EDTA	ethylenediaminetetraacetic acid
ECL	enhanced chemiluminescence
EGFP	enhanced green fluorescent protein
ExPASy	Expert Protein Analysis System
FB33	<i>Helicoverpa zea</i> fat body cell line 33
FITC	fluorescein isothiocyanate
g	gravitation force
gRNA	genomic ribonucleic acid
GST	glutathione S-transferase

Abbreviations

IPTG	isopropyl β -D-1-thiogalactopyranoside
Ig	immunoglobulin
Kb	kilobase (s)
kDa	kiloDalton
LB	Lysogeny Broth
MG8	Helicoverpa zea midgut cell line 8
min	minutes
MOI	multiplicity of infection
NES	nuclear exporting signal
NCBI	National Centre for Biotechnology Information
NLS	nuclear localization signal
nt(s)	nucleotide (s)
OD ₆₀₀	optical density at 600 nm
ORF	open reading frame
P1	passage 1
P2	passage 2
P3	passage 3
P4	passage 4
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PEI	polyethylenimine
Pfu	plaque forming units
RdRp	RNA-dependant RNA polymerase
RdDp	RNA-dependent DNA polymerase
RT	reverse transcriptase
rpm	revolutions per minute
Sf9	<i>Spodoptera frugiperda</i> cell line 9
Se-1	<i>Spodoptera exigua</i> cell line 1
ss	single stranded
SDS	sodium dodecyl sulphate
SDS-PAGE	sodium dodecyl sulphate polyacrylamide gel electrophoresis
SFM	serum free cell culture medium
TAE	tris acetate EDTA buffer
TBST	tris buffered saline buffer

TE buffer	tris EDTA buffer
Tris	tris-2-amino-2-(hydroxymethyl)-1,3-propandiol
VPg	genome linked viral protein
vRNA	viral ribonucleic acid
wt	wild type
X-Gal	5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside

VIRUS NAMES ABBREVIATIONS

AeV	<i>Antheraea eucalypti</i> virus
BYDV	Barley yellow dwarf virus
BYCV	Beet yellows clostero virus
CIRV	Carnation italian ringspot virus
CNV	Cucumber necrosis virus
CMoMV	Carrot mottle mimic virus
CymRSV	Cymbidium ringspot virus
DAV	Drosophila A virus
EeV	<i>Euprosterna elaeasa</i> virus
EMCV	Encephalomyocarditis virus
GRV	Groundnut rosette virus
HaSV	<i>Helicoverpa amigera</i> stunt virus
IPNV	Infectious pancreatic necrosis virus
IBDV	Infectious bursal disease virus
MNSP	Melon necrotic spot virus
MCMV	Maize chlorotic mottle virus
N β V	<i>Nudaurelia capensis</i> β virus
N ω V	<i>Nudaurelia capensis</i> ω virus
N ζ V	<i>Nudaurelia capensis</i> ζ virus
N ψ V	<i>Nudaurelia capensis</i> ψ virus
OCSV	Oat chlorotic stunt virus
OPMV	Opium poppy mosaic virus
PCRVP	Pelargonium chlorotic ring pattern virus
PEMV-2	Pea enation mosaic virus-2
PFBV	Pelagornium flower bleak virus

Abbreviations

PMV	Panicum mosaic virus
PrV	Providence virus
PoLV	Pothos latent virus
RCNMV	Red clover necrotic mosaic virus
TaV	<i>Thosea asigna</i> virus
TMV	Tobacco mosaic virus
TNVD	Tobacco necrosis virus D
TBSV	Tomato bushy stunt virus
TBTv	Tobacco bushy top virus
TCV	Turnip crinkle virus
TYMV	Turnip yellow mosaic virus

LIST OF TABLES AND FIGURES

Table 1. 1: Conserved motifs found within the replicase domains of (+ve) ss RNA viruses	3
Table 1. 2: General properties of tetraviruses with well characterized genomes	20
Table 1. 3: Replication proteins and mode of expression of the <i>Tombusviridae</i> replicases.....	29
Table A. 1: Plasmids used in this study	155
Figure 1. 1: Schematic representation of different types of viral replicases.....	2
Figure 1. 2: Orientation of the secondary structures within the RdRp motifs IV,V,VI and VII.	4
Figure 1. 3: The right-hand like structure of different types of polymerases.	5
Figure 1. 4: Schematic representation of the crystal structure of Polio virus RdRp domain (3D ^{pol}).	6
Figure 1. 5: Three-dimensional reconstruction of NβV, NωV, PrV and HaSV particles using cryo-electron microscopy.	16
Figure 1. 6: Icosahedral capsids with triangulation numbers of 1, 3 and 4.	17
Figure 1. 7: The structure and quasi-equivalence of NωV capsid.	19
Figure 1. 8: Schematic representation of the genomic organisation of the <i>Alphatetraviridae</i>	21
Figure 1. 9: Schematic representation of the genomic organisation of the <i>Permutotetraviridae</i>	23
Figure 1. 10: Schematic representation of the genomic organisation of the <i>Carmotetraviridae</i>	25
Figure 1. 11: Schematic representation of the genome organization of the <i>Tombusviridae</i>	28
Figure 1. 12: Schematic representation of the prototypic tombusvirus genome represented by <i>Tomato bushy stunt virus</i> (TBSV).	31
Figure 1. 13: Schematic replication cycle of TBSV.	33
Figure 1. 14: Genome organization of groundnut rosette virus.	38

Figure 2. 1: Schematic representation of the construction of pRN12.....	45
Figure 2. 2: Graphic representation of the restriction map of pRN18.	46
Figure 2. 3: PCR verification of transposition of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ expression cassette into Bacmid DNA.....	52
Figure 2. 4: Expression of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ in Sf9 cells from six recombinant baculovirus.....	53
Figure 2. 5: Optimization of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ expression in Sf9 cells.	54
Figure 2. 6: The solubility of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ expressed in Sf9 cells infected with BVp55 (P4).	55
Figure 2. 7: Solubilization of 6xHis-p104 ₍₄₃₇₋₉₁₅₎	56
Figure 2. 8: The Hopp and Woods plot of p104 showing the region between 889 - 902 amino acids selected for antiserum production.....	57
Figure 2. 9: Validation of anti-p104 ₍₈₈₉₋₉₀₂₎ antiserum	58
Figure 2. 10: Western blot analysis to detect p104 in PrV infected MG8 cells.....	59
Figure 3. 1: Schematic representation of the cloning strategy for the construction of pRN58.....	64
Figure 3. 2: Schematic representation of the cloning strategy for the construction of pRN59.....	66
Figure 3. 3: Detection of replication complexes in PrV infected MG8 cells using immunofluorescence microscopy.	69
Figure 3. 4: Immunofluorescence of PrV infected MG8 cells.....	71
Figure 3. 5: Detection of EGFP-tagged PrV replicase in <i>S. frugiperda</i> Sf9 cells.	73
Figure 3. 6: Subcellular distribution of EGFP-tagged PrV p40 and p104.....	74
Figure 4. 1: Validation of conjugation of anti-p104 ₍₈₈₉₋₉₀₂₎ antibodies to magnetic dynabeads.....	80
Figure 4. 2: Detection of 6xHis-p104 ₍₄₃₇₋₉₁₅₎ in Sf9 cell free extracts by immunoprecipitation and western blot analysis.....	81
Figure 4. 3: Detection of p40 and p104 in PrV infected MG8 by immunoprecipitation and western blot analysis.	82
Figure 4. 4: Validation of conjugation of anti-GFP antibody to magnetic dynabeads.	84

Figure 4. 5: Detection of EGFP-p40 and p40 by immunoprecipitation and western blot analysis.	85
Figure 5. 1: Schematic diagram of p104 sequence showing the putative catalytic palm subdomain of the RdRp domain.	89
Figure 5. 2: Multiple sequence alignment of the RdRp domain of PrV p104 with tombusvirus and umbravirus replicases.	92
Figure 5. 3: Identification of hydrophilic and surface exposed characteristics of PrV arginine-rich motif.	93
Figure 5. 4: Identification of a potential RNA-binding domain in PrV p104.....	95
Figure 5. 5: Comparing hydrophilic and surface exposed properties of the interaction domain of PrV p40 and TBSV p33.....	96
Figure 5. 6: Sequence alignment of PrV p40 and TBSV p33 interaction domain.....	97
Figure 5. 7: Identification of potential transmembrane domains in PrV p104.	99
Figure 5. 8: Identification of the antigenic regions selected for antibody production in relation to the hydrophobic profile of PrV p104.....	100
Figure 5. 9: The p104 sequence showing potential domains.	101
Figure 5. 10: The two transmembrane model of PrV p104 and TBSV p92.	103
Figure 5. 11: Schematic map showing potential phosphorylation sites in PrV and TBSV replicase.	105
Figure 5. 12: Identification of the surface exposed properties of the potential PrV p130 RNA binding site.	107
Figure 5. 13: Identification of potential disordered regions within PrV p130 and GRV p27.	109
Figure 5. 14: Schematic representation of putative domains within PrV p130.....	110
Figure 5. 15: Summary of domains and motifs found within ORFs encoded by PrV, GRV and TBSV genomes.....	112
Figure 6. 1: Schematic representation of the proposed model of PrV replication...	118
Figure A. 1: Graphic representation of the expression plasmid pCW58.	157
Figure B. 1: Model of the potential topology of PrV p104 with four transmembrane helices.	159

Figure B. 2: Model of the potential topology of p104 with three transmembrane helices.	160
---	-----

CHAPTER 1

LITERATURE REVIEW

1.1 INTRODUCTION

RNA viruses need to encode their own polymerase for the multiplication of their RNA genomes. There are two main types of polymerase; RNA-dependent RNA polymerase (RdRp) and RNA-dependent DNA polymerase (RdDp) also known as reverse transcriptase. The former are involved in an RNA life cycle that has no DNA stage, whilst the latter is employed in a viral life cycle that involves conversion of RNA to DNA; which is later integrated in the host genome (Baltimore, 1971). The RdRp is involved in the replication of single-stranded (ss) plus (+ve), minus (-ve) or double-stranded (ds) RNA viruses while the RdDp is involved in the replication of retroviruses, transposable integrated elements (non-viral and viral retroposons) and some DNA viruses (hepadnaviruses) (Weiner *et al.*, 1986; Ishihama and Nagata, 1988). Despite the diversity of genomes replicated by these polymerases, the RdRp show significant structural and sequence conservation and their sole function is to produce negative, positive and subgenomic RNA strands during replication (Joyce and Steiz, 1995).

1.2 THE REPLICASES OF POSITIVE SENSE RNA VIRUSES

The replicase of (+ve) ss RNA viruses encode several domains with conserved sequences found in proteins that mediate genome replication and expression. The three most prominent domains are the methyltransferase (MET), helicase (HEL) and RdRp domains (Figure 1.1). The MET is involved in capping genomic RNA while the HEL and the RdRp domains are responsible for unwinding intramolecular secondary structures and RNA polymerization respectively (Karmer and Argos, 1984; Gorbalenya *et al.*, 1989; Rozanov *et al.*, 1992). Only the RdRp domain is present in all (+ve) ss RNA viruses (Koonin, 1991) while the other two domains are present in the replicases of viruses within supergroup III (Figure 1.1) (Koonin, 1991).

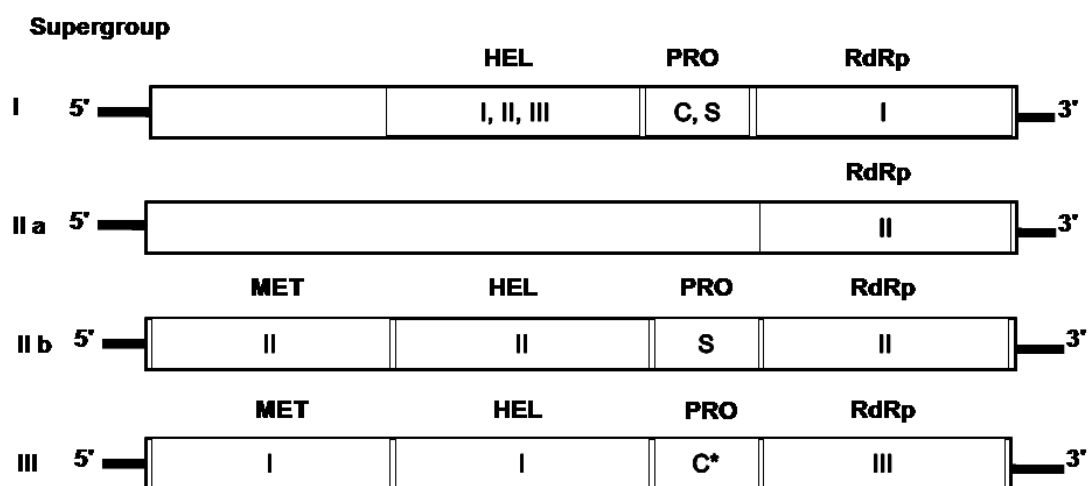


Figure 1. 1: Schematic representation of different types of viral replicases. Supergroup I (picorna-like lineage) consists of helicase (HEL) (type I, II or III), proteinase (PRO) with cysteine (C) or serine (S) residue and RdRp (type I) domains. Supergroup IIa (carbo-like lineage) has only a type II RdRp domain. Supergroup IIb (flavi-like lineage) consists of methyltransferase (MET) (type II), helicase (HEL) (type II), proteinase (PRO) with a conserved serine (S) residue and RdRp (type II) domains. Supergroup III (alpha-like lineage) consists of methyltransferase (MET) (type I), helicase (HEL) (type I), papain-like proteinase (PRO) indicated by a star, a cysteine (C) residue and RdRp (type III) domain (Koonin, 1991; Koonin and Dolja; 1993).

The MET and HEL domains encompass distinct conserved amino acids (Table 1.1). The importance of the conserved amino acids within the MET domain has been demonstrated by mutational analysis (Wang *et al.*, 1996; Ahola *et al.*, 1997; Ahola and Ahlquist, 1999). Substitution of H80A, D106A and R136A abolished the RNA capping activity of methyltransferase. This suggests that the MET domain is important in genome replication. Mutation analysis has revealed the importance of the conserved amino acids within the HEL domain. Substitution of K691A, D249A and G290S abolished ATPase activity and ability to unwind nucleic acid (Ahola *et al.*, 2000; Graves-Woodward *et al.*, 1997). This suggests that the HEL domain has an important role in genome replication.

The canonical GDD motif, originally proposed by Kamer and Argos (1984), is considered the hall mark for the viral RdRp domain (Table 1.1). Poch *et al.* (1989) identified four highly conserved motifs namely A, B, C, D also designated motifs IV, V, VI and VII by Koonin (1991) within the RdRp palm subdomain (Table 1.1).

Motif IV (Dx_4-5D), where x denotes any amino acid and Motif VI (GDD) are spatially arranged with the D residues correctly positioned for binding divalent ions (Mn^{2+} or Mg^{2+}) that are required for the catalysis of the polynucleotide polymerisation reaction (Poch *et al.*, 1989). Motif V ((S/T) Gx_2-3Tx_3N (S/T)) consists of an essential N residue that allows incorporation of ribonucleic triphosphate groups to the growing RNA and determines whether DNA or RNA is synthesized by discriminating between dNTPs and NTPs (Poch *et al.*, 1989; Hansen *et al.*, 1997). Motif VII (LKR) is a basic motif that determines correct and efficient incorporation of single nucleotides to the growing RNA (Cameron *et al.*, 2009; Yang *et al.*, 2012).

Table 1. 1: Conserved motifs found within the replicase domains of (+ve) ss RNA viruses

Domain	Conserved amino acids			
MET	H	DxxR	Y	
HEL	I (GxGKT/S)	II (DE)	III (GDxxQ)	
RdRp	IV (Dx_4-5D)	V (S/T) Gx_2-3Tx_3N (S/T)	VI (GDD)	VII (LKR)

Methyltransferase conserved amino acids H, D, R, Y (Rozanov *et al.*, 1992; Koonin and Dolja, 1993); Helicase motifs I-III (Koonin and Dolja, 1993); RdRp motifs IV, V, VI, VII (Kamer and Argos, 1984; Poch *et al.*, 1989; Koonin, 1991).

According to Poch *et al.* (1989), the orientation of these motifs within the RdRp palm subdomain is conserved in all viral RdRps and provides the geometrical arrangement required for the substrate molecules and metal ions required for nucleic acid polymerization. Motif IV has a β -turn- α conformation with the D residues located within the turn region while motif V consists of a β -turn followed by a long α helical stretch with the highly conserved G residue located in the turn region (Figure 1.2). Similarly, motif VII has its K residue embedded in the α helix structure while the motif VI is flanked on either side by hydrophobic residues creating a β -turn- β conformation with the conserved DD residues in the turn region (Figure 1.2). The location of these highly conserved residues in the turn region is required for proper orientation for cation binding, template specificity and polymerase activity (Poch *et al.*, 1989). Mutational studies have revealed the importance of motif C in

RNA replication. In fact viral RNA replication is reduced or abolished when the GDD sequence of the RdRp is mutated to GED, ADD, GAD or RDD (Longstaff *et al.*, 1993), suggesting its importance for the survival of (+ve) ssRNA viruses.

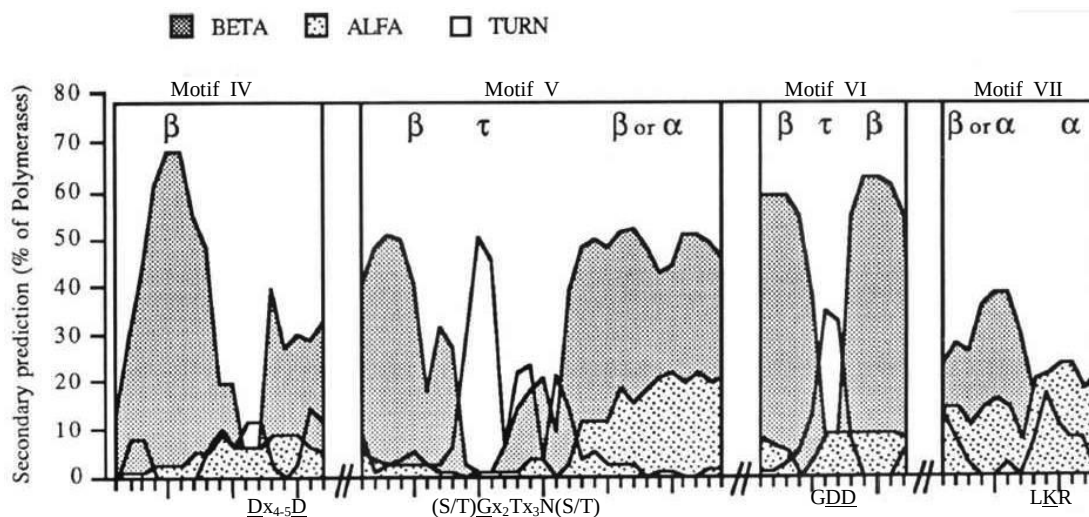


Figure 1. 2: Orientation of the secondary structures within the RdRp motifs IV, V, VI and VII. The grey curves represent β strands; white curves represent a turn structure while dotted curves represent α helix. The position of the underlined D residues of motif IV, G residue of motif V and D residues of motif VI are located within the turn region. The K residue (underlined) of motif VII is found in the α helix region. Reproduced from Poch *et al.* (1989).

1.2.1 Shape, structure and function of the RdRp

Viral RdRps display a basic right-hand like structure similar to DNA-dependent RNA polymerases (DdRp), reverse transcriptases (RTs) and DNA-dependent DNA polymerases (DdDp) (Ollis *et al.*, 1985; Kohlstaedt *et al.*, 1992; Sousa *et al.*, 1993; Hansen *et al.*, 1997) (Figure 1.3). The right-hand morphology of the polymerases allows correct positioning of the substrate molecules and metal ions at the active site for catalysis (Joyce and Steitz, 1995). The conserved structure of the viral RdRp was determined by studying a large number of animal and bacteriophage RdRps (Hansen *et al.*, 1997; Ago *et al.*, 1999; Ng *et al.*, 2002; Tao *et al.*, 2002; Koivunen *et al.*, 2008). This structural data together with studies on cellular polymerases confirmed that all polymerases share similarities in structure and not surprisingly in the mechanism of nucleic acid catalysis (Joyce and Steitz, 1995).

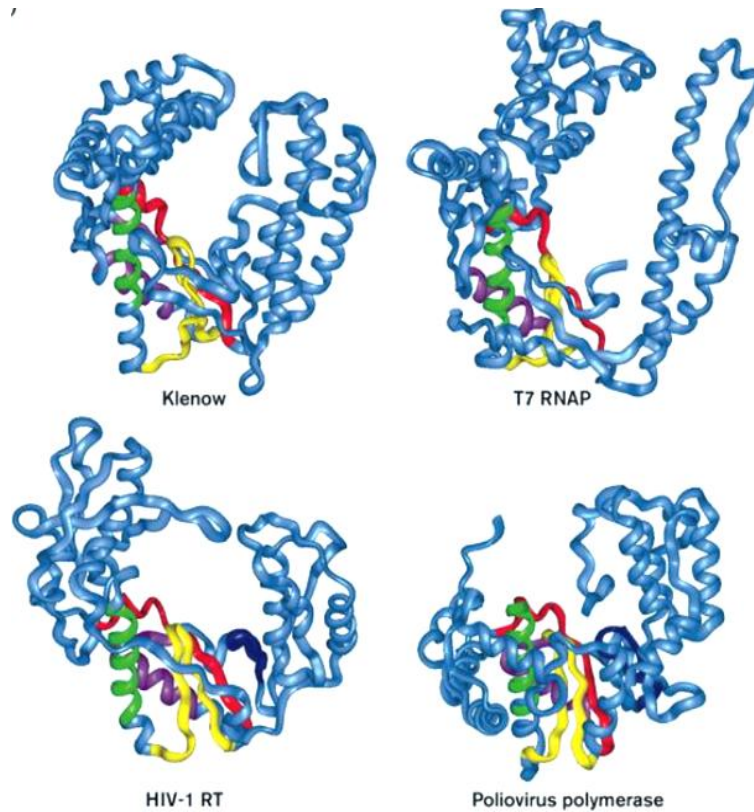


Figure 1. 3: The right-hand like structure of different types of polymerases. Top left; Klenow polymerase (DdDp). Top right; T7 RNA polymerase (DdRp). Bottom left; HIV-1 reverse transcriptase (RdDp). Bottom right; Poliovirus polymerase (RdRp). The fingers are positioned at the top left, palm at the bottom and thumb at the top right. The motifs IV to VII are represented as IV (Red), V (Green), VI (Yellow) and VII (Violet). Images reproduced from Hansen *et al.* (1997).

The polymerase consists of three subdomains namely palm, thumb and finger domain (Figure 1.4) with the most conserved motifs located within the palm subdomain of the RdRp (Poch *et al.*, 1989). Although the palm subdomain motifs are arranged in an ordered sequence of A-B-C, a permuted arrangement of C-A-B has been identified among double stranded birnaviruses such as infectious pancreatic necrosis virus (IPNV) and infectious bursal disease virus (IBDV), two tetraviruses that include *Thosea asigna* virus (TaV) and *Euprosteria elaeasa* virus (EeV) (Gorbalenya *et al.*, 2002). Other viruses include *Drosophila A* virus and Grapevine virus Q (Ambrose *et al.*, 2009; Sabanadzovic *et al.*, 2009). Gorbalenya *et al.* (2002), demonstrated that such an arrangement is compatible with the secondary structure conformation that allows the correct orientation for substrate molecules and metal ions required for polymerase activity.

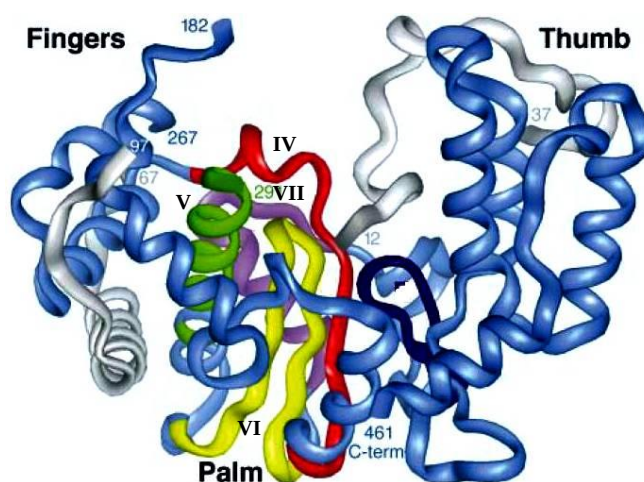


Figure 1. 4: Schematic representation of the crystal structure of Polio virus RdRp domain (3D^{pol}). A ribbon representation with the fingers (top left), palm (bottom) and thumb (top right) subdomains. The motifs IV to VII are represented as IV (Red), V (Green), VI (Yellow) and VII (Violet). Reproduced from Hansen *et al.* (1997).

Viral RdRps are involved in transcription, replication of viral RNA, recombination and evolution of viruses (Lai, 1992; 1998; Kao *et al.*, 2001). Most viral RdRps can initiate RNA synthesis through different mechanisms such as primer dependent RNA initiation and/or *de novo* RNA initiation with the latter being the more common method (Kao *et al.*, 2001). Therefore regulatory elements responsible for RNA replication must be able to recruit the RdRp and promote complementary RNA synthesis from the initiating nucleotide, which has been demonstrated in several (+ve) ssRNA viruses to be a GTP (Blumenthal, 1980; Kao and Sun, 1996).

1.2.2 Classification of (+ve) ss RNA viral replicases

Within the RNA domain, the (+ve) ss RNA viruses comprise the largest and diverse class of the RNA viruses and are classified into three lineages; picornavirus-like, alphavirus-like and flavivirus-like based on phylogenetic analysis of the RdRp (Koonin, 1991). Apart from the RdRp, genes that encode superfamily 3 helicase (S3H) and chymotrypsin-like protease are shared among the picornavirus-like superfamily while the capping enzyme and superfamily 1 helicase (S1H) are conserved among the alphavirus-like superfamily. The flavivirus-like superfamily share only related RdRps (Reviewed in Dolja and Koonin, 2011).

As mentioned above, the RdRps of (+ve) ssRNA viruses are classified into three supergroups based on sequence conservation between the RdRp domain and the presence or absence of other associated domains that include the MET, HEL and proteinase (PRO) (Koonin 1991; Koonin and Dolja, 1993; Van der Heijden and Bol, 2002). Comparison of plant viruses, animal viruses and bacteriophages revealed the presence of conserved motifs I to VIII scattered over a wide region with motifs IV (A), V (B) and VI (C) conserved throughout the viruses (Koonin *et al.*, 1989; Koonin, 1991; Koonin and Dolja, 1993).

The RdRp supergroup 1 consists of viruses of the picorna-like lineage that infect animals, plants, fungi, excavates and alveolates (Koonin and Dolja, 1993; Dolja and Koonin, 2011). The genomes of the picorna-like viruses encode a type I RdRp domain and a conserved cysteine or serine proteinase required for post-translational processing of the replicase (Koonin and Dolja, 1993). In addition, they may encode a type I, II or III HEL but lack the MET domain because the majority of these viruses use the nucleotidylation (VPg) signal to initiate positive and negative RNA synthesis (Wimmer *et al.*, 1993).

Supergroup II is divided into two major lineages that include the carmo-like and flavi-like viruses. The carmo-like viruses comprise mainly of plant viruses whose genome encodes a type II RdRp but lacks HEL, MET and proteinase domains. In addition, they encode a small replication protein with no identifiable replication signature motifs. The RdRp domain is produced as a C-terminal fusion over a readthrough stop codon or by ribosomal frameshift mechanism (Dorrington *et al.*, 2011; Rochon *et al.*, 2011; Ryabov *et al.* 2011). The flavi-like viruses comprise animal viruses and the genomes encode a type II RdRp, HEL and MET, along with a serine proteinase (Koonin, 1991).

The genomes of the supergroup III viruses encode a type III RdRp with type I HEL and MET domain (Dolja and Koonin, 2011; Van der Heijden and Bol, 2002). The supergroup III is made up of the alpha-like viruses such as animal viruses including *Togaviridae* and *Hepeviridae*, insect viruses belonging to *Tetraviridae* and plant viruses belonging to *Bromoviridae*, *Tymoviridae*, *Closteroviridae* and *Virgaiviridae* (Van der Heijden and Bol, 2002; Dolja and Koonin, 2011). The replicase of the

animal alpha-like viruses is characterized the presence of the X-domain (Koonin and Dolja, 1993) that is conserved with an adenosine binding pocket that allows binding of ADP ribose binding proteins (Malet *et al.*, 2009). The insect alpha-like viruses express their proteins from either a monopartite or bipartite genome (reviewed in Dorrington and Short, 2010) and the plant alpha-like viruses use different strategies to express their replication proteins (reviewed in Van der Heijden and Bol, 2002). Some of the plant alpha-like viruses have the MET, HEL and RdRp expressed as a single polypeptide from a single open reading frame (Martelli and Jelkmann, 1998) while in others; the polypeptide is proteolytically cleaved into the MET, HEL and RdRp domain (Kadaré *et al.*, 1995). Others use the readthrough to express the three domains as a full-length product (Osman and Buck, 1996).

1.2.3 Translation strategies for the expression of the replicase among (+ve) ssRNA viruses

Positive sense RNA viruses have evolved unconventional strategies to allow expression of one or more proteins from their small RNA genomes (Reviewed in Dreher and Miller, 2006; Thiébeauld *et al.*, 2007). The most studied strategies for expression of the replicase include overlapping genes, ribosomal frameshifting, leaky scanning and readthrough stop codon mechanism (Reviewed in Drugeon *et al.*, 1999; Dreher and Miller, 2006; Thiébeauld *et al.*, 2007). Overlapping genes are a common feature of (+ve) ssRNA viruses and serve as a mechanism to optimize the coding potential of their small genomes. Several examples have been demonstrated among the (+ve) ssRNA with the RdRp expressed via the full-length open reading frame (Reviewed in Buck, 1996).

The frameshift strategy allows the production of the replicase in either -1 or +1 direction. An example of a -1 ribosomal frameshift strategy is provided by Red clover necrotic mosaic virus RNA 1 (Xiong *et al.*, 1993; Kim and Lommel, 1994). The p88 replicase is translated by -1 ribosomal frameshift from the N-terminal ORF1 (27 kDa) through to the 3' of ORF2 (57 kDa). The event occurs when the heptanucleotide signals form a slippery site that allows ribosome tRNAs to slip and skip a number of nucleotides during translation resulting in a shift in the reading frame (Kim and Lommel, 1994; 1998; Brierley *et al.*, 1992). In addition, a stable stem loop structure downstream of the heptanucleotide signal is presumed to pause

the elongating ribosomes over the slippery site allowing more time for correct positioning of the ribosomal A and P sites (Somogyi *et al.*, 1993; Kim and Lommel, 1998; Lopinski *et al.*, 2000). The +1 ribosomal frameshift uses a similar mechanism as that described for the -1 frameshift for expression of the replicase. An example of the + 1 ribosomal frameshift is demonstrated in beet yellows closterovirus (Agranovsky *et al.*, 1994; Klassen *et al.*, 1995).

Leaky scanning occurs when scanning ribosomes bypass the first start codon in a suboptimal context and initiate translation at downstream start codon that resides in an optimal context, resulting in an extended polypeptide or different proteins (Thiébeauld *et al.*, 2007). An example of the leaky scanning mechanism is provided by turnip yellow mosaic virus (TYMV) with overlapping ORFs that encode the p69 movement protein (ORF1) and p206 replicase (ORF2) (Morch *et al.*, 1988; Branson *et al.*, 1995). For the expression of the replicase, translation is initiated at the first start codon within an optimal context (GUAAUGG) and continues as ribosomes bypass the second start codon in a suboptimal context (CAAAUGA) (Kozak, 1991; Matsuda *et al.*, 2004).

Several other translation strategies are adopted by (+ve) ssRNA viruses for efficient replication (Dreher and Miller, 2006; Thiébeauld *et al.*, 2007; Martínez-Salas *et al.*, 2008). For example the cap-independent internal initiation mechanism is used by encephalomyocarditis virus and whose genome lacks a 5' cap structure resulting in the use of an internal ribosome entry site (IRES) RNA sequence that forms complexes with eukaryotic initiation factors (eIFs) resulting in the subsequent recruitment of the ribosome to the internal regions of mRNA (Pestova *et al.*, 1996; Hellen and Sarnow, 2001). The poly (A) independent initiation mechanism is another translation strategy used by (+ve) ssRNA viruses whose genomes lack a 3' poly (A) tail but possess a 5' cap structure such as TYMV (Matsuda *et al.*, 2004). The 3' UTR of TYMV is used as a substitute for the poly (A) tail and contains a tRNA-like structure (TLS) that interacts and forms complexes with eIFs. The TLS-eIF complex forms a closed loop with the 5' cap and this conformation facilitates protein translation (Matsuda *et al.*, 2004). For (+ve) ssRNA viruses that lack both a 5' cap and 3' poly (A) tail within the genome use the translation enhancer domains (TED) at the 3' UTR and which recruits the eIFs to the 3' end (Meulewaeter *et al.*, 1998a;

1998b). The TED-eIFs complex is joined to the 5' UTR forming a closed loop format (Gazo *et al.*, 2004) that facilitates the loading of the ribosome at the 5' start codon and thereby initiating protein translation in a cap and poly (A) independent initiation process (Gazo *et al.*, 2004; Wang *et al.*, 1997).

The readthrough stop codon strategy involves misreading of an in-frame stop codon as a sense codon by the ribosomes resulting in the synthesis of a C-terminally extended polypeptide with novel functions. This phenomenon was first reported for the plant virus, Tobacco mosaic virus (TMV) and later observed in several (+ve) ssRNA plant and animal viruses (Pelham, 1978). For most viruses, the process allows the generation of a putative RdRp or elongated coat protein that is involved in replication, virus assembly and transmission (Zaccomer *et al.*, 1995).

The readthrough mechanism requires three critical elements in order to be functional; the codon context surrounding the stop codon, cellular suppressor transfer RNAs and 3' RNA structural element adjacent to the stop codon. The codon context is a major determinant known to affect the efficiency of translation termination codons and extends to the hexanucleotide sequence at the 3' end of the readthrough stop codon (Skuzeski *et al.*, 1991; Beier and Grimm, 2001). Mutational experiments within this region have either abolished or reduced suppression of the readthrough stop codon suggesting its importance in RNA replication (Kopelowitz *et al.*, 1992). Interestingly, the codon context does not have to be located at the 3' end of the stop codon as recoding signals located 700 bases away from the stop codon can still promote readthrough mechanism (Brown *et al.*, 1996). More recent studies have shown that the presence of 3' RNA structures located eight nucleotides downstream of the stop codon enhance readthrough efficiency (Firth *et al.*, 2011; Napthine *et al.*, 2012).

Comparative analyses of the codon context surrounding the stop codon in several plant and animal viruses revealed sequence homology at the 3' end. This resulted in the classification into three groups, namely I, II, III (Beier and Grimm, 2001). The type 1 readthrough region is found in plant viral RNAs that share the six nucleotides with a sequence consensus of CAAUYA (Y = pyrimidine) at the 3' side of the suppressed UAG or UAA codons. Mutations of bases within this region either

abolish or reduce UAG suppression *in vitro* (Skuskei *et al.*, 1991; Zerfass and Beier 1992a; Beier and Grimm, 2001). The type II readthrough stop is characterized with a CGG or CUA codon at the 3' end of the suppressed UGA or UAA codon (Beier and Grimm, 2001). Type III is a more complex signal that consists of a purine-rich spacer between the 3' end of the termination codon and a pseudoknot that spans a number of nucleotides consisting of guanines (Feng *et al.*, 1992; Wills *et al.*, 1994; Beier and Grimm, 2001). Both type II and III readthrough stop codons are common in plant and animal viruses.

The second major determinant that affects the efficiency of translation termination codon is the availability of naturally occurring suppressor eukaryotic tRNAs (Beier and Grimm, 2001). These have the ability to suppress the readthrough stop codon by insertion of an appropriate amino acid. Several studies have demonstrated the significance of suppressor tRNAs in promoting readthrough in (+ve) ssRNA viruses (Beier and Grimm, 2001). These tRNAs can be classified as UAG/UAA and UGA suppressors. The first group of eukaryotic UAG/UAA suppressors is recognized by cytoplasmic tRNA^{Tyr}, tRNA^{Gln} and tRNA^{Leu} resulting in the insertion of a tyrosine, glutamine or leucine amino acid respectively at the stop codon site (Beier and Grimm, 2001). Substantial data obtained from studies with cytoplasmic tRNA^{Tyr} demonstrated its ability to recognize and naturally suppress in-frame UAG stop codons (Beier *et al.*, 1984a; Beier *et al.*, 1984b; Barciszewski *et al.*, 1985; Beier *et al.*, 1987).

Cytoplasmic tRNA^{Tyr} carries a GΨA anticodon sequence shown to function as an active UAG suppressor *in vitro* and therefore a promoter of the readthrough mechanism (Beier *et al.*, 1984b; Zerfass and Beier, 1992a). The pseudouridine (Ψ) base at the second position of the anticodon is a unique feature of all eukaryotic tRNA^{Tyr} and critical for the suppression of the UAG stop codon (Beier *et al.*, 1984a; Beier *et al.*, 1984b; Suter *et al.*, 1986). Zerfass and Beier (1992a) demonstrated that substituting the tRNA^{Tyr} GΨA with a GUA anticodon resulted in the inability of this tRNA to suppress the UAG codon, indicating that the Ψ is necessary for the unconventional codon reading. The pseudouridine base within the GΨA sequence base pairs with the A of the UAG whilst the A of the anticodon base pairs with U of

the stop codon forming a stable codon-anticodon interaction (Beier *et al.*, 1984b). Base pairing of the guanine within the GΨA anticodon to the G of the UAG is also important for the suppression of the UAG stop codon. Modifications within the GΨA anticodon have been reported to have significant impact on UAG suppression. These modifications involve the substitution of the guanine to queuosine base (Q) and have been isolated in several plants and animals (Beier *et al.*, 1987; Johnson *et al.*, 1985). The modified Q nucleoside is synthesized as a free queuosine base that is then inserted into tRNA^{Tyr} by a transglycosylase in place of guanine (Okada *et al.*, 1979). The presence of this hypermodified Q at the first anticodon position of tRNA^{Tyr} does not suppress the UAG stop codon and results in a lack of stimulation of the UAG readthrough (Bienz and Kubli, 1981; Junker *et al.*, 1997).

Cytoplasmic tRNAs^{Gln} carries either a CUG or U*UG anticodon (where * denotes a modified nucleoside). These two isoacceptors exist in all prokaryotes and eukaryotes and suppress UAG and UAA stop codons by the insertion of a glutamine at the site of the stop codon (Yoshinaka *et al.*, 1985; Weiss and Friedberg, 1986; Sprinzl *et al.*, 1998). The ability to stimulate readthrough over the UAG codon has been demonstrated for TMV RNA *in vitro* with the U*UG anticodon as a less efficient UAG suppressor than tRNA^{Gln} with a CUG anticodon (Kuchino *et al.*, 1987; Grimm *et al.*, 1998). The cytoplasmic tRNA^{Leu} carries CAA and CAG anticodons and these have been shown to suppress the UAG stop codon both *in vitro* and *in vivo*, with the insertion of a leucine residue at the site of the stop codon (Valle *et al.*, 1987).

The second group which comprises the eukaryotic UGA suppressors are recognized by cytoplasmic tRNA^{Trp}, tRNA^{Cys} and tRNA^{Arg} that allow the insertion of tryptophan, cysteine or arginine at the site of the stop codon. The cytoplasmic tRNA^{Trp} with CmCA anticodon sequence promotes UGA readthrough *in vitro* with efficient activity observed for chloroplast tRNA^{Trp} compared to cytoplasmic tRNA^{Trp} (Zerfass and Beier, 1992b). Such differences are attributed to specific base modifications located at position 24 of the D-stem of the chloroplast tRNA structure. Mutations at this position are presumed to permit unconventional base pairing and in turn result in greater suppression efficiency of the UGA codon (Raferty *et al.*, 1986; Smith and Yarus, 1989). The tRNAs^{Cys} isolated from chloroplast and cytoplasm have a GCA anticodon and directs the incorporation of cysteine at the site of the UGA

codon (Feng *et al.*, 1990). Readthrough activity of both isoacceptors has been demonstrated *in vitro* with efficient activity observed with chloroplast tRNAs^{Cys} than its cytoplasmic counterpart (Urban and Beier, 1995). Lastly, tRNA^{Arg} with U*CG, and to a minor extent tRNA^{Arg} with the ICG anticodon, promote UGA readthrough with the insertion of an arginine at the site of UGA stop codon (Baum and Beier, 1998).

1.3 THE TETRAVIRUSES

The tetraviruses were first described by Hendry *et al.* (1968) with five structurally distinct small RNA viruses isolated from infected *Nudaurelia cytherea capensis* larvae and designated *Nudaurelia capensis* γ , δ , α , ε , β virus according to their electrophoretic migration in sucrose gradient (Juckes, 1970). Of the five, *Nudaurelia capensis* β virus (N β V) is the most extensively characterized with structural analysis of the virus particles revealing their unusual $T = 4$ icosahedral symmetry (Finch *et al.*, 1974). N β V became the type virus for a new group of viruses known as the *Nudaurelia* β -like viruses (Matthews, 1982). Earlier reports suggested serological relatedness between *Antheraea eucalypti* virus (AeV) from *Antheraea eucalypti* and N β V and so AeV became the second member of the *Nudaurelia* β -like virus group (Grace and Mercer, 1965; 1976). The characteristic $T = 4$ quasi-symmetry of their capsids resulted in renaming the group the *Tetraviridae* (Francki *et al.*, 1991). In 1985, Hendry *et al.* isolated another virus named *Nudaurelia capensis* ω virus (N ω V) with no serological relationship to N β V, despite similar biophysical and host range characteristics. Although both displayed the same $T = 4$ icosahedral capsid symmetry, N ω V contained a bipartite genome as opposed to the monopartite genome of N β V (Hendry *et al.*, 1985; Finch *et al.*, 1974). Based on this N ω V became the type member of a second genus, *Nudaurelia* ω -like tetravirus and later on were joined by *Helicoverpa amigera* stunt virus (HaSV) that was discovered in a collapsed *Helicoverpa armigera* (cotton bollworm) colony (Hanzlik *et al.*, 1993).

1.3.1 Host range and pathology

The host range of tetraviruses is limited to one or a few closely related species of the order Lepidoptera with the midgut tissue as the primary site of infection (Moore, 1991; Bawden *et al.*, 1999; Dorrington and Short, 2010; Dorrington *et al.*, 2011). Tetraviruses are pathogenic and cause significant mortality in infected insect larvae (Grace and Mercer, 1965; Tripcone, 1970; Greenwood and Moore, 1984; Hanzlik *et al.*, 1993). The symptoms of infected insect larvae include discolouration of the larvae, lethargy, flaccidity, regurgitation and diarrhea (Juckes, 1970; Hendry *et al.*, 1985; Walter *et al.*, 2008). Infection of *H. amigera* larvae with HaSV revealed that the virus particles were confined to the cytoplasm of the columnar, goblet and regenerative cells of the midgut tissue, suggesting that the replication cycle of this virus is exclusively cytoplasmic (Moore, 1991; Bawden *et al.*, 1999; Brooks *et al.*, 2002). The morphological changes of the infected midgut cells induced by HaSV infection were characterized by accumulation of lysosomes, increased cytoplasmic vacuolation, endoplasmic vesiculation and increased shedding of cells into the midgut lumen (Brooks *et al.*, 2002).

In the absence of tissue culture cell line susceptible to infection, Short *et al.* (2010) and Short and Dorrington, (2012) expressed the HaSV viral replicase in the absence of replication in Sf9 cells and HeLa cells to determine subcellular localization of the protein. The replicase-EGFP punctate structures partially overlapped with late endosomes and this led to proposal that the viral replicase was likely to associate with membranes derived from the endocytic pathway. The EGFP-labelled punctate structures were observed to undergo fusion and fission events (Short *et al.*, 2010) and similar events are common with endocytic vesicles (Reviewed in Mayor and Pogano, 2007). This observation further supports association of HaSV replicase with the endosomes. In persistently PrV infected MG8 cells, the virus particles were found confined to the cytoplasm and in some cases within a structure that appeared membrane bound (Pringle *et al.*, 2003).

1.3.2 Taxonomic classification of tetraviruses

The tetraviruses were originally classified into the *Tetraviridae* as non-enveloped insect RNA viruses with the characteristic $T = 4$ capsid symmetry and then into two genera based on the bipartite vs monopartite genome organisation (Finch *et al.*, 1974; Dorrington *et al.*, 2011). The discovery of several new betatetraviruses has resulted in the successful proposal to International Committee on Taxonomy of Viruses (ICTV) to re-classify the viruses into three families based upon their replicases, including *Alphatetraviridae*, *Permutotetraviridae* and *Carmotetraviridae* (Dorrington *et al.*, 2011).

1.3.3 Capsid structure, assembly and maturation

Cryo-electron microscope studies with N ω V revealed a novel $T = 4$ icosahedral lattice symmetry with each capsid consisting of 240 identical protein subunits arranged as four Y-shaped trimers per triangular face (Finch *et al.*, 1974). With the advent of high resolution processing techniques, the structure of N β V virions was solved to 32 Å showing three distinct pits on each triangular face separated by a deep groove (Olson *et al.*, 1990) (Figure 1.5A). The crystal structure of N ω V was solved at 2.8 Å showing similar structural features to those of N β V, although N ω V capsid structure was smaller and compact compared to N β V and displayed no pits or grooves between each triangular face (Figure 1.5C) (Munshi *et al.*, 1996; Helgestrand *et al.*, 2004). The three dimensional structures of HaSV and PrV particles have also been determined with the crystal structure of PrV solved at a resolution of 3.8 Å (Figure 1.5 D and B) (Speir and Johnson, 2008; Speir *et al.*, 2010).

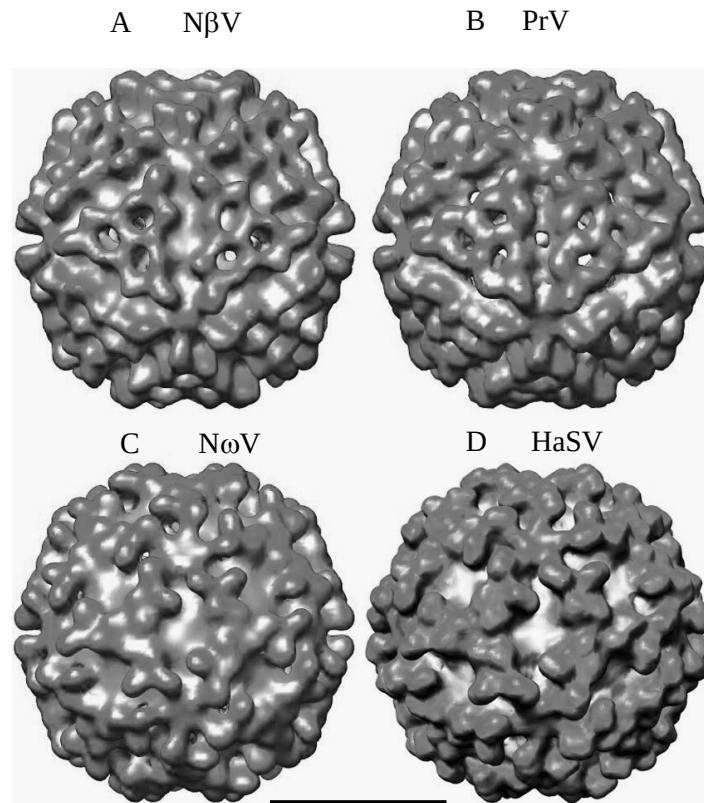


Figure 1. 5: Three-dimensional reconstruction of N β V, N ω V, PrV and HaSV particles using cryo-electron microscopy. The structures are at 25-30 Å resolution and viewed down their icosahedral two-fold axes. The bar represents 40 nm for *Nudaurelia capensis* ω virus (N ω V), Providence virus (PrV), *Helicoverpa amigera* stunt virus (HaSV) and *Nudaurelia capensis* β virus (N β V). Reproduced from Speir and Johnson (2008).

The unique $T = 4$ icosahedral architecture of tetravirus particles distinguishes them from other non-enveloped viruses. According to Caspar and Klug (1962), the use of a large triangulation (T) number implies that the protein subunits are asymmetrically arranged with respect to the edge of each face which creates uneven spacing between all protein subunits. This affects the assembly of the protein subunits. To overcome this problem, Casper and Klug (1962) proposed the idea of quasi-equivalence that suggests that nearly equivalent bonds are formed between protein subunits within the same local environment allowing self-assembly of icosahedral capsids from multiple subunits. Therefore assembly of a particle with $T = 4$ symmetry would require formation of trimers that would associate as groups of four trimers on each icosahedral face to collectively form the icosahedral shell (Finch *et al.*, 1974) (Figure 1.6C).

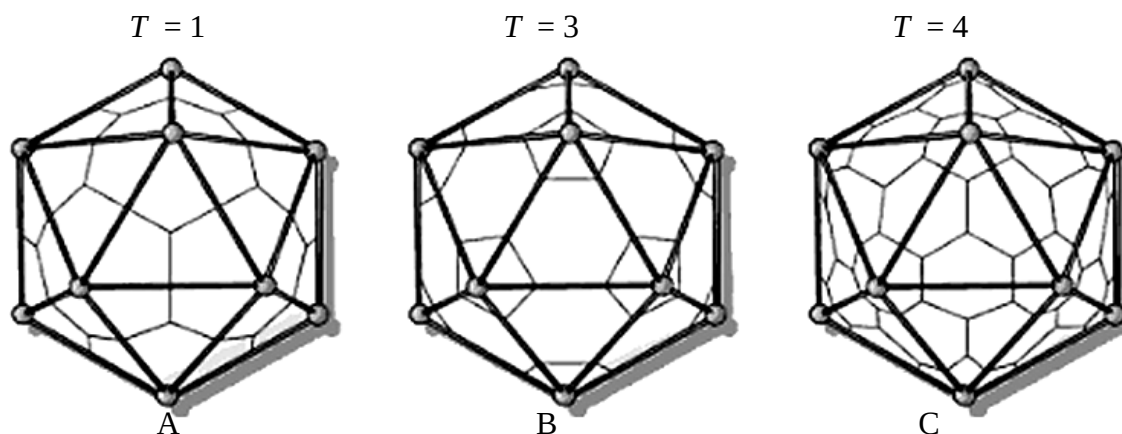


Figure 1. 6: Icosahedral capsids with triangulation numbers of 1, 3 and 4. A $T = 1$. B $T = 3$. C $T = 4$. At $T = 1$ and $T = 3$ icosahedral architecture, the protein subunits are arranged symmetrically with respect to the edge of each face. **C** $T = 4$, the protein subunits are asymmetrically arranged with respect to the edge of each face. Reproduced from Cann, (2005).

Nudaurelia capensis ω is the tetravirus whose structure has been studied in the most detail. Each of the 240 identical protein subunits of N ω V consists of an external immunoglobulin-like domain, a core jelly roll domain and an internal helical domain comprising the N- and C-terminus of the protein subunit (Figure 1.7, Panel A). Both the jelly roll domain and the internal helical domain display significant similarity to the $T = 3$ nodaviruses (Munshi *et al.*, 1996; Helgestrand *et al.*, 2004). The positively-charged internal helical domain has been implicated in RNA binding and particle stabilization during assembly whilst the jelly roll β barrel structures form the contiguous shell of the capsid that protects the viral RNA from the outer environment (Munshi *et al.*, 1996; Helgestrand *et al.*, 2004). The immunoglobulin (Ig)-like domain located outside the capsid shell is proposed to function in host cell recognition during infection (Munshi *et al.*, 1996; Helgestrand *et al.*, 2004).

Prior to capsid maturation, the protein subunits of N ω V undergo self-association with each subunit adopting different conformations of A, B, C and D with nearly equivalent bonding with each other and resulting in self-assembly of the procapsid (α) (Figure 1.7, Panel B) (Agrawal and Johnson, 1996; Canady *et al.*, 2000). The procapsid undergoes autoproteolytic cleavage at the C-terminus resulting into a major (β) protein subunit and minor (γ) peptide that form mature viral capsids during

maturation (Agrawal and Johnson, 1995). The cleavage of the γ -peptide results in the stabilization of mature virus particle and is proposed to be important for infectivity (Munshi *et al.*, 1996; Taylor *et al.*, 2002; Helgstrand *et al.*, 2004). The autoproteolytic sites at the C-terminus of N ω V procapsids are located at the N₅₇₀ and F₅₇₁ of the internal helical domain N ω V (Taylor, 2005). The architecture of the N ω V capsid is controlled by a molecular switch formed by the C-termini of the γ -peptide that is ordered in the C and D subunits. The helical regions of the C and D subunit are wedged between ABC and DDD asymmetrical units resulting in the formation of a flat contact at this location (Figure 1.7, Panel B). The lack of order of the γ -peptide in the A or B subunits results in the bent morphology of the interface between two ABC units (Munshi *et al.*, 1996; Helgstrand *et al.*, 2004).

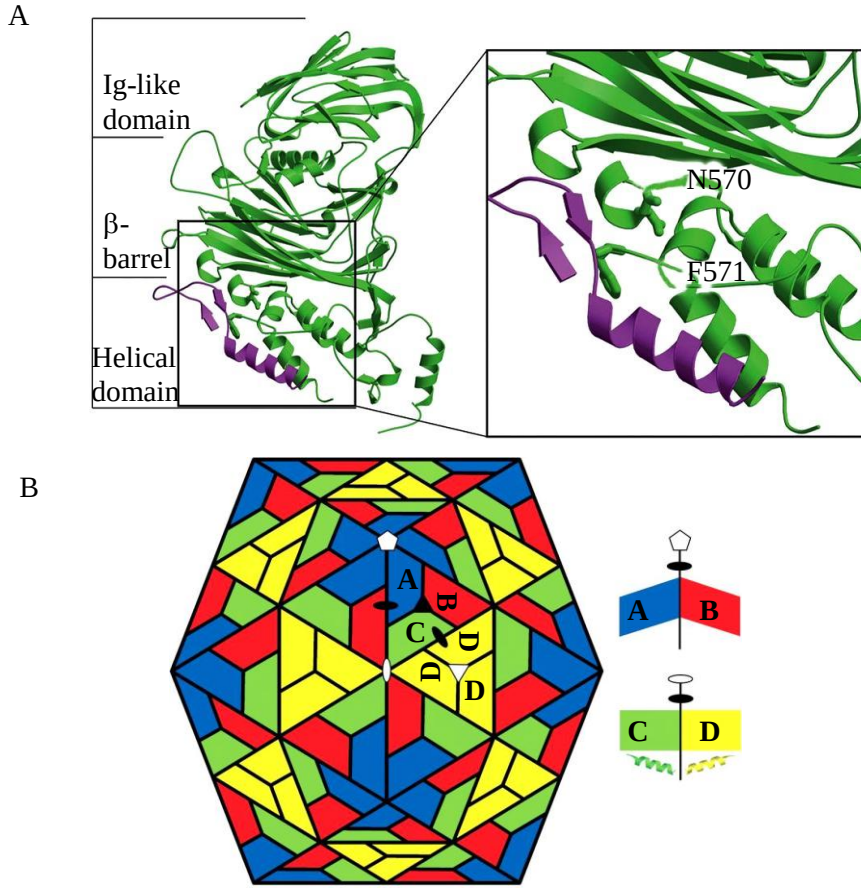


Figure 1. 7: The structure and quasi-equivalence of NøV capsid. **A** Ribbon diagram of the C subunit protein showing three structural domains. The top represents the external Ig-like domain while the middle represents the jelly roll β barrel. The internal helical domain is at the bottom and an enlarged portion showing residues N₅₇₀ and F₅₇₁ autoproteolytic cleavage sites. **B** Schematic diagram of the positions of the four subunits A (blue), B (red), C (green) and D (yellow) in $T = 4$ capsid symmetry. White pentagons, triangles and ellipses represent 5-fold, 3-fold and 2 fold icosahedral symmetry axes while black ellipses and triangles represent quasi-2-fold and quasi-3-fold axes. The bent A-B or flat C-D conformations result from absence or presence of an ordered C-terminal helix. Reproduced from Speir and Johnson, (2008).

A combination of cryoEM and X-ray crystallography analysis of NøV VLPs has shown that the procapsid and capsids are structurally distinct with the procapsid larger and more rounded than mature capsids (Canady *et al.*, 2000, 2001). The procapsid to capsid transition is pH dependent with large stable particles formed at neutral pH while small and mature compact capsid structures are formed at low pH (Taylor *et al.*, 2002).

Table 1. 2: General properties of tetraviruses with well characterized genomes

Family	Type virus	Host	Length of genome (nts)	Replicase (kDa)	VCAP (kDa)	Accession number	References
<i>Alphatetraviridae</i> Genus: <i>Omegatetravirus</i>	<i>Helicoverpa armigera</i> stunt virus	<i>Helicoverpa armigera</i>	RNA 1(5312) RNA 2(2478)	187	71	NC_001981.1 NC_001982.1	Hanzlik <i>et al.</i> , 1993 Gordon <i>et al.</i> , 1995 Hanzlik <i>et al.</i> , 1995
	<i>Dendrolimus punctatus</i> tetravirus	<i>Dendrolimus punctatus</i>	RNA 1(5492) RNA 2(2490)	180	70	NC_005898.1 NC_005899.1	Yi <i>et al.</i> , 2005
	<i>Nudaurelia capensis</i> ω virus	<i>Nudaurelia capensis</i>	RNA 1 (5300) RNA 2 (2445)	180	70	S43937.1	du Plessis <i>et al.</i> , 2005 Hendry <i>et al.</i> (1985)
Genus: <i>Betatetravirus</i>	<i>Nudaurelia capensis</i> β virus	<i>Nudaurelia capensis</i>	6625	215	66	NC_001990.1	Gordon <i>et al.</i> , 1999 Juckes, 1970
<i>Permutotetraviridae</i>	<i>Thosea asigna</i> virus	<i>Thosea asigna</i>	5714	140	82	AF282930.1	Pringle <i>et al.</i> , 1999
	<i>Euprosterina elaeasa</i> virus	<i>Euprosterina elaeasa</i>	5698	140	82	NC_003412.1	Zeddam <i>et al.</i> , 2010
<i>Carmotetraviridae</i>	Providence virus	<i>Helicoverpa zea</i> midgut cell line 8	6155	104	68	NC_014126.1	Pringle <i>et al.</i> , 2003 Walter <i>et al.</i> , 2010

The Alphetetraviridae

The *Alphetetraviridae* consist of two genera that include the *Omegatetravirus* and *Betatetravirus* with the genome defined as bipartite or monopartite respectively (Table 1.2). The complete bipartite genome sequence for members of the *Omegatetravirus* has been determined for HaSV, *Dendrolimus punctatus* tetravirus (DpTV) and only RNA 2 for NβV (Gordon *et al.*, 1995; Hanzlik *et al.*, 1995; Yi *et al.*, 2005; Agrawal and Johnson, 1992; du Plessis *et al.*, 2005). The replicase is encoded by RNA1 (Gordon *et al.*, 1995; Yi *et al.*, 2005). In addition, three additional polypeptides that consist of 11 kDa, 15 kDa, 8 kDa are encoded at the 3' end of the RNA 1 (Gordon *et al.*, 1995; Yi *et al.*, 2005) (Figure 1.8). RNA 2 encodes the capsid protein precursor (VCAP) (70 kDa - 71 kDa) and a smaller overlapping ORF producing a polypeptide with a predicted molecular mass of 17 kDa that is encoded at the 5' proximal end of RNA 2 (Hanzlik *et al.*, 1995; du Plessis *et al.*, 2005; Yi *et al.*, 2005). Both RNA 1 and RNA 2 molecules are capped at the 5' end and are non-polyadenylated at the 3' end (Dorrington *et al.*, 2011).

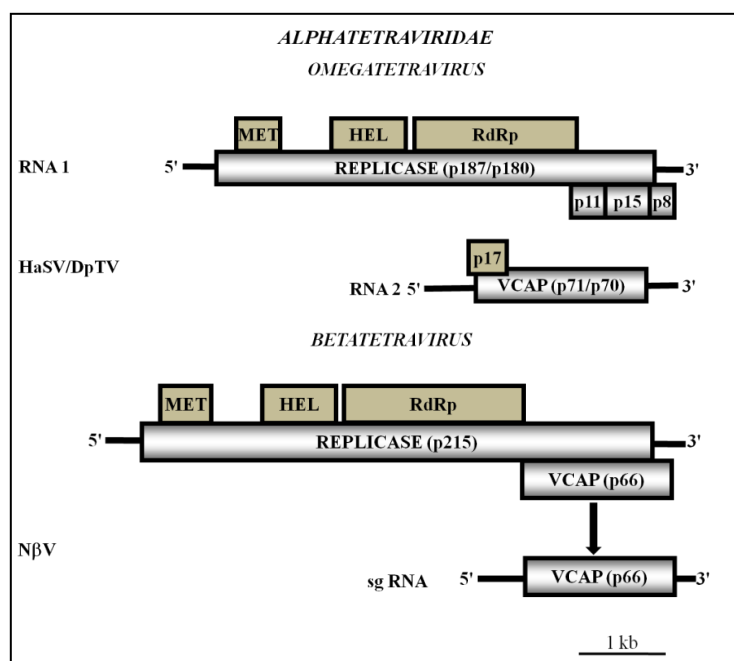


Figure 1. 8: Schematic representation of the genomic organisation of the *Alphetetraviridae*. Molecular masses are in kDa. The conserved domains of the replicase are denoted as MET, HEL and RdRp. Genomic RNA is denoted as RNA while subgenomic RNA is indicated as sgRNA. HaSV- *Helicoverpa armigera* stunt virus, DpTV- *Dendrolimus punctatus* tetravirus and NβV- *Nudaurelia capensis* beta virus.

Within the genus *Betatetravirus* is the species NβV. The monopartite genome encodes two overlapping ORFs with the viral replicase located at the 5' end and the VCAP at the 3' end. The VCAP ORF is translated from a subgenomic RNA originating upstream of the VCAP coding sequence (Gordon *et al.*, 1999).

The replicases encoded by the alphatetraviruses (HaSV, DpTV and NβV) consist of a MET, HEL and RdRp functional domain in an amino-to-carboxyl terminal arrangement (Gordon *et al.*, 1999; 1995; Yi *et al.*, 2005). There is a higher degree of sequence similarity between HaSV and DpTV and less between these two viruses and NβV. The HaSV and DpTV display 67 %, 63 % and 76 % while HaSV and NβV exhibit 45 %, 42 % and 33 % identity between MET, HEL and RdRp domains respectively. In addition, less sequence similarity is observed for DpTV and NβV with 38 % and 36 % sequence similarity between MET and RdRp (Gordon *et al.*, 1999; Yi *et al.*, 2005). Although several insect-borne animal (alphaviruses) and plant viruses (marafiviruses) have been classified as alpha-like viruses, the alphatetraviruses; HaSV, DpTV and NβV are the only insect-specific viruses within the supergroup whose replicases do not undergo proteolytic cleavage processes (Gordon *et al.*, 1995; Gordon *et al.*, 1999; Yi *et al.*, 2005). This is a characteristic shared with only a small group of plant viruses that includes the unclassified *Foveavirus* and the *Potexvirus* genera (Martelli and Jelkmann, 1998; Huisman *et al.*, 1988).

Subcellular localization studies of HaSV replicase revealed that the N- and C-terminal region are required for targeting to membranes derived from the endocytic pathway (Short *et al.*, 2010; Short and Dorrington, 2012). Deletion of the first 44 amino acids at the N-terminus and last 400 amino acids at the C-terminus altered the distribution of punctate fluorescence of the full-length HaSV replicase. The HaSV replicase amino acid sequence contains a putative helix-loop-helix conformation located at 6 – 44 amino acids of the N-terminus. This region that is conserved among the replicases of HaSV, DpTV and NβV suggests a possibility of a common localization domain (Short and Dorrington, 2012).

The *Permutotetraviridae*

The *Permutotetraviridae* family contains only two species *Thosea asigna virus* (TaV) and *Euprosterina elaeasa virus* (EeV), whose genomes are almost identical but which infect different species (Reinganum *et al.*, 1978; Pringle *et al.*, 1999; Gorbalenya *et al.*, 2002; Zeddarn *et al.*, 2010). The monopartite genome of TaV and EeV encodes the viral replicase at the 5' and the VCAP at the 3' end (Figure 1.9). The TaV VCAP is translated from a subgenomic RNA originating upstream of the VCAP coding sequence (Pringle *et al.*, 1999). The presence of a 2A-like processing site located near the N-terminus of the TaV VCAP coding sequence is predicted to generate a putative 17 kDa N-terminal peptide as well as the capsid protein precursor of approximately 65 kDa (Pringle *et al.*, 1999). The latter is then autoproteolytically cleaved at the C-terminus during particle maturation to produce p58 and p7 that are found in mature capsids (Pringle *et al.*, 1999). The EeV VCAP ORF encodes a protein with a predicted molecular mass of 82.5 kDa and predicted to undergo co-translational processing at an N-terminus via a 2A-like site that is conserved in the same position as in the TaV VCAP ORF (Zeddarn *et al.*, 2010).

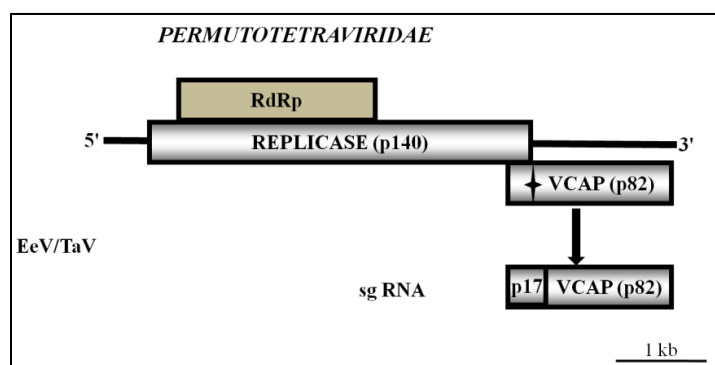


Figure 1. 9: Schematic representation of the genomic organisation of the *Permutotetraviridae*. Molecular masses are in kDa. The star represents a 2A-like site. sgRNA refers to subgenomic RNA. EeV- *Euprosterina elaeasa virus* and TaV- *Thosea asigna virus*.

The picorna-like replicases of TaV and EeV contain an internal permutation of the replicase and share 68 % sequence identity over the 1257 amino acid protein sequence that are predicted to generate a protein with a molecular mass of 140 kDa (Gorbalenya *et al.*, 2002; Zeddarn *et al.*, 2010). TaV and EeV replicase lack a MET and HEL domain however both contain a putative VPg signal that is present

upstream of the RdRp domains and believed to serve as a protein primer for RNA synthesis (Gorbalenya *et al.*, 2002; Zeddam *et al.*, 2010). The possession of a VPg distinguishes the *Permutotetraviridae* from the *Alphatetraviridae* and *Carmotetraviridae*.

The Carmotetraviridae

The focus of this study is Providence virus (PrV), the only member of the *Carmotetraviridae* and the only tetravirus that is able to replicate in tissue culture (Dorrington *et al.*, 2011). The virus was discovered by Pringle *et al.* (2003) as a persistent infection in MG8 cells of a laboratory colony *Helicoverpa zea* (*H. zea*) at the 8th passage. Goodman and McIntosh, (1995) established the MG8 cell line and found that the cell line can grow in culture medium lacking foetal calf serum and can be cultured through a number of passages. This led to the conclusion that the cell line is likely to be derived from the basal regenerative cells of the midgut tissue and therefore supporting persistent infection of PrV. In addition, high titres of PrV particles are required for productive infection and this led to the conclusion that the virus is not highly infectious (Pringle *et al.*, 2003).

PrV has a monopartite genome of 6155 nts that, unlike the other tetraviruses, encodes three ORFs, a unique feature of PrV. These include p130, VCAP and p104 (also the viral replicase) (Walter *et al.*, 2010). The p130 and VCAP ORFs are located at the 5' and 3' end of the genome respectively while p104 ORF is located upstream of the VCAP coding sequence and overlaps almost entirely with the p130 ORF (Figure 1.10). There are three putative 2A-like processing sites. PrV-2A₁ is located at the N-terminus of p130 and whose activity is predicted to produce two translation products of 17 kDa and 113 kDa respectively. *In vitro* studies showed that PrV-2A₁ sequence is functional (Luke *et al.*, 2008). The second (PrV-2A₂) and third (PrV-2A₃) 2A-like processing sites are located at the N-terminus of the VCAP ORF (Figure 1.10) and these sequences are functional *in vitro* studies (Luke *et al.*, 2008). The activity of PrV-2A₃ is predicted to produce a 13.4 kDa translation product however the protein was not detected *in vivo* instead *in vitro* experiments detected a 15 kDa protein using anti-VCAP antibodies suggesting that PrV-2A₃ is functional (Pringle *et al.*, 2003; Walter *et al.*, 2010).

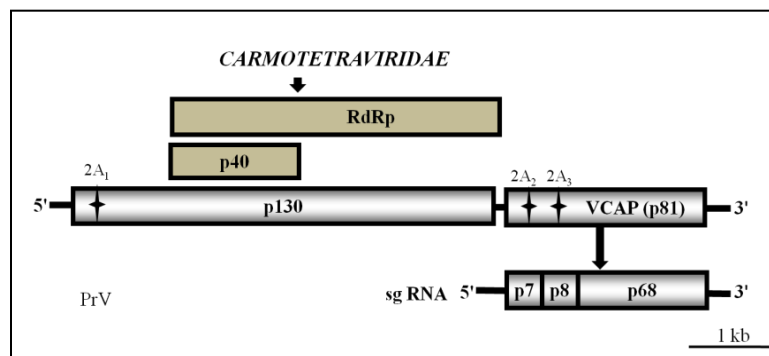


Figure 1. 10: Schematic representation of the genomic organisation of the *Carmotetraviridae*. The short black arrow indicates the readthrough stop signal while the star indicates the 2A-like processing sites. Molecular masses are in kDa. PrV- Providence virus

During particle maturation, the capsid protein precursor of 68 kDa is autoproteolytically cleaved during maturation to form 60 kDa and 7.4 kDa respectively (Taylor *et al.*, 2006). The surface morphology of PrV capsid resembles that of NβV capsid and each capsid subunit of PrV has three domains similar to those observed in NωV (Olson *et al.*, 1990; Speir *et al.*, 2010). Unlike NωV, the molecular switch in PrV is formed by the N-termini of the β-peptide and the groove between the ABC and DDD units is filled by the ordered N-termini of the C and D subunits forming a flat contact (Speir *et al.*, 2010). The molecular switch in PrV is more similar to *T* = 3 nodaviruses and therefore PrV may structurally represent an ancestral tetravirus from which NωV evolved (Speir *et al.*, 2010).

The replicase of PrV lacks the MET and HEL domains, and shares similarities with members of the carmo-like superfamily (Walter *et al.*, 2010). The RdRp domain of PrV forms a distinct cluster with the replicases of (+ve) ssRNA plant viruses belonging to the *Tombusviridae* and *Umbraviridae* that are classified within the carmo-like supergroup II (Walter *et al.*, 2010; Dorrington *et al.*, 2011). A further distinction between PrV and other tetraviruses replicases is the presence of a readthrough stop codon that results in the production of a major (p40) and a minor (p104, containing the RdRp domain) replication proteins required for the establishment of infection (Walter *et al.*, 2010).

The identification of PrV RdRp motifs (Walter *et al.*, 2010) conserved within the replicases of (+ve) ssRNA plant viruses of the carmo-like superfamily suggest that PrV replicase is phylogenetically most closely related to members of the family *Tombusviridae* and *Umbraviridae*. The translational control system for the expression of PrV replicase resembles that employed by tombusviruses and this further suggests that PrV replicase might share functional similarities with tombusvirus replicase.

1.4 THE *TOMBUSVIRIDAE*

The family *Tombusviridae* comprises nine genera that include *Carmovirus*, *Necrovirus*, *Dianthovirus*, *Luteovirus*, *Machlomovirus*, *Aureusvirus*, *Avenavirus*, *Panicovirus* and *Tombusvirus* (Rochon *et al.*, 2011). The viral particles are small (28-34 nm) with a non-enveloped capsid that exhibits $T = 3$ icosahedral symmetry. Tombusvirus particles encapsidate a 3.5 - 5.7 kb (+ve) ssRNA monopartite genome which is neither 3' polyadenylated nor 5' capped (Rochon *et al.*, 2011). However, machlomoviruses and dianthoviruses possess a 5' cap structure while the luteoviruses have a Vpg protein that is covalently linked to the 5' end of the genome (Nutter *et al.*, 1989; Xiong and Lommel, 1989; Murphy *et al.*, 1998). Functions performed by the 5' cap are substituted for by the possession of a translation enhancer domain (TED) within the 3' untranslated region in viruses lacking the 5' cap structure (Wang *et al.*, 1997). Unlike the rest of the tombusviruses, the genus *Dianthovirus* consists of a bipartite genome with RNA 1 (3.7 - 4 kb) and RNA 2 (1.3 - 1.4 kb) that are encapsidated in separate viral particles (Rochon *et al.*, 2011).

1.4.1 Genome organisation

The genomes of the tombusviruses encode three to five ORFs with three ORFs encoded by *Dianthovirus* and *Avenavirus* while *Necrovirus* and *Panicovirus* express five ORFs (Rochon *et al.*, 2011). The *Carmovirus*, *Luteovirus*, *Machlomovirus*, *Aureusvirus*, *Panicovirus* and *Tombusvirus* encode four ORFs. The 5' proximal ORFs, ORF1 and ORF1-RT or ORF1-FS (Figure 1.11) are transcribed from the genomic RNA while the internal ORF2 and 3' proximal ORF3 and ORF4 are transcribed from subgenomic RNAs (Rochon *et al.*, 2011). The ORF1 encodes the

accessory protein (28 - 50 kDa) while ORF1-RT or ORF1-FS encodes full-length replicase (88 - 112 kDa) produced via the readthrough signal or -1 ribosomal frame shift (-1 FS) (Table 1.3). The ORF2 encodes the capsid protein while the ORF3 and ORF4 encode the movement protein and suppressor protein respectively (Rochon *et al.*, 2011). The full-length replicase contains signature RdRp motifs (I - VIII) within its non-overlapping C-terminal region and functions as the polymerase (Rochon *et al.*, 2011). On the other hand, the capsid protein encapsidates genomic RNA, mediates long distance movement and establishes infection (Dalmay *et al.*, 1992; Qu and Morris, 2002; Reade *et al.*, 2010). The movement protein is required for virus movement between cells while the suppressor protein is responsible for suppressing the RNA silencing host defence mechanism (Scholthof *et al.*, 1995a; Huppert *et al.*, 2002; Ahn *et al.*, 2011; Hao *et al.*, 2011).

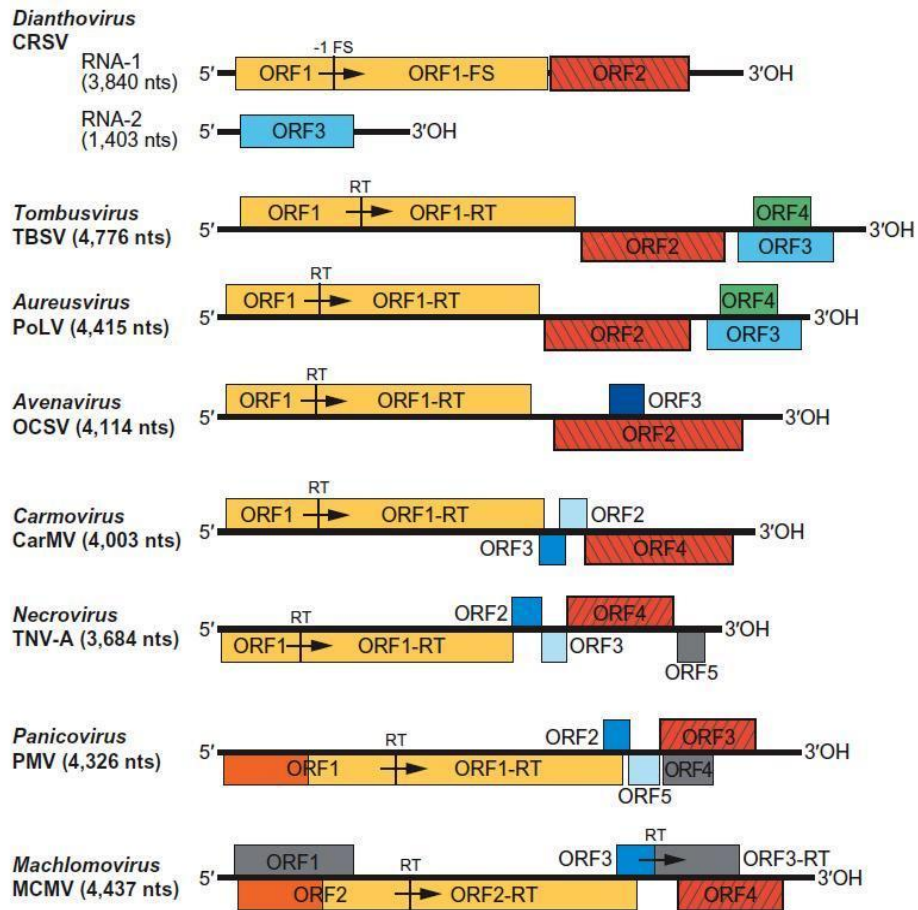


Figure 1. 11: Schematic representation of the genome organization of the *Tombusviridae*. The length of each genome (nts) for the genera and type species of the *Tombusviridae* is shown at left. The yellow boxes (■) represent replicase ORF. Red-hatched boxes (■) represent capsid protein (CP) ORF. Blue boxes represent viral movement proteins (MPs). Light-blue shaded boxes represent MPs conserved in carmoviruses, necroviruses and panicoviruses (■). Dark blue represent MP conserved in avenaviruses (■). The green shaded boxes (■) represent silencing suppressor proteins exhibited in tombusviruses and aureusviruses. The orange shaded (■) region in *Panicovirus* and *Machlomovirus* is conserved among the two genera but not with other members of the *Tombusviridae*. The gray boxes (■) represent ORFs that encode unique proteins with no sequence similarity with other viral proteins among the *Tombusviridae*. RT denotes readthrough signal. The -1 ribosomal frameshift is denoted by -1 FS. All are uncapped genomes except for *Machlomovirus* and *Dianthovirus*. Reproduced from Rochon *et al.* (2011).

Table 1. 3: Replication proteins and mode of expression of the *Tombusviridae* replicases

Genera	Type virus	Replication proteins	Expression mode of RdRp	References
<i>Tombusvirus</i>	<i>Tomato bushy stunt virus</i> (TBSV)	p33 and p92	RT	Scholthof <i>et al.</i> , 1995b
	<i>Cucumber necrosis virus</i> (CNV)			Johnston and Rochon, 1990; Rochon and Tremaine, 1989; Panaviene <i>et al.</i> , 2004
	<i>Cymbidium ringspot virus</i> (CymRSV)			Lupo <i>et al.</i> , 1994; Kollár and Burgyán, 1994
	<i>Carnation italian ringspot virus</i> (CIRV)	p36 and p95		Pantaleo <i>et al.</i> , 2004; Rubino <i>et al.</i> , 2004
	<i>Turnip crinkle virus</i> (TCV)	p28 and p88		Carrington <i>et al.</i> , 1989; Dougherty <i>et al.</i> , 1981
<i>Machlomovirus</i>	<i>Maize chlorotic mottle virus</i> (MCMV)	p50 and p111	RT	Nutter <i>et al.</i> , 1989
<i>Necrovirus</i>	<i>Tobacco necrosis virus D</i> (TNVD)	p22 and p82	RT	Molnár <i>et al.</i> , 1997
<i>Aureusvirus</i>	<i>Pothos latent virus</i> (PoLV)	p25 and p84	RT	Rubino <i>et al.</i> , 1995
<i>Avenavirus</i> ,	<i>Oat chlorotic stunt virus</i> (OCSV)	p23 and p84	RT	Boonham <i>et al.</i> , 1995
<i>Panicovirus</i>	<i>Panicum mosaic virus</i> (PMV)	p48 and p112	RT	Turina <i>et al.</i> , 1998
<i>Carmovirus</i>	<i>Pelagornium flower bleak virus</i> (PFBV)	p27 and p86	RT	Rico and Hernandez, 2004
<i>Dianthovirus</i>	<i>Red clover necrotic mosaic virus</i> (RCNMV)	p27 and p88	FS	Kim and Lommel, 1994
<i>Luteovirus</i> Subgroup 1	<i>Barley yellow dwarf virus</i> (BYDV)	p39 and p99	FS	Miller <i>et al.</i> , 1988

Key: RT = Readthrough stop; FS = Frameshift

1.4.2 The tombusvirus replicase and replication complexes

The replication of tombusviruses is well demonstrated in *Tomato bushy stunt virus* (TBSV) (Scholthof *et al.*, 1995b; Oster *et al.*, 1998; Rajendran and Nagy, 2004; Gursinsky *et al.*, 2009). The replicase gene at the 5' end of TBSV genomic RNA, encodes the p33 and p92 replication proteins with the p92 produced via the ribosomal readthrough over the p33 termination amber codon (Scholthof *et al.*,

1995b). The C-terminus of p92 contains the RdRp signature motifs (Rochon *et al.*, 2011) indicating that the replication protein functions as the polymerase (Oster *et al.*, 1998; Gursinsky *et al.*, 2009). Unlike p92, no known functional motifs have been identified within p33 (Scholthof *et al.*, 1995b; Oster *et al.*, 1998). Mutational analysis of p33 has shown that the replication protein is necessary for TBSV replication since the substitution of the UAG stop codon at the end of p33 sequence with UAC or UAU abolishes TBSV replication (Scholthof *et al.*, 1995b).

The replication complex of TBSV consists of p33, p92 and viral RNA and several host factors (Rajendran and Nagy, 2006; Wang and Nagy, 2008; Li *et al.*, 2008; Li *et al.*, 2009). The p33 replication protein within the replication complex is produced approximately 10-20 fold in excess compared to p92 in virus infected cells and in yeast (Scholthof *et al.*, 1995b; Rajendran and Nagy, 2006). The presence of the viral RNA template within the replication complex increases the activity of the replicase, which in turn results in the accumulation of the RNA template (Rajendran and Nagy, 2006). Host factors that support the replication of TBSV include heat shock protein 70 chaperones (Hsp70, Ssa1/2p in yeast), Pex19p, glyceraldehyde-3-phosphate dehydrogenase (GAPDH, encoded by *TDH2* and *TDH3* in yeast), eukaryotic translation elongation factor 1A (eEF1A) and respective functions will be discussed in detail below (Pogany *et al.*, 2008; Pathak *et al.*, 2008; Wang and Nagy, 2008; Li *et al.*, 2009).

1.4.3 Replication of viral RNA

The tombusvirus gRNA carries regulatory elements required for RNA replication. Two short promoters, namely minus strand initiation promoter (gPR) at the 3' end of the plus strand and complementary promoter (cPR) located at the 3' end of the minus strand, are essential for minus and plus strand synthesis respectively (Figure 1:12). Additional regulatory elements within the (+ve) gRNA include the p33 recognition element (p33RE) and a recognition silencer element (RSE) (Panavas *et al.*, 2003). While p33RE is required for genomic RNA selection and recruitment into the replication complex, the RSE together with gPR and p33RE are required for the formation of the functional replicase complex both *in vivo* and *in vitro* (Panaviene *et al.*, 2005; Panaviene *et al.*, 2004; Pogany and Nagy, 2008; Pogany *et al.*, 2008). A T-

shaped structure at the 5' non-coding region of the genomic RNA has been implicated in viral RNA replication although its function is not essential (Wu and White, 1998; Wu *et al.*, 2001). Two distinct enhancer RNA elements, namely the stronger enhancer element (RIII (-) REN) and the weaker enhancer element (PPE-REN), stimulate plus RNA strand synthesis by approximately 2-20 fold *in vitro* and *in vivo* (Panavas and Nagy, 2003; Ray and white, 2003). The host factor, GAPDH binds to an AU-rich element at the 3' end of the minus RNA strand (Figure 1.12) and this interaction is involved in retaining the minus strand RNA within the replicase complex (Wang and Nagy, 2008). Overall, these three regulatory elements are responsible for the asymmetric synthesis of RNA that results in a greater number of plus strand viral progeny over the intermediate minus RNA strands (Nagy and Pogany, 2008; Wang and Nagy, 2008).

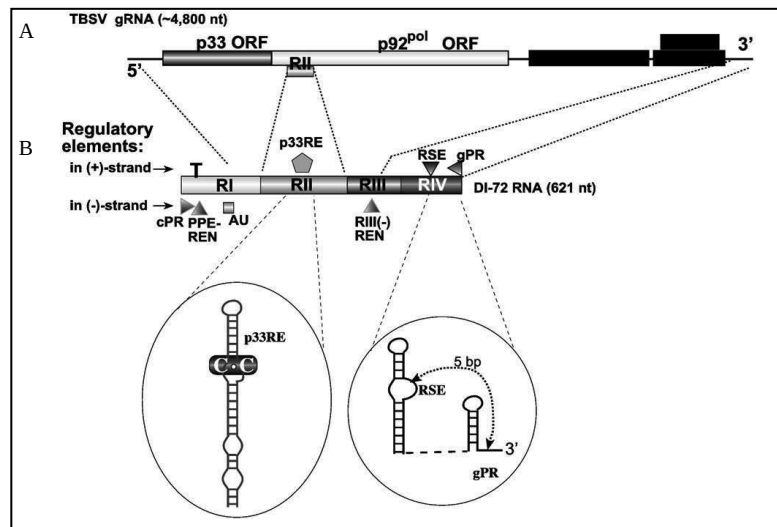


Figure 1. 12: Schematic representation of the prototypic tombusvirus genome represented by *Tomato bushy stunt virus* (TBSV). **A** TBSV RNA of approximately 4800 nts with regions from which the non-contiguous regions of DI RNA are derived. **B** Positions of the *cis*-acting elements within the DI RNA. The T-shape domain is indicated as T, p33RE represents p33 recognition element, RSE refers to replication enhancer element, gPR denotes genomic promoter and (-) strand promoter is indicated as cPR. The stronger enhancer element is denoted as RIII (-) REN while weaker enhancer element is denoted as PPE-REN. The AU rich sequence is represented as AU. Adapted from Jiang *et al.* (2010a).

TBSV has developed mechanisms to repair 5' and 3' sequences of viral RNA that have been lost during replication (Jiang *et al.*, 2010b). To compensate for the loss of

regulatory sequences, three mechanisms are utilized by TBSV including initiation of RNA synthesis from internal sequences, addition of non-viral nucleotides by TBSV replicase and RNA recombination. Firstly, internal sequences similar to cPR and PPE-REN sequences in the minus strand are used by TBSV replicase for initiation of plus strand RNA synthesis. Alternatively, the internal cPR and PPE-REN like sequences in the minus RNA strand are repositioned to the 3' proximal when the 5' end is partially deleted and result in the inhibition of RNA synthesis by TBSV replicase (Jiang *et al.*, 2010a). Secondly non-viral nucleotides are introduced at the 3' end of the minus strand or 5' end of plus strand RNA by TBSV replicase resulting in the synthesis of viable RNA (Jiang *et al.*, 2010a). Finally, recombination between partially degraded TBSV viral RNA results in the repositioning of cPR/PPE-REN like sequences to 3' terminal locations (Jiang *et al.*, 2010a).

1.4.4 Chaperone activity of tombusvirus replication proteins

Viruses belonging to the *Tombusviridae* family do not encode RNA helicases that facilitate proper folding of viral RNA during replication. Instead these viruses employ alternative means such as RNA chaperones, as a substitute for the function of the helicases (Reviewed in Russell, 2008). RNA chaperones possess unique abilities that include cooperative binding to RNA, which enhances the folding process, interactions between RNAs and accessibility of the RNA to other proteins (Russell, 2008; Zuniga *et al.*, 2009). The p33 replication protein of TBSV contains a short p33:p33/p92 interaction domain important for mediating protein-protein interaction between p33 and other p33 molecules and with p92 (Panavas *et al.*, 2005; Rajendran and Nagy, Rajendran and Nagy, 2004; 2006). The abundance of TBSV p33 (approximately 10-20 fold excess compared to p92) (Scholthof *et al.*, 1995b) within the replicase complex allows cooperative binding of p33 to RNA, which facilitates unwinding of secondary structures within the genomic RNA (Rajendran and Nagy, 2006; Stork *et al.*, 2011). When in high concentration, p33 is able to separate dsRNA or dsDNA in the absence of ATP *in vitro* and facilitate RNA cleavage by stimulating the activity of ribozymes (Stork *et al.*, 2011). In separating RNA strands, p33 facilitates accessibility of the cPR promoter and the AU rich region in minus strand thereby allowing binding of the RdRp to these sequences resulting in the eventual plus strand synthesis (Stork *et al.*, 2011). The RPR motif within TBSV p33

replication protein, containing the sequence RPRRRP is essential for RNA chaperone activity. Mutants lacking the RPR RNA binding motif don't separate dsRNA strands suggesting that p33 must interact with the RNA template in order to function as a chaperone (Stork *et al.*, 2011).

1.4.5 *Tombusviridae* replication complex-host interactions

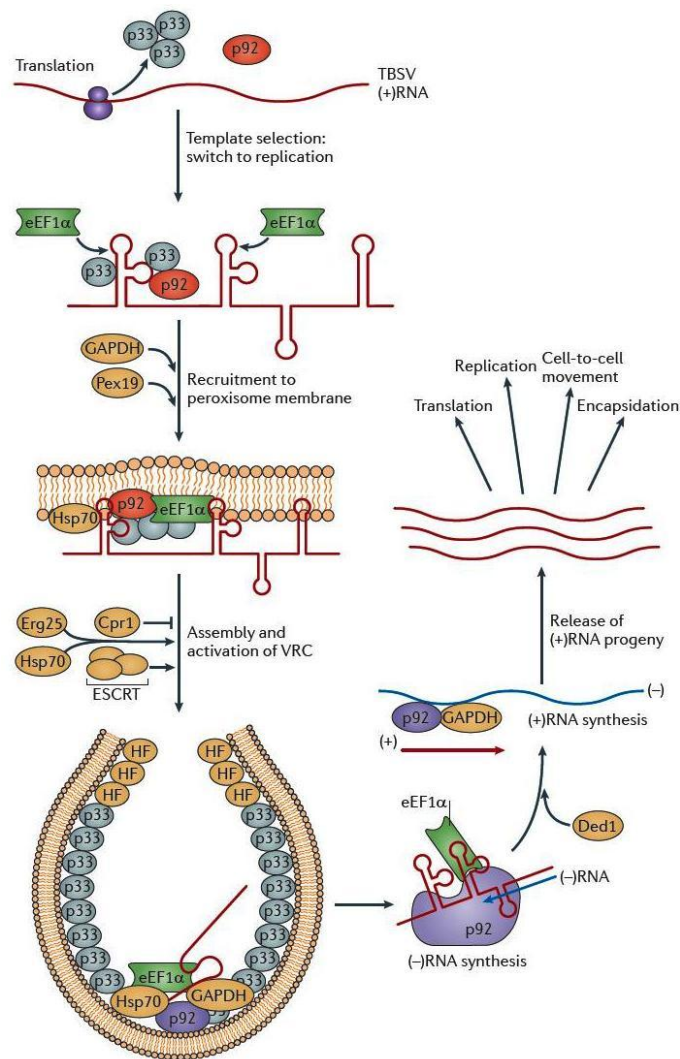


Figure 1. 13: Schematic replication cycle of TBSV. The viral proteins involved in TBSV replication are shown as grey and red for p33 and p92 respectively while host factors are denoted as green and yellow. Plus stranded RNA is shown as a red line while the blue line refers to the minus strand viral RNA. The peroxisome membrane is shown as circles. Reproduced from Nagy and Pogany, (2012).

As mentioned above, there are several host factors that are required for tombusvirus replication. Pogany *et al.* (2008) used a yeast experimental system to identify host factors required for TBSV replication. Mutant yeast strains lacking the *ssa1/2* genes were transformed with expression plasmids that encode TBSV p33 and p92. Purified Ssa1/2p from mutant yeast strains inhibited the assembly of p33, p92 and RNA template and the subsequent synthesis of TBSV RNA. This suggests that Hsp70 or Ssa1/2p (homologue of Hsp70 in yeast) ensures correct folding of the replication proteins during the assembly of TBSV replication complex. In addition, Pathak *et al.* (2008) also reported that Pex19p, a cellular transport protein, is required for recruitment of p33 replication protein to the peroxisome membranes.

The GAPDH host protein is a metabolic enzyme present in abundance within eukaryotic cells and in yeast. Tdh2 and Tdh3 are isozymes of GAPDH found in yeast. Upon TBSV infection, GAPDH is shuttled from the cytosol to the peroxisome membrane surface and thereafter is incorporated into the replicase complex. Wang and Nagy, (2008) showed inhibition of TBSV replication in *TDH2* and *TDH3* yeast mutants, suggesting the importance of GAPDH in TBSV replication. GAPDH promotes asymmetric RNA synthesis by selectively retaining minus strand RNA and creating an excess of plus strand RNA over minus strand RNA, a feature that is characteristic of viral RNA replication. Inhibition of GAPDH results in equivalent levels of plus and minus strand (Wang and Nagy, 2008).

Using yeast as a model host, Li *et al.*, (2010) showed that the translation initiation factor, eEF1A, is directly involved with the assembly of the replication complex during TBSV replication. Inhibition of eEF1A activity prior to the assembly of TBSV replication complex prevents the formation of replication complex and subsequent RNA synthesis. Inhibition of eEF1A activity after assembly does not halt the formation of the replication complex. The inhibition of the eEF1A activity prevents the recruitment of the plus strand RNA into the replication complex. This step is critical because the template forms the platform from which the viral replicase complex is assembled. Once recruitment of the viral plus RNA is inhibited, then the assembly of the viral replicase cannot occur and this results in the abolishment of TBSV replication in yeast and *in vitro* (Li *et al.*, 2010).

While most host factors have been shown to up-regulate viral RNA during replication, others down-regulate the replication process. For example, nucleolin protein has an inhibitory role in TBSV replication (Jiang *et al.*, 2010b). The overexpression of nucleolin inhibits the accumulation of TBSV RNA by approximately 10-fold after interfering with the assembly of the replication complex both *in vitro* and *in vivo*. The inhibitory effect of nucleolin is significant during the early stages rather than the late stages of TBSV replication. This observation is supported by ability of nucleolin to inhibit recruitment of template RNA by preferentially binding to the 3' UTR of the template RNA that has the replication enhancer sequence during the early stage of infection (Jiang *et al.*, 2010b). Other host proteins that inhibit the accumulation of TBSV RNA during replication include cyclophilin (Cpr1) and parvulin prolyl isomerases (Ess1) (Mendu *et al.*, 2010).

Apart from host proteins, host lipids such as sterols play a significant role in the replication of TBSV. Sterols are found abundantly in the cell and have essential roles that include regulation of membrane rigidity, fluidity and permeability. Sharma *et al.* (2010) reported that inhibition of sterol biosynthesis reduced TBSV replication. Silencing the *ERG25* gene (involved in sterol biosynthesis pathway) in yeast and the orthologous *SMO1* and *SMO2* genes in *N. benthamiana* or inactivation of the *ERG25* protein with chemical inhibitors in both yeast and plant cells showed comparable inhibitory effects on TBSV RNA accumulation. It is proposed that sterols regulate TBSV replication by affecting the lipids and proteins of the subcellular membranes supporting TBSV replication. The replication proteins are more stable in a sterol-rich environment since exposure to cytosolic proteases in sterol-poor environments results in their subsequent premature degradation (Sharma *et al.*, 2010).

1.4.6 Subcellular localization of tombusvirus replication proteins

The replication of tombusviruses occurs either on the surface of the peroxisome membranes (TBSV, CNV and CymRSV), endoplasmic reticulum as an alternative site to peroxisomes (TBSV) or mitochondria (CIRV) (Rubino *et al.*, 2001; Weber-Lotfi *et al.*, 2002; Pantaleo *et al.*, 2004; McCartney *et al.*, 2005; Jonczyk *et al.*, 2007; 2007; Pathak *et al.*, 2008). These cellular membranes are thought to facilitate the building of viral factories, promoting high concentrations of membrane-bound viral

proteins and protecting intermediate dsRNA against the RNA silencing host defensive system, cellular nucleases and proteases (Novoa *et al.*, 2005; Denison, 2008). The TBSV replication proteins p33 and p92 are anchored to the surface of the peroxisomal membrane, the site for assembly of the replication complex and viral RNA synthesis (Jonczyk *et al.*, 2007; McCartney *et al.*, 2005).

Mutational analysis of the N-terminal region of TBSV p33 revealed three peroxisome targeting sequences (K₁₁K₁₂), (K₇₆R₇₇R₇₈R₈₀), (R₁₂₄,K₁₂₉,K₁₃₀) and a peroxisome endoplasmic reticulum (pER) signal (K₅R₆), all of which are responsible for targeting the replication protein to the peroxisomes and the pER via peroxisome derived vesicles (McCartney *et al.*, 2005). In addition to the peroxisome and pER targeting signals, the N-terminal region of TBSV p33 contains two transmembrane domains (TMD1) and (TMD2) that anchor the replication protein to peroxisomes (McCartney *et al.*, 2005). In a later study, Jonczyk *et al.* (2007) showed that TBSV replicated on the endoplasmic reticulum (ER) membrane in the absence of peroxisomes and with comparable efficiency to the replication that occurred on the peroxisomal membrane surface. This suggests that the replication proteins target the ER as an alternative membrane organelle using the pER signal.

The CNV gRNA encodes p33 and p92 replication proteins that localize to the peroxisomes (Panaviene *et al.*, 2004; Panavas *et al.*, 2005). The two transmembrane domains TMD1 and TMD2 at the N-terminus of p33 are implicated in targeting the replication protein to the peroxisomes (Panavas *et al.*, 2005). Interestingly, deletion of eleven critical residues within the p33:p33/p92 interaction domain altered localization of p33 and p92 to the ER with a small portion of the mutant replication proteins accumulating to the peroxisomes (Rajendran and Nagy, 2004). Therefore the interaction domain of the CNV replication proteins is essential for localization (Panavas *et al.*, 2005).

The CymRSVp33 replication protein has been shown to target the protein to the peroxisomes using the motif (K₇₆R₇₇R₇₈) found within the N-terminus of p33 (Grieco *et al.*, 1989; Kollár and Burgyán, 1994; Navarro *et al.*, 2004). In addition, an arginine rich motif (R₂₁₃R₂₁₄R₂₁₅R₂₁₆) located downstream of the motif (K₇₆R₇₇R₇₈) has also been found to target the replication protein to the peroxisomes in the absence of the

motif (K₇₆R₇₇R₇₈), suggesting its use as an alternative peroxisomal targeting signal (Navarro *et al.*, 2004). Similar to TBSV replication proteins, the CymRSV p33 is also localized to the ER in the absence of the peroxisomes (Navarro *et al.*, 2004). Two hydrophobic transmembrane segments at the N-terminus of CymRSV p33 are responsible for anchoring the protein to the peroxisomal membrane (Navarro *et al.*, 2004).

In CIRV the p36 and p95 replication proteins localize and anchor to the mitochondrial membranes (Pantaleo *et al.*, 2004; Rubino *et al.*, 2004; Weber-Lotfi *et al.*, 2002). The mitochondrial targeting signal resides within the sequence that consists of two transmembrane domains TMD1₍₁₀₂₋₁₁₉₎ and TMD2₍₁₆₆₋₁₈₈₎ separated by a hydrophilic sequence between amino acid residues 120 - 164 (Weber-Lotfi *et al.*, 2002; Hwang *et al.*, 2008).

1.4.7 Establishment of persistent infections

Defective interfering particles (DIs), that are responsible for the establishment of persistent infections in various cell cultures, have been detected in association with (+ve) ssRNA viruses (Hillman *et al.*, 1987; Li *et al.*, 1989; Burgyán *et al.*, 1992). The DIs are generated via two mechanisms that include introduction of mutations and RNA recombination within the genome. The latter is the most frequent mechanism used to drive RNA virus evolution with more dramatic changes as it involves joining non-contiguous genomic RNA segments which accumulate *de novo* after high serial passages of the parent virus at high multiplicity of infection (Holland and Levine, 1978; Lai, 1992). The viral replicase has been implicated in the generation of DIs through the ligation of genomic RNA segments forming a non-contiguous RNA by switching between templates during complementary RNA synthesis (Lai, 1992; Nagy and Simon, 1998). The nature of DI RNAs renders them unable to replicate in the absence of the parental virus. They interfere with the replication of the parental virus by competing for limited replication resources that include the replicase and capsid proteins, resulting in suppression of parental virus accumulation, attenuation of symptoms and persistent infections (White and Morris, 1999). A typical TBSV DI RNA is composed of four non-contiguous regions, RI, RII, RIII and RIV, derived from both terminal and internal portions of the viral

genome, with regions RI, RII and RIV essential for replication of DI RNA (Chang *et al.*, 1995; Jiang *et al.* 2010a). Association with satellite viruses has also been detected with necroviruses and panicoviruses (Andriessen *et al.*, 1995; Zhang *et al.*, 1991). Just like DI RNA satellites depend on the parental viral genome for replication however no known function has been established yet (Zhang *et al.*, 1991).

1.5 THE UMBRAVIRIDAE

The family *Umbraviridae* consists of (+ve) ssRNA plant viruses with monopartite genome that lack a poly A tail and a 5' cap structure (Halk *et al.*, 1979; Ryabov, *et al.*, 2011). These viruses lack a capsid gene and therefore move from one host plant to another by being packaged by helper virus capsids (Talianky *et al.*, 2000). The complete genome sequence has been determined for *Carrot mottle mimic virus* (CMoMV), *Groundnut rosette virus* (GRV), *Pea enation mosaic virus-2* (PEMV-2) and *Tobacco bushy top virus* (TBTv) (Demler *et al.*, 1993; Gibbs *et al.*, 1996; Talianky *et al.*, 1996; Mo *et al.*, 2003).

The genome of umbraviruses is organised into four overlapping ORFs (Figure 1.14). ORF1 encodes a putative 31 – 37 kDa accessory protein while the ORF2 encodes a putative 63 – 65 kDa protein (RdRp) that lacks an AUG initiation codon at the 5' end (Ryabov, *et al.* 2011). However, a seven nucleotide sequence immediately before the stop codon at the 3' end of ORF1 is associated with a -1 ribosome frameshift expression strategy that results in the translation of ORF2 protein, the putative replicase of 94 – 98 kDa (Ryabov *et al.*, 2011).

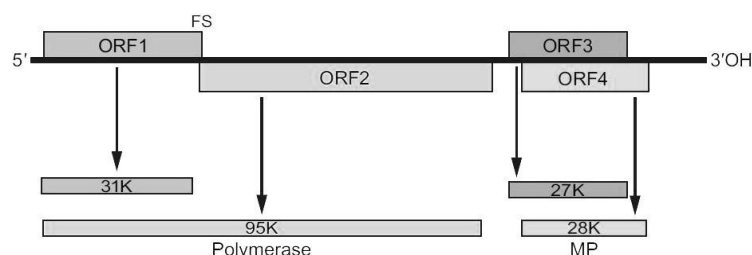


Figure 1. 14: Genome organization of groundnut rosette virus. Molecular masses are in kDa. FS refers to frameshift. MP refers to movement protein. Reproduced from Ryabov *et al.* (2011).

The predicted amino acid sequence of the ORF2 translation product contains signature motifs similar to tombusviral RdRps including all eight conserved motifs of the RdRp of (+ve) ss RNA viruses (Koonin and Dolja, 1993; Ryabov *et al* 2011). The accessory protein (ORF1) and replicase (ORF2) have not been studied in detail and therefore limited data exists about their functions. The translation products ORF3 and ORF4, which are almost completely overlapping in different reading frames, have been better studied and each ORF yields a 26 - 27 kDa product which is translated from a subgenomic RNA (Ryabov *et al.*, 2011). The ORF3 protein is important for RNA chaperone activities, protection of viral RNA and mediating long distance movement through the phloem (Ryabov *et al.*, 1999; Ryabov *et al.*, 2001; Taliansky *et al.*, 2003). The ORF4 movement protein is involved in cell to cell movement of the virus (Ryabov *et al.*, 1998). Umbraviruses lack the ability to produce conventional virus particles in infected plants (Taliansky and Robinson, 2003; Taliansky *et al.*, 2003, Ryabov *et al* 2011). This has been the reason for the focus of studies on umbravirus ORF3 using GRV as the model virus to elucidate the function of this protein.

Evidence that the ORF3 translation product is involved in facilitating the long distance movement of the virus was provided from gene replacement studies that generate hybrids between TMV and GRV. The gene that encodes the TMV capsid was replaced by GRV *ORF3* and the chimeric TMV RNA was used to infect *N. benthamiana*. The outcome of this experiment showed that GRV RNA requires ORF3 protein in order to spread systemically and mediate long distance movement through the phloem (Ryabov *et al.*, 1999). Later work done by Taliansky *et al.* (2003) revealed that GRV ORF3 protein and viral RNA were distributed throughout the cells of the infected *N. benthamiana* and therefore confirming the protein's role in mediating long distance movement.

Another important function of GRV ORF3 protein is its role as an RNA chaperone. Analysis of the properties of recombinant GRV ORF3 protein expressed from TMV expression vector in *N. benthamiana* showed that it bound to RNA cooperatively (Taliansky *et al.*, 2003) and also revealed a strong tendency to aggregate and form stable dimers, trimers and higher order oligomers (Taliansky *et al.*, 2003).

The role of GRV ORF3 protein in protecting viral RNA was investigated by studying the intracellular localization of GRV ORF3 associated cytoplasmic inclusions (Ryabov *et al.*, 1998). The GRV ORF3 cytoplasmic inclusions associate with TMV specific amorphous X bodies or viroplasms to form complexes that vary in size from 300 nm to 4 µm (Taliensky *et al.*, 2003; Kim *et al.*, 2007). The cytoplasmic inclusions within the X-bodies contain filamentous structures that consist of ORF3 protein and viral RNA, in complexes called ribonucleoprotein (RNP) (Taliensky *et al.*, 2003; Kim *et al.*, 2007) that are proposed to protect viral RNA against plant's defensive RNA silencing system. The RNPs are used by umbraviruses as alternatives to classic capsid proteins to shuttle viral RNA via long distance movement through the phloem and to establish systemic viral infection (Nelson and van Bel, 1998; Taliensky *et al.*, 2003; Kim *et al.*, 2007).

The participation of host factors has been implicated in the intracellular localization of GRV ORF3 protein. Fibrillarin, a major nucleolus protein interacts with GRV ORF3 protein via the N-terminal glycine and arginine-rich domain of fibrillarin and localizes GRV ORF3 protein to the nucleolus (Kim *et al.*, 2007). In addition the interaction between fibrillarin and GRV ORF3 is necessary to mediate the formation of an RNP complex with GRV RNA (Kim *et al.*, 2007). Plants inoculated with the RNP complex developed symptoms similar to plants infected with wild-type and this led to the conclusion that the RNPs are infectious. The RNP is also resistant to RNase and which further suggests the role of RNPs in protection of viral RNA (Kim *et al.*, 2007).

The intracellular localization of GRV ORF3 protein has been investigated (Ryabov *et al.*, 1998; Kim *et al.*, 2007). The expression of GRV ORF3 as a fusion protein of the green fluorescent protein (GFP) from TMV expression vectors accumulated in the cytoplasmic inclusion bodies and nucleoli in *N. benthamiana* (Ryabov *et al.*, 1998). Analysis of the predicted amino acid sequence of GRV ORF3 revealed an arginine-rich (Rx₂Rx₃Rx₆R) and leucine-rich domain (LxxLL) that is conserved among umbraviral ORF3 sequences (Taliensky and Robinson, 2003; Ryabov *et al.*, 2004). Localization studies of GRV ORF3-GFP fusion protein revealed that the arginine-rich nuclear localizing signal (NLS) and in addition to the L₁₄₉ of the leucine-rich domain are essential for targeting GRV ORF3 protein to the nucleolus (Ryabov *et*

al., 2004; Kim *et al.*, 2007). Prior to targeting the nucleolus, the GRV ORF3 protein causes the reorganization of cajal bodies into multiple cajal body like structures that fuse with the nucleolus thereby allowing entry of GRV ORF3 into the nucleolus. The leucine-rich nuclear export signal (NES) is responsible for localizing the ORF3 protein to large cytoplasmic inclusions (Ryabov *et al.*, 2004; Kim *et al.*, 2007).

Even with the lack of a capsid protein, umbraviruses have been found to replicate, accumulate and spread efficiently within infected plants with the employment of two alternative mechanisms (Murant *et al.*, 1969; Reddy and Ansa, 1985). Firstly is their dependence on a helper virus, usually from the *Luteoviridae* family, for transmission between plants, using aphid vectors. The capsids of luteovirus package umbraviral RNA forming hybrid virus particles that are transmitted by aphids or mechanical inoculation (Talianky *et al.*, 2000). Once within the infected cells, protection of the vRNA from nuclease attack and the RNA silencing machinery is afforded by cytoplasmic RNP particles that serve to encapsidate the viral RNA (Talianky *et al.*, 2003; Kim *et al.*, 2007).

1.6 RESEARCH MOTIVATION

The discovery of the tetravirus, PrV, and its unique ability to replicate in tissue culture cells, provided an opportunity to study tetravirus biology *in vivo*. Despite similarities in terms of the capsid structure and host range, PrV is different from other tetraviruses in that it carries a carmo-like replicase with a functional readthrough stop codon as well as a non-structural protein, p130, whose function is unknown (Walter *et al.*, 2010). This raises important research questions relating to the expression strategies and functions of the PrV non-structural replication proteins in replication and whether PrV is more similar to the tombusviruses and umbraviruses than to other tetraviruses. Thus the focus of the research described in this thesis was on the expression and function of the replication proteins of PrV in the life cycle of PrV.

1.7 OBJECTIVES

The following objectives were developed to answer questions relating to the function of PrV replication proteins in replication. These include;

1. To develop experimental systems for studying PrV replication
2. To investigate the subcellular localization of PrV replication proteins
3. To characterize potential protein-protein interaction between PrV replication proteins
4. To analyse PrV non-structural protein sequences for potential functional domains using bioinformatic tools.

CHAPTER 2

PRODUCTION OF ANTISERUM FOR THE DETECTION OF PrV REPLICASE

2.1 INTRODUCTION

The overall objective of this research project was to study the expression and subcellular localization of the PrV replication proteins, p40 and p104. The proteins are expressed from the same ORF via the activity of a Type I readthrough stop codon. In order to establish the function of p40 and p104 in replication, it is important to generate antibodies against these proteins. In previous studies, expression of the replication proteins was detected using a polyclonal antibody raised against the C-terminus of p40 (amino acids 248 – 331), which cross-reacts with both p40 and p104 (Walter *et al.*, 2010). This antiserum was also found to be suitable for immunofluorescence studies (Walter, 2008). However a drawback of this antibody was that it is unable to distinguish between p40 and p104. It was decided to generate an antibody to detect p104 alone. Therefore the objective of the research described in this chapter was to produce a specific antibody directed against the C-terminus of p104.

2.2 MATERIALS AND METHODS

2.2.1 Bacterial strains

Escherichia coli strains DH5 α was used for plasmid propagation while DH10BacTM (Hanahan, 1983; Invitrogen Life Technologies) was used to construct the recombinant bacmid by homologous recombination. The *E. coli* strain BL21 (DE3) was used for expression of recombinant protein using T7 promoter (Studier and Moffatt, 1986). The *E. coli* strains DH5 α and BL21 (DE3) were grown in Lysogeny Broth (LB) and selected using ampicillin at a final concentration of 100 μ g/ml. The DH10BacTM were grown in super optimal broth with catabolite repression (SOC) and selected using 50 μ g/ml kanamycin, 7 μ g/ml gentamycin, 10 μ g/ml tetracycline, 300

ug/ml 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside (X-Gal) and 50 ug/ml isopropyl β -D-1-thiogalactopyranoside (IPTG).

2.2.2 Construction and expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in *E. coli*

The region corresponding to the C-terminal region between amino acid residues 437 and 915 of p104 in plasmid, pRN3, was fused to a 6xHis tag using restriction analysis and the sequence named 6xHis-p104₍₄₃₇₋₉₁₅₎. The presence of the 6xHis tag was to facilitate nickel-affinity purification of the truncated p104₍₄₃₇₋₉₁₅₎ peptide. The amino acid residues 437 - 915 of p104 are located seventy six amino acid residues down stream of the readthrough stop codon and p40. This region was chosen because it contains potential antigenic sites suitable for generation of antiserum against the C-terminal region of p104.

The protocol used for construction of a plasmid expressing 6xHis-p104₍₄₃₇₋₉₁₅₎ involved digestion of the plasmid, pRN3 with *Bam* HI and *Age* I and ligated to the 32 base pair double-stranded linker, (RN6F 5'-GAT CCA TGG AGC ATC ACC ATC ACC ATC ACT TA-3') and RN6F 5'-CCG GTA AGT GAT GGT GAT GGT GAT GCT CCA TG-3') creating pRN11. Next, pRN11 was digested with *Bam* HI and *Pvu* I and the fragment was subcloned into a *Bam* HI/*Pvu* I digested pT7-7 vector to yield the final expression vector, pRN12 carrying the sequence 5'-GAA GGA GAT ATA CAT ATG GCT AGA ATT CGC GCC CGG GGA TCC ATG GAG CAT CAC CAT CAC CAT CAC TTA CCG GTA ACC CCG -3' that contains the RNA binding site (RBS), *Bam* HI, start codon (AUG), 6xHis residues and *Age* I site respectively. This construct encoded 6xHis-p104₍₄₃₇₋₉₁₅₎ coding sequence downstream of the T7 promoter (Figure 2.1). The integrity of the 5' end and the 6xHis fusion was confirmed by Sanger DNA sequencing.

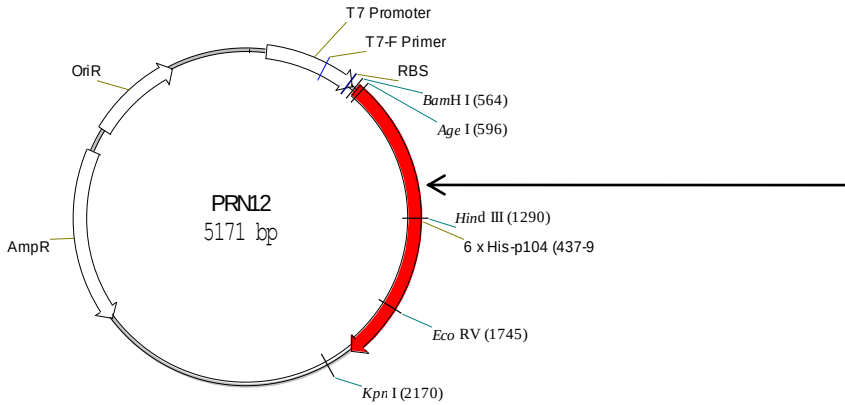


Figure 2. 1: Schematic representation of the construction of pRN12. Expression plasmid pRN12 with fusion sequence above showing RBS, start codon, 6xHis tag, restriction enzymes and first 20 amino acids of p104₍₄₃₇₋₉₁₅₎. *Bam* HI and *Age* I sites used for restriction analysis are indicated on plasmid map. The arrow indicates where the sequence was inserted. Plasmid features include origin of replication (*oriR*), β -lactamase gene (AmpR), T7 bacteriophage RNA polymerase promoter (T7 promoter).

For protein expression of 6xHis-p104₍₄₃₇₋₉₁₅₎, 5 ml log phase overnight culture from a single colony harbouring pRN12 was inoculated into 100 ml of LB with ampicillin at a final concentration of 100 μ g/ml and agitated at 250 rpm. The culture was allowed to grow for 4 hours at 37 °C until an OD_{600nm} of 0.6 was attained. Next, expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ was induced using IPTG to a final concentration of 1 mM and further incubated for 4 hours at 37 °C. To monitor protein expression, 1 ml of culture was collected prior to induction and post-induction at a time interval of 1 hour. The cells were harvested by centrifugation at 6000 x g for 10 min and pellets re-suspended in lysis buffer (100 mM NaCl, 100 mM Tris-HCl pH 7.6) and sonicated using a hand-held 3 mm probe (Vibra cell™ sonicator) at 400 mHz for 10 second intervals on ice until viscosity disappeared. The recombinant protein was examined using SDS-PAGE and visualized by Coomassie staining. The identity of the protein was confirmed with western analysis using an anti-His mouse monoclonal antibody (anti-6xHis antibodies) (Roche).

Since the expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in BL21 (DE3) resulted in no detectable product of approximately 55 kDa protein by SDS-PAGE and western analysis, the plasmid pRN12 was transformed into BL21 (DE3) containing the plasmid, pRIG that encodes tRNA genes *argU*, *ileX*, and *glyT* (Novagen). Co-expression with pRIG was carried out to improve expression of 6xHis-p104₍₄₃₇₋₉₁₅₎. The procedure for protein expression carried out is similar to that mentioned above.

2.2.3 Baculovirus-mediated expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in *Spodoptera frugiperda* (Sf9) cells

Construction of recombinant baculovirus expressing 6xHis-p104₍₄₃₇₋₉₁₅₎

The plasmid pRN11, containing 6xHis-p104₍₄₃₇₋₉₁₅₎ coding sequence was digested with *Bam* HI and *Sac* I and subcloned into the *Bam* HI/*Sac* I digested pFastBacTM Dual vector (Invitrogen), downstream of the polyhedron promoter, *polh*, to generate pRN18 (Figure 2.2).

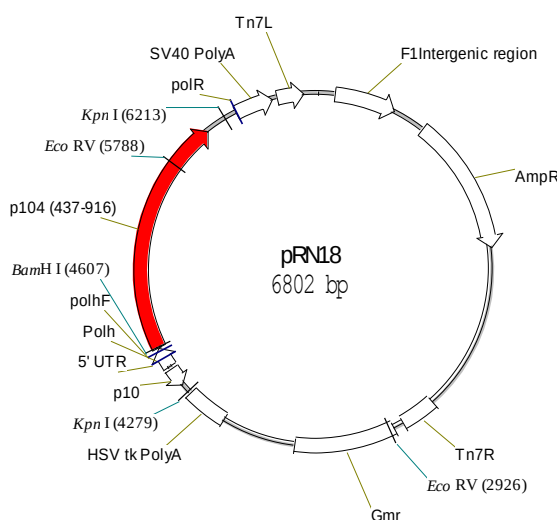


Figure 2. 2: Graphic representation of the restriction map of pRN18. The plasmid, pRN18 contains the coding sequence p104₍₄₃₇₋₉₁₅₎ fused to the 6xHis tag at the N-terminus. Other features of the plasmid include, β -lactamase gene (AmpR), gentamycin resistance gene (Gmr), Herpes Simplex Virus type 1 thymidine kinase (HSVtk polyA), Promoters p10 and polh, Simian Virus 40 polyadenylation signal (SV40 polyA), mini-Tn7 left element (Tn7L), 5' untranslated sequence (UTR) and mini-Tn7 right element (Tn7R).

The pFastBacTM Dual vector has been designed in that the 5'untranslated region within the polyhedrin promoter of the pFastBacTM Dual vector drives expression of

the recombinant protein (Invitrogen). For proper expression of recombinant protein, the gene to be cloned into pFastBacTM Dual vector must contain a translation initiation start codon and stop codon (Invitrogen). The sequence of 6xHis-p104₍₄₃₇₋₉₁₅₎ in pRN18 was verified with sequence primers polhF (5'- ATT CAT ACC GTC CCA CCA TCG GG -3') and polR (5'- ACC TCT ACA AAT GTG GTA TGG CTG AT-3'). Next, pRN18 was transformed into competent DH10BacTM and cells were cultured on LB agar plates supplemented with 50 μ g/ml kanamycin, 7 μ g/ml gentamycin, 10 μ g/ml tetracycline, 300 μ g/ml X-Gal and 50 μ g/ml IPTG to select for recombinant bacmids.

Candidate white colonies were selected and recombinant bacmid DNA was isolated according to the manufacturer's instructions (Bac-to-Bac[®] manual, Invitrogen). PCR analysis was used to confirm the presence of the pRN18 insert in recombinant bacmids, using pUCF/M13F (5'-CCC AGT CAC GAC GTT GTA AAA CG-3') and pUCR/M13R (5'-AGC GGA TAA CAA TTT CAC ACA GG-3') primers which bind to DNA sequence flanking the *lacZ α* complementation region.

To generate a recombinant baculovirus, a 6-well tissue culture plate was seeded with 8×10^5 *S. frugiperda* Sf9 cells per well in 2 ml of Grace's insect serum-free cell culture medium (Sigma Aldrich) supplemented with penicillin (50 μ g/ml) and streptomycin (50 μ g/ml) (Gibco). The cells were allowed to settle for 2 hrs at 28 °C after which they were transfected with 800 ng of recombinant bacmid DNA using FUGENE[®] HD transfection reagent (Roche) as described previously (Short *et al.*, 2010). The cells were incubated for 72 hrs at 28 °C. The recombinant baculovirus (referred to as BVp55 hereafter) was harvested as passage 1 (P1) virus stock from the transfected cell supernatant after 72 hrs.

Next the recombinant baculovirus was purified by plaque assay. Six-well tissue culture plate was seeded with 8.5×10^5 Sf9 cells in 2 ml of Grace's insect complete medium and incubated at 28 °C for 1 hr to allow cell attachment. The cells were then infected with a 10-fold serial dilution of BVp55 (P1) (10^{-2} to 10^{-8}) and incubated at 28 °C with gentle agitation. After 1 hr, the medium was removed and replaced with 2 ml of molten agarose in Grace's insect complete medium at a ratio of 1:1; plates were incubated for 6 days at 28 °C. To visualize plaques, the cells were stained with

0.1 % thiazolyl blue tetrazolium bromide in phosphate buffered saline (PBS, pH 7.4). A total of six single plaques were isolated and re-suspended in 0.5 ml of Grace's insect complete medium. The viral suspension from each plaque was used to infect Sf9 cells as previously described to produce passage 2 (P2).

To verify the expression of recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎, Sf9 cells infected with BVp55 (P2) were harvested 72 hrs post-infection by centrifugation at 1000 x *g* for 5 minutes. The supernatant containing BVp55 (P2) was stored at 4 °C and subsequently used to produce P3 stocks. The cell pellet was washed three times in PBS (pH 7.4) and re-suspended in 100 μ l PBS (pH 7.4) and resolved by SDS-PAGE followed by western analysis with anti-6xHis antibody using Roche chemiluminescent western blotting kit according to the manufacturer's instructions. Two BVp55 clones that expressed 6xHis-p104₍₄₃₇₋₉₁₅₎ efficiently were selected and used to make P3 and P4 virus stocks. The titre of P4 was determined by plaque assay as described previously. The concentration of infectious virus particles (plaque forming units (pfu/ml)) was determined using the plaque assay method described above and calculated using the following formula provided by the manufacturer:

$$\text{pfu/ml} = \frac{\text{Number of plaques formed (pfu)}}{\text{Dilution factor} \times \text{ml of inoculum}}$$

Baculovirus expression of 6xHis-p104₍₄₃₇₋₉₁₅₎

Expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ from the P4 stock was optimised with respect to multiplicity of infection (MOI) and the duration of expression. The optimal MOI was determined by seeding a six-well tissue culture plate with 1×10^6 cells in 2 ml of Grace's medium followed by incubation at 28 °C for 1 hour. The cells were then infected with recombinant virus at MOIs of 0.1, 1, 2, 5 and 10. The optimal time course expression analysis was achieved by infecting a 6-well tissue culture plate seeded with 1×10^6 Sf9 cells at an MOI of 0.1. Recombinant baculovirus infected cells were harvested after different incubation times (24, 48, 72, 96, 120 and 144 hrs) post-infection. Expression of the recombinant protein was analyzed using SDS-PAGE and Coomassie staining to visualise resolved proteins. The identity of the protein was confirmed by western analysis using anti-6xHis antibodies.

Solubility studies

To determine whether p104 was soluble, a T75 tissue culture flask was seeded with 1.6×10^7 Sf9 cells, infected with baculovirus at MOI of 0.1, and then incubated for 96 hrs. The cells were collected by centrifugation at $1000 \times g$ for 2 min, washed three times in PBS (pH 7.4) and then disrupted by passing through a 25 gauge needle 40 times in 0.8 ml of disruption buffer (8 % sucrose, 1 mM EDTA, 50 mM Tris-HCl pH 7.5, protease inhibitor cocktail tablet (Roche)). Soluble protein was separated from insoluble fraction by centrifugation at $17,000 \times g$ for 20 min at 4 °C. In order to extract more protein, the pellet fraction was re-suspended in 1/2 the volume of the original disruption buffer, incubated on ice for 30 min, passed through a 25 gauge needle 40 times and re-centrifuged at $17,000 \times g$ for 20 min. The supernatant from both fractions was pooled and an equal volume of disruption buffer was added to the pellet. All fractions were resolved by SDS-PAGE and subjected to western analysis to determine the solubility of 6xHis-p104₍₄₃₇₋₉₁₅₎ protein. To solubilize 6xHis-p104₍₄₃₇₋₉₁₅₎, baculovirus-infected cells were harvested and washed as described above, re-suspended in PBS (pH 7.4) supplemented with different solubilizing agents including 8 M urea, 100 mM Na₂CO₃ pH 11.0, 2 M NaCl, 1 % Triton, 0.1 % Polyethylenimine, 7.5 % Sarkosyl or 1 % NP40 followed by incubation on ice for 15 minutes. The cells were then disrupted as described before.

2.3 RESULTS

2.3.1 Expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in *E. coli*

The aim of the research described in this chapter was to produce an antibody that would be able to distinguish between p40 and p104, corresponding to amino acids 437 – 915, downstream of the readthrough stop codon, which encode a peptide with a predicted molecular weight of approximately 55 kDa. The first strategy was to construct an *E. coli* vector expressing p104₍₄₃₇₋₉₁₅₎ fused to a 6xHis tag at the N-terminus. However, repeated attempts to produce this peptide in *E. coli* were unsuccessful, with no detectable expression of the 55 kDa peptide by Coomassie staining or western blot analysis with anti-6xHis antibody (data not shown). The lack of expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ was attributed to the presence of a tandem arginine repeat (AGG AGG AGG) located between nucleotides (nts) 2455 – 2463 within the p104₄₃₇₋₉₁₅ coding sequence. Tandem arginine (AGG) repeats have been

shown to inhibit protein expression in *E. coli* by mimicking the Shine-Dalgarno sequence and resulting in the production of truncated peptides; these in turn are degraded by the SsrA peptide tagging system (Quaggio *et al.*, 1993; Chen and Inouye, 1994). Removal of the three arginine AGG codons by site directed mutagenesis did not improve expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ (data not shown).

It was possible that the codon usage may have been a factor in the lack of detectable expression. The coding sequence of p104 contains high levels of arginine (AGA, AGG and CGA), and isoleucine (ATA) codons whose corresponding tRNAs are found only at low levels in *E. coli*. However, co-transformation with the pRIG plasmid (Novagen), that expresses rare tRNAs for arginine, isoleucine and glycine did not improve expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ (data not shown). Therefore these codons were not the cause for the lack of expression of 6xHis-p104₍₄₃₇₋₉₁₅₎. In view of the lack of expression in *E. coli* cells, it was decided to express the antigen in insect cells using recombinant baculovirus expression system.

2.3.2 Baculovirus-mediated expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in Sf9 cells

Generation of recombinant bacmid expressing 6xHis-p104₍₄₃₇₋₉₁₅₎

The Bac-to-Bac Baculovirus expression system is based upon the ability to generate recombinant baculovirus for expression of recombinant protein in insect cells (Invitrogen Life Technologies). The gene of interest is placed downstream of the AcMNPV polyhedrin promoter, polh or p10 in the pFastBacTM donor plasmid and the recombinant plasmid is transformed into DH10BacTM. The pFastBacTM donor plasmid carries a Tn7 element containing gentamycin resistance gene, the polyhedron promoter, multiple cloning site and SV40 polyadenylation signal. The DH10BacTM carries a baculovirus shuttle vector (bacmid) that contains a mini-F replicon, kanamycin resistance gene, an attachment Tn7 (attTn7) site and the *lacZα* complementation element. The helper plasmid provides proteins required to transpose the gene of interest in pFastBacTM to the bacmid. Site specific transposition of the gene of interest in pFastBacTM occurs via the Tn7L and Tn7R transposition elements in pFastBacTM and the attachment Tn7 sites of the bacmid. The recombinant bacmid is isolated and used to transfect insect cell culture, resulting in the production of infectious recombinant baculoviruses.

The pFastBacTM-Dual expression cassette carrying *6xHis-p104*₍₄₃₇₋₉₁₅₎ cDNA (pRN18) and downstream of *polh*, was transposed into bacmid DNA in DH10BacTM and recombinant bacmids selected by blue-white screening on LB agar plates supplemented with 300 μ g/ml X-gal, 50 μ g/ml IPTG, 50 μ g/ml Kanamycin, 7 μ g/ml gentamycin and 10 μ g/ml tetracycline. The presence of the *6xHis-p104*₍₄₃₇₋₉₁₅₎ cDNA within the *lacZ α* complementation region was verified by PCR using primers pUC/M13F and pUC/M13R, which flank the expression cassette within the bacmid. An amplification product corresponding to the predicted size of 4145 bp of the *6xHis-p104*₍₄₃₇₋₉₁₅₎ gene sequence was detected in recombinant bacmids (Figure 2.3C, Lane 3). A PCR product of 2560 bp that corresponds to the region flanked with Tn7L and Tn7R transposition elements in the parental vector without the p104 coding sequence (Figure 2.3C, Lane 2) was detected in a recombinant bacmid generated with pFastBacTM-Dual. A 273 bp PCR product that is consistent with the *lacZ α* complementation region was detected in the native bacmid DNA (Figure 2.3C, Lane 1). This suggests that the expression cassette carrying *6xHis-p104*₍₄₃₇₋₉₁₅₎ cDNA was transposed into bacmid DNA.

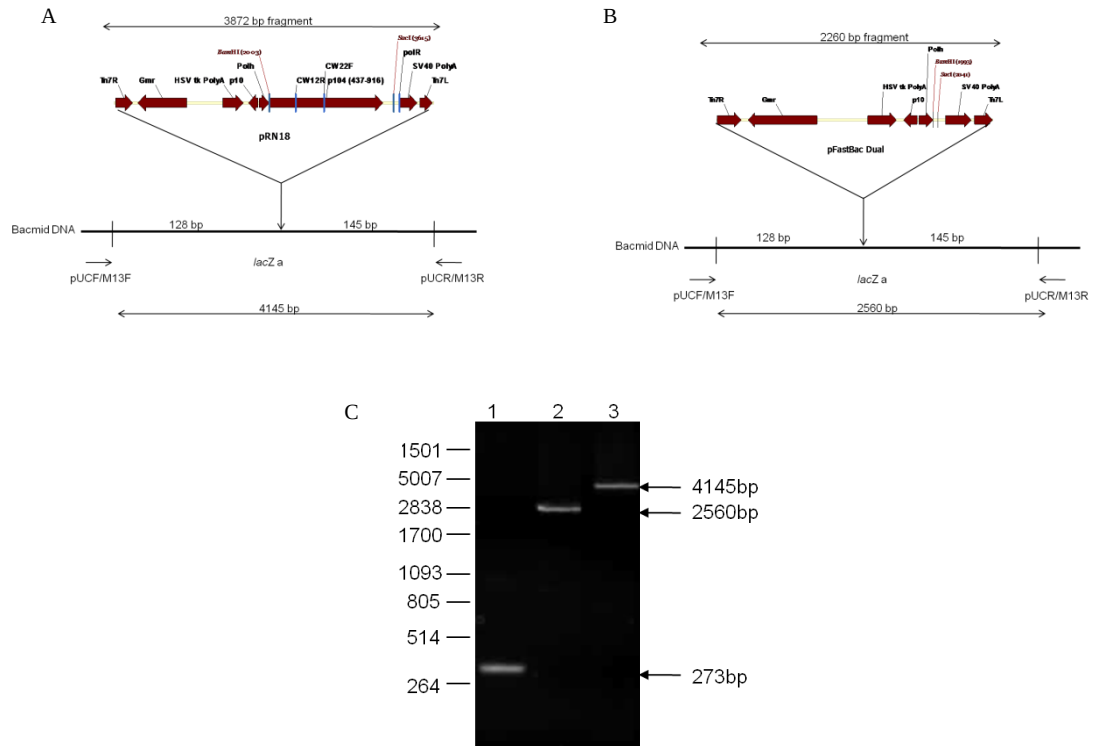


Figure 2. 3: PCR verification of transposition of 6xHis-p104₍₄₃₇₋₉₁₅₎ expression cassette into Bacmid DNA. Panel A and B: Schematic representation of the transposition event resulting in the recombinant bacmid containing p104₍₄₃₇₋₉₁₅₎ and bacmid DNA without p104₍₄₃₇₋₉₁₅₎ coding sequence. Panel C: Analysis of PCR products. Lane 1: native bacmid. Lane 2: bacmid transposed with pFastBac™ Dual vector. Lane 3: Bacmid DNA transposed with 6xHis-p104₍₄₃₇₋₉₁₅₎. The selected bacmid containing the 6xHis-p104₍₄₃₇₋₉₁₅₎ expression cassette was used to generate recombinant baculoviruses that were screened for protein expression. The primers pUC/M13F and pUC/M13R used flank the *lacZα* and produce 4145 bp 6xHis-p104₍₄₃₇₋₉₁₅₎ expression cassette within the *lacZα* gene while 2560 bp contain expression cassette without p104₍₄₃₇₋₉₁₅₎ coding sequence. The relative molecular weights are indicated on the left hand side.

Expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in Sf9 cells

Recombinant baculovirus stock P1, resulting from transfection of the recombinant bacmid into *S. fugiperda* Sf9 cells, was used to infect fresh Sf9 cells. Single plaques were purified and six isolates selected for screening. These were used to infect fresh cells that were screened for the expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ by analysing cell-free extracts by SDS-PAGE followed by western analysis using anti-6xHis antibody. A protein of approximately 55 kDa was detected in cells infected with all six selected recombinant baculoviruses (Figure 2.4A, Lanes 3 – 8). This band was not present in

uninfected Sf9 cells or in Sf9 cells infected with wild type (wt) baculovirus (Figure 2.4A, Lanes 1 - 2). Western analysis confirmed that the 55 kDa protein band corresponded to 6xHis-p104₍₄₃₇₋₉₁₅₎ (Figure 2.4B, Lane 3 – 8). The recombinant virus (2) (referred to as BVp55 (P2-2)) was selected for further virus amplification.

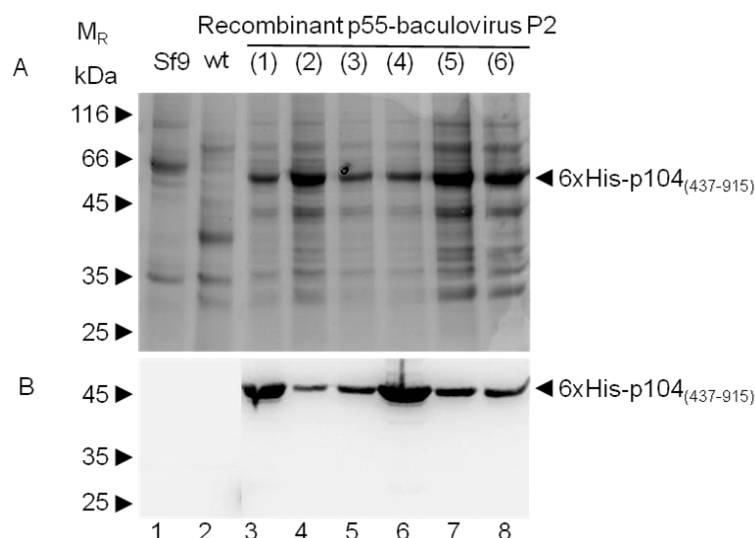


Figure 2. 4: Expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in Sf9 cells from six recombinant baculovirus. Panel A: Coomassie-stained 12 % SDS-polyacrylamide gel used to resolve cell-free protein extracts. Panel B: Western analysis of protein extracts from recombinant baculovirus-infected cells using anti-6xHis antibody to detect the 6xHis tag. Lane 1: uninfected Sf9 cells: Lane 2: bacmid: Lanes 3 - 8: Isolated recombinant baculoviruses expressing 6xHis-p104₍₄₃₇₋₉₁₅₎. Relative molecular weights are indicated on the left in kDa. The 6xHis-p104₍₄₃₇₋₉₁₅₎ is indicated on the right (arrow).

Optimization of 6xHis-p104₍₄₃₇₋₉₁₅₎ expression in Sf9 insect cells

The optimum expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ was determined by infecting Sf9 cells with BVp55 (P4) at MOIs varying between 0.1 and 10. SDS-PAGE analysis of crude extracts revealed that maximum protein expression was achieved when cells were infected at a MOI of 0.1 (Figure 2.5, Panel A, Lane 1) with protein expression decreasing as the MOI increased (Figure 2.5, Panel A, Lane 2 - 5). No protein corresponding to p55 was observed in uninfected Sf9 cells or cells infected with wild-type baculovirus (Figure 2.5, Panel A, Lane 6 - 7).

The optimal time for expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ was determined over a period of six days. Protein expression was detectable on Coomassie-stained SDS-PAGE after 72 hrs and increased up to 144 hrs post-infection with the highest levels at 144 hrs (Figure 2.5, Panel B, Lane 3 - 6). It was decided to harvest the cells after 96 hrs to avoid the formation of insoluble protein that results from protein aggregate formation or loss of translated recombinant protein after cell lysis.

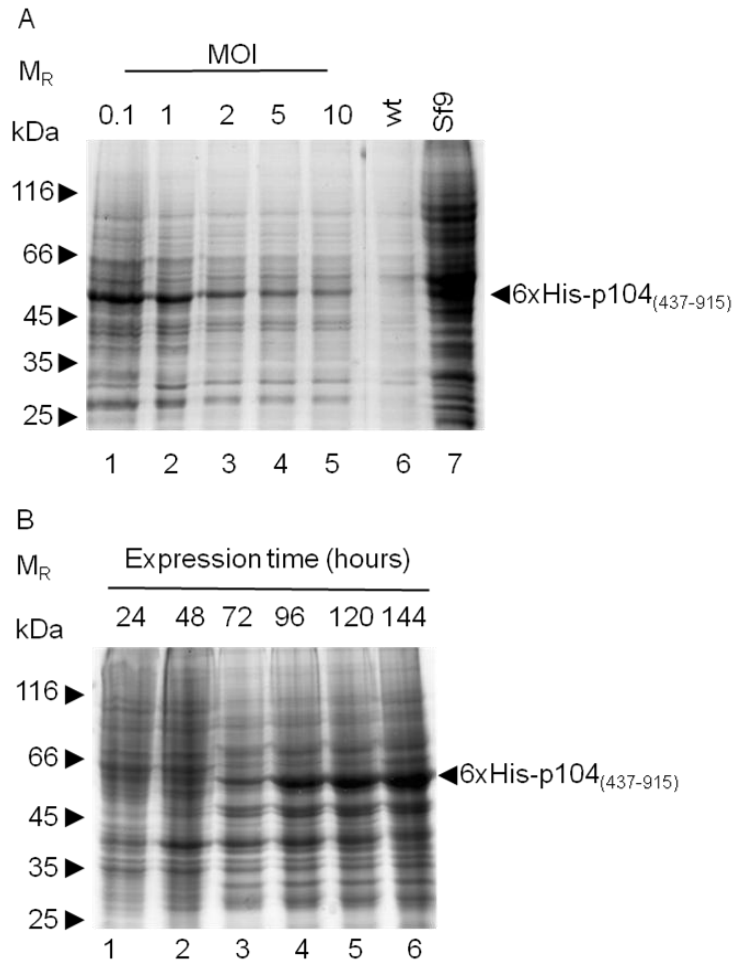


Figure 2. 5: Optimization of 6xHis-p104₍₄₃₇₋₉₁₅₎ expression in Sf9 cells. Panel A: SDS-PAGE analysis of 6xHis-p104₍₄₃₇₋₉₁₅₎ expression on 12 % Coomassie-stained gel. The Sf9 cells were infected at varying MOIs of 0.1, 1, 2, 5 and 10 with BVp55 (P4) and harvested 96 h.p.i. Lane 1 - 5: Infected Sf9 crude extracts at MOI 0.1, 1, 2, 5 and 10 respectively: Lane 6: Sf9 crude extracts of wt baculovirus at MOI of 0.1. Lane 7: Crude extracts from Sf9 cells. Panel B: SDS-PAGE analysis of recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎ expression on a 12 % Coomassie-stained gel. Lane 1 - 6: Sf9 crude extracts harvested at 24, 48, 72, 96, 120 and 144 hrs post infection and at an MOI of 0.1. Recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎ is shown on the right (arrow). Relative molecular weights are indicated on the left.

As high levels of protein expression are often burdened with problems of protein solubility and the HaSV His-REP-V5 had been found to be insoluble (Short and Dorrington, 2012), the solubility of 6xHis-p104₍₄₃₇₋₉₁₅₎ was examined. Cells were lysed and soluble protein was separated from the insoluble fraction by centrifugation. The fraction was analysed by SDS-PAGE and western blot analysis using anti-6xHis antibody and a 55 kDa protein corresponding to 6xHis-p104₍₄₃₇₋₉₁₅₎ was present predominantly in the pellet fraction of these cell lysates as opposed to the soluble fraction (Figure 2.6, Panel A and B, Lane 2 vs Lane 3).

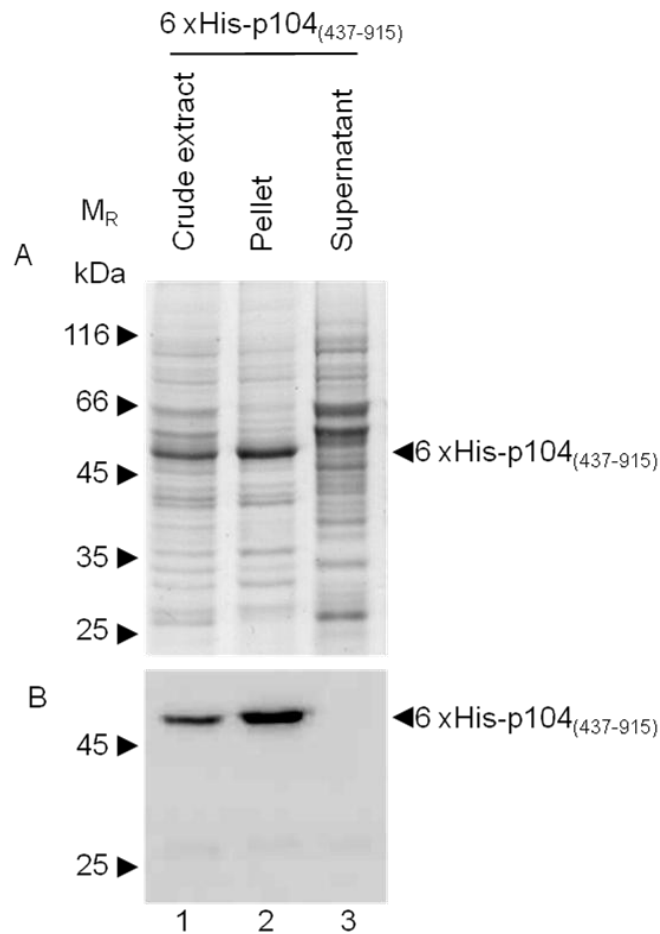


Figure 2. 6: The solubility of 6xHis-p104₍₄₃₇₋₉₁₅₎ expressed in Sf9 cells infected with BVp55 (P4). Panel A: SDS-PAGE Coomassie-stained 12 % gel showing solubility of expressed recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎. Lane 1: Crude extracts: Lane 2: Insoluble fraction (pellet): Lane 3: Soluble fraction (Supernatant). Panel B: Western blot analysis of protein extracts in panel A probed with anti-6xHis antibodies. Lane 1: Crude extracts. Lane 2: Insoluble fraction (pellet): Lane 3 Soluble fraction (Supernatant). Recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎ is shown on the right (arrow). Relative molecular weights are indicated on the left in kDa.

Solubilization of 6xHis-p104₍₄₃₇₋₉₁₅₎

The expression of 6xHis-p104₍₄₃₇₋₉₁₅₎ in the baculovirus system resulted in the production of an insoluble protein. Attempts were made to solubilize the protein by treating the pellet fractions from Sf9 crude extracts with a range of solubilization agents to determine the one which best released p55 into soluble fraction. The pellets from Sf9 crude extracts were re-suspended in buffer containing detergents (1 % Triton, 1 % NP40, 7.5 % Sarkosyl, 0.1 % Polyethylenimine), denaturant (8 M Urea), high pH (NaCO₂ pH 11.1) or high salt (1 M NaCl). After this treatment the proteins were resolved by SDS-PAGE. The purpose of using these buffers was to disrupt non-covalent associations of membrane bound proteins. The detergents disrupt hydrophobic bonds while high pH and high salt disrupt ionic and hydrogen bonds of integral and peripheral membrane proteins. The denaturants disrupt non-covalent interactions of peripheral membrane proteins. None of the agents resulted in the release of 6xHis-p104₍₄₃₇₋₉₁₅₎ into the soluble fraction and so the protein could not be detected by western analysis using the anti-6xHis antibody (Figure 2.7). As a result, the insoluble peptide could not be used to raise antibodies.

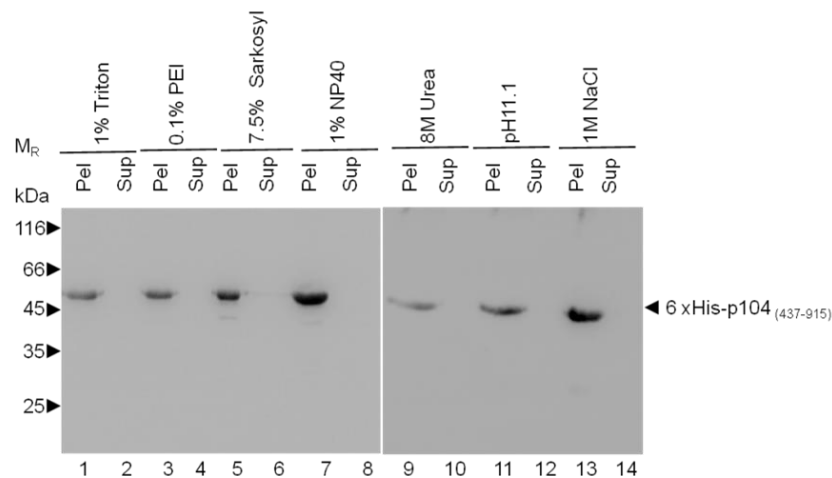


Figure 2. 7: Solubilization of 6xHis-p104₍₄₃₇₋₉₁₅₎. Western blot analysis of Sf9 infected crude extracts containing 6xHis-p104₍₄₃₇₋₉₁₅₎ using anti-6xHis antibody. Lane 1 - 2: Pellet and supernatant fraction treated with 1 % Triton X-100. Lane 3 - 4: Pellet and supernatant fraction treated with 0.1 % Polyethylenimine. Lane 5 - 6: Pellet and supernatant fraction treated with 7.5 % Sarkosyl. Lane 7 - 8: Pellet and supernatant fraction treated with 1 % NP40. Lane 9 - 10: Pellet and supernatant fraction treated with 8 M Urea. Lane 11 - 12: Pellet and supernatant fraction treated with NaHCO₂ pH 11.1. Lane 13 - 14: Pellet and supernatant fraction treated with 1 M NaCl. Recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎ is shown on the right (arrow). Relative molecular weights are indicated on the left in kDa.

Production of anti-p104₍₈₈₉₋₉₀₂₎ antiserum

The inability to produce soluble 6xHis-p104₍₄₃₇₋₉₁₅₎ for antibody production prompted a change of approach to using an antigenic peptide derived from the C-terminal region of p104. Bioinformatic analysis using the Hopp and Woods program to predict potential antigenic region indicated that the region from 889 - 902 amino acids was likely to be both soluble and antigenic (Figure 2.8) and this sequence was used to synthesize a peptide that was used to raise rabbit polyclonal antiserum by a commercial company (GenScript USA).

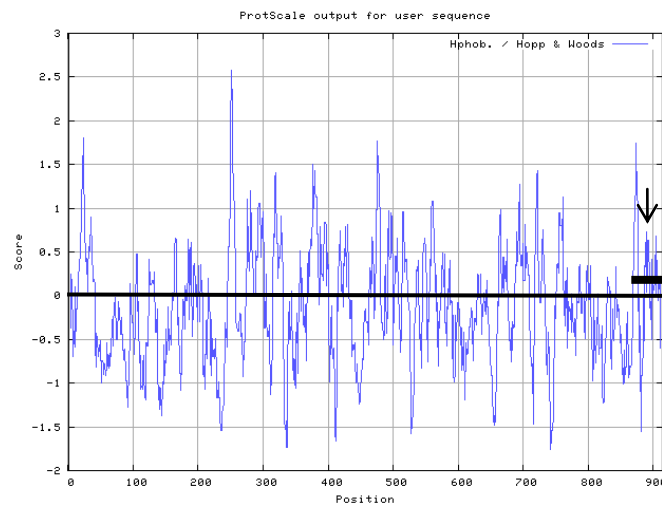


Figure 2. 8: The Hopp and Woods plot of p104 showing the region between 889 - 902 amino acids selected for antiserum production. A window size of 5 was used for prediction of putative antigenic site 889 - 902. The black line at 0 represents the threshold. The black bar and arrow represent the antigenic region of p104 from which the peptide 889 - 902 was derived. The hydrophilic antigenic sites are above zero.

Crude extracts from Sf9 cells infected with the recombinant baculovirus BVp55 (P4) were subjected to western analysis using the pre-immune serum and peptide antibody at varying concentrations to determine whether the antiserum was specific. The pre-immune antiserum (Day 0) antiserum did not cross-react with proteins in uninfected or baculovirus-infected cells (Figure 2.9, Lanes 1, 2) while a 55 kDa or 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein was detected in infected Sf9 cells by the antiserum from the immunized rabbit (Figure 2.9, Lanes 4 and 6). The 55 kDa band was not present in the uninfected cells (Figure 2.9, Lane 3, 5 and Lane 7) indicating that the antiserum was able to detect p104. The presence of several bands in both uninfected

and baculovirus-infected Sf9 cells suggested that there was some non-specific interaction with unknown cellular proteins (Figure 2.9, Lane 3, 4, 5, 6, 7 and 8).

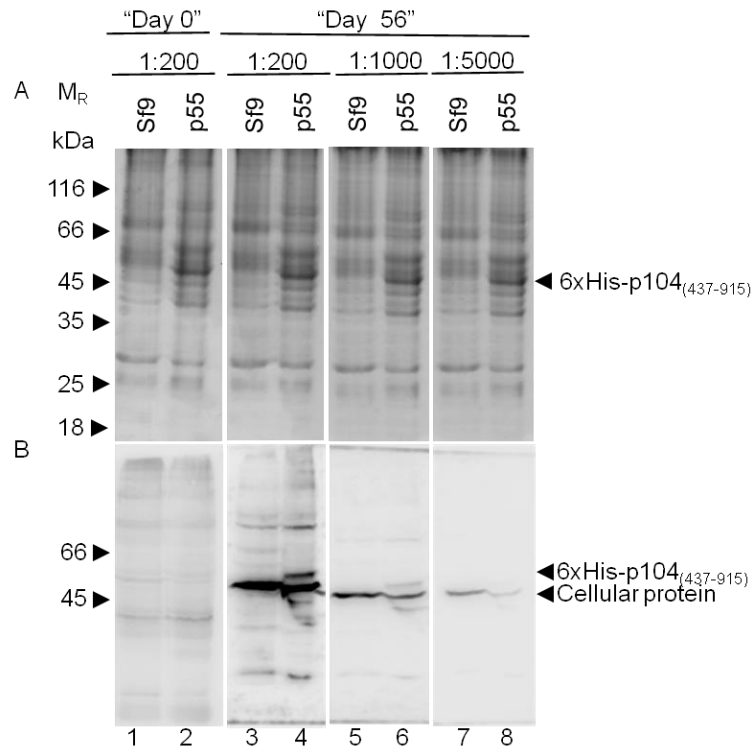


Figure 2. 9: Validation of anti-p104₍₈₈₉₋₉₀₂₎ antiserum . Panel A: Coomassie-stained 12 % SDS-PAGE gel showing respective resolved proteins. Lane 1, 3, 5, 7: Crude extracts from Sf9 cells. Lane 2, 4, 6, 8: Crude extracts from BVp55 (P4) infected Sf9 cells. Panel B: Western blot analysis of extracts probed with anti-p104₍₈₈₉₋₉₀₂₎ antiserum. Lane 1 and 2: Sf9 cells and BVp55 (P4) infected Sf9 cells probed with pre-immune antiserum or “Day 0” at 1:200 dilution. Lane 3 and 4: Sf9 and BVp55 (P4) infected Sf9 crude extracts probed with anti-p104₍₈₈₉₋₉₀₂₎ antiserum at 1: 200 dilution. Lane 5 and 6: Sf9 and BVp55 (P4) infected Sf9 crude extracts probed with anti-p104₍₈₈₉₋₉₀₂₎ antiserum at 1: 1000 dilution. Lane 7 and 8: Sf9 and BVp55 (P4) infected Sf9 crude extracts probed with anti-p104₍₈₈₉₋₉₀₂₎ antiserum at 1: 5000 dilution. Recombinant 6xHis-p104₍₄₃₇₋₉₁₅₎ is shown on the right (arrow). Relative molecular weights are indicated on the left in kDa.

Next the anti-p104₍₈₈₉₋₉₀₂₎ antiserum was used to detect PrV p104 in PrV infected MG8 cells. No protein bands corresponding to p104 were detected (Figure 2.10). In contrast, a band corresponding to p55 or 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein and which served as a positive control, was detected in baculovirus-infected cells. This suggests

that the peptide antibody has low affinity for the low expression levels of p104 in MG8 cells.

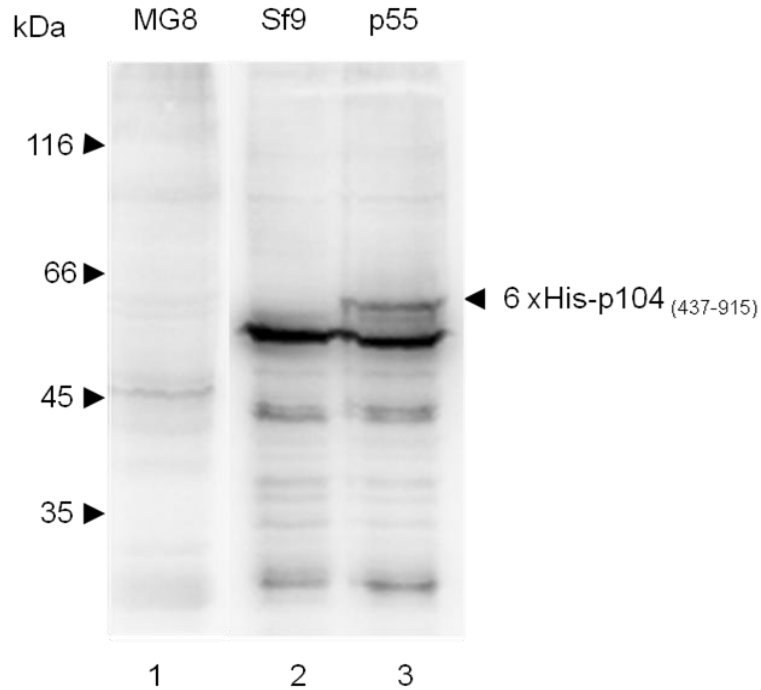


Figure 2. 10: Western blot analysis to detect p104 in PrV infected MG8 cells. Lane 1: PrV infected MG8 crude extracts. Lane 2: Crude extracts of uninfected Sf9 cells. Lane 3: Crude extracts of BVp55 (P4) infected Sf9 cells.

2.4 DISCUSSION

The objective of the research reported in this chapter was to produce an antibody raised against the C-terminal region of p104 that would distinguish between p40 and p104 in immunofluorescence studies. Attempts to express 6xHis-p104₍₄₃₇₋₉₁₅₎ in *E. coli* were unsuccessful; no detectable protein was detected by western analysis. The antigen was however expressed in insect cells using recombinant baculovirus expression. The 6xHis-p104₍₄₃₇₋₉₁₅₎ was found to be insoluble. A range of solubilization treatments did not improve protein solubility and this may be attributed to expression of a truncated peptide that is misfolding because of the absence of the N-terminal region that solubilizes it. Short and Dorrington, (2012) also found that HaSV replicase was insoluble and native PrV replicase detergent resistant. It is likely that PrV replicase may also be associated with detergent resistant membranes.

The insoluble nature of 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein prompted the use of a short and hydrophilic antigenic peptide derived from the C-terminal region of p104. Antiserum raised against the sequence p104₍₈₈₉₋₉₀₂₎ was able to detect 6xHis-p104₍₄₃₇₋₉₁₅₎ in Sf9 cells but not native p104 in PrV infected MG8 cells. This could be due to the very low levels of p104 expression in MG8 cells and also possibly low affinity of the peptide antibody to the protein. Therefore this antibody cannot be used for western analysis but may be useful for immunofluorescence or immunoprecipitation studies (discussed in detail in Chapter 4).

CHAPTER 3

SUBCELLULAR LOCALIZATION OF PrV REPLICATION

3.1 INTRODUCTION

The ability of PrV to replicate in cell culture has offered a unique opportunity to study tetravirus replication *in vivo*. Despite the differences between the PrV replicase and those of tetraviruses belonging to the *Alphatetraviridae* and *Permutotetraviridae* families, there is some evidence that the tetraviruses may have a common site of replication in their host cells. Flotation assays have shown that the replicase of the alphatetravirus, HaSV is membrane-associated, leading to the proposal that HaSV replication was associated with detergent-resistant membranes derived from the endocytic pathway (Short *et al.*, 2010; Short and Dorrington, 2012). Fluorescence microscopy and biochemical fractionation analyses have demonstrated that the PrV replicase (detected by the anti-p40₍₂₄₈₋₃₃₁₎ antiserum) also associates with detergent-resistant punctate cytosolic structures, which are very similar to those formed by the replicase of the alphatetravirus, HaSV (Short and Dorrington, 2012), raising the possibility that the site of PrV replication may also be on membranes derived from the endocytic pathway.

The aim of the research described in this chapter was to further examine the subcellular localization of PrV replication in persistently infected *H. zea* MG8 cell line and in *Spodoptera frugiperda* Sf9 cells expressing EGFP-replicase fusion proteins. Data from this chapter has been published in Short, J. R., Nakayinga, R., Walter, C., Hughes, G. E., and Dorrington, R. A. (2013). Providence virus (family: *Carmotetraviridae*) replicates vRNA in association with the Golgi apparatus and secretory vesicles. *J. Gen. Virol.* 94, 1073 - 1078 (Appendix C).

3.2 MATERIALS AND METHODS

3.2.1. Cell culture

The *H. zea* MG8 cell line (Goodman and McIntosh, 1995), which is persistently infected with PrV, was used for virus replication studies. *S. frugiperda* pupal ovarian Sf9 cells (cell line 9 CRL 1711) were used for subcellular localization of EGFP-replicase fusion proteins (Vaughn *et al.*, 1977). MG8 cells were grown in ExCell 420 complete insect medium (SAFC Biosciences™) containing 1 % penicillin/streptomycin at 28 °C without CO₂ or foetal calf serum. The Sf9 cells were maintained in Grace's insect medium supplemented with 10 % foetal calf serum (Invitrogen) and 1 % penicillin/streptomycin (Gibco) at 28 °C without CO₂.

3.2.2 Immunofluorescence staining of PrV infected MG8 cells

MG8 cells were stained with antibodies directed against p40/p104 and dsRNA as follows. Tissue culture plate containing a single glass coverslip in each well was seeded with 2.5×10^5 MG8 cells in 1 ml complete EX-CELL™ 420 medium (SAFC Biosciences™), after which the cells were gently agitated at 100 rpm and then allowed to attach to coverslips at 28 °C. Next, the medium was removed from the cells, the cells were washed three times in PBS (pH 7.6) and fixed with 4 % paraformaldehyde in 0.4 ml PBS (pH 7.6) for 10 min. The fixation process was followed by three washes in PBS (pH 7.6) and the coverslips were incubated with 0.4 ml permeabilization buffer (1 % Triton X-100, 5 % goat serum, 10 % sucrose in PBS pH 7.6) for 20 min at room temperature with agitation at 100 rpm.

The anti-p40₍₂₄₈₋₃₃₁₎ rabbit antiserum (Walter, 2008) and mouse monoclonal anti-dsRNA antibodies (clone K1, English and Scientific Consulting) were diluted 1:7500 and 1:5000, respectively into 0.4 ml permeabilization buffer and added to the cells, which were then incubated for 1 hr. Next, the cells were washed three times in PBS-Tween (1 % Tween 20 in PBS, pH 7.6) and incubated further for 1 hr in secondary antibody. The Fluorescein isothiocyanate (FITC) conjugated goat anti-rabbit secondary antibody (Zymax™) (diluted 1:350 in 0.4 ml permeabilization buffer), was used to detect p40 and p104 while the Alexa Fluor 546-goat anti-mouse antibody, diluted 1:700 (Invitrogen) in 0.4 ml permeabilisation buffer, was used to detect dsRNA. The cells were then briefly rinsed three times in PBS-Tween followed

by three 10 minute washes also in PBS-Tween. DAPI (4', 6-diamidino-2-phenylindole) (500 ug/ml) was added to the second wash to visualise nuclei. Coverslips were mounted on glass slides using fluorescent mounting medium (DakoCytomation) and analysis was conducted on the Zeiss LSM 510 META confocal microscope using the 405 nm (blue diode), 488 nm (argon) and 543 nm (helium/neon) laser lines.

3.2.3 Construction of native and mutant EGFP-tagged PrV replicase expression plasmids

The replication-independent subcellular distribution of the PrV replicase proteins was investigated in Sf9 cells using recombinant plasmids expressing p104 and/or p40 fused at their N-terminus to the C-terminus of EGFP. The EGFP-fusion proteins were expressed under the control of the *Autographa californica* nucleopolyhedrovirus enhanced promoter (*hr:AcMNPV*) (Ayres *et al.*, 1994). Plasmids pCW57 and pCW56 encode EGFP-p40 and the EGFP CDS respectively (Walter, 2008). Plasmids expressing EGFP fused to the full-length p104 ORF were constructed as shown in Figure 3.1. First, pRN39 that encodes the full-length replicase ORF, was digested with *Eco* RI, producing a fragment of 2560 nts, and this fragment was inserted into the *Eco* RI site of the vector pEGFP-C1, generating the intermediate construct pRN54. This plasmid contained the amino acids 1 - 847 of p104, located downstream of the cytomegalovirus (CMV) promoter and followed by the simian virus 40 poly(A) signal.

Next, pRN39 was digested with *Eco* RV and *Kpn* I and an insert of 425 nts, representing amino acids 848 to 915 of p104, was subcloned into the *Eco* RV/*Kpn* I sites of pRN54. This resulted in the second intermediate plasmid, pRN56, which was digested with *Bgl* II and *Kpn* I to release a fragment of 2926 nts carrying the full-length p104 ORF, which was subcloned into the same restriction sites in pCW57 (containing the EGFP ORF fused to the N-terminus of p40 downstream of the *hr:AcMNPV* promoter) (Walter, 2008). The resulting recombinant plasmid, pRN58, carries the EGFP-tagged wild-type p104 CDS with the UAG readthrough stop codon. The integrity of the 5' end of the EGFP-fusion protein and the *hr:AcMNPV* promoter was confirmed by DNA sequencing using the primer JRS32 F (5'-CCG CGT TGG

TTT TAG AGG GC-3') while the 3' end of the fusion protein was confirmed using primer CW15 R (5'-GGT GGG ACA TGC ACG TGC -3').

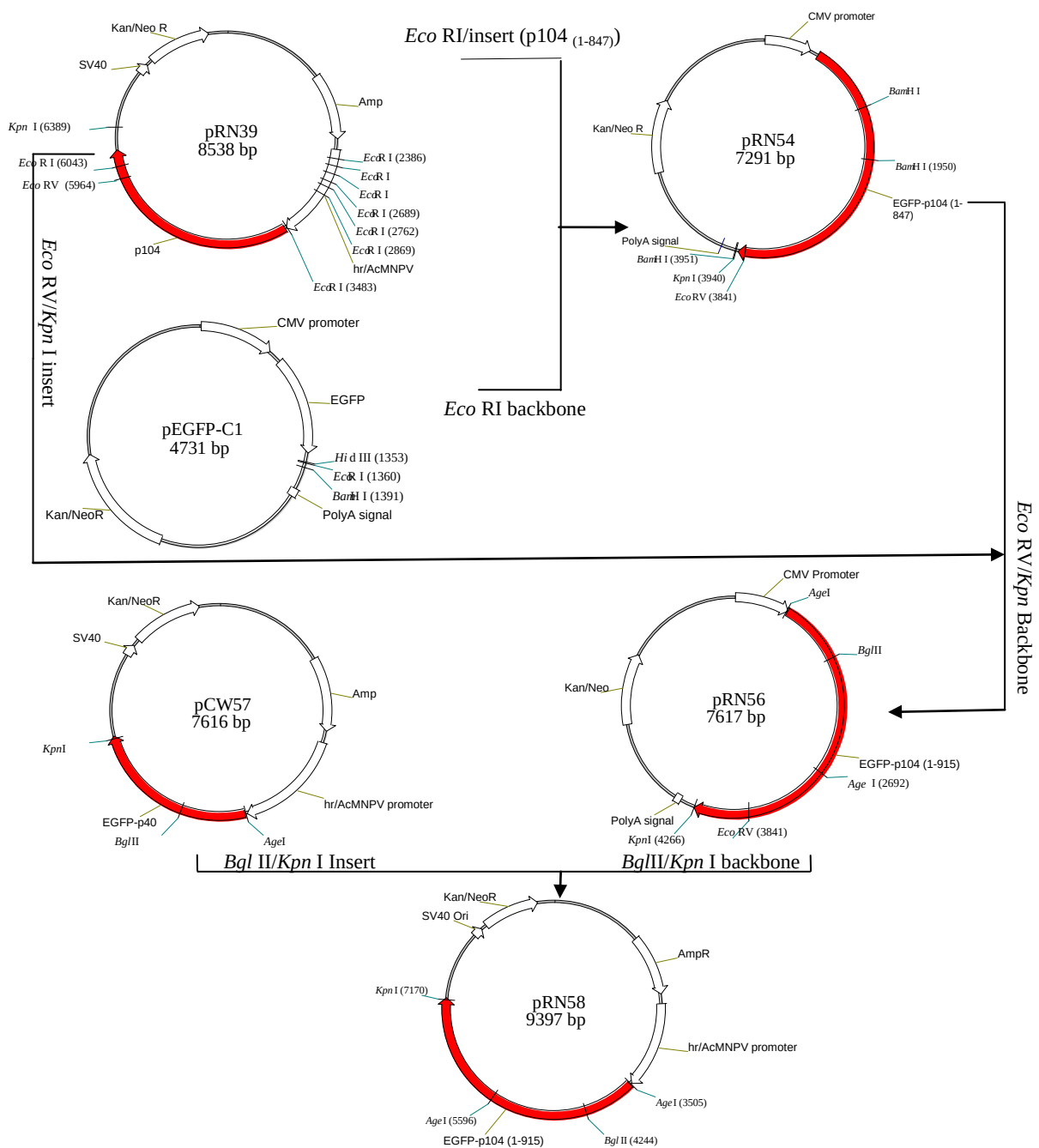


Figure 3. 1: Schematic representation of the cloning strategy for the construction of pRN58. The plasmid encodes the β -lactamase gene (AmpR), EGFP-p104, hybrid insect promoter (hr/AcMNPV promoter), Neomycin phosphotransferase (Kan/Neo R), Simian Virus 40 origin of replication (SV40 Ori).

To construct a recombinant plasmid (pRN59) expressing the p104 ORF with the UAG readthrough stop codon mutated to leucine (CUU), the plasmid pCW41 that encodes the full-length p104, was subjected to site-directed mutagenesis PCR (Quickchange kit, Stratagene) and the UAG readthrough stop codon at positions 1078 - 1080 of the p104 CDS were mutated to a leucine codon (CTT), resulting in the plasmid pRN3 (Figure 3.2). The primers RN1F (5'-GGA AGA ACA ATC CCG ACA AAG CTT CAA CTA AAG TGG GTC-3') and RN1R (5'-GAC CCA CTT TAG TTG AAG CTT TGT CGG GAT TGT TCT TCC-3') were used to introduce the leucine codon and a *Hind* III restriction site (underlined) for screening purposes. The identity of the mutation within the resultant plasmid, pRN3, was confirmed by DNA sequencing using CW17 F (5'-GGA GGA GCA CAT GAA GCC C-3'). The plasmid pRN3 was then digested with *Eco* RI and a fragment of 2548 nts was subcloned into the *Eco* RI site of the linearized vector pEGFP-C1 creating the intermediate plasmid pRN55. This plasmid contained amino acids 1 - 847 of p104, located downstream of the cytomegalovirus (CMV) promoter and followed by simian virus 40 poly(A) signal. The JRS32 F (5'-CCG CGT TGG TTT TAG AGG GC-3') was used to confirm whether EGFP sequence was in-frame with the p104 ORF

Next, pRN3 was digested with *Eco* RV and *Kpn* I and the resulting 425 nts fragment carrying amino acids 848 to 915 of p104, was subcloned into pRN55 digested with *Eco* RV and *Kpn* I, creating a second intermediate plasmid pRN57 (Figure 3.2). Thereafter pRN57 was digested with *Bgl* II and *Kpn* I and the 2926 nts fragment carrying the full-length p104 ORF was subcloned between the *Bgl* II and *Kpn* I sites of pCW57 (contains EGFP fused to the N-terminus of p40 and downstream of the *hr:AcMNPV* promoter) (Walter, 2008). The resultant recombinant plasmid pRN59 (Figure 3.2) carries the mutant EGFP-tagged p104(CTT). The primers used for DNA sequencing of pRN59 were similar to those described for pRN58.

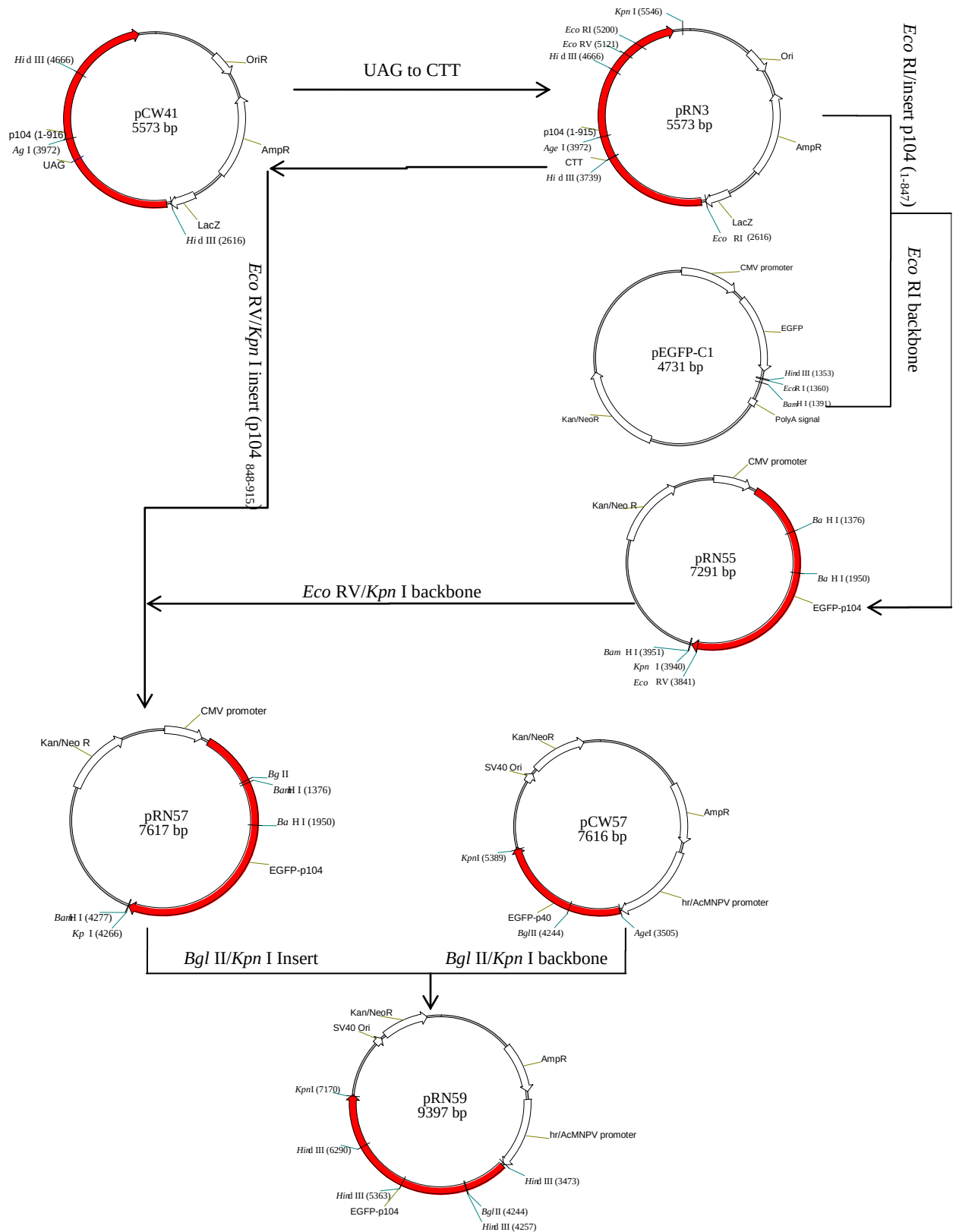


Figure 3. 2: Schematic representation of the cloning strategy for the construction of pRN59. The plasmid encodes the β -lactamase gene (*AmpR*), EGFP-p104 (CTT), hybrid insect promoter (*hr/AcMNPV* Promoter), Neomycin phosphotransferase (*Kan/Neo R*), Simian Virus 40 origin of replication (*SV40 Ori*).

3.2.4 Expression of EGFP-tagged PrV replicase plasmids pRN58, pRN59 and pCW57 in *S. fugiperda* Sf9 cells

To express EGFP-replicase fusion proteins in *S. fugiperda* Sf9 cells, a tissue culture plate containing a single glass coverslip in each well was seeded with 2.5×10^5 of *S. fugiperda* Sf9 cells in 1 ml of complete Grace's insect medium supplemented with 1 % penicillin/streptomycin (Gibco). The cells were allowed to settle for 1 hour at 28 °C. For larger plasmids such as pRN58 (9.4 kb), pRN59 (9.4 kb) and pCW57 (7.6 kb), approximately 300 μ g of each plasmid DNA was used to transfect Sf9 cells. For the smaller plasmid, pCW56 (6.4 kb), 250 μ g of DNA was used for transfection of Sf9 cells using FUGENE[®] HD transfection reagent (Roche) as described previously (Short *et al.*, 2010). The cells were incubated at 28 °C for 30 min and thereafter a volume of 80 μ l Grace's insect complete medium was added followed by incubation for 3 hours at 28 °C. An additional volume of 900 μ l Grace's insect complete medium was added and the cells incubated for 72 hours at 28 °C.

For confocal analysis, the coverslips were washed three times in PBS (pH 7.6) and fixed with 4 % paraformaldehyde in 0.4 ml PBS (pH 7.6) for 10 min. Next, the cells were briefly rinsed three times in PBS-Tween followed by three 10 minute washes also in PBS-Tween. DAPI (500 μ g/ml) was added to the second wash to visualise nuclei. Coverslips were mounted on glass slides using fluorescent mounting medium (DakoCytomation) and analysis was conducted on the Zeiss LSM 510 META confocal microscope using the 405 nm and 488 nm laser lines. To detect protein expression, the EGFP-tagged replicase fusion protein was detected using western analysis with anti-p40₍₂₄₈₋₃₃₁₎ antiserum at a dilution of 1:50,000 and secondary antibody prepared according to the kit instructions (BM Chemiluminescence western blotting kit, Roche).

3.3 RESULTS

The experimental approach was to employ fluorescence microscopy to identify the site of PrV replication in MG8 cells by staining the cells with a mouse antibody directed against dsRNA to localize replicating vRNA and a rabbit polyclonal anti-p40₍₂₄₈₋₃₃₁₎ antiserum (Walter *et al.*, 2010) to localize p40 and p104. However, since

the antibody detects both p104 and p40, it cannot distinguish between the full-length (p104) and the short (p40) proteins. The anti-p104₍₈₈₉₋₉₀₂₎ peptide antibody produced in this study (Chapter 2) did not detect p104 in MG8 cells and could not be used for immunofluorescence studies (data not shown). Therefore it was decided to use the rabbit anti-p40₍₂₄₈₋₃₃₁₎ antibody for immunofluorescence studies despite its inability to distinguish between p40 and p104.

3.3.1 Immunofluorescence co-localization of PrV replication proteins and replicating vRNA in *H. zea* MG8 cells

PrV-infected MG8 cells were stained with anti-p40₍₂₄₈₋₃₃₁₎ antiserum and mouse anti-dsRNA antibodies and the distribution of fluorescence was analyzed using confocal microscopy. A total of 2207 cells were surveyed and on average 11 % of the cells stained positive for both p40/p104 and dsRNA (Figure 3.3). Amongst the remainder of the cells, 0.8 % stained positive for p40/p104 alone (Figure 3.3, open arrow) while dsRNA alone was present in a further 0.9 % of the cells (Figure 3.3, indicated with the thick arrow). These observations suggested that in the MG8 cell line, only a relatively small population of cells was infected with PrV. This meant that more than 89 % of the cells were likely to be uninfected.

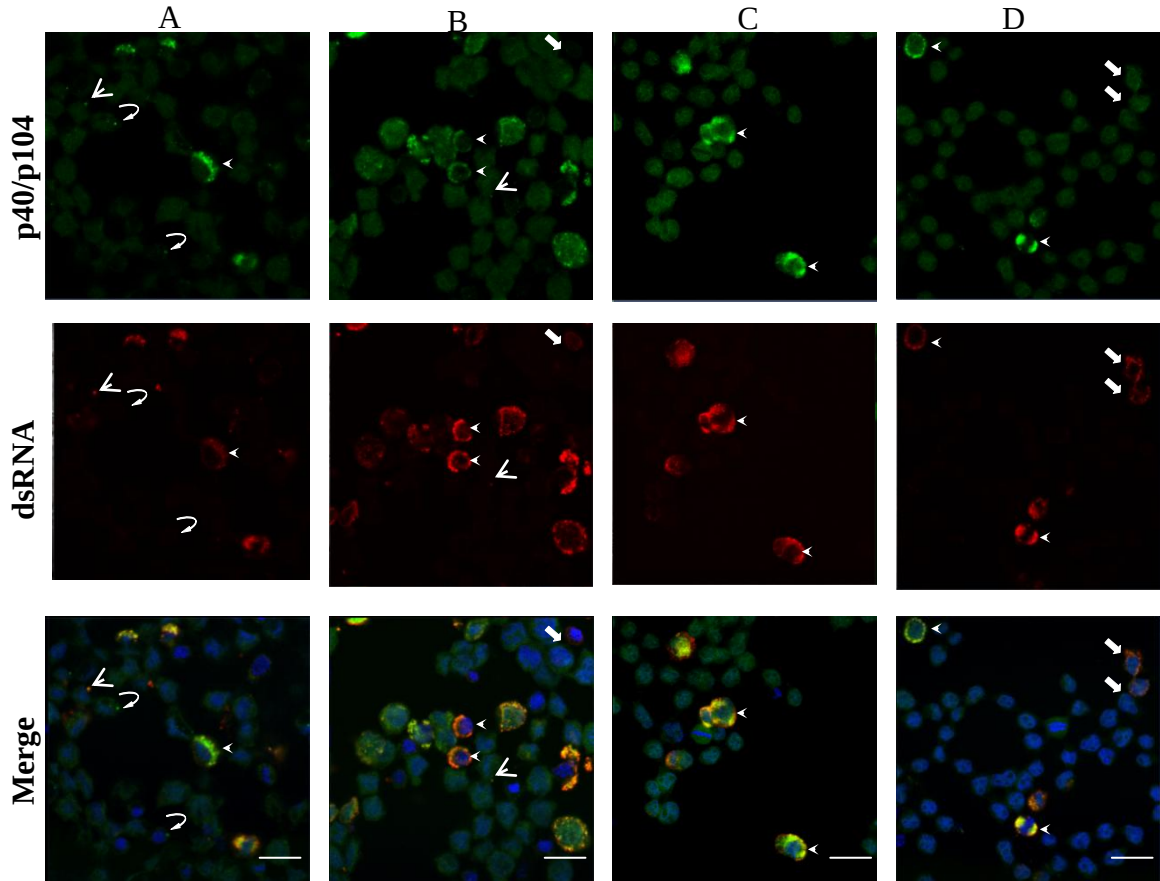


Figure 3. 3: Detection of replication complexes in PrV infected MG8 cells using immunofluorescence microscopy. Top; cells stained with anti-p40₍₂₄₈₋₃₃₁₎ antiserum, middle; cells stained with anti-dsRNA antibodies, bottom; merged images with DAPI. Single punctate structures with overlapping p40/p104 and dsRNA are denoted with an arrow with a tail (↵). Aggregated structures of p40/p104 and dsRNA are denoted with an arrow head (◄). Single punctate structures with p40/p104 alone are shown with an open arrow (↗) while dsRNA containing aggregate punctate structures are shown with a thick arrow (↘). Scale bars represent 10 μ m.

Further examination of confocal microscopy images of the infected MG8 cells stained for p40 and dsRNA revealed four different phenotypes of the punctate structures that are likely to represent the different stages of PrV infection (Figure 3.3). The first phenotype, which corresponds to p40 single punctate structures, might represent the initial stage of PrV infection and setting up of the replication factory. The second phenotype which corresponds to single punctate structures with overlapping p40/p104 and dsRNA fluorescence signals is represented approximately in 0.01 % of the cells and likely to represent early stages of vRNA

replication. The third phenotype is represented by cells with aggregated punctate structures containing overlapping p40/p104 and dsRNA fluorescence signals, distributed throughout the cytoplasm. The fourth phenotype is represented by cells containing dsRNA aggregate punctate structures scattered throughout the cytoplasm and likely to represent full viral RNA replication during the late stages of PrV infection. The different types and distribution of these punctate structures may suggest that the MG8 cells are being re-infected continually. Alternative explanations for the different staining patterns observed could also be a result of different phases of the cell cycle on the distribution of PrV-occupied organelles and differential impacts of cellular anti-viral responses on p40/p104 accumulation.

Closer examination of MG8 cells stained with p40/p104 and dsRNA fluorescence revealed that not all cells have the same phenotypes. In some cells, the two signals overlapped entirely (Figure 3.4A) while in others there was partial overlap with some signal only for dsRNA (Figure 3.4B and C). In cells that exhibited overlapping signals of p40/p104 and dsRNA, the p40/p104 was on the perinuclear side while the dsRNA appeared to extend beyond the p40/p104 signal into the cytoplasm (Figure 3.4C). This suggests that p40/p104 may form a replication complex with the vRNA being replicated on the cytoplasmic face, away from the nucleus.

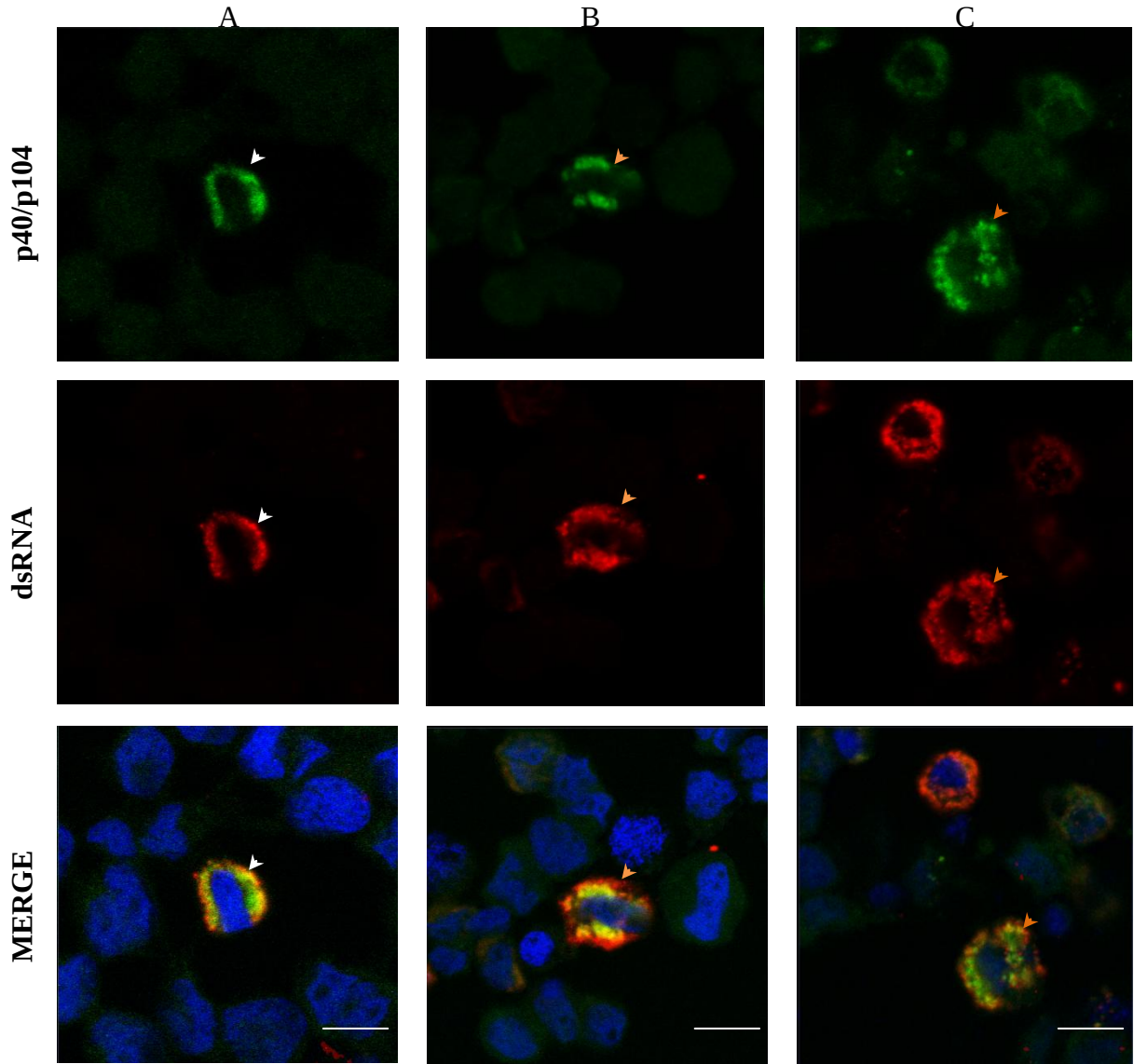


Figure 3. 4: Immunofluorescence of PrV infected MG8 cells. Selected PrV infected MG8 cells stained with anti-p40₍₂₄₈₋₃₃₁₎ antiserum (Top), mouse anti-dsRNA antibodies (middle), merged (bottom). All cells stained with DAPI. The cell displaying aggregated structures with overlapping signal for p40/p104 and dsRNA is denoted with an arrow head (white) while cells with aggregated structures with some signal only for dsRNA are shown with an arrow head (Orange). Scale bars represent 20 μm .

3.3.2 Expression of PrV replicase in *S. fugiperda* Sf9 cells

The next step was to investigate the mechanism by which p40 and p104 are trafficked to the site of replication using EGFP-fusion plasmids, to observe protein trafficking. The approach was to construct N-terminal EGFP fusions that would produce only EGFP-p40, EGFP-p40 together with EGFP-p104 (the wild-type p104

CDS containing the readthrough stop codon) or only EGFP-p104 with the readthrough stop replaced by CUU encoding leucine (Leu). The subcellular distribution of these fusion proteins was examined in Sf9 cells using confocal microscopy.

Expression of the EGFP-fusion proteins in *S. fugiperda* Sf9 cells transfected with pRN58 (EGFP-p40/p104), pRN59 (EGFP-p104(CTT)) and pCW57 (EGFP-p40) was confirmed by subjecting cell-free protein extracts to western analysis using anti-p40₍₂₄₈₋₃₃₁₎ antiserum. A 67 kDa protein, corresponding to the predicted size of EGFP-p40, was detected in cells transfected with pCW57 (Figure 3.5, Lane 4). This band was also present in cell-free extracts expressing EGFP-p40/p104 (pRN58) (Figure 3.5, Lane 2, indicated with an arrow), but not in uninfected Sf9 cell extracts (Figure 3.5, Lane 1). The full-length EGFP-p104 and EGFP-p104(Leu) proteins, with a predicted size of approximately 131 kDa, was not detected, likely due to the low expression levels of the full-length fusion protein (Figure 3.5, Lanes 2 - 3). Although the expression of EGFP-p104 could not be confirmed by western analysis, the presence of EGFP-p40, produced by the activity of the readthrough stop within the p104 CDS indicated that expression of EGFP-p104 was likely to be occurring but at very low levels. The lower expression of EGFP-p40/p104 in the p104 fusion as opposed to EGFP-p40 (pCW57) is likely due to differences in transformation efficiency due to the different sizes of the two plasmids. This may also result from the activity of the readthrough stop codon within the full-length p104 ORF.

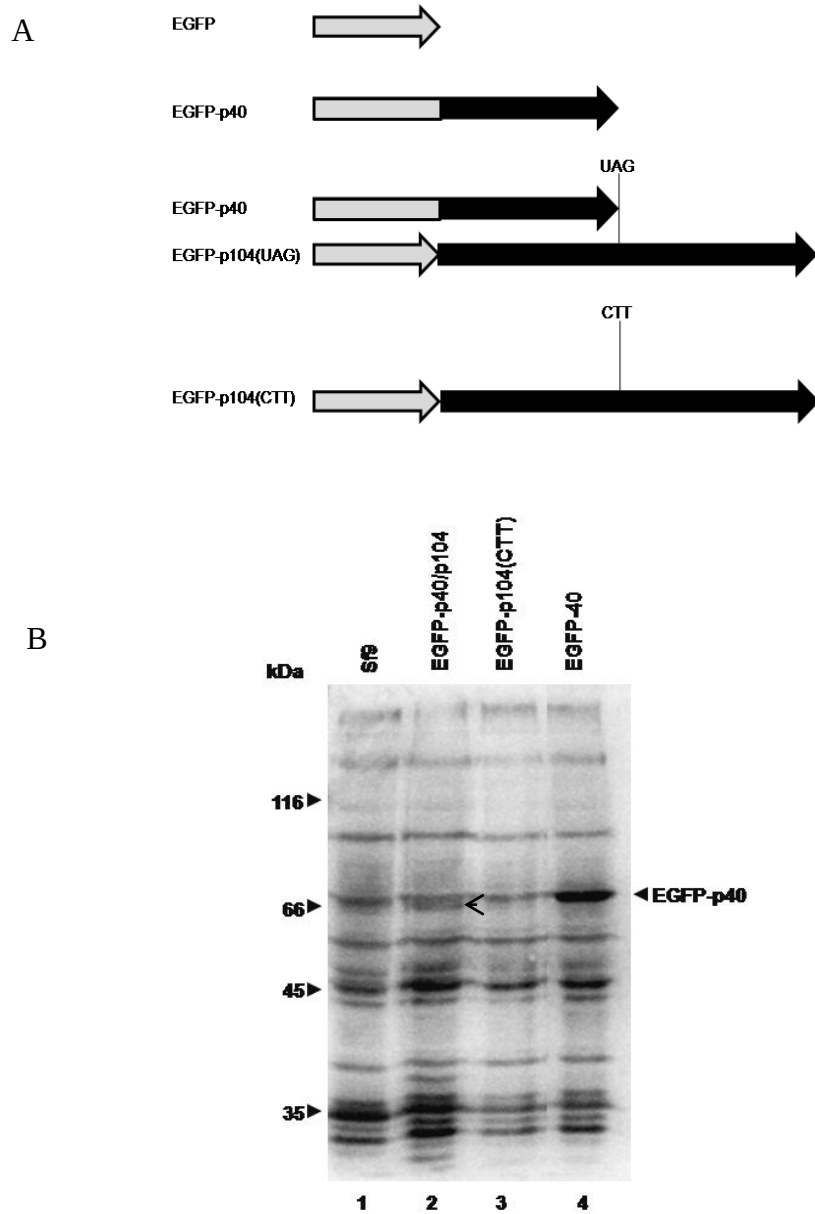


Figure 3. 5: Detection of EGFP-tagged PrV replicase in *S. frugiperda* Sf9 cells. A. Schematic representation of the plasmids used. The grey region represents EGFP. The grey/black boxed arrows represent EGFP-tagged p40 or p104. The readthrough stop codon is indicated as UAG and is replaced by the CTT in EGFP-p104(CTT). **B.** Western blot analysis of cell free extracts from Sf9 cells transfected with expression plasmids using anti-p40₍₂₄₈₋₃₃₁₎ antiserum. Lane 1: Uninfected Sf9 cells. Lane 2: Sf9 cells transfected with pRN58 (EGFP-p40/p104, indicated with an arrow). Lane 3: Sf9 cells transfected with pRN59 (EGFP-p104(CTT)). Lane 4: Cells transfected with pCW57 (EGFP-p40). Relative molecular weights are indicated on the left in kDa while proteins and respective molecular weights are shown on the right.

3.3.3 Subcellular localization of EGFP-replicase fusion proteins

The plasmids pCW56 (EGFP), pCW57 (EGFP-p40), pRN58 (EGFP-p40/p104) and pRN59 (EGFP-p104(CTT)) were transfected into *S. frugiperda* Sf9 cells and the distribution of EGFP fluorescence was analyzed using confocal microscopy. The general distribution of native EGFP was diffused, throughout the cytoplasm with accumulation in the nucleus (Figure 3.6A). In contrast, fluorescence associated with EGFP-p40 was localized to numerous punctate structures that occupied the entire cytoplasm but were excluded from the nucleus (Figure 3.6B). The *S. frugiperda* Sf9 cells transfected with EGFP-p40/p104 (Figure 3.6C) and EGFP-p104(CTT) (Figure 3.6D) also contained fluorescent cytoplasmic structures of variable sizes but there were far fewer of these structures compared with cells expressing EGFP-p40 (Figure 3.6B). These observations imply that p40 and p104 localize independently to cytosolic punctate structures, probably using localization signals likely to be present within the N-terminal, p40 sequence. The similarity in distribution suggests that the EGFP tag at the N-terminus of p40/p104 does not affect trafficking of the proteins.

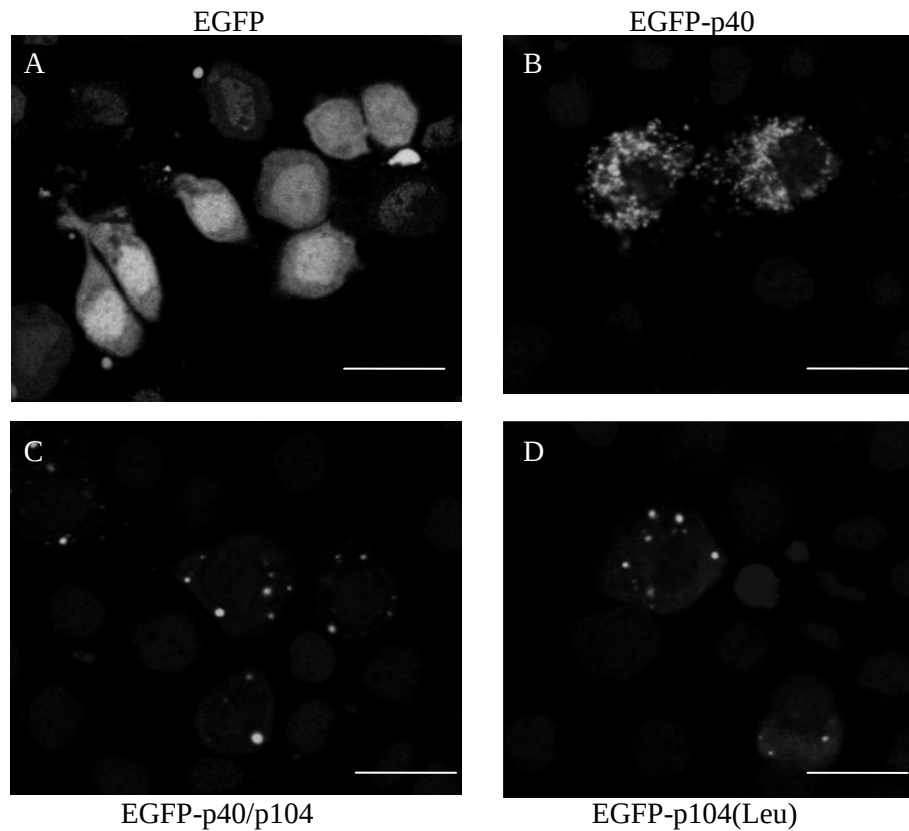


Figure 3. 6: Subcellular distribution of EGFP-tagged PrV p40 and p104. Confocal fluorescence images of *S. frugiperda* Sf9 cells expressing EGFP (A), EGFP-p40 (B), EGFP-p104 (wt) (C), EGFP-p104(Leu) (D). Scale bars represent 20 μm.

3.4 DISCUSSION

In this study, two approaches were used to investigate the subcellular localization of PrV p40 and p104. The first was to use immunofluorescence microscopy to detect whether p40/p104 co-localized with replicating virus in infected MG8 cells. In the second the replicase proteins were tagged with EGFP to study their subcellular localization in the absence of viral replication, in *S. frugiperda* Sf9 cells.

The most important finding from the immunofluorescence studies was that p40/p104 co-localizes with dsRNA in cytosolic punctate structures. The presence of the dsRNA signal suggests that p104, which contains the RdRp, is the viral replicase within the replication complex. Furthermore, different phenotypes of the punctate structures were observed that could represent the different stages of PrV replication. The initial stage of replication was characterized by single p40 punctate structures in 0.8 % of the infected cells and these may represent early steps in the formation of the viral replication complex. A total of 11 % of the infected cells contained replicating viral RNA. Of these, 0.01 % of the cells contained overlapping p40/p104 and dsRNA single punctate structures likely to represent viral replication complexes containing replicating RNA formed in the early stage of replication. The remainder of the cells were characterized by overlapping p40/p104 and dsRNA aggregated structures that might represent multiple viral replication complexes formed after several rounds of viral replication. Within the infected MG8 cells that stained both p40/p104 and dsRNA, the p40/p104 fluorescence signal overlapped with the signal for dsRNA in cytoplasmic structures and these observations are consistent with data obtained from similar experiments by JR Short and reported in Short *et al.* (2013). A small proportion of the cells (0.9 %) exhibited dsRNA aggregated structures that were distributed throughout the cytoplasm and these were likely to represent full viral RNA replication in the late stage of replication. Since p104 contains the RdRp domain, the presence of dsRNA punctate structures implies that p104 may be present although this cannot be verified with immunofluorescence microscopy.

The results of EGFP-tagged PrV replication proteins showed that PrV p40 and p104 localized independently as EGFP-fusion proteins to cytoplasmic structures that were similar in distribution to the structures observed with direct staining of virus-infected

cells with anti-p40₍₂₄₈₋₃₃₁₎ antiserum. This suggests that the sequence responsible for subcellular localization to the site of replication must be located within the N-terminus of the p40 sequence. The presence of p40 within the replication complex is likely since EGFP-p40 accumulates to similar cytoplasmic structures as EGFP-p104. In addition, co-localization of p40 and dsRNA may suggest that p104 is present in the replication complex.

In conclusion, data reported from this chapter suggests that p104 is the viral replicase. A small proportion of cells are infected and within these cells, the initial, early and late stages of replication were demonstrated. The p40 and p104 are likely to localize to the same replication site. This then raises the question whether p40 and p104 interact and whether both are present in the same complex, as observed in TBSV. Additionally, what domains are present within the replicases that are responsible for subcellular localization?

CHAPTER 4

PROTEIN-PROTEIN INTERACTIONS BETWEEN PrV p40 AND p104

4.1 INTRODUCTION

Thus far, results from this study lend support for the hypothesis that p40 and p104 localize to small cytoplasmic structures that co-localize with dsRNA in PrV infected MG8 cells (Chapter 3), suggesting that both proteins are likely to be involved in establishing virus replication factories. In the tombusviruses, the (smaller) accessory protein forms multi-protein complexes with itself and with the full-length replicase via the interaction domain at the C-terminal region of the accessory protein which results in the establishment of the virus replication factory (Panavas *et al.*, 2005; Rajendran and Nagy, Rajendran and Nagy, 2004; 2006, Stork *et al.*, 2011). The similarities between the PrV and tombusvirus replicases raise the possibility that PrV p40 may also be involved in similar multi-protein interactions with itself and that both p40 and p104 are targeted to the membrane to interact and form a replication complex that initiates viral RNA synthesis. The research described in this chapter tests this hypothesis by characterizing potential interactions that result in the formation of the PrV replication complex, focusing on the potential interactions between p40 and p104.

4.2 MATERIALS AND METHODS

4.2.1 Conjugation of antibodies to magnetic dynabeads

The anti-p104₍₈₈₉₋₉₀₂₎ and anti-GFP antiserum were coupled to the M-270 epoxy magnetic dynabeads according to manufacturer's instructions (Dynabeads® Co-Immunoprecipitation Kit, Invitrogen). The conjugated beads were re-suspended in 1 ml of the buffer supplied with the kit to a final concentration of 20 mg antibody-coupled beads/ml and stored at 4 °C. To verify binding of the antibody to beads, 5 µl of the original sample containing antibody and the depleted supernatant obtained after antibody-bead coupling reaction were re-suspended in an equal volume of 2 x SDS-PAGE loading buffer (100 mM Tris-HCL pH 6.8, 200 mM dithiothreitol, 4 %

SDS, 0.2 % bromophenol blue and 20 % glycerol), heated to 95 °C for 5 min and resolved on 12 % SDS-polyacrylamide gel. The primary antibodies were detected by western analysis. The antibody coupled to beads was detected using the dot blot technique, by applying 5 μ l of the uncoupled beads, antibody or antibody-bead complex to a nitrocellulose membrane and detection using anti-mouse IgG-POD/ anti-rabbit IgG-POD antibody.

To validate the p104 immunoprecipitation assay, cell-free lysates containing 6xHis-p104₍₄₃₇₋₉₁₅₎ (baculovirus-infected Sf9 cells) were mixed with the antibody-bead complex and incubated for 1 hr at 4 °C. The protein complex was separated from the beads using the buffer supplied by the kit and resolved on a 12 % SDS-polyacrylamide gel followed by the detection of the 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein by western analysis using anti-p104₍₈₈₉₋₉₀₂₎ peptide antibody at 1:200 dilution and anti-mouse IgG-POD/ anti-rabbit IgG-POD antibody.

4.2.2 Co-immunoprecipitation of PrV replication proteins

A confluent monolayer of PrV infected MG8 cells were grown in 15 ml of EX-CELLTM 420 medium (SAFC BiosciencesTM) and harvested by centrifugation at 5000 x *g* for 5 min at 4 °C. The cells were lysed using a buffer supplied in the kit and disrupted by passing through a 25 gauge needle 40 times. Next, the soluble fraction was mixed and incubated with anti-p104₍₈₈₉₋₉₀₂₎ antibody-bead complex for 1 hr at 4 °C. The proteins were resolved on a 12 % SDS-polyacrylamide gel and detected using biotinylated primary anti-p40₍₂₄₈₋₃₃₁₎ antiserum at 1: 13,000 and secondary streptavidin horse radish peroxidase at 1:7500. The secondary detection system does not bind the heavy or light chain of the primary antibody but to the biotin via streptavidin.

4.2.3 Co-immunoprecipitation of PrV replication proteins in infected *S. fugiperda* Sf9 cells.

Two tissue culture flasks (25 cm²) containing confluent Sf9 cells were transfected separately with pCW57 (EGFP-p40) (Chapter 3, subsection 3.2.3) in one flask and pCW58 (p40) (Appendix A) in another, using a concentration of plasmid DNA of 500 μ g. The cells were then disrupted by passing through a 25 gauge needle twenty times followed by a centrifugation step at 5000 x *g* for 5 min at 4 °C. Equal volumes

of 13 μ l of the soluble fraction containing EGFP-p40 or p40 were mixed, incubated for 5 hr at 4 °C to allow the formation of protein complexes between EGFP-p40 and p40 and followed by immunoprecipitation with anti-GFP antibody coupled beads. The mixture was incubated overnight with agitation at 50 rpm at 4 °C. The protein complexes were separated from the beads as described above and detected by western analysis using anti-p40₍₂₄₁₋₃₃₁₎ antiserum at 1:50,000 dilution.

4.3 RESULTS

The experimental approach for detecting protein-protein interactions between p40 and p104 was to use the anti-p104₍₈₈₉₋₉₀₂₎ peptide antibody, which detects the C-terminal region of p104, but not p40, to immunoprecipitate p104. Thereafter, the immunoprecipitated proteins could be subjected to western blot analysis using the anti-p40₍₂₄₈₋₃₃₁₎ antiserum that could detect both p104 and p40.

4.3.1 Conjugation of anti-p104₍₈₈₉₋₉₀₂₎ antibodies to magnetic dynabeads

The conjugation of the anti-p104₍₈₈₉₋₉₀₂₎ antibody to the magnetic beads was verified by western analysis using anti-rabbit secondary antibodies. Comparison of the sample containing the antibody and the depleted supernatant collected after antibody-bead coupling reaction showed a reduction in the 50 kDa and 25 kDa bands (corresponding to heavy and light chains respectively) in the depleted supernatant. SDS-PAGE analysis followed by Coomassie staining showed the presence of the antibody heavy chain (50 kDa) and light chains (25 kDa) (Figure 4.1, Panel A, Lane 2 and 1) and this was confirmed by western analysis (Figure 4.1, Panel B, Lane 2 and 1). The antibody conjugated to beads was confirmed by dot blot technique (data not shown).

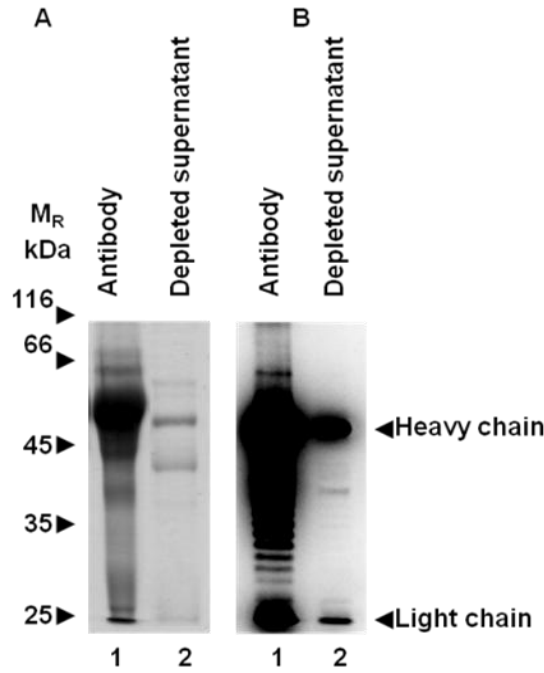


Figure 4. 1: Validation of conjugation of anti-p104₍₈₈₉₋₉₀₂₎ antibodies to magnetic dynabeads. **A** SDS-PAGE analysis using 12 % SDS-polyacrylamide gel. Lane 1: antibody sample before antibody-bead coupling reaction (or antibody). Lane 2: depleted supernatant after antibody-bead coupling reaction (or depleted supernatant). **B** Western analysis. Lane 1: antibody sample before antibody-bead coupling reaction (or antibody). Lane 2: depleted supernatant after antibody-bead coupling reaction (depleted supernatant). Relative molecular weights are indicated on the left in kDa. Bands corresponding with the heavy and light antibody chains are indicated on the right (arrow).

4.3.2 Experiment aimed at validating the immunoprecipitation protocol

Since the anti-p104₍₈₈₉₋₉₀₂₎ peptide antibody was not able to detect p104 in infected MG8 cells, the C-terminal 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein, expressed via recombinant baculovirus BVp55 (P4) in *S. frugiperda* Sf9 cells, was used to optimise the immunoprecipitation assay. To verify whether anti-p104₍₈₈₉₋₉₀₂₎ antibody was able to immunoprecipitate 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein, Sf9 cell-free extracts containing 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein were mixed with anti-p104₍₈₈₉₋₉₀₂₎ antibody-bead complex and the protein attached to the antibody-coupled beads was resolved by SDS-PAGE. Western analysis showed the presence of the fusion protein in the whole cell fraction (Figure 4.2, Lane 2, indicated with a dot), which was absent in the uninfected Sf9 cells (Figure 4.2, Lane 1) and low amounts in lane 3. A possible explanation for the presence of the fusion protein in lane 3 is that only a fraction of

the expressed protein was recovered using the beads. Beads coupled with anti-p104₍₈₈₉₋₉₀₂₎ antibody immunoprecipitated a 55 kDa protein from the Sf9 cell lysate (Figure 4.2, Lane 6 or indicated as E). Uncoupled beads did not pull down this protein from the cell free lysates (Figure 4.3, Lane 7 or indicated as B) and it was therefore concluded that the anti-p104₍₈₈₉₋₉₀₂₎ antibody was specific for the C-terminus of p104 and could be used for further immunoprecipitation assays.

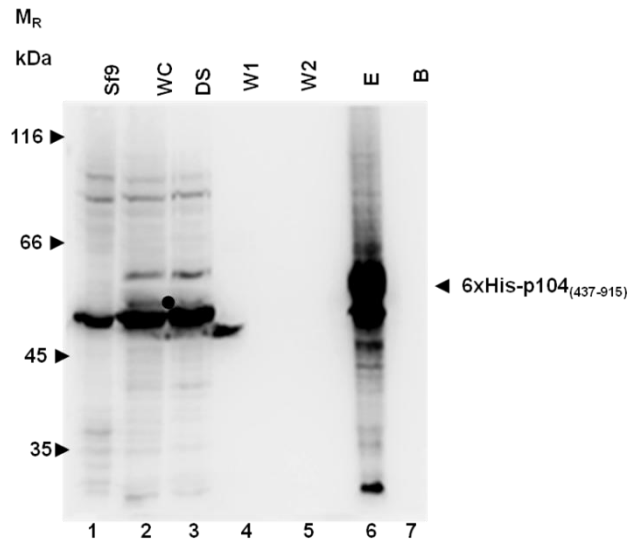


Figure 4. 2: Detection of 6xHis-p104₍₄₃₇₋₉₁₅₎ in Sf9 cell free extracts by immunoprecipitation and western blot analysis. Cell-free extracts mixed with anti-p104₍₈₈₉₋₉₀₂₎ antibody-coupled beads and detection with western blot using anti-p104₍₈₈₉₋₉₀₂₎ antibody. Lane 1: Uninfected Sf9 cells. Lane 2: Whole cell lysate (WC). The fusion protein in lane 2 is shown with a dot. Lane 3: Fraction of lysate after the formation of bead-antibody-protein complex or depleted supernatant (DS). Lane 4 and 5: washes (W1 and W2). Lane 6: Elution fraction containing protein complex (E). Lane 7: Elution fraction of the immunoprecipitation with uncoupled beads (or B). The 6xHis-p104₍₄₃₇₋₉₁₅₎ fusion protein is shown on the right (arrow). Relative molecular weights are indicated on the left in kDa.

4.3.3 Co-immunoprecipitation of PrV replication proteins from virus-infected cells

PrV p104-p40 interactions *in vivo*

To determine whether p104 and p40 interact *in vivo*, the anti-p104₍₈₈₉₋₉₀₂₎ antibody conjugated beads were mixed with cell free lysates derived from PrV infected MG8 cells and the immunoprecipitate was subjected to western blot analysis using

biotinylated anti-p40₍₂₄₈₋₃₃₁₎ antiserum and secondary streptavidin horse radish peroxidase conjugate. Western analysis of the immunoprecipitate showed the presence of a protein band corresponding to p104 in the MG8 cell lysates (Figure 4.3, Lane 5) as well as a band of 40 kDa, corresponding to p40. The relative intensities of the p40 and p104 bands was consistent with the high level of p40 expression relative to p104 in the crude lysates (Figure 4.3, compare Lane 5 with Lane 1). Both bands were absent in the control immunoprecipitation with uncoupled beads (Figure 4.3, Lane 6). This data shows that immunoprecipitation of p104 results in co-immunoprecipitation of p40, which suggests that these two proteins might interact to form the PrV replication complex. An alternative explanation is that there might be indirect interaction via a common host cell factor or that both proteins are localized to the same subcellular membranes resulting in co-precipitation of both proteins with the anti-p104₍₈₈₉₋₉₀₂₎ antibodies.

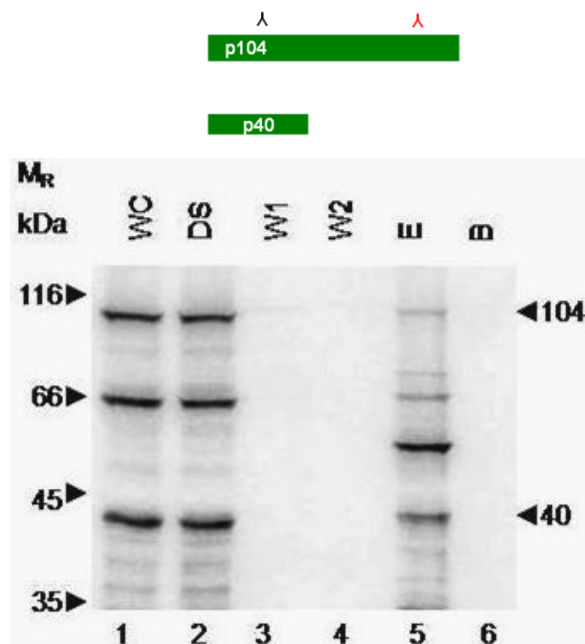


Figure 4. 3: Detection of p40 and p104 in PrV infected MG8 by immunoprecipitation and western blot analysis. Cell-free extracts from PrV infected MG8 cells immunoprecipitated with anti-p104₍₈₈₉₋₉₀₂₎ and detection with western blot using biotinylated anti-p40₍₂₄₈₋₃₃₁₎ antiserum. Lane 1: Whole cell lysate (WC), Lane 2: Fraction of lysate after the formation of antibody-bead-protein complex or depleted supernatant (DS), Lane 3 and 4: washes (W1 and W2), Lane 5: Elution fraction containing protein complex (E), Lane 6: Elution fraction from beads only (also indicated as B). Above the western blot images are schematic representations of p40 and p104 with the precipitating antibody indicated in red and antibody for detection indicated in black.

PrV p40 self interaction *in vitro*

Subcellular localization studies in Sf9 cells (Chapter 3) reveal that EGFP-p40 fusion proteins localized to cytoplasmic structures that were similar to those observed for the native protein in PrV-infected MG8 cells. This result also demonstrated that the presence of EGFP at the N-terminus of p40 did not appear to affect its trafficking or association with the punctate structures. Therefore it was decided to use this system to investigate whether p40 self interacts by determining whether immunoprecipitation of EGFP-p40 with anti-GFP antibodies would result in the co-immunoprecipitation of p40. To do this experiment two proteins p40 and EGFP-p40 were expressed separately, mixed and incubated gently at 4 °C on a rocker for 5 hrs, subjected to immunoprecipitation with anti-GFP antibodies. The ability to of anti-GFP antibody to pull down p40 provides evidence for self-interaction. First, the efficient coupling of the anti-GFP antibody to the magnetic beads was verified by western blot analysis using anti-rabbit secondary antibody. The sample containing the antibody was compared with the depleted supernatant collected after antibody-bead coupling reaction. A reduction in the 50 kDa (heavy chain) and 25 kDa (light chain) bands was observed (Figure 4.4, Panel A, Lane 2 vs Lane 1) and detected by western analysis (Figure 4.4, Panel B, Lane 2 vs Lane 1). The anti-GFP antibody conjugated to beads was confirmed by dot blot technique (data not shown).

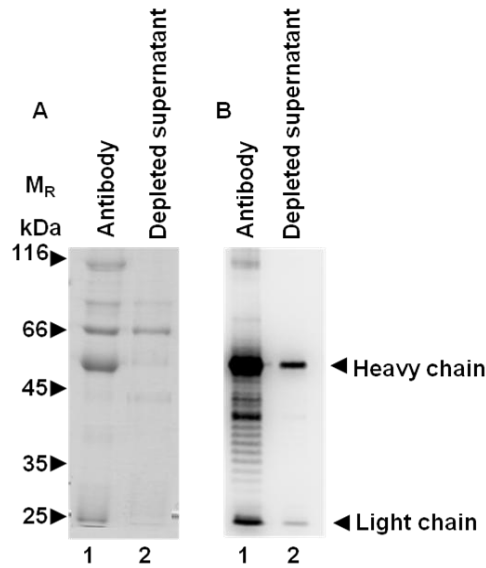


Figure 4. 4: Validation of conjugation of anti-GFP antibody to magnetic dynabeads.

A. Coomassie-stained 12 % SDS-polyacrylamide gel. Lane 1: sample before antibody-bead coupling reaction (or antibody). Lane 2: depleted supernatant after antibody-bead coupling reaction (or depleted supernatant). **B.** Western analysis. Lane 1: sample before antibody-bead coupling reaction (or antibody). Lane 2: depleted supernatant after antibody-bead coupling reaction (depleted supernatant). Relative molecular weights are indicated on the left in kDa. The heavy and light chains are indicated on the right (arrow).

Cell-free extracts containing EGFP-p40 and p40 were mixed together to allow for interaction between the two proteins, after which EGFP-p40 was immunoprecipitated using the anti-GFP antibody-coupled beads. Western analysis of the immunoprecipitate captured by the anti-GFP coupled beads showed the presence of protein bands migrating at 67 kDa (EGFP-p40) and 40 kDa (p40) (Figure 4.5, Lane 6) that were absent in Sf9 cell-free lysates treated with uncoupled beads (Figure 4.5, Lane 7). The immunoprecipitation of p40 by the EGFP antibody could only be explained by interaction between EGFP-p40 and p40 *in vitro*. Interestingly in contrast to the crude cell extract, there was more p40 relative to EGFP-p40 in the immunoprecipitate (Figure 4.5, compare Lane 6 to Lane 3). This suggests that p40 interacts with other p40 molecules. The presence of EGFP-p40 only in wash fractions (Figure Lanes 4 and 5) was unexpected however it is possible that anti-GFP antibody is not a high affinity antibody resulting in the elimination of unbound EGFP-p40 each wash.

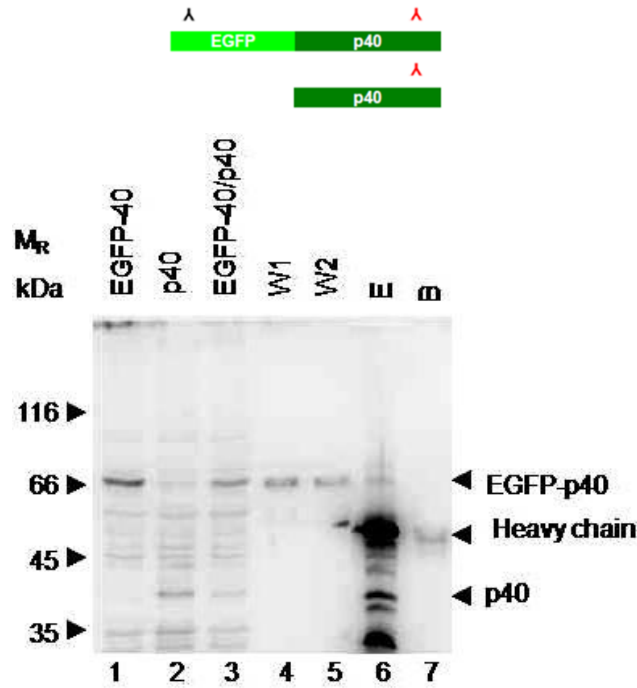


Figure 4. 5: Detection of EGFP-p40 and p40 by immunoprecipitation and western blot analysis. Western blot analysis using anti-p40₍₂₄₈₋₃₃₁₎ antiserum. Lane 1: Whole cell lysate from Sf9 cells expressing EGFP-p40 fusion protein. Lane 2: Whole cell lysate from Sf9 cells expressing p40. Lane 3: Mixed lysate containing EGFP-p40 fusion protein and p40. Lane 4 and 5: washes (W1 and W2). Lane 6: Elution fraction containing protein complex (E). Lane 7: Elution fraction from Sf9 cells treated with uncoupled beads (B). Above the western blot images are schematic representations of p40 and p104 with the precipitating antibody indicated in black and antibody for detection indicated in red.

4.4 DISCUSSION

The aim of the research described in this chapter was to characterize potential interactions between p40 and p104. Findings from immunoprecipitation experiments showed that PrV p40 likely interacts with p104 because the anti-p104₍₈₈₉₋₉₀₂₎ antibody pulled down p40 along with p104 from infected PrV MG8 cell lysates. This supports the data obtained from subcellular localization studies that suggested that p40 and p104 might be targeted to the same site (Chapter 3). The higher concentrations of p40 relative to p104 in the immunoprecipitate likely reflect higher concentrations of p40 relative to p104 in the whole cell fractions in infected MG8 cells. This then raised the question of whether p40 and p104 are interacting directly or indirectly via an intermediate host factor. Further experiments showed that p40

self interacts *in vitro* since anti-GFP antibodies used to immunoprecipitate EGFP-tagged p40 also pulled down p40. This is evidence that direct interaction exists between the p40 domains and therefore it is likely that p104 and p40 also interact via a domain that must be located within the p40 sequence upstream of the readthrough stop codon. Immunoprecipitation with anti-p104₍₈₈₉₋₉₀₂₎ antiserum pulled down p104 in lower quantities compared to p40. This data confirms the higher expression levels of p40 vs p104 that Walter *et al.* (2010) reported, however the actual ratio within the protein complex could not be concluded.

In conclusion, data from this study shows that PrV p40 interacts with itself and with p104 via an as yet unidentified domain within the p40 sequence. This motivates for bioinformatics analysis to identify the putative regions within the p40 sequence, which might be important for mediating protein-protein interactions between PrV replication proteins.

CHAPTER 5 IDENTIFICATION OF POTENTIAL FUNCTIONAL DOMAINS IN THE PrV REPLICASE AND p130
--

5.1 INTRODUCTION

PrV shares the characteristic $T = 4$ capsid symmetry with other tetraviruses however Speir *et al.*, (2010) found that PrV capsid architecture is controlled by a molecular switch similar to that of $T = 3$ nodaviruses. The non-structural proteins of PrV are very different from other tetraviruses. The carmo-like replicase has a readthrough stop codon and PrV is the only virus to have a third large non-structural protein (p130) with unknown function.

The focus of this study has been on the PrV non-structural proteins and in particular on the function of p104 in the viral life cycle. Subcellular localization studies in PrV-infected MG8 cells demonstrated that p40/p104 co-localize with dsRNA to small cytoplasmic structures presumed to be viral replication complexes (Chapter 3, section 3.3.1), confirming the proposed function for p40 and p104 in viral replication. Furthermore, EGFP-tagged p40 and p104 localize independently to similar punctate structures (Chapter 3, section 3.3.3) suggesting that they might carry the signals required for trafficking to the site of virus replication. Since the p40 sequence is common to both proteins, it is reasonable to suppose that the localization domain is present in the N-terminal region of p104, upstream of the readthrough stop signal. Finally, immunoprecipitation demonstrated that p40 self-interacts *in vitro* and interacts with p104 *in vivo* (Chapter 4). This data provides strong evidence of protein-protein interaction between p40 and p104 that is mediated through a domain in the N-terminal region corresponding to p40.

The aim of the research described in this chapter was to examine the sequence of PrV p104 using a range of bioinformatic tools to identify functional domains. In addition, a preliminary bioinformatic analysis of the p130 sequence was carried out to identify putative functional sequence motifs and domains that shed light on the role of this protein in the PrV life cycle. Finally, comparison of the PrV non-structural proteins

with those of the well-studied non-structural proteins belonging to the tombusvirus, TBSV and umbravirus, GRV, was performed to identify conserved sequence motifs.

5.2 MATERIALS AND METHODS

Potential functional domains within PrV replicase amino acid sequence were identified by bioinformatic analysis. The conserved domains within the sequence were identified by NCBI Conserved Domain (CD) program (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>; last accessed on 22/06/2009) operated by the Basic Local Alignment Search Tool (BLAST) (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>). Protein homologues of PrV replicase were identified by position specific iterative BLAST (PSI-BLAST) against Non-Redundant protein (NR) database (<https://www.ncbi.nlm.nih.gov>; accessed on 22/06/2009) operated by BLAST. Establishment of evolutionary relationships was performed by generating a multiple sequence alignment of the protein sequences using ClustalW (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>; accessed on 27/06/2009). Potentially hydrophilic domains and antigenic sites of the PrV replicase were generated by the Protscale program hosted by the Expert Protein Analysis System (ExPASy) (<http://web.expasy.org/protscale/>; accessed on 27/04/2012) based on parameters described by Hopp and Woods (1981). The surface exposed regions were generated using Kyte and Doolittle (1982) using a window size of 5 (<http://web.expasy.org/protscale/>; accessed on 27/04/2012). Putative transmembrane helices were generated using TMPred (Hofmann and Stoffel, 1993; http://www.ch.embnet.org/software/TMPRED_form.html; accessed on 27/04/2012). The Kyte and Doolittle (1982) was also used to predict putative transmembrane helices using a window size of 19 (<http://web.expasy.org/protscale/>; accessed on 27/04/2012). Potential disordered regions were identified using IUPRED (Dosztanyi *et al.*, 2005) (<http://iupred.enzim.hu/>; accessed on 09/08/2012) while secondary structures were identified using the PHD method (Rost and Sander 1993; 1994) operated by the Predict Protein programme (<http://www.predictprotein.org/>; accessed on 17/08/2012). Potential phosphorylation sites were predicted by the NetPhos 2 programme (Blom *et al.*, 1999) (<http://www.cbs.dtu.dk/services/NetPhos/>; accessed on 10/10/2012). The SignalP 4.0 (Petersen *et al.*, 2011)

(<http://www.cbs.dtu.dk/services/SignalP/>; accessed on 27/04/2012) was used to identify putative signal peptides within p104 and p130 sequence.

5.3 RESULTS

5.3.1 Identification of the putative RdRp domain within p104

In a previous study, Walter, (2008) conducted an analysis of the PrV p104 amino acid sequence, identifying a putative RdRp domain within the C-terminus of p104. The analysis suggested a permuted arrangement of the conserved motifs C, A, B within the p104 RdRp domain similar to that described for TaV and EeV by Gorbalenya *et al.* (2002). However, a phylogenetic analysis showed that p104 was more related to the replicases of viruses from the family *Umbraviridae* and *Tombusviridae*. Since the replicases of the *Umbraviridae* and *Tombusviridae* do not encode a permuted RdRp domain, it was decided in this study, to first revisit the identification of the PrV RdRp domain.

Sequence analysis using the NCBI CD programme identified a putative catalytic palm subdomain in the C-terminus of the protein sequence, downstream of the readthrough stopcodon, at amino acid positions 584 - 589 (motif IV), 640 - 650 (motif V), 722 - 724 (motif VI) and 766 - 768 (motif VII) (Figure 5.1). The profile output of the NCBI CD programme showed motif IV and VI as the putative metal binding sites while V as the putative nucleic acid binding site (Figure 5.1).

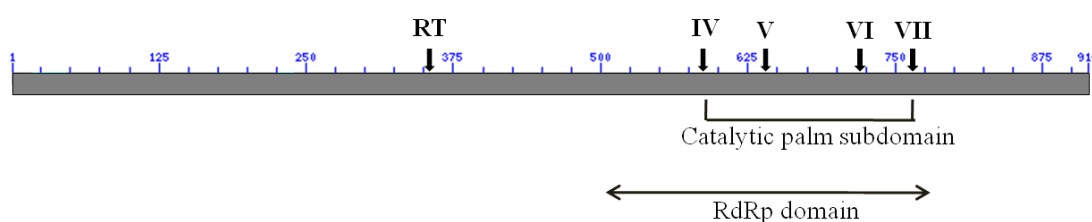


Figure 5. 1: Schematic diagram of p104 sequence showing the putative catalytic palm subdomain of the RdRp domain. The catalytic palm subdomain has motifs IV, V, VI and VII indicated by black arrows above sequence. The RdRp domain is shown as a black line below sequence. Readthrough stop codon (RT) is indicated as RT above the sequence.

According to this prediction, the putative catalytic palm subdomain of p104 is located at the C-terminus and shares motifs IV, V, VI and VII with (+ve) ss RNA RdRps (O'Reilly and Kao, 1998; Koonin 1991; Koonin and Dolja, 1993). The RdRp domain of PrV p104 is closely related to viruses from the family *Umbraviridae* and *Tombusviridae* (Walter *et al.*, 2010). Thus the organisation of the putative p104 RdRp domain was further analysed by aligning the predicted p104 C-terminal amino acid sequence with nine selected plant viral RdRps from both families using ClustalW.

Seven of the eight conserved motifs were identified within the PrV p104 sequence (Figure 5.2). Motif I (KMEK vs the consensus sequence, KxEK) was found at amino acids 505 - 508, next is motif II (R) at amino acid position 523, while motif III (KG) lies at amino acid positions 558 - 559. Motif IV or A (DLSRWD vs the consensus sequence Dx₄₋₅D) was at amino acid position 584 - 589, while motif V or B (SGDMMTGLGN vs the consensus sequence (S/T) Gx₂₋₃Tx₃N) lies at amino acid position 641 - 650. Motif VI or C (GDD) was at amino acid position 722 - 724 while motif VII or D (FCQ) lies at amino acid position 766 - 768. The motif VIII (K/R) found in (+ve) ss RNA viruses that belong to the carmo-like supergroup II (Koonin, 1991; Koonin and Dolja, 1993) and which should be at amino acid position 782 in p104 RdRp domain, appears to be absent (Figure 5.2). Unique to PrV RdRp is an additional motif IV or A₁ (DGTAED vs the consensus sequence Dx₄₋₅D) at amino acid position 755 - 760 (Figure 5.2). Walter (2008) showed that the PrV RdRp was permuted with the motifs arranged in C, A, B manner. In this analysis, motif C is followed by motif A₁, consistent with Walter (2008) observation however the motif B (NGATRLTMVPN) (Walter, 2008) at amino acid position 774 - 783 does not follow the consensus sequence of (S/T)Gx₂₋₃Tx₃N conserved in RdRps of viruses belonging to the carmo-like supergroup II (Koonin and Dolja, 1993). Therefore it is possible that motif B at amino acid position 774 - 783 may not be present in p104. Sequence similarity between the RdRp of PrV and that of the members of *Tombusviridae* and *Umbraviridae* resulted in the formation of a distinct cluster classified within the carmo-like supergroup II (Koonin, 1991; Koonin and Dolja, 1993). The arrangement of the conserved RdRp motifs in TBSV RdRp show similar organisation with PrV and the tombusviruses listed in the legend of Figure 5.2.

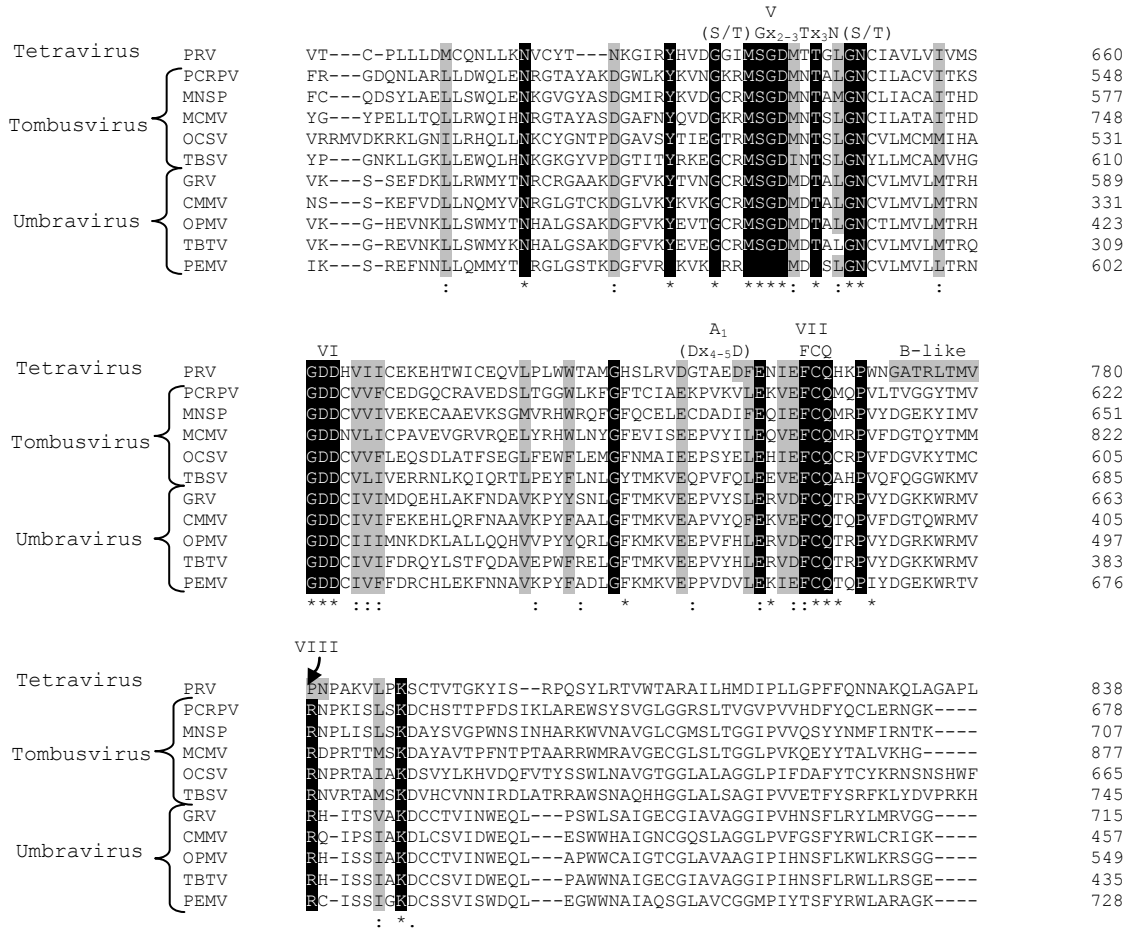


Figure 5. 2: Multiple sequence alignment of the RdRp domain of PrV p104 with tombusvirus and umbravirus replicases. Shaded black boxes represent the conserved residues also indicated by an asterisk (*). Grey shaded boxes indicate semi conserved amino acids also indicated by a colon (:). The cross (□) indicates the position of the readthrough stop within the PrV replicase (p104). Signature motifs and consensus sequences are indicated above the sequence according to Kamer and Argos (1984), Poch *et al.*, (1989), Koonin (1991) and Koonin and Dolja (1993). Motif A₁ is underlined and motif B-like sequence is shaded in grey and underlined. Motif VIII conserved in supergroup II virus is shown with an arrow. Providence virus (PRV, YP_003620397), Pelargonium chlorotic ring pattern virus (PCRVP, YP_052925), Melon necrotic spot virus (MNSV, BAG13035), Maize chlorotic mottle virus (MCMV, NP_619718), Oat chlorotic stunt virus (OCSV, NP_619751), Tomato bushy stunt virus (TBSV, NP_062897), Groundnut rosette virus (GRV, NP_619714), Carrot mottle mimic virus (CMMV, NP_054006), Opium poppies mosaic virus (OPMV, ABV66257), Tobacco bushy top virus (TBTv, NP_733848), Pea enation mosaic virus-2 (PEMV, AEM45994).

5.3.2 Other potentially functional domains of p104

Within the p104 sequence and downstream of the readthrough stop codon is a putative arginine-rich motif containing the sequence, RRRRYA, at amino acids 476 - 481. The PrV arginine rich motif contains amino acids R, Y, A, also found in the TBSV p33 arginine/proline-rich motif (RPRRRPYA or RPR motif or RNA binding region I (RBR 1)) at amino acids 213 – 220 and important for binding RNA (Rajendran and Nagy, 2003). Sequence analysis of the RBR 1 in TBSV p33 lies within the hydrophilic region and is surface exposed (Rajendran and Nagy, 2003). In order to establish whether PrV p104 arginine-rich motif is potentially hydrophilic and surface exposed, the Hopp and Woods programme and Kyte and Doolittle programme were used for analysis. The returned output of the analysis revealed that the PrV arginine-rich motif (referred to as PrV RBR 1) lies within the hydrophilic region (score of 2.5 above threshold score of 0) and is surface exposed (score of 4.2 below threshold score of 0) (Figure 5.3). This may imply that PrV RBR 1 is exposed to the cytosol.

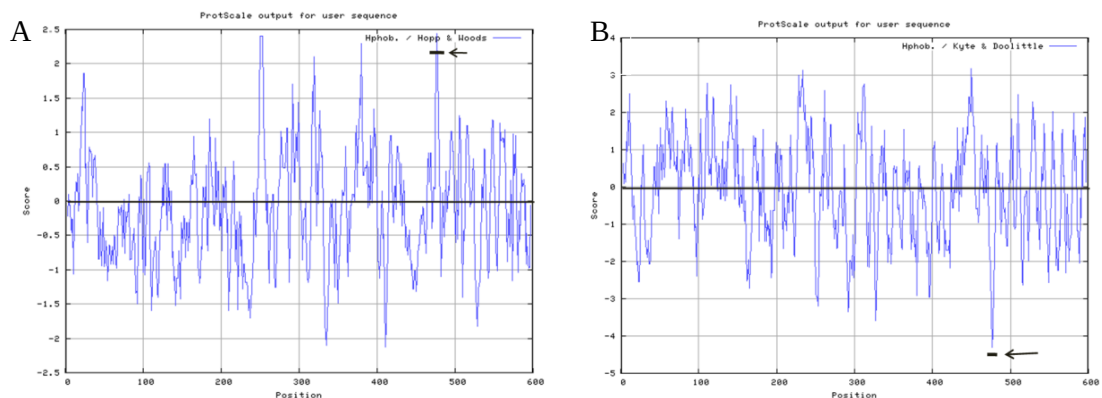


Figure 5. 3: Identification of hydrophilic and surface exposed characteristics of PrV arginine-rich motif. A. Hopp and Woods programme using a window size of 5. Hydrophilic region of PrV RBR I indicated with a black bar and arrow. Regions above 0 are hydrophilic **B.** Kyte and Doolittle programme using a window size of 5. The surface exposed region is shown as a black bar and arrow. Regions below 0 are potentially surface exposed. Black line at 0 represents threshold value. Y axis represents scores while X axis represents position of the amino acid sequence.

The similarity in amino acid composition, hydrophilic and surface exposed properties raises the possibility that PrV RBR 1 is a potential RNA binding site. It is worth

noting that the location of the putative PrV RBR 1 differs from that of TBSV RBR 1. In PrV p104, the putative RBR 1 is found at the N-terminus of p64 (C-terminal sequence of p104) while the TBSV RBR 1 is found at the C-terminal region of TBSV p33. The difference in location adds motivation for experimental verification of p104 RBR 1 as a functional RNA binding site.

Apart from TBSV p33 RBR 1, two TBSV RNA binding regions namely; RBR 2 and RBR 3 at amino acid position 441 - 486 and 688 – 818 of the C-terminal region of TBSV p92 bind RNA (Rajendran and Nagy, 2003). Sequence alignment of PrV p104 and TBSV p92 revealed 26 % similarity with amino acids D, F, V, K, E, K, P, R, I, Q, R, E within TBSV p92 RBR 2 also conserved in PrV p104 between amino acids 495 and 540 (Figure 5.4A). Deletion of the region containing these amino acids abolished the RNA binding activity of TBSV p92 RBR 2 (Rajendran and Nagy, 2003). Sequence conservation of TBSV RBR 2 and PrV implies that amino acids between the position 495 and 540 within p104 may represent a potential RNA binding site (referred to as putative PrV p104 RBR 2 hereafter). The putative PrV p104 RBR 2 is located at the C-terminal region of p104 and this is similar to the location of the TBSV p92 RBR 2 found at the C-terminus. The similarity in location may add some support to the presence of RBR 2 in PrV p104. Sequence alignment of PrV p104 and TBSV p92 RBR 3 revealed 10 % similarity between both regions (Figure 5.4B). It is likely that the putative PrV p104 RBR 3 at amino acid position 782 – 915 is not an RNA binding site or lack of homology may reflect that PrV interacts with an RNA structural element that is specific to this virus.



Figure 5. 4: Identification of a potential RNA-binding domain in PrV p104. Sequence alignment of (A) TBSV p92 and PrV p104 RBR 2 (B) TBSV p92 and PrV p104 RBR 3. The solid line above the sequence represents RNA binding regions, RRB 2 and RBR 3. Conserved residues are shaded black and have an asterisk (*) below the sequence. Colon (:) represent semi conserved amino acids. ClustalW parameters were set at 10, 0.2 and 5 for gap open, gap extension and gap distance respectively. The positions of the amino acids are shown on the right. Providence virus (PRV, YP_003620397), Tomato bushy stunt virus (TBSV, NP_062897).

While the RdRp consensus sequence was located in the p104 C-terminal region, neither the NCBI CD programme nor the BLAST identified any protein homologues or conserved domains within p40 amino acid sequence, upstream of the readthrough stop codon. Given the functional similarities between the PrV and tombusvirus, TBSV replicase, a comparative analysis was performed using the sequence of the interaction domain (ID) within TBSV p33 (amino acid residues 241 - 282) that mediates p33 self-interaction between p33 and p92 (Rajendran and Nagy, Rajendran and Nagy, 2004; 2006).

The TBSV p33 ID was examined for its hydrophilic and surface exposed properties using Hopp and Woods programme and Kyte and Doolittle programme. The data from the analysis was used to identify similar regions within p40 sequence with comparable characteristics. The C-terminal amino acids 241 – 282 of TBSV are located in the hydrophilic region (score of 1.3 above threshold value 0) and are surface exposed (score of -3.0 below threshold value 0) (Figure 5.5.). This implies that the TBSV p33 ID is exposed to the cytosol and as a result is able to mediate protein-protein interactions between TBSV replication proteins. Analysis of PrV p40

sequence revealed that the region 292 – 332 at the C-terminus of p40 lies within the hydrophilic region (score of 1.4 above threshold value 0) and is surface exposed (score of -2.5 below threshold value 0 using) (Figure 5.5). These properties are comparable to TBSV p33 ID and may suggest that the p40 sequence at amino acid position between 292 and 332 is likely to be exposed to the cytosol similar to TBSV p33 ID. Both the p40 sequence (292 – 332) and TBSV p33 ID are located at the C-terminus of these proteins and which may add some support to the predicted function of p40 as an interaction domain.

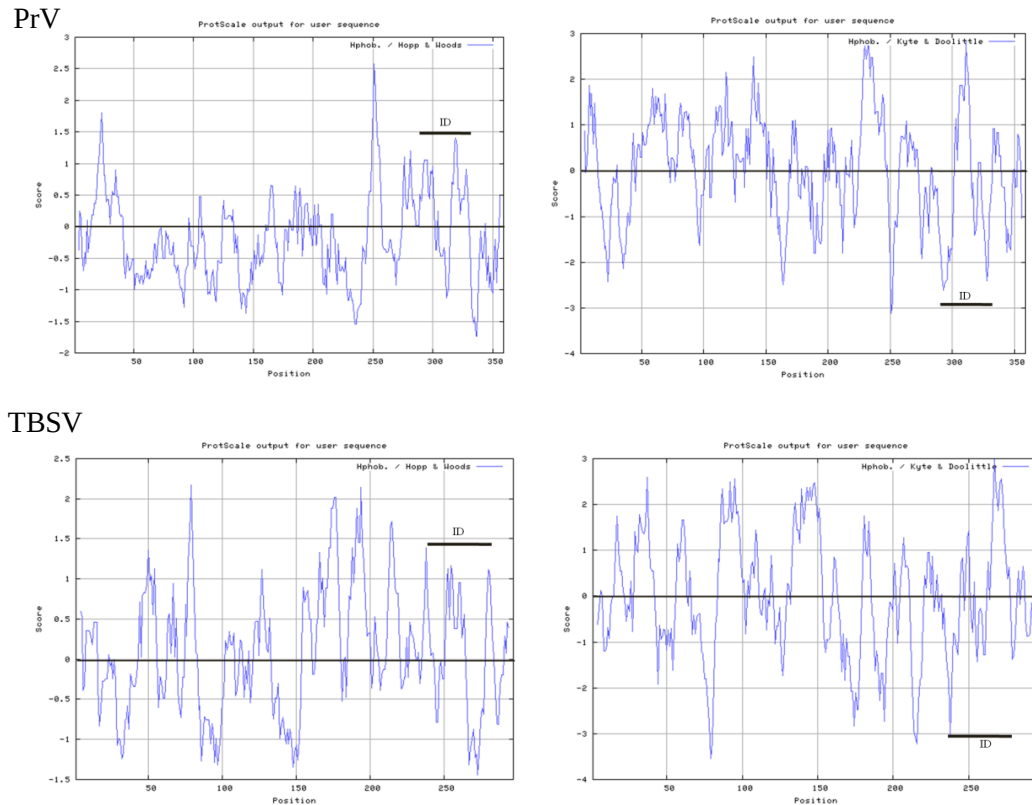


Figure 5. 5: Comparing hydrophilic and surface exposed properties of the interaction domain of PrV p40 and TBSV p33. The Hopp and Woods plot (window size 5) showing hydrophilic regions. The Kyte and Doolittle analysis plot (window size 5) showing surface exposed regions. **Top, Left.** PrV p40 interaction domain denoted as ID and underlined with a black line. **Top, Right.** PrV p40 interaction domain denoted as ID and underlined with a black line. **Below, Left.** TBSV p33 interaction domain denoted as ID and underlined with a black line. **Below, Right.** TBSV p33 interaction domain denoted as ID and underlined with a black line. Black line at 0 represents the threshold value. Hydrophilic regions are above 0 with Hopp and Woods plot while surface exposed regions are below 0 with Kyte and Doolittle plot. Y axis represents scores while X axis represents position of the amino acid sequence.

Since the hydrophilic and surface exposed properties are comparable in both PrV p40 (292 – 332) and TBSV p33 ID and that both PrV p40 (292 – 332) and TBSV p33 ID (241 – 282) are located at the C-terminus of the accessory proteins, the next step was to identify amino acids conserved in both regions. Sequence alignment of the surface exposed region of p40 (292 - 332) and TBSV p33 ID (241 – 282) using ClustalW revealed 13 % similarity of conserved amino acids in both protein sequences (Figure 5.6). A number of polar amino acids with similar charge i.e. positive or negative charge or uncharged amino acids were conserved in both TBSV p33 ID and PrV p40 (292 - 332). Based on the similarity of hydrophilic and surface exposed characteristics, location at the C-terminus of the accessory proteins and amino acid conservation, PrV p40 (292 – 332; referred to as putative PrV p40 ID hereafter) may be a potential interaction domain.

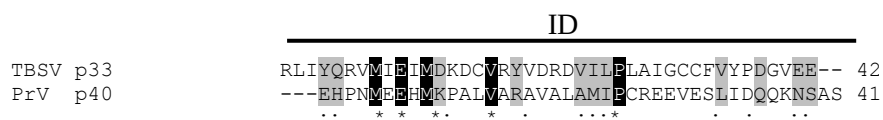


Figure 5. 6: Sequence alignment of PrV p40 and TBSV p33 interaction domain. Sequence alignment of the TBSV p33 interaction domain (ID) and PrV p40 surface exposed sequence. Conserved residues are shaded black and denoted as an asterisk (*). Grey shaded regions denoted by a colon (:) represent conserved amino acids with same charge. The parameters for the ClustalW programme were set at 10, 0.2 and 5 for the gap open, gap extension and gap distance respectively. The positions of the amino acids are shown on the right. Providence virus (PRV, YP_003620397), Tomato bushy stunt virus (TBSV, NP_062897).

5.3.3 Potential transmembrane domains and topology of p104

Generally (+ve) ss RNA virus replication is associated with membranes (Denison, 2008), a feature also found for HaSV replicase (Short and Dorrington, 2012). Data generated in this study (Chapter 2, section 2.3.2 and Chapter 3, section 3.3.3) provides evidence that PrV p104 might also be membrane associated and that the domain that is responsible for interacting with target membranes may be present within the p104 CDS. The TMPred program was used to identify putative transmembrane (TM) domain (Figure 5.7A). The returned output of the analysis displayed four distinct peaks at amino acid positions 52 - 72 (TM1), 225 - 247

(TM2), 437 - 456 (TM3) and 647 - 661 (TM4) with scores of 1101, 1949, 1213 and 1638 respectively (Figure 5.7A).

To verify results generated by TMPred, the Kyte and Doolittle programme was used to identify putative transmembrane domains within p104. The analysis of p104 displayed three distinct peaks at amino acids 225 - 247 (TM2), 433 - 456 (TM3) and 647 - 661 (TM4) with scores of 1.989, 1.642 and 1.779 respectively (Figure 5.7B). The putative TM1 was not detected by Kyte and Doolittle programme reflecting differences between algorithms. Overall, the peaks that represent TM2, TM3 and TM4 are well conserved and occupy the same position in both TMPred programme and Kyte and Doolittle programme. This analysis may suggest that p104 is a membrane protein and the putative transmembrane helices traverse the lipid membrane. Further examination of p104 using Kyte and Doolittle programme revealed that the hydrophobic regions (above zero) are intercepted with intervening hydrophilic regions (below zero) (Figure 5.7B). The hydrophilic regions are likely to be exposed to the cytoplasmic environment while the hydrophobic amino acids are buried inside protein. In addition, analysis of p104 sequence with Kyte and Doolittle programme mapped the putative PrV ID, RBR 1 and RBR 2 within the hydrophilic regions (Figure 5.7B) and which means that these domains are located in the cytosol.

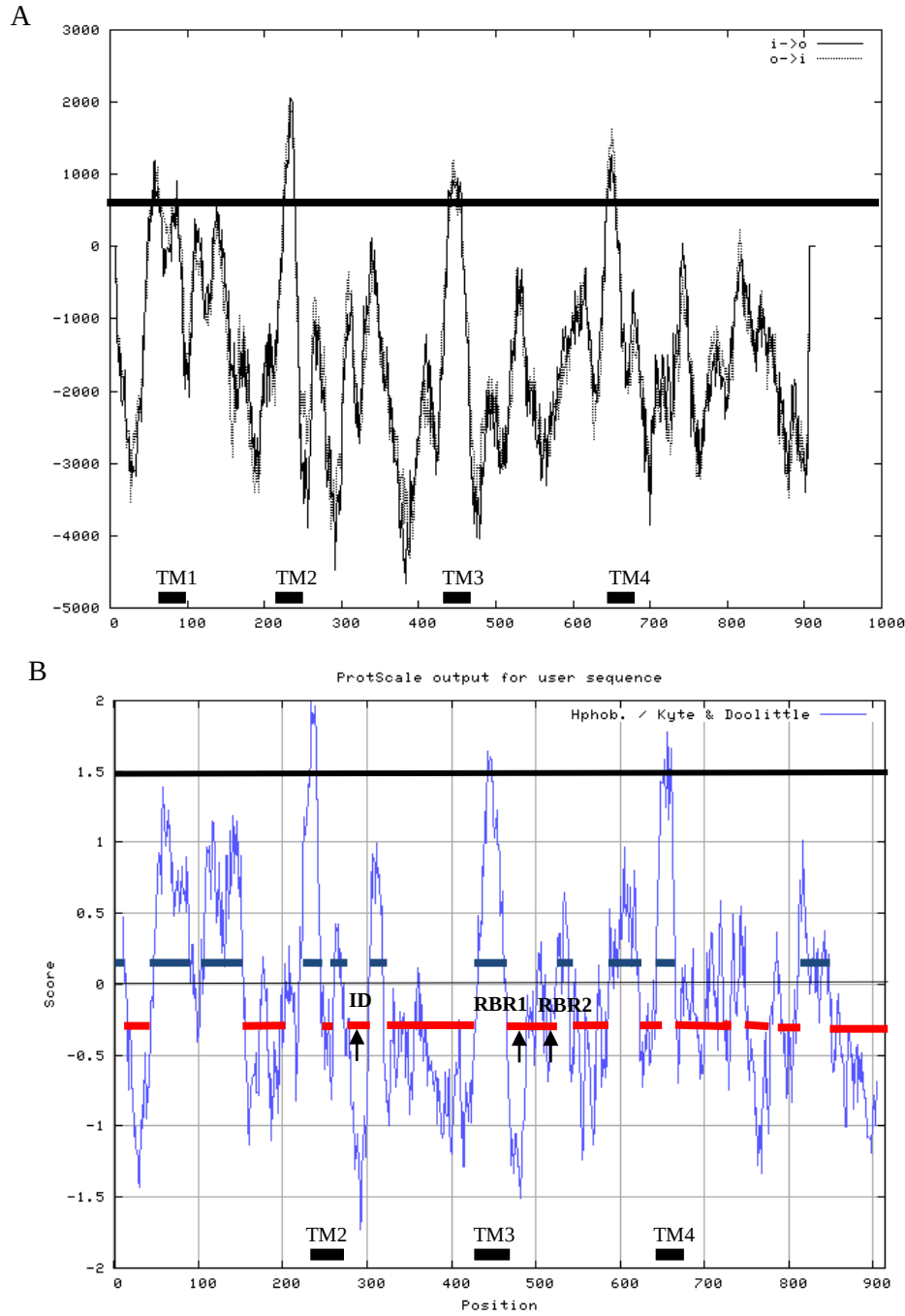


Figure 5. 7: Identification of potential transmembrane domains in PrV p104. A. Hydrophobicity plot generated by TMpred with scores above 500 considered statistically significant for transmembrane domain prediction. **B.** Kyte-Doolittle hydropathy plot using a window scale of 19. The black thick line at a score of 1.5 represents the threshold value for predicting transmembrane helices. The transmembrane domains; TM1 (generated by TMpred), TM2, TM3 and TM4 (generated by TMpred and Kyte-Doolittle) are shown as black bars. In the Kyte-Doolittle plot, regions rich in hydrophobic amino acids are indicated as blue bars above the threshold score of 0 while hydrophilic regions are indicated in red bars below the 0. The interaction domain (ID), RNA binding sites (RBR 1 and RBR 2) are shown with arrows.

The antigenic sites within p104 were examined by Hopp and Woods programme. Thirteen putative antigenic sites were identified as sequences with a high likelihood of being exposed in the cytosol and therefore potentially available for antibody recognition (Figure 5.8). The antigenic site between amino acids 248 and 331 and which covers the regions designated I, II, III is the region that was used to generate anti-p40₍₂₄₈₋₃₃₁₎ antiserum while 888 and 906 amino acids (designated IV) were used to generate anti-p104₍₈₈₉₋₉₀₂₎ antiserum (Figure 5.8). This means that the region that contains amino acids 248 to 331 used for the production of anti-p40₍₂₄₈₋₃₃₁₎ antiserum and 888 to 906 used for the production of anti-p104₍₈₈₉₋₉₀₂₎ antiserum must be in the cytosol. All the thirteen putative antigenic sites lie within the hydrophilic regions of p104 (Figure 5.9) and it is possible that these sites are exposed to the cytoplasm.

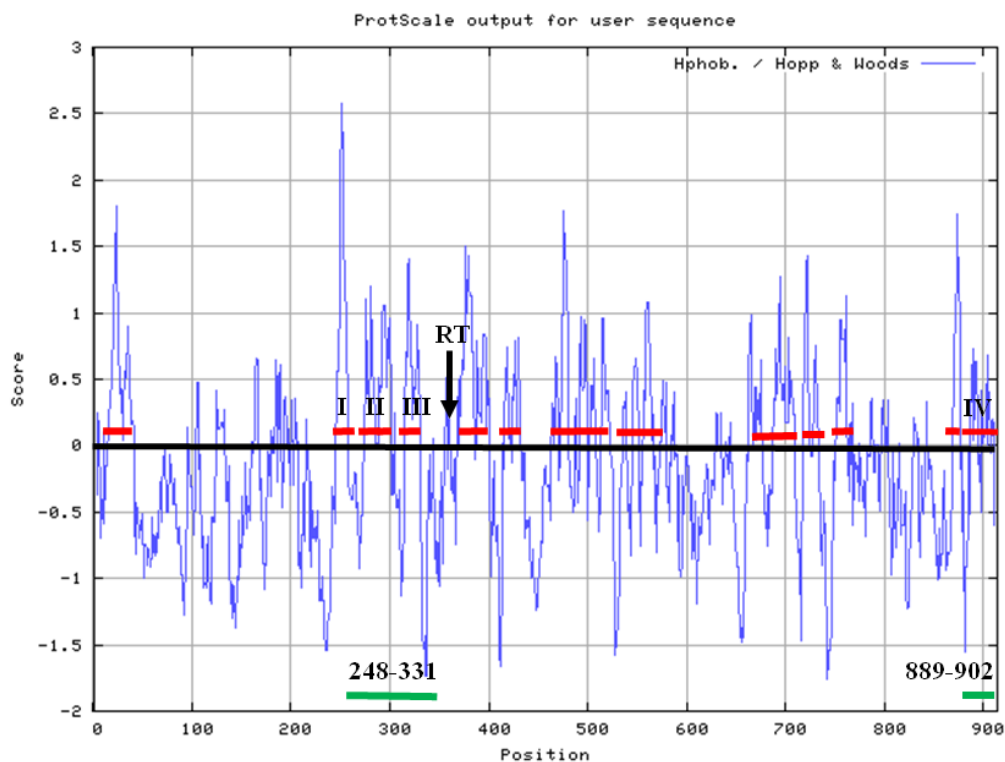


Figure 5. 8: Identification of the antigenic regions selected for antibody production in relation to the hydrophobic profile of PrV p104. Hopp and woods plot using a window size of 5 for the identification of potential antigenic sites. Putative antigenic sequences are shaded red. The green bars above the x-axis and region I, II, III and IV represent sequences used to raise anti-p40₍₂₄₈₋₃₃₁₎ antiserum and anti-p104₍₈₈₉₋₉₀₂₎ antiserum. The position of the readthrough stop codon is denoted as RT and a thick black arrow. The black line at 0 represents the threshold value with hydrophilic regions > 0 and hydrophobic regions < 0.

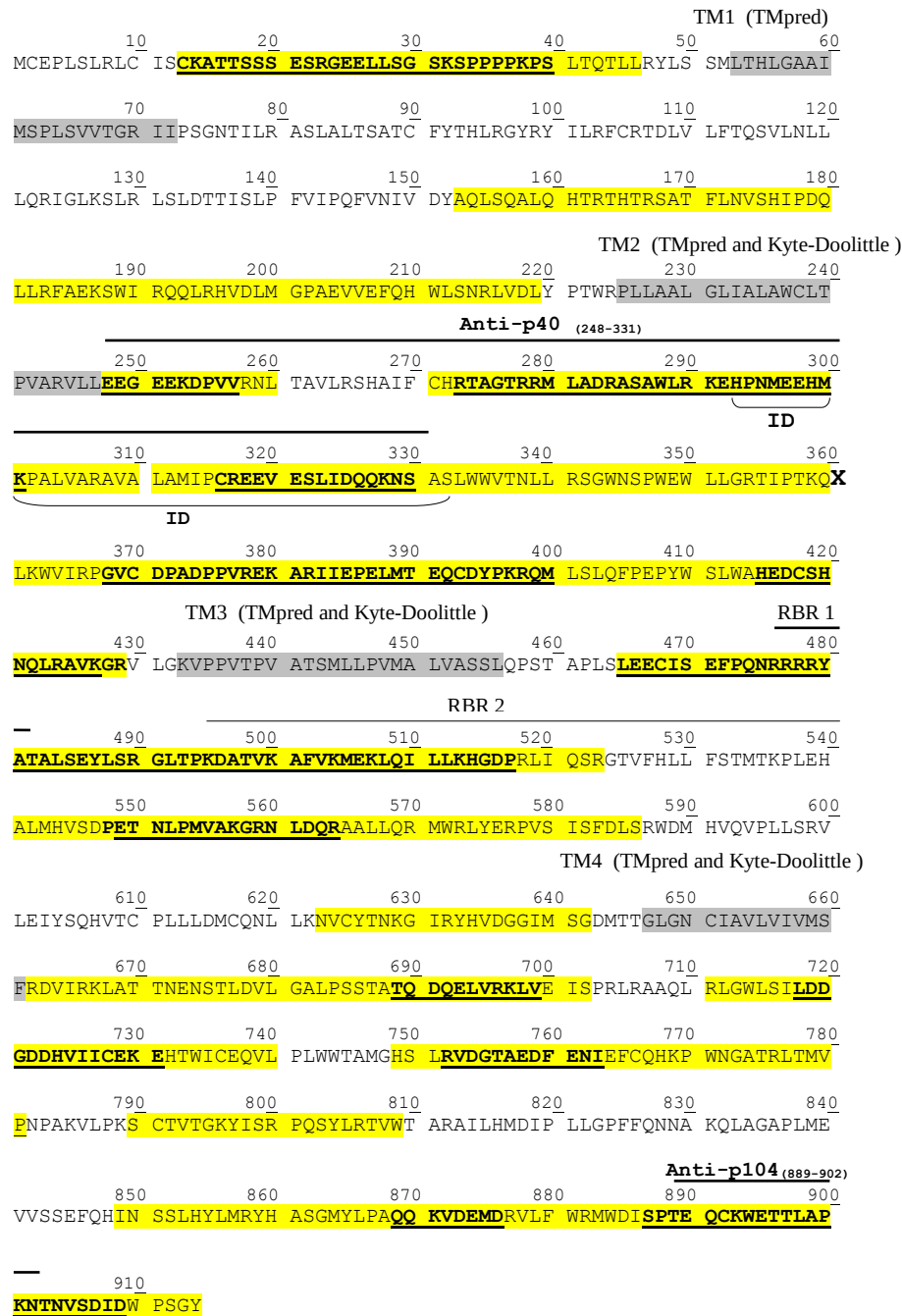


Figure 5. 9: The p104 sequence showing potential domains. Transmembrane domains are shaded grey and denoted as TM1 (predicted by TMpred), TM2, TM3 and TM4 (predicted by TMpred and Kyte-Doolittle). Hydrophilic sequences are shaded yellow. Antigenic amino acids are in bold, underlined and within the yellow shaded region. Hydrophobic regions are not shaded. Anti-p40₍₂₄₈₋₃₃₁₎ and anti-p104₍₈₈₉₋₉₀₂₎ refer to regions used to generate antiserum. ID refers to putative interaction domain. RBR 1 and RBR 2 refer to RNA binding sites. The readthrough stop is indicated with the cross (x) sign. The positions of the amino acids are above the sequence.

Three possible models for the transmembrane topology of p104 have been considered based on the number of transmembrane domains predicted by TMpred programme and Kyte and Doolittle programme. The topology of p104 based on the four transmembrane domains identified by TMpred (Appendix B.2) and three by TMpred programme and Kyte and Doolittle programme (Appendix B.3) were eliminated because the RdRp has to be on the cytoplasmic face of the membrane in these topology models.

The two transmembrane model for p104 was generated based on two putative N-terminal transmembrane domains upstream of the readthrough stop codon, TM1 (52 - 72, identified by TMpred) and TM2 (225 - 247, identified by both TMpred and Kyte and Doolittle). The two transmembrane domains would traverse the hypothetical lipid membrane twice and this would ensure that the C-terminal region which also consists of the RdRp domain (505 - 768), readthrough stop signal, putative PrV p104 ID, RBR 1 and RBR 2 are positioned within the cytoplasm (Figure 5.10A). This model allows the RdRp to be exposed to the cytoplasmic face of the membrane. In addition, the putative antigenic sites including those used to generate anti-p40₍₂₄₈₋₃₃₁₎ and anti-p104₍₈₈₉₋₉₀₂₎ antiserum would be located within the cytoplasm. The predicted topology of PrV p104 was compared with the TBSV p33 topology based on two transmembrane domains. The putative transmembrane domains in TBSV p33 were predicted using TMHMM and aligned with Cymbidium ringspot virus transmembrane domains (McCartney *et al.*, 2005). The N-terminus transmembrane domain of TBSV p33 at amino acid positions 83 - 98 (TM1) and 132 - 154 (TM2) allow the RdRp domain, readthrough stop codon, ID, RBR 1, RBR 2, RBR 3 and peroxisome targeting signal (1 – 81) to be positioned within the cytoplasm (Figure 5.10 B). Identification of potential localization signals using SignalP within p104 was unsuccessful but the similarity of the PrV p104 topology model with the TBSV p33 model might suggest that the cytosolic amino acid position between 1 and 52 of p104 might contain membrane targeting signals.

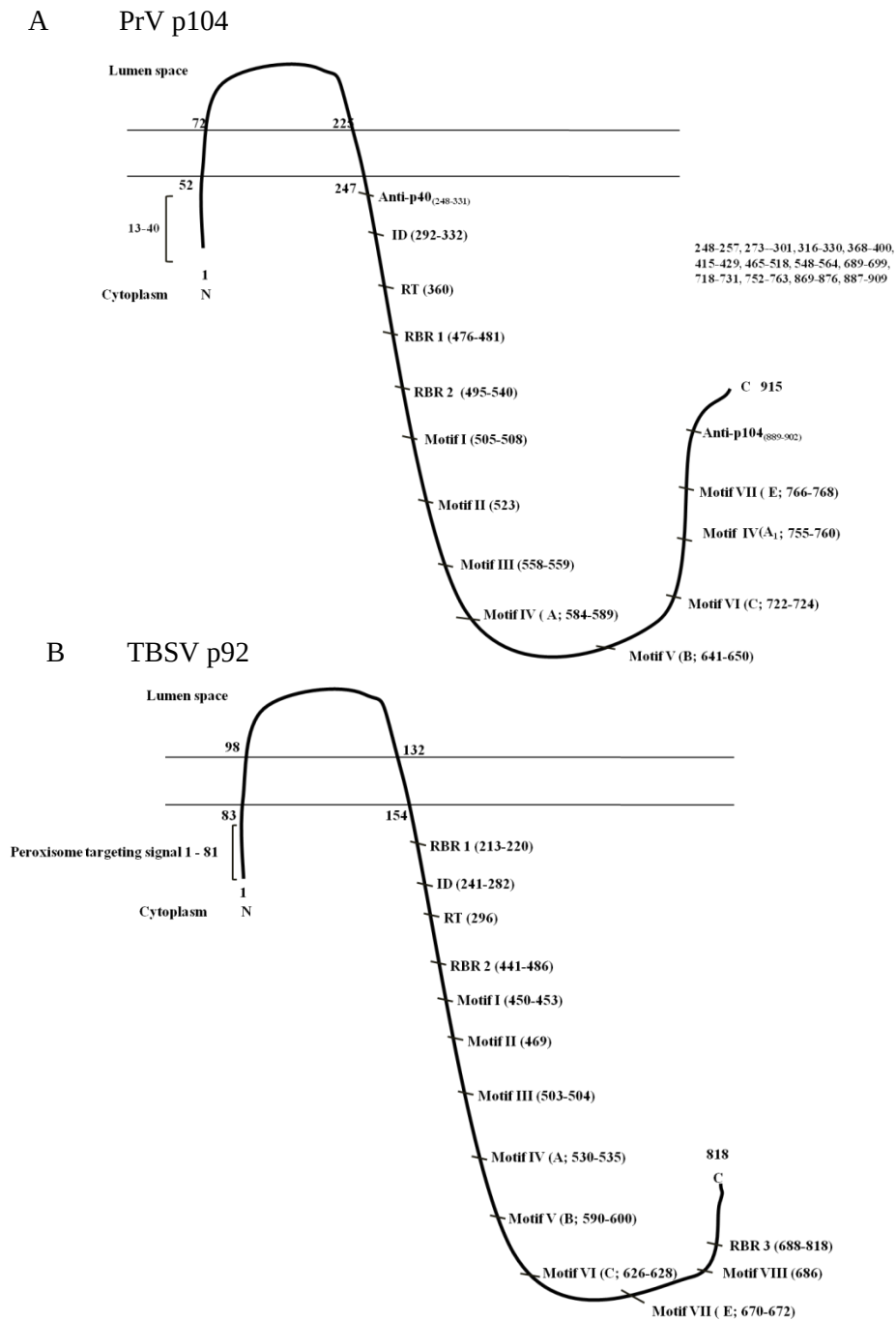


Figure 5. 10: The two transmembrane model of PrV p104 and TBSV p92. **A.** PrV p104. **B.** TBSV p92. The amino acid stretch of each transmembrane helice is annotated at either side of the lipid bilayer represented by a black curve. The RdRp motifs are labeled on the sequence, each with its amino acid position relative to PrV p104 and TBSV p92. Anti-p40 (248-331) and anti-p104 (889-902) refer to regions used to generate antiserum. ID refers to interaction domain. RBR 1 and RBR 2 reffer to RNA binding sites. Peroxisome targeting signal of TBSV p92 is shown on the left of the lipid bilayer. The readthrough stop codon is labeled as RT. Positions of the putative antigenic sites of p104 are indicated on the sides of the lipid bilayer.

5.3.4 Potential phosphorylation sites within p104

The importance of phosphorylation of non-structural viral proteins has been demonstrated in tombusvirus replication. Phosphorylation has been shown to promote efficient binding of the replication proteins to viral RNA and the release of viral RNA from the replication proteins during the replication cycle of TBSV (Shapka *et al.*, 2005; Nagy and Pogany, 2006). The sequence conservation between the RdRp domains of PrV p104 and TBSV replicase raised the possibility that phosphorylation may play a similar role in PrV replication so the NetPhos 2 program was used to identify potential phosphorylation sites within PrV p104 and these results were compared with putative phosphorylation sites in the TBSV replicase.

A total of ten putative serine phosphorylation sites (with a probability of > 0.97 of being phosphorylated in eukaryotic cellular environment) that span the entire p104 sequence, were identified in p104 at amino acid positions 18, 20, 22, 322, 464, 580, 586, 803, 887 and 906 (Figure 5. 11). Three of serine phosphorylation sites at amino acid position 18, 20 and 22 were concentrated at the N-terminus of the p40/p104 coding sequence (Figure 5.11). In addition, a putative threonine phosphorylation site at amino acid position 493 was found adjacent to p104 RBR 2 (Figure 5. 11). As expected, given the sequence conservation of the RdRp regions of the two replicases, the S586 in PrV p104 and S531 in TBSV p92 are conserved within motif A of the RdRp of both replicases (Figure 5.11).

Previous reports have shown that phosphorylation of serine and threonine residues adjacent to the RNA binding site in TBSV p33 are important for viral RNA replication (Shapka *et al.*, 2005). Phosphorylation of the TBSV p33 serine residue at position 210 and threonine at positions 205 and 211 located adjacent to the RBR 1 abolished RNA binding and the eventual release of viral RNA during replication. When compared to PrV p104, a serine residue at position 464 was found adjacent to the putative PrV RBR 1 (Figure 5.11). In addition, a threonine at position 493 was found adjacent to the RBR 2 region of PrV p104. The presence of potential phosphorylation sites adjacent to p104 putative RNA binding sites may suggest that PrV RNA binding sites are likely to be present and are susceptible to phosphorylation however this adds motivation for experimentation verification.

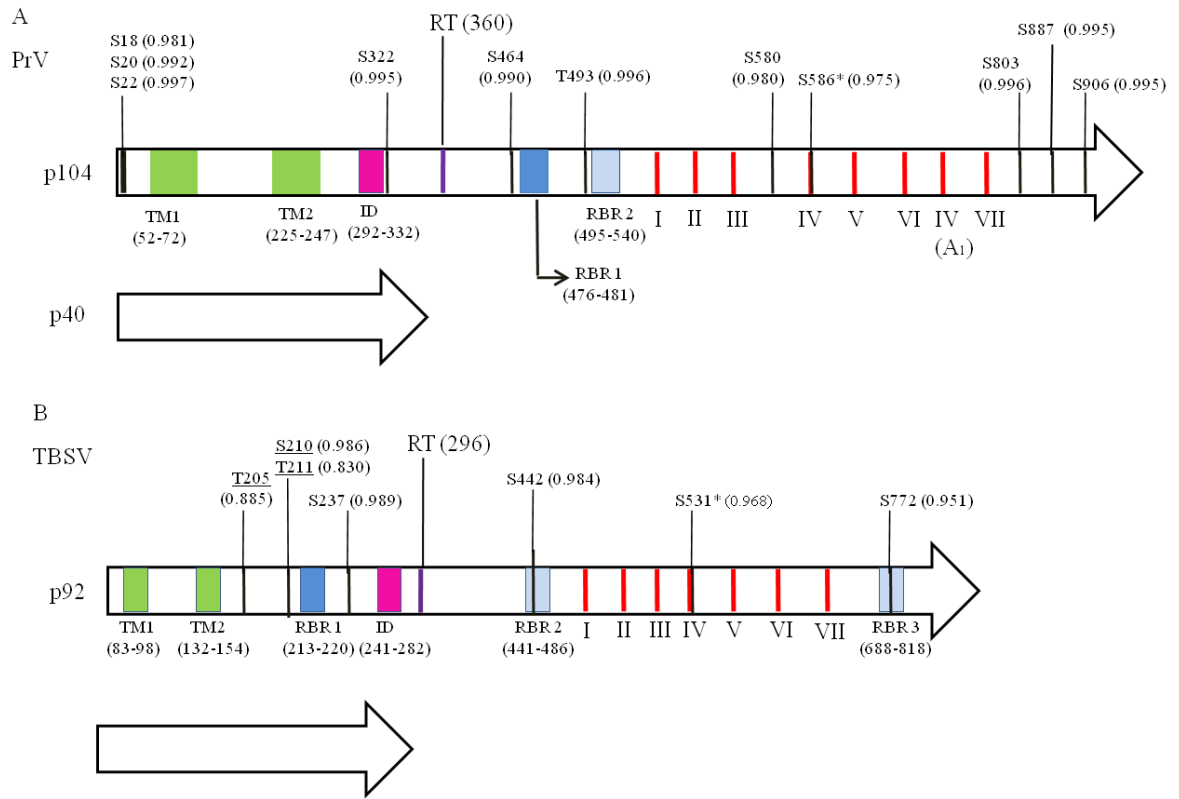


Figure 5. 11: Schematic map showing potential phosphorylation sites in PrV and TBSV replicase. A. Putative serine (S) and Threonine (T) phosphorylation sites within PrV p104. **B.** Putative serine (S) and Threonine (T) phosphorylation sites within TBSV p92. Underlined S and T residues are phosphorylated in TBSV. Putative phosphorylation sites with star (*) are conserved in motif A of PrV and TBSV replicases. Each phosphorylation site is annotated by a single amino acid code letter which stands for the phosphorylated residue and its amino acid number relative to the amino acid sequence. The statistical scores of the phosphorylation sites are bracketed. TM1 and TM2 refer to transmembrane domains and shaded as green boxes. The RdRp motifs within PrV p104 and TBSV p92 sequence are labeled as red bars. RT represents readthrough stop codon with a purple bar. RBR 1 (Blue bar), RBR 2 and RBR 3 (light blue bars) refer to RNA binding region 1, 2 and 3 respectively. ID refers to interaction domain and shown as a pink bar. Amino acid positions of domains are shown below.

5.3.5 Identification of putative motifs and domains within p130

The phylogenetic relatedness of the PrV RdRp with those of the tombusviruses and umbraviruses led to the search for the existence of any other relevant motifs or domains that may be conserved within p130 and the non-structural proteins of

umbraviruses and tombusviruses. Apart from the RdRp, other non-structural proteins encoded by the umbravirus genome are p27 (ORF3) and p28 (ORF4) (Ryabov *et al.*, 2011) while p22 (ORF3) and p19 (ORF4) are encoded by tombusviruses (Rochon *et al.*, 2011).

Within p130 sequence is an arginine-rich region (RRRRRPRDNLR) at amino acids 1047 – 1057 that resembles the arginine-rich domain (RPRRRAGRSGGMDPR) at amino acid positions 108 – 122 of umbravirus GRV p27 that binds RNA (Taliensky *et al.*, 2003) and transports viral RNA through the phloem (Kim *et al.*, 2007). Substitution of the arginine residues, R₁₀₈ R₁₁₀ R₁₁₁ R₁₁₂ R₁₁₅ R₁₂₂ to alanine abolished the function of p27 as an RNA binding protein (Kim *et al.*, 2007). The similarity in amino acid composition of the p130 arginine-rich domain with GRV p27 arginine-rich domain may suggest that the p130 arginine-rich domain is a potential RNA binding site. Since the p130 arginine-rich domain is also similar to the arginine/proline-rich RBR 1 of TBSV p33, it was decided that the hydrophilic and surface exposed properties of p130 arginine-rich domain are investigated. The Hopp and Woods programme displayed a strong peak with a score of 2.571 (above 0) at amino acids 1047 – 1057 (Figure 5.12A). In addition, the Kyte and Doolittle programme revealed a strong peak with a score of -4.086 (below 0) at amino acids 1047 – 1057 (Figure 5.12B). This analysis suggests that the p130 arginine-rich region lies within the most hydrophilic region of the protein and that it is surface exposed. The hydrophilic and surface exposed properties of p130 are similar to findings reported for TBSV RBR 1 (Rajendran and Nagy, 2003) and this may possibly suggest that p130 arginine-rich region is likely to be an RNA binding site (referred to as RNA binding domain; RBD hereafter).

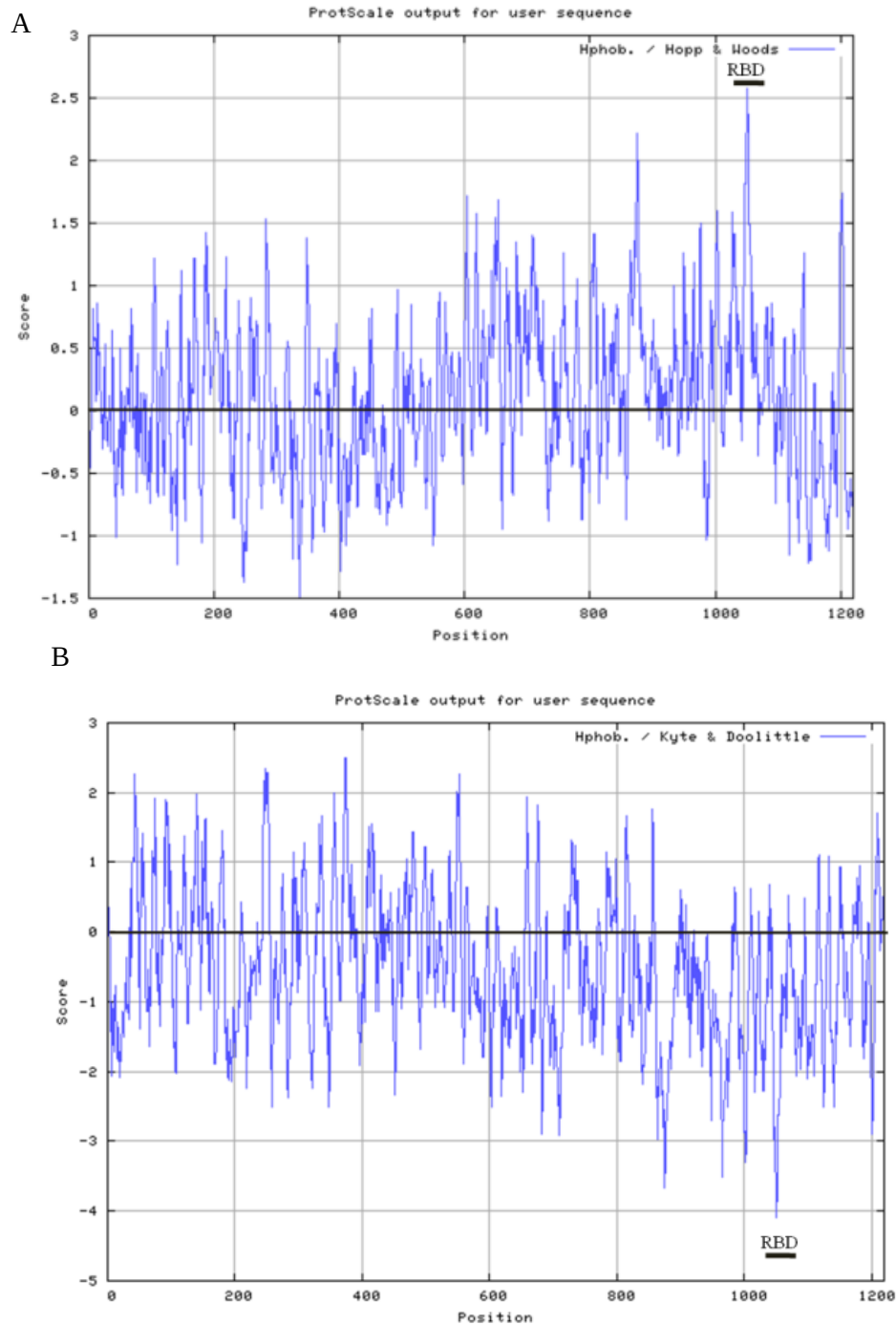
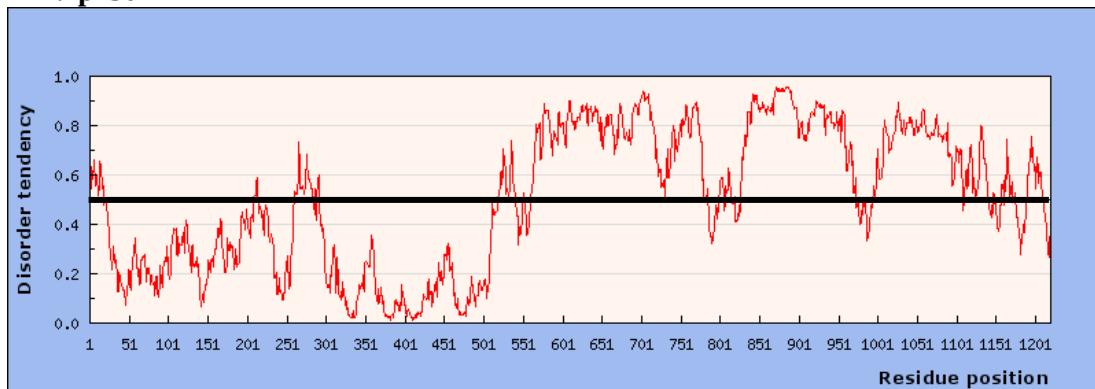


Figure 5. 12: Identification of the surface exposed properties of the potential PrV p130 RNA binding site. **A.** The Hopp and Woods plot using window size 5 to predict most hydrophilic regions. **B.** The Kyte and Doolittle plot using window size 5 to predict surface exposed regions. Potential hydrophilic and surface exposed RNA binding domain is denoted as RBD. Black line at 0 represents the threshold. In the Hopp and Woods plot, regions above 0 are hydrophilic. In the Kyte and Doolittle plot, regions below 0 are surface exposed regions. Y axis represents scores while X axis represents position of the amino acids within p104 sequence.

Further analysis of the p130 sequence using the IUPRED programme identified a putative disordered region with a score of 0.9 (above the threshold value 0.5 for statistically significant regions). The analysis suggested that p130 may consist of a disordered C-terminal domain that spans amino acids 517 and 1220 (Figure 5.13A). The secondary structure of the disordered C-terminal region was further examined with Predict protein programme and the analysis revealed that disordered region has mainly turns which make up 74 % of the secondary structure composition and which would result in a flexible and unstable region. For comparison, the GRV p27, a long distance movement protein, was analysed by IUPRED programme and the analysis showed that the entire sequence was disordered with a score of 0.9 above the threshold value 0.5 (Figure 5.13B). The predicted secondary structure of GRV p27 using Predict protein programme revealed that the protein is composed mainly of turn structures that account for 77% of the total secondary structure composition. GRV p27 is an RNA chaperone (Taliensky *et al.*, 2003) and Tompa and Csermely, (2004) reported that proteins that display RNA chaperone activity possess disordered regions which interact and loosen the misfolded RNA structure without use of ATP. The identification of putative disordered regions within PrV p130 and GRV p27 may indicate a conserved feature that might have an important role in RNA chaperone activity.

PrV p130



GRV p27

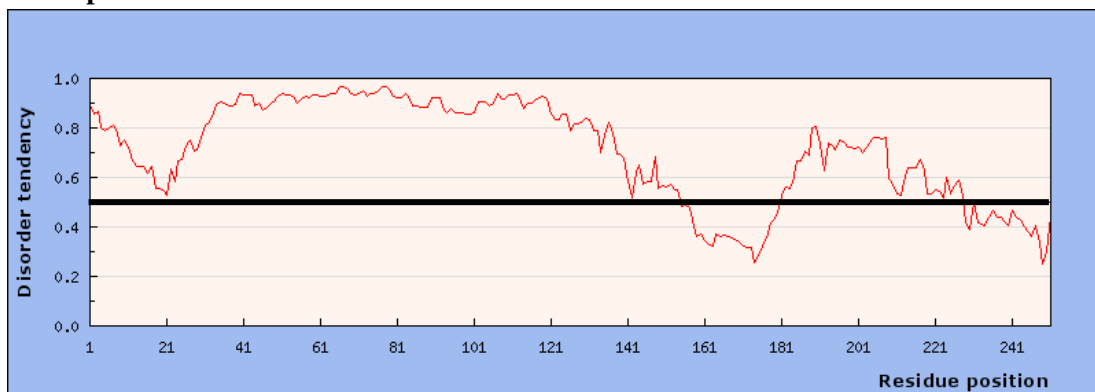


Figure 5. 13: Identification of potential disordered regions within PrV p130 and GRV p27. The plots created by IUPRED show scores above or below the threshold value 0.5 (indicated as black line). Regions above 0.5 are disordered while regions below 0.5 are ordered. Y axis represents scores while X axis represents position of the amino acids within the sequence.

Apart from the putative arginine-rich RBD and disordered domain, a putative 2A-like processing site is present within p130 sequence (Walter *et al.*, 2010) (Figure 5.14). This motif has not been reported in GRV and TBSV. Therefore PrV differs from umbraviruses and tombusviruses in that it uses a different mechanism to regulate gene expression.

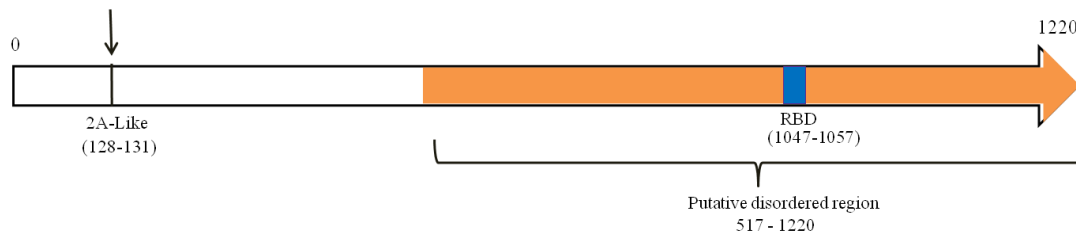


Figure 5. 14: Schematic representation of putative domains within PrV p130. The orange shaded region represents the disordered region. The blue shaded region represents the RNA binding domain (RBD). The arrow and black line represents the 2A-like site. Numbers represent amino acid position of sequence (above) or motifs and domain (below).

5.4 DISCUSSION

The bioinformatic analysis of PrV non-structural proteins has revealed several putative sequence motifs that may be important in the function of p104 and p130 in the PrV life cycle. Within the p104 sequence, two putative transmembrane helices, an arginine-rich RBR 1 site, RBR 2 site, interaction domain (ID) and RdRp domain were identified and these are similar to tombusviruses (Figure 5.15). Within the p130 sequence is a putative arginine-rich RNA binding domain and a disordered region within the C-terminus (Figure 5.15).

Sequence comparison of PrV p104 and replicases belonging to members of the tombusvirus and umbravirus revealed amino acid conservation in the RdRp domain raising the possibility that PrV RdRp is of a plant virus origin. The putative PrV RBR 1 shares amino acids, hydrophilic and surface exposed properties with TBSV p33 RBR 1 and could be a potential RNA binding site. In addition, the putative PrV p104 RBR 2 shows significant amino acid conservation with the TBSV p92 RBR 2 suggesting that the putative RBR 2 in PrV p104 may possibly be involved in viral RNA binding. The presence of the phosphorylation sites adjacent to the PrV RBR 1 and RBR 2 sequence may add some support for the presence of the RNA binding sites within p104. There was no significant amino acid similarity between the RBR 3 domain of TBSV p92 and that of PrV p104 raising the possibility that the putative RBR 3 domain in p104 may be required for interaction with RNA structural elements to specific to PrV.

The putative PrV p40 ID showed sequence conservation with TBSV ID. The hydrophilic and surface exposed properties of PrV p40 ID are similar to TBSV p33 ID. In both proteins, the ID is located at the C-terminus. There's a possibility that the putative PrV p40 ID is a potential protein-protein interaction domain.

The two based transmembrane domain model of p104 suggests that the transmembranes domains at 1 – 52 and 225 – 247 may serve to anchor the replicase to target membranes. The two transmembrane topology model of p104 positions the RdRp domain, readthrough stop codon, ID, RBR 1 and RBR 2 within the cytoplasmic environment and which is similar to the TBSV p92 topology model. Similar to p104, p40 also contains putative transmembrane domains, a possibility that both proteins intergrade in the membrane.

The PrV p130 arginine-rich region is similar to the arginine-rich region of GRV p27 whose function is to bind viral RNA and mediate long distance movement through the phloem. The hydrophilic and surface exposed properties of p130 arginine-rich region are similar those belonging to TBSV p33 arginine-rich RBR 1. Such similarities may possibly suggest that the arginine-rich region within p130 is a potential RNA binding site. The C-terminal region of PrV p130 contains a putative disordered region that is similar to GRV p27 and which might have an important function in both proteins.

Lastly, the p130 does not contain sequences similar to the nuclear localizing signal (NLS) and nuclear export signal (NES) found in GRV p27 (Figure 5.15) and responsible for targeting to the nucleolus and exportation back to the cytoplasm (Ryabov *et al.*, 2004; Kim *et al.*, 2007). Therefore, p130 is likely to use different signals for subcellular localization.

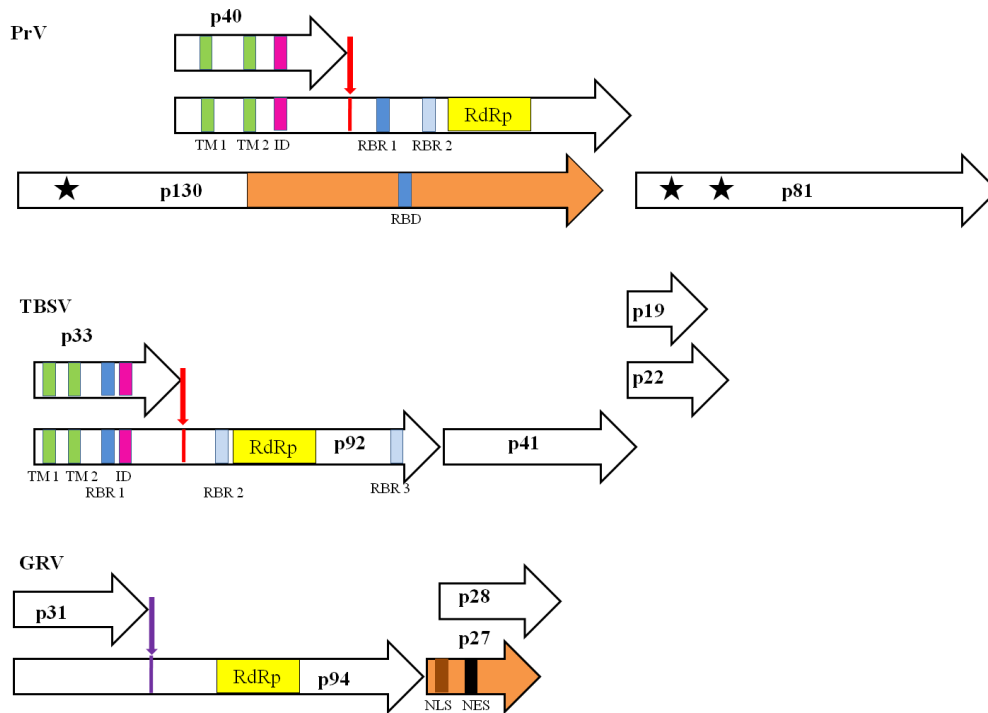


Figure 5. 15: Summary of domains and motifs found within ORFs encoded by PrV, GRV and TBSV genomes. Transmembrane helices (■), arginine- rich RBR 1 and RBD (■), RBR 2 and RBR 3 (■), RdRp domain (■), interaction domain (■), disordered region (■), NLS (■), NES (■), Readthrough stop codon (■), -1 ribosome frameshift sequence (■). 2A-like sites (★). The translated protein of each ORF is indicated inside the open arrow while the molecular masses are represented as the value “p” that precedes the protein. PrV-*Providencia virus*, GRV-*Groundnut rosette virus*, TBSV-*Tomato bushy stunt virus*

In conclusion, data from this chapter has shown that PrV non-structural proteins have potential motifs and domains that are similar to functional domains found in tombusviruses and umbraviruses. These sequences may have important functions in the replication of PrV. The similarity in the motifs and domains within the non-structural proteins of PrV, tombusviruses and umbraviruses suggest that PrV is closely to tombusviruses and umbraviruses.

CHAPTER 6

GENERAL DISCUSSION AND CONCLUSIONS

Studies into the replication biology of tetraviruses have been hampered by the lack of a suitable cell line that can support tetravirus replication. The discovery of Providence virus, a unique tetravirus that can replicate in cell culture (Pringle *et al.*, 2003), provided for the first time an experimental system for studying fundamental aspects of tetravirus replication. Prior to this project, research focused on the structural biology and genome organization of PrV and the mechanism by which the virus replicates was yet to be elucidated. Therefore the principal aim of the research described in this thesis was to study virus replication, focusing on the subcellular localization and potential interactions between the PrV replication proteins. Additionally, bioinformatic analysis was used to identify putative functional domains.

6.1 SUBCELLULAR LOCALIZATION OF THE SITE OF PrV REPLICATION

As a first step towards understanding the viral replication cycle, the subcellular localization of Providence virus replication proteins and vRNA was investigated in virus-infected MG8 cells. Cells stained with antibodies against p40 (and by inference also p104) and dsRNA showed co-localization of p40 and viral RNA in punctate structures within the cytoplasm of infected cells, which are likely the sites of viral replication. In the absence of an antibody able to distinguish between p40 and p104, it was not possible to confirm that p104 was also present in these punctate structures, but this was presumed to be the case since the RdRp domain required for viral RNA replication is located in the C-terminal half of p104. Studies with EGFP-p40 and EGFP-p104 fusion proteins in *S. frugiperda* Sf9 cells showed that both replication proteins localize independently to these punctate structures, which are now known to be targeted to the Golgi apparatus and secretory pathway (Short *et al.*,

2013). Short *et al.* (2013) suggested that p40 and p104 are targeted to the site of replication on membranes derived from the Golgi and/or secretory vesicles. The presence of putative transmembrane domains TM1 (52 - 72) and TM2 (225 - 247) within p40 suggests that PrV p40 and p104 are likely to anchor the viral replication complex to target membranes belonging to the Golgi apparatus and secretory vesicles. In this study, attempts to identify putative localization signals within p104 sequence and using bioinformatic tools were unsuccessful, however the suggested two transmembrane topology model of p104 showed that the sequence between amino acids 1 and 52 lies within the cytosolic side of the membrane (Chapter 5), which is similar to the cytosol position of the peroxisome targeting sequences (1 – 81) of the TBSV p92 topology model. It is possible that the localization signals of PrV replication proteins are at the N-terminus of p40 between amino acids 1 and 52 and therefore future work is required to verify amino acids within this region responsible for membrane targeting. In addition, the putative transmembrane helices should be tested to verify whether they are functional using deletion analysis or mutagenesis of the EGFP constructs.

In this study, immunoprecipitation studies demonstrated interaction between p104 and p40 and also that p40 self-interacts. Since p40 is expressed at approximately 10 fold higher levels than p104 (Walter *et al.*, 2010; this study) then the PrV replication complex is formed mostly through p40 self-interaction and the recruitment of p104. The potential domain responsible for this interaction is at the C-terminus of p40 between 292 and 332 and shows sequence conservation with the TBSV p33 interaction domain (ID) and has similar hydrophilic and surface exposed properties with TBSV p33 ID. It is recommended that further research be undertaken to verify whether the putative PrV p40 ID is functional.

6.2 A COMPARISON BETWEEN PRV, TOMBUSVIRUSES AND UMBRAVIRUSES

The replicase of PrV, tombusviruses and umbraviruses share similar features including sequence homology and expression strategies. Firstly, the RdRp domain of p104 consists of motifs that are conserved in tombusvirus and umbravirus replicases. Interestingly, motif VIII is present in tombusvirus and umbravirus replicases but absent in PrV while motif A is duplicated in PrV replicase and absent in tombusvirus

and umbravirus replicases. Putative RNA binding sites are present in p104 and share amino acids with TBSV p92 RNA binding sites. The number of RNA binding sites differ with two of these found within PrV p104 and three within TBSV p92. In addition, there are differences in the location of RBR 1 within PrV p104 and TBSV p92. In PrV, the RBR 1 is found at the N-terminus of p64 while in TBSV p92 the RBR 1 is located at the C-terminus of p33.

All three virus families utilise a readthrough stop codon (UAG) to regulate replicase translation. The 3' nucleotide sequence adjacent to the readthrough stop codon differs in PrV and tombusviruses. The Type I readthrough region in p104 has the sequence CAA CUA AAG while the Type III readthrough in TBSV (tombusvirus) replicase consists of the sequence GGA GGC CUA (Beier and Grimm, 2001). The readthrough stop codon (UAG) of PrV resembles that of the tombusvirus replicase. The presence of the readthrough stop codon results in expression of p40 and p104 in PrV (Walter *et al.*, 2010) and p33 and p92 in TBSV (Scholthof *et al.*, 1995b). The expression strategy of umbraviruses differs from that of PrV and tombusviruses in that a heptanucleotide sequence (AAAUUUU) associated with the -1 ribosome frameshift expression strategy is used to produced the replicase (Ryabov *et al.*, 2011).

Second, in terms of function the replication proteins of both PrV and tombusvirus form multiprotein complexes. In PrV p40 interacts with itself and with p104 while TBSV p33 interacts with other p33 molecules and with p92. Within the protein complex, p40 is more abundant than p104, an observation similar to the 10:1 ratio of p33:p92 found within the TBSV replication complex. Lastly, PrV and TBSV have two transmembrane domains located at the N-terminus of PrV p40 and TBSV p33.

Despite the similarities in their replicase, PrV, tombusviruses and umbraviruses differ in several ways. The subcellular localization site of PrV replication is the Golgi apparatus and secretory pathway (Short *et al.*, 2013), while that of TBSV is the peroxisomes (McCartney *et al.*, 2005). This implies that the targeting signals encoded on their respective replicases will be different and probably accounts for the lack of sequence conservation between p40 and TBSV p33.

Bioinformatic analysis of PrV p130 and umbravirus GRV p27 revealed domains that are similar within both proteins. The putative disordered region within the C-terminus of p130 is also found in GRV p27. In addition, the arginine-rich region within p130 resembles the arginine-rich region within GRV p27 that binds viral RNA. The p130 sequence differs from GRV p27 in that it contains a putative 2A-like processing site (Walter *et al.*, 2010) and which is absent in GRV p27.

6.3 A MODEL FOR THE REPLICATION OF PrV

TBSV is by far the best studied tombusvirus and serves as the model for understanding the replication biology of this virus family. The model developed for TBSV replication consists of six separate steps (Nagy and Pogany, 2006). These include entry and genome translation, formation of protein-protein complexes, selection of viral RNA template for replication, recruitment of the viral replication complex to the site of replication, synthesis of viral RNA progeny and release of the viral progeny from replication.

Once TBSV enters target host cell, the genome is released from the viral capsid and is translated into p33 and p92 replication proteins (Scholthof *et al.*, 1995b). The TBSV p33 forms protein-protein complexes with other p33 molecules and with p92 via an interaction domain located at the C-terminus of p33 (Rajendran and Nagy, 2004). The genome is then selected by p33 via the p33 recognition element (p33RE) an internal *cis*-acting stem loop element with a signature C.C mismatch in the p92 RdRp coding region of genomic RNA (Monkewich *et al.*, 2005; Pogany *et al.*, 2005). The viral replication complex (VRC) is then recruited to the peroxisomal membrane by p33 via the peroxisomal targeting signal (McCartney *et al.*, 2005). In addition, host factors such as eEF1 α (RNA binding protein) and Pex19p (cellular transport protein) have been implicated in the recruitment of the VRC to the peroxisomes (Li *et al.*, 2010). Recruitment of the VRC is further facilitated by the N-terminus transmembrane domains within p33 and which anchor the VRC to peroxisome membranes (McCartney *et al.*, 2005).

Prior to synthesis of the plus progeny viral RNA, p33 molecules bind cooperatively to viral RNA via the RBR 1 and this unwinds the secondary structure of genomic

RNA there by allowing the RdRp to bind the initiation promoter (gPR) promoter located at the 3' end of the plus strand RNA (Stork *et al.*, 2011). This results in the production of full-length minus strand complementary RNA which are used as template for the synthesis of full-length viral progeny RNA using the complementary promoter (cPR) (Panavas *et al.*, 2003). Release of the viral progeny RNA from the site of replication requires phosphorylation of threonine and serine residues adjacent to the RBR 1 in p33 (Shapka *et al.*, 2005; Nagy and Pogany, 2006). The released plus strand progeny RNA performs additional roles including a new cycle of translation, replication, viral RNA encapsidation and cell-to-cell movement of viral RNA. The phosphorylation of p33 and p92 is proposed to play a role in disassembly since non-phosphorylated mutants supported replication better at late time points than phosphorylation of competent wild-type virus (Shapka *et al.*, 2005).

In light of the results of this study, a preliminary model for the early stages of PrV replication has been developed based on data generated in this study together with the work reported by several authors. This model for PrV replication is built around the same six separate steps that describe the early stages of TBSV replication. These include entry and early translation of replication proteins, formation of protein-protein complexes, template selection, recruitment of the VRC to the site of replication, assembly of the VRC, minus/ plus strand RNA synthesis and release of plus stranded RNA progeny. The late stage of replication consists of further rounds of translation and replication of progeny RNA (Figure 6.1).

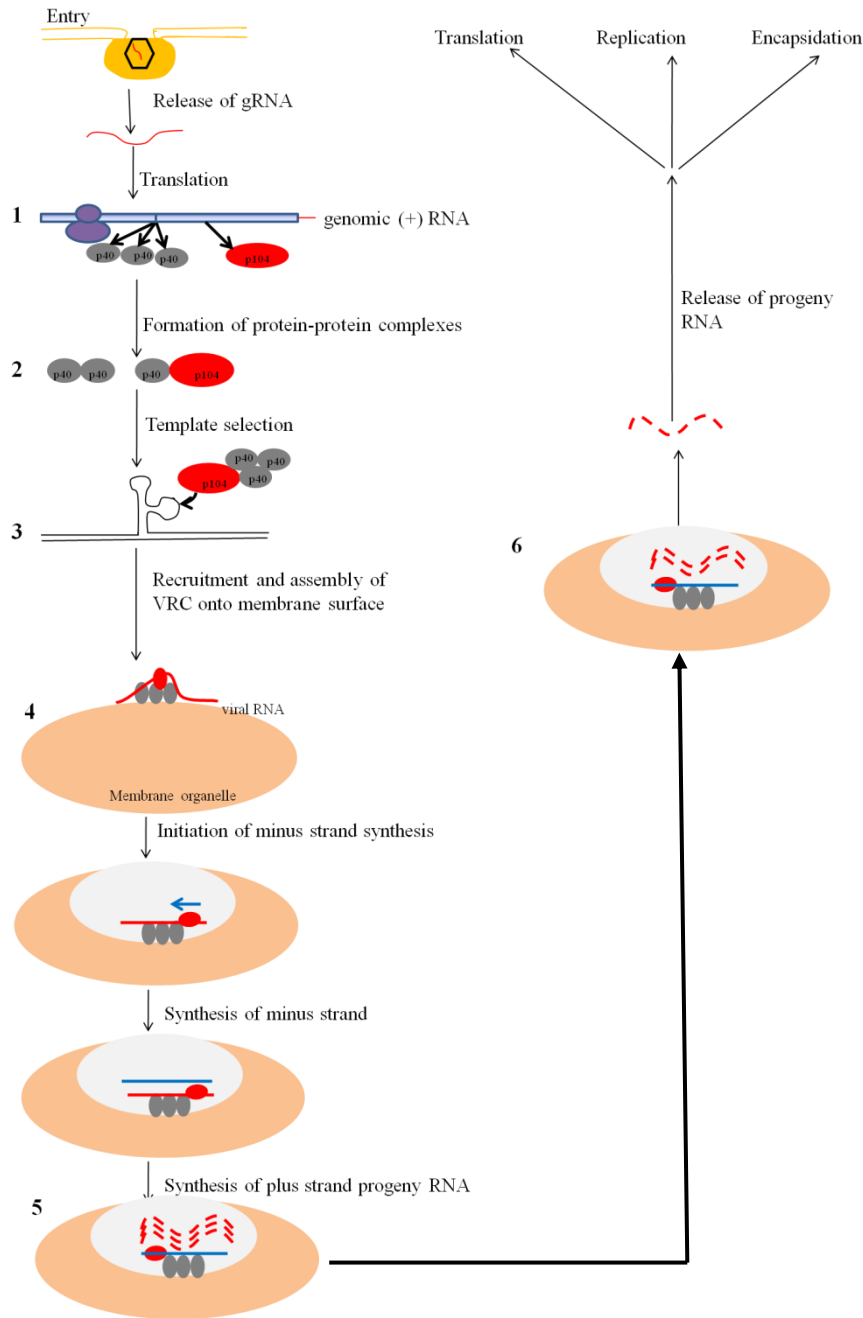


Figure 6. 1: Schematic representation of the proposed model of PrV replication. The six proposed steps of PrV replication. 1. Entry and genome translation. 2. Formation of protein-protein complexes. 3. Selection of viral genome. 4. Recruitment of VRC. 5. Synthesis of progeny viral RNA. 6. Release of viral RNA. Target cell is shown as yellow circle while organelle membrane is denoted as orange ovals. p104 is shown as red circles. p40 is shown as grey circles. Ribosome is shown as purple circles. Viral gRNA is represented as red line (not dotted). Minus strand is shown as blue line. Plus strand progeny RNA is represented as dotted red lines.

Entry and genome translation

To begin with, time course experiments of PrV infection reported by Pringle *et al.* (2003) revealed that input genomic RNA was observed at 6 and 12 hrs post infection (h.p.i), which represents the first two steps in the PrV infection cycle. First is the entry of PrV into the target cell and secondly, the release of genomic RNA for the initial translation of the replication proteins. Data from this study (Chapter 3) showed single p40 punctate structures, likely representing the initial translation of p40 and which compares with TBSV p33 that is expressed early in infection from the viral genome (Scholthof *et al.*, 1995b). The full-length product, p104, which contains the RdRp domain, is translated by readthrough stop codon at the 3' end of p40 ORF (Walter *et al.*, 2010), a translation strategy similar to TBSV (Scholthof *et al.*, 1995b). The major challenge with the subcellular localization studies (Chapter 3) was the lack of sensitivity of the anti-p104₍₈₈₉₋₉₀₂₎ peptide antibody to detect p104 (C-terminus) (Chapter 2). The lack of sensitivity of anti-p104₍₈₈₉₋₉₀₂₎ to detect p104 could be a result of low antigenicity generated by the small sized peptide that consisted of 14 amino acids.

Formation of protein-protein complexes between PrV replication proteins

The findings from chapter 4 demonstrate that protein-protein complexes are formed between p40 and itself and with p104 via a probable domain within the p40 sequence. The p40 was found to be more abundant than p104 within the PrV replication complex and this further supports self interaction of p40. This data is similar to the VRC of TBSV that consists of p33 approximately 10-20 fold in excess compared to p92 (Scholthof *et al.*, 1995b; Rajendran and Nagy, 2006). The putative p40 ID at amino acids position 292 – 332 is a potential candidate for mediating protein-protein interaction between p40:p40 and p40:p104 as it shares similar hydrophilic surface exposed characteristics with TBSV p33 ID and contains amino acids that are also conserved in TBSV p33 ID. Follow-up studies are recommended in future to verify the mechanism of protein-protein interaction between PrV replication proteins using experimental approaches such as surface Plasmon resonance (SPR) and fluorescence resonance energy (FRET).

Selection of PrV genome for replication

The putative PrV RBR 1 (RRRRYA) at the N-terminus of p64 (C-terminal region of p104) resembles the RBR 1 (RPRRRPYA) of TBSV p33 at the C-terminus which binds RNA. A putative RBR 2 within p104 contains amino acids conserved in TBSV p92 RBR 2 and which binds RNA. This means that the PrV replicase may bind the genome via the putative RBR 1 and RBR 2.

Recruitment of the viral replication complex to target membranes

The putative transmembrane helices, TM1 (52 - 72) and TM2 (225 - 247) within p104 might function to anchor the viral replication complex to membranes derived from the Golgi and/or secretory vesicles. According to the two transmembrane topology model of p104, the amino acids at positions 1 and 52 are found within cytosolic side of the membrane and this orientation resembles that of TBSV p92 which positions the peroxisome targeting signal 1 – 81 within the cytosolic side of the membrane. Therefore signals important for targeting Golgi and/or secretory vesicles are likely to be found within the cytosolic amino acid residues between 1 and 52 of p104 sequence. The role of host factors such as Pex19 and eEF1A have been implicated in the recruitment of TBSV viral replication complex (Pathak *et al.*, 2008; Li *et al.*, 2010). An important area for further study would require investigations into whether the same host factors are required for PrV replication.

Synthesis of progeny viral RNA

In this study, single punctate structures were detected with overlapping fluorescence signal of p40/p104 and dsRNA, likely representing initial synthesis of viral RNA. In aggregated structures of overlapping p40/p104 and dsRNA signal, the fluorescence for p40/p104 was on the perinuclear side while that for dsRNA extended into the cytoplasm. This suggests that viral RNA is synthesized and accumulates into the cytoplasm during replication.

Release of viral progeny RNA from replication

The release of plus strand progeny RNA from replication is possible since the serine and threonine phosphorylation sites were found adjacent to the putative RBR 1 and RBR 2 of p104. Experimental evidence is required to confirm whether the putative phosphorylation sites perform similar functions as reported for tombusvirus, TBSV.

Late stage of replication

The late phase of PrV replication was characterized by aggregated structures of overlapping p40/p104 and dsRNA fluorescence signal; likely representing accumulated virus factories formed after multiple rounds of translation of progeny viral RNA. Similarly, dsRNA aggregated structures were detected throughout the cell suggesting the occurrence of multiple rounds of plus strand progeny RNA replication in the late phase of PrV replication. Similar findings were reported by Pringle *et al.* (2003) when maximum viral RNA synthesis was detected during the late phase of infection and this coincided with the expression of the VCAP. This means during PrV replication, the plus strand progeny RNA is further replicated to produce more viral RNA with some translated into capsid proteins that encapsidates viral RNA during virion assembly.

Role of p130 in PrV replication

The identification of the putative arginine-rich region within p130 and which resembles the arginine-rich region of GRV p27 might suggest that the arginine-rich region is a potential RNA binding site. Furthermore, the arginine-rich region of p130 revealed hydrophilic and surface exposed characteristics that are similar to TBSV p33 RBR 1. This adds to the view that the arginine-rich region within p130 might function as an RNA binding site during replication. The C-terminal region of p130 is potentially disordered, a feature also present within GRV p27, an RNA chaperon (Taliensky *et al.*, 2003). This feature may add some support to the view that p130 may function as an RNA chaperon in replication.

6.4 THE ORIGIN OF PROVIDENCE VIRUS

With such remarkable diversity amongst tetravirus replicases, the origin of PrV is difficult to trace and can only be speculated on based on structural characteristics and comparative genomics. PrV shares the host range, capsid morphology and genome organisation of other tetraviruses, all evidence for an evolutionary relationship with tetraviruses. Despite the similarities with the capsid morphology of NβV capsid and NωV capsids (Speir *et al.*, 2010), the mechanisms of capsid assembly differ between PrV and other tetraviruses. PrV uses a molecular switch that is more similar to the switch described for the *T = 3* nodavirus, *Flock House virus* (FHV) rather than that

of NωV (Speir *et al.*, 2010). The capsid architecture of PrV and FHV is controlled by the N-termini of the β subunit while the C-termini of the γ -peptide controls the NωV capsid architecture (Speir *et al.*, 2010). This led Speir *et al.*, (2010) to propose that a monophyletic origin for PrV is possible i.e. PrV may have descended from a single primordial $T = 4$ particle derived presumably from a $T = 3$ ancestor and from which NωV evolved.

The replicase of PrV shares conserved RdRp motifs with plant viruses of the carmo-like super group II, leading to the conclusion that PrV has a plant virus-like replicase system that it shares with a common ancestor belonging to the carmo-like super group II viruses. This then raises questions on whether the replicase was acquired from a plant virus during evolution or alternatively that the PrV replicase evolved by co-evolution as a result of the close relationship between plant and insect hosts (Dolja and Koonin, 2011). Interestingly, within the RdRp is an additional motif A and this possibly reflects divergence towards permuted tetraviral RdRps, raising some interesting questions about its evolutionary origin. It is possible that RNA recombination may have occurred between a tetraviral and plant viral RdRp. The resemblance of the expression strategy of PrV replicase to tombusviruses and not tetraviruses is additional evidence of the evolutionary relationships between PrV and plant viruses of the carmo-like supergroup II. The readthrough signals of both PrV and TBSV replicase are of plant virus origin and this adds further support that PrV replicase might be of a plant virus origin.

The role of an insect vector in the evolution of PrV is likely. The umbraviruses lack a capsid protein and therefore depend on a helper virus from the *Luteoviridae* to encapsidate umbraviral RNA, transmitted by aphid vectors between plants and resulting in efficient replication within infected plants (Murant *et al.*, 1969; Reddy and Ansa, 1985). It is possible that a similar scenario might have happened with PrV were a tetravirus was accidentally picked up by a plant virus from the *Tombus* or *Umbraviridae*, forming hybrid virus particles that might have been transmitted by an insect vector. This type of evolution via the horizontal virus transfer between plants and animals has been suggested by Dolja and Koonin, (2011) and is characterized with high similarity between protein sequences and genome architectures of plant

and animal viruses. Usually, vehicles for transfer are available and such viruses have the capacity to replicate in both plant and insect systems (Dolja and Koonin, 2011).

Overall, findings from this research show that the replication system of PrV is related to tombusviruses and umbraviruses. The domains within PrV non-structural proteins are similar to the functional domains of tombusviruses and umbraviruses and therefore might perform similar roles in the replication of PrV.

REFERENCES

Ago, H., Adachi, T., Yoshida, A., Yamamoto, M., Habuka, N., Yatsunami, K. and Miyano, M. (1999). Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. *Structure*. **7**, 1417-1426.

Agranovsky, A. A., Koonin, E. V., Boyko, V. P., Maiss, E., Frötschl, R., Lunina, N. A. and Atabekov, J. G. (1994). Beet yellows closterovirus: complete genome structure and identification of a leader papain-like thiol protease. *Viol.* **198**, 311-324.

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

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

Ahola, T. and Ahlquist, P. (1999). Putative RNA capping activities encoded by Brome mosaic virus: methylation and covalent binding of guanylate by replicase protein 1a. *J. Gen. Virol.* **73**, 10061-10069.

Ahola, T., Laakkonen, P., Vihinen, H. and Kääriäinen, L. (1997). Critical residues of Semliki forest virus RNA capping enzyme involved in methyltransferase and guanyltransferase-like activities. *J. Gen. Virol.* **71**, 392-397.

Ahola, T., den Boon, J., A. and Ahlquist, P. (2000). Helicase and capping enzyme active site mutations in Brome mosaic virus protein 1a cause defects in template recruitment, negative-strand RNA synthesis and viral RNA capping. *J. Gen. Virol.* **74**, 8803-8811.

- Ahn, J. W., Lee, J. S., Davarpanah, S. J., Jeon, J. H., Park, Y. I., Liu, J. R. and Jeong, W. J. (2011). Host-dependent suppression of RNA silencing mediated by the viral suppressor p19 potato. *Planta*. **234**, 1065-72.
- Ambrose, R. L., Lander, G. C., Maaty, W. S., Bothner, B., Johnson, J. E. and Johnson, K. N. (2009). Drosophila A virus is an unusual RNA virus with a $T = 3$ icosahedral core and permuted RNA-dependant RNA polymerase. *J. Gen. Virol.* **90**, 2191-2200.
- Andriessen, M., Meulewaeter, F. and Cornelissen, M. (1995). Expression of tobacco necrosis virus Open Reading Frames 1 and 2 is sufficient for the replication of satellite tobacco necrosis virus. *Virol.* **212**, 222-224.
- Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M. and Possee, R. D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology*. **202**, 586-605.
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriol. Rev.* **35**, 235-241.
- Barciszewska, J., Barciszewska, B., Suter, B. and Kubli, E. (1985). Plant tRNA suppressors: *in vivo* readthrough properties and nucleotide sequence of yellow lupin seeds tRNA^{Try}. *Plant Sci.* **40**, 193-196.
- Baum, M. and Beier, H. (1998). Wheat cytoplasmic arginine tRNA isoacceptor with a U*CG anticodon is an efficient UGA suppressor *in vitro*. *Nucleic Acids Res.* **26**, 1390-1395.
- Bawden, A. L., Gordon, K. H. J. and Hanzlik, T. N. (1999). The specificity of *Helicoverpa armigera* stunt virus infectivity. *J. Inv. Pathol.* **74**, 156-163.
- Beier, H. and Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res.* **29**, 4767-4782.

- Beier, H., Barciszewska, M. and Sickinger, H. D. (1984a). The molecular basis for the differential translation of TMV RNA in tobacco protoplasts and wheat germ extracts. *EMBO. J.* **3**, 1091-1096.
- Beier, H., Barciszewska, M., Krupp, G., Mitnacht, R. and Gross, H. J. (1984b). UAG readthrough during TMV RNA translation: isolation and sequence of two tRNA^{Try} with suppressor activity from tobacco plants. *EMBO. J.* **3**, 351-356.
- Beier, H., Zech, U., Zubrod, E. and Kersten, H. (1987). Queuine in plants and plant tRNAs: differences between embryonic tissue and mature leaves. *Plant Mol. Biol.* **8**, 345-353.
- Bienz, M. and Kubli, E. (1981). Wild type tRNA^{Try} (G) reads the TMV RNA stop codon, but Q base modified tRNA^{Try} (Q) does not. *Nature*, **294**, 188-190.
- Blom, N., Gammeltoft, S. and Brunak, S. (1999). Sequence and structure based prediction of eukaryotic protein phosphorylation sites. *J. Mol. Biol.* **294**, 1351-1362.
- Blumenthal, T. (1980). Q β replicase template specificity: different templates require different GTP concentrations for initiation. *Proc. Natl. Acad. Sci. U.S.A.* **77**, 2601-2605.
- Boonham, N., Henry, C. M. and Roger Wood, K. (1995). The nucleotide sequence and proposed genome organization of oat chlorotic stunt virus, a new soil borne virus of cereals. *J. Gen. Virol.* **76**, 2025-2034.
- Bransom, K. L., Weiland, J. J., Tsai, C. H. and Dreher, T. W. (1995). Coding density of the turnip yellow mosaic virus genome: Roles of the overlapping coat protein and p206-readthrough coding regions. *Virology*. **206**, 403-412.
- Brierley, I., Jenner A, J., Inglis, S. C. (1992). Mutational analysis of the “slippery-sequence” component of a coronavirus ribosomal frameshifting signal. *J. Mol. Biol.* **227**, 463-479.

- Brooks, E. M., Gordon, K. H. J., Dorrian, S. J., Hines, E. R. and Hanzlik, T. N. (2002). Infection of its lepidopteran host by the *Helicoverpa armigera* stunt virus (Tetraviridae). *J. Invertebr. Pathol.* **80**, 97–111.
- Brown, C. M., Dinesh-Kumar, S. P. and Miller, W. A. (1996). Local and distant sequences are required for efficient readthrough of the barley yellow dwarf virus-PAV coat protein gene stop codon. *J. Gen. Virol.* **70**, 5884-5892.
- Buck, K. W. (1996). Comparison of the replication of positive stranded RNA viruses of plants and animals. *Adv. Virus. Res.* **47**, 159-251.
- Burgyán, J., Dalmay, T., Rubino, L. and Russo, M. (1992). The replication of cymbidium ringspot tomosvirus defective interfering-satellite RNA hybrid molecules. *Virol.* **190**, 579–586.
- Cameron, C. E., Moustafa, I. M. and Arnold, J. J. (2009). Dynamics: The missing link between structure and function of the viral RNA-dependent RNA polymerase? *Curr. Opin. Struc. Biol.* **19**, 768-774.
- Cann, J. A. (2005). Principles of Molecular Virology fourth edition. Elsevier Academic press, pp 36.
- 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 (N ω V). *J. Mol. Biol.* **299**, 573-584.
- Canady, M. A., Tsuruta, H. and Johnson, J. E. (2001). Analysis of rapid, large-scale protein quaternary structural changes: Time-resolved X-ray solution scattering of *Nudaurelia capensis* ω virus (N ω V) maturation. *J. Mol. Biol.* **311**, 803-814.
- Carrington, J. C., Heaton, L. A., Zuideman, D., Hillman, B. I. and Morris, T. J. (1989). The genome structure of turnip crinkle virus. *Virology.* **170**, 219-226.

- Caspar, D. L. D. and Klug, A. (1962). Physical principles in the construction of regular viruses. Cold Spring Harbour Symp. *Quant. Biol.* **27**, 1-24.
- Chang, Y. C., Borja, M., Scholthof, H. B., Jackson, A. O. and Morris, T. J. (1995). Host effects and sequences essential for accumulation of defective interfering RNAs of cucumber necrosis and tomato bushy stunt tombusviruses. *Virology*. **210**, 41-53.
- Chen, G. T. and Inouye, M. (1994). Role of the AGA/AGG codons, the rarest codons in global gene expression in *Escherichia coli*. *Genes Dev.* **8**, 2641-2652.
- Dalmay, T., Rubino, L., Burgyán, J. and Russo, M. (1992). Replication and movement of a coat protein mutant of cymbidium ringspot tombusvirus. *Mol Plant Microbe Interact.* **5**, 379-383.
- Demler, S.A., Rucker, D.G., de Zoeten, G. A. (1993). The chimeric nature of the genome of pea enation mosaic virus: the independent replication of RNA 2. *J. Gen. Virol.* **74**, 1-14.
- Denison, M. R. (2008). Seeking membranes: positive strand RNA virus replication complexes. *PLOS Biol.* **6**, e270.
- Dolja, V. V. and Koonin, E. V. (2011). Common origins and host-dependent diversity of plant and animal viromes. *Curr. Opin. Virol.* **1**, 322-331.
- Dorrington, R. A., Gorbalenya, A. E., Gordon, K. H. J., Lauber, C. and Ward, V. K. (2011). *Tetraviridae*. In: King, A. M. Q., Adams, M. J., Carstens, E. B. and Lefkowitz, E. J. (eds). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, Oxford, pp. 1091-1102.
- Dorrington, R. A. and Short, J. R. (2010). Tetraviruses. In: Asgari, S. and Johnson, K. (eds). *Insect Virology*. Caister Academic Press, Norfolk, U.K, pp 238-305.

- Dosztanyi, Z., Csizmok, V., Tompa, P. and Simon, I. (2005). IUPred: web server for the prediction of intrinsically unstructured regions of proteins based on estimated energy content. *Bioinformatics*. **21**, 3433-3434.
- Dougherty, W. G. and Kaesberg, P. (1981). Turnip crinkle virus RNA and its translation in rabbit reticulocyte and wheat embryo extracts. *Virology*. **115**, 45-56.
- Dreher, T. W. and Miller, W. A. (2006). Translation control in positive strand RNA plant viruses. *Virology*. **344**, 185-197.
- Dugeon, G., Urcuqui-inchima, S., Milner, M., Kadaré, G., Valle, R.P.C., Voyatzakis, A., Haenni, A. L. and Schirawski, J. (1999). The strategies of plant virus gene expression: Models of economy. *Plant Sci*. **148**, 77-88.
- 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.
- Fauquet, C. M., Mayo, M. A., Maniloff, J., Desselberger, U. and Ball, L. A. (2005). Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses. Academic Press.
- Feng, Y. X., Copeland, T. D., Oroszlan, S., Rein, A. and Levin, J. G. (1990). Identification of amino acids inserted during suppression of UAA and UGA termination codons at the *gag-pol* junction of moloney murine leukemia virus. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 8860-8863.
- Feng, Y. X., Yuan, H., Rein, A. and Levin, J. G. (1992). Bipartite signal for readthrough suppression in murine leukemia virus mRNA: an eight nucleotide purine-rich sequence immediately downstream of the gag termination codon followed by an RNA pseudoknot. *J. Virol.* **66**, 5127-5132.

- Firth, A. E., Wills, N. M., Gesteland, R. F., Atkins, J. F. (2011). Stimulation of stop codon readthrough: frequent presence of an extended 3' RNA structural element. *Nucleic Acids Res*, **39**, 6679-6691.
- Finch, J. T., Crowther, R. A., Hendry, D. A. and Struthers, J. K. (1974). The structure of *Nudaurelia capensis* β virus: the first example of a capsid with icosahedral surface symmetry $T = 4$. *J. Gen. Virol.* **24**, 191-200.
- Francki, R. I. B., Fauquet, C. M., Knudson, D. L. and Brown, F. (1991). Classification and nomenclature of viruses: Fifth Report of the International Committee on Taxonomy of Viruses. (Springer-Verlag, Vienna).
- Gazo, B. M., Murphy, P., Gatchel, J. R., Browning, K. S. (2004). A novel interaction of cap-binding protein complexes eukaryotic initiation factor (eIF) 4F and eIF(iso)4F with a region in the 3'-untranslated region of satellite tobacco necrosis virus. *J. Biol. Chem.* **279**, 13584-13592.
- Gibbs, M. J., Cooper, J. I., Waterhouse, P. M. (1996). The genome organisation and affinities of an Australian isolate of carrot mottle umbravirus. *Virol.* **224**, 310-313.
- Goodman, C. L. and McIntosh, A. H. (1995). Characteristics of midgut-derived insect cell lines. *In vitro Cell. Dev. Biol.* **31**, 85A.
- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P. and Blinov, V. M. (1989). Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.* **17**, 4713-4730.
- Gorbalenya, A. E., Pringle, F. M., Zeddam, J. L., Luke, B. T., Cameron, C. E., Kalkmakoff, 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. *Viol.* **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 genome organization. *Viol.* **258**, 42-53.
- Grace, T. D. C. and Mercer, E. H. (1965). A new virus of the *Saturniid anthereae eucalypti* Scott. *J. J. Invertebr. Pathol.* **7**, 241-244.
- Grace, T. D. C. and Mercer, E. H. (1976). Serological relations between twelve small RNA viruses of insects. *J. Gen. Virol.* **31**, 131-134.
- Graves-Woodward, K. L., Gottlieb, J., Challberg, M. D. and Weller, S. K. (1997). Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis, DNA unwinding and coupling between hydrolysis and unwinding. *J. Biol. Chem.* **272**, 4623-4630.
- Greenwood, L. K. and Moore, N. F. (1984). Determination of the location of an infection in *Trichoplusia ni* larvae by a small RNA-containing virus using enzyme-linked immunosorbent assay and electron microscopy. *Microbiologica.* **7**, 97-102.
- Grieco, F., Burgyán, J. and Russo, M. (1989). The nucleotide sequence of cymbidium ringspot virus RNA. *Nucleic Acids Res.* **17**, 6383.
- Grimm, M., Nass, A., Schüll, C. and Beier, H. (1998). Nucleotide sequences and functional characterization of two tobacco UAG suppressor tRNA^{Gln} isoacceptors and their genes. *Plant Mol. Biol.* **38**, 689-697.
- Gursinsky, T., Schulz, B. and Behrens, S. E. (2009). Replication of tomato busy stunt virus RNA in a plant *in vitro* system. *Virology.* **390**, 250-260.

- Halk, E. L., Robinson, D. J. and Murrant, A. F. (1979). Molecular weight of the infective RNA from leaves infected with carrot mottle virus. *J. Gen. Virol.* **45**, 383-388.
- 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. D. (1993). A novel small RNA virus isolated from 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. J. (1995). Sequence of RNA 2 of the *Helicoverpa armigera* stunt virus (*Tetraviridae*) and bacterial expression of its genes. *J. Gen. Virol.* **76**, 799-811.
- Hao, X., Lu, A., Sokal, N., Bhagwat, B., Leung, E., Mao, R., Reade, R., Wu, Y., Rochon, D. and Xiang, Y. (2011). Cucumber necrosis virus p20 is a viral suppressor of RNA silencing. *Virus Res.* **155**, 423-32.
- 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. H. V. (1968). A non inclusion virus of the pine emperor moth *Nudaurelia cytherea capensis* Stoll. *S. Afr. Med. J.* **42**, 117.
- Hendry, D. A., Hodgson, V., Clark, R. and Newman, J. (1985). Small RNA viruses co-infecting the pine emperor moth (*Nudaurelia cytherea capensis*). *J. Gen. Virol.* **66**, 627-632.

- Hellen, C. U. T and Sarnow, P. (2001). Internal ribosome entry sites in eukaryotic mRNA molecules. *Genes and Development*. **15**, 1593-1612.
- Hillman, B. I., Carrington, J. C. and Morris, T. J. (1987). A defective interfering RNA contains a mosaic of plant RNA genome. *Cell*. **51**, 427-433.
- Hofmann, K. and Stoffel, W. (1993). TMbase-a database of membrane spanning protein segments. *J. Biol. Chem. Hoppe-Seyler*. **347**, 166.
- Holland, J. J. and Levine, A. (1978). Mechanisms of viral persistence. *Cell*. **14**, 447-452.
- Hopp, T. P. and Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3824-3828.
- Huisman, M. J., Linthorst, H. J. M., Bol, J. F., Cornelissen, B. J. C. (1988). The complete nucleotide sequence of potato virus X and its homologies at the amino acid level with various plus stranded RNA viruses. *J. Gen Virol*. **69**, 1789-1798.
- Huppert, E., Szilassy, D., Salánki, K., Divéki, Z., Balázs, E. (2002). Hetelogenous movement protein strongly modifies the infection phenotype of cucumber mosaic virus. *J. Gen. Virol*. **76**, 3554-7.
- Hwang, Y. T., McCartney, A. W., Gidda, S. K. and Mullen, R. T. (2008). Localization of the carnation Italian ringspot virus replication protein p36 to the mitochondrial outer membrane is mediated by an internal targeting signal and the TOM complex. *BMC Cell Biology*. **9**, 54.
- Ishihama, A. and Nagata, K. (1988). Viral RNA polymerases. *CRC. Rev. Biochem*. **23**, 27-76.
- International committee for virus taxonomy: <http://ictvonline.org>

- Jiang, Y., Cheng, C. P., Serviène, E., Shapka, N. and Nagy, D. P. (2010a). Repair of lost 5' terminal sequences in tombusviruses: Rapid recovery of promoter-and enhancer-like sequences in recombinant RNAs. *Virology*. **404**, 96-105.
- Jiang, Y., Li, Z. and Nagy, P. D. (2010b). Nucleolin/Nsr1p binds to the 3' noncoding region of the tombusvirus RNA and inhibits replication. *Virol.* **396**, 10-20.
- Johnston, J. C. and Rochon, D. M. (1990). Translation of cucumber necrosis virus RNA *in vitro*. *J. Gen. Virol.* **71**, 2233-41.
- Johnson, G. D., Pirtle, I. L. and Pirtle, R. M. (1985). The nucleotide sequence of tyrosine tRNA^{QΨA} from bovine liver. *Arch. Biochem. Biophys.* **236**, 448-453.
- 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.
- Joyce, C. M. and Steitz, T. A. (1995). Polymerase structures and function: variations on a theme? *J. Bacteriol.* **177**, 6321-6329.
- Jukes, I. R. M. (1970). Viruses of the Pine Emperor Moth. *Bull. S. Afr. Soc. Plant Pathol. Microbiol.* **4**, 18.
- Junker, V., Teichmann, T., Hekele, A., Fingerhut, C. and Beier, H. (1997). The tRNA^{Try} isoacceptor and their genes in the ciliate *tetrahymena thermophila*: Cytoplasmic tRNA^{Try} has QΨA anticodon and is coded by multiple intron-containing genes. *Nucleic Acids Res.* **25**, 4194-4200.
- Kadaré, G., Rozanov, M., Haenni, A. L. (1995). Expression of the turnip yellow mosaic virus proteinase in *Escherichia coli* and determination of the cleavage site within the 206 kDa protein. *J. Gen. Virol.* **76**, 2853-2857.

- Kao, C. C. and Sun, J. H. (1996). Initiation of minus strand RNA synthesis by the brome mosaic virus RNA-dependent RNA polymerase: use of oligoribonucleotide primers. *J. Virol.* **70**, 6826-6830.
- Kao, C. C., Singh, P. and Ecker, D. J. (2001). *De novo* initiation of viral RNA-dependent RNA synthesis. *Virol.* **287**, 251-260.
- Karmer, G. and Argos, P. (1984). Primary structural comparison of RNA-dependent polymerases from plant, animal and bacterial viruses. *Nucleic Acids Res.* **12**, 7269-7282.
- Kim, K. H. and Lommel, S. A. (1994). Identification and analysis of the site of -1 ribosomal frameshifting in red clover necrotic mosaic virus. *Virol.* **200**, 574-582.
- Kim, K. H. and Lommel, S. A. (1998). Sequence element required for efficient -1 ribosomal frameshifting in red clover necrotic mosaic dianthovirus. *Virol.* **250**, 50-59.
- Kim, S. H., Ryabov, E. V., Kalinina, N. O., Rakitina, D. V., Gillespie, T., MacFarlane, S., Haupt, S., Brown, J. W. and Taliansky, M. (2007). Cajal bodies and the nucleolus are required for systemic infection of a plant virus. *EMBO.* **26**, 2169-2179.
- Klaassen, V. A., Boeshore, M. L., Koonin, E. V., Tian, T. and Falk, B. W. (1995). Genome structure and phylogenetic analysis of lettuce infectious yellows virus, a whitefly-transmitted, bipartite closterovirus. *Virology.* **208**, 99-110.
- Kohlstaedt, L. A., Wang, J., Friedman, J. M., Rice, P. A. and Steitz, T. A. (1992). Crystal structure of 3.5 Å resolution of HIV-1 reverse transcriptase complexed with an inhibitor. *Science.* **256**, 1783-1790.
- Koivunen, M. R. L., Sarin, L. P., Bamford, D. H. (2008). Structure-function insights into the RNA-dependent RNA polymerase of the dsRNA bacteriophage ϕ6. In: Patton, J. T. (eds). Segmented double stranded RNA viruses: Structure and Molecular Biology. Caister Academic Press, Norfolk, pp. 239-258.

- Kollár, A. and Burgyán, J. (1994). Evidence that ORF 1 and 2 are the only virus-encoded replicase genes of cymbidium ringspot tombusvirus. *Virology*. **201**, 169-172.
- Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive stranded RNA viruses. *J. Gen. Virol.* **72**, 2197-2206.
- Koonin, E. V., Chumakov, K. M. and Gorbalenya, A. E. (1989). Tentative identification of the RNA-dependent RNA polymerases of dsRNA viruses. *FEBS Letters*. **252**, 42-46.
- Koonin, E. V. and Dolja, V. V. (1993). Evolution and taxonomy of positive-strand RNA viruses: implications of comparative analysis of amino acid sequences. *Crit. Rev. Biochem. Mol. Biol.* **28**, 375-430.
- Kopelowitz, J., Hampe, C., Goldman, R., Reches, M. and Engelberg-Kulka, H. (1992). Influence of codon context on UGA suppression and readthrough. *J. Mol. Biol.* **225**, 261-269.
- Kozak, M. (1991). Structural features of eukaryotic mRNAs that modulate the initiation of translation. *J. Biol. Chem.* **266**, 19867-19870.
- Kuchino, Y., Beier, H., Akita, N. and Nishimura, S. (1987). Natural UAG suppressor glutamine tRNA is elevated in mouse cells infected with Moloney murine leukemia virus. *Proc. Natl. Acad. Sci. U.S.A.* **84**, 2668-2672.
- Kyte, J. and Doolittle, R., F. (1982). A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.
- Lai, M. M. (1992). RNA recombination in animal and plant viruses. *Microbiol. Rev.* **56**, 61-79.

- Lai, M. M. (1998). Cellular factors in the transcription and replication of viral RNA genomes: A parallel to DNA dependent RNA transcription. *Viol.* **244**, 1-12.
- Li, X. H., Heaton, L. A., Morris, T. J. and Simon, A. E. (1989). Turnip crinkle virus defective interfering RNAs intensify viral symptoms and are generated *de novo*. *Proc. Natl. Acad. Sci. U.S.A.* **86**, 9173-9177.
- Li, Z., Barajas, D., Panavas, T., Herbst, D. A., Nagy, P. D. (2008). Cdc34p ubiquitin-conjugating enzyme is a component of the tombusvirus replicase complex and ubiquitinates p33 replication protein. *J. Virol.* **82**, 6911-6926.
- Li, Z., Pogany, J., Panavas, T., Xu, K., Esposito, A. M., Kinzy, T. G. and Nagy, P. D. (2009). Translation elongation factor 1A is a component of the tombusvirus replicase complex and affects the stability of the p33 replication co-factor. *Viol.* **385**, 245-260.
- Li, Z., Pogany, J., Tupman, S., Esposito, A. M., Kinzy, T. G., Nagy, P. D. (2010). Translation elongation factor 1A facilitates the assembly of the tombusvirus replicase and stimulates minus-strand synthesis. *PLOS Pathog.* **6**, e1001175.
- Longstaff, M., Brigneti, G., Boccard, F., Chapman, S. and Baulcombe, D. (1993). Extreme resistance to potato virus X infection in plants expressing a modified component of the putative viral replicase. *EMBO. J.* **12**, 379-386.
- Lopinski, J. D., Dinman, J. D. and Bruenn, J. A. (2000). Kinetics of ribosomal pausing during programmed -1 translational frameshifting. *Molecular and Cellular Biology.* **20**, 1095-1103.
- Luke, G. A., de Felipe, P., Lukashev, A., Kallioinen, S. E., Bruno, E. A. and Ryan, M. D. (2008). Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. *J. Gen. Virol.* **89**, 1036-1042.
- Lupo, R., Rubino, L. and Russo, M. (1994). Immunodetection of the 33K/92K polymerase proteins in cymbidium ringspot virus-infected and in transgenic plant tissue extracts. *Arch Virol.* **138**, 135-42.

- Malet, H., Coutard, B., Jamal, S., Dutartre, H., Papageorgiou, N., Neuvenon, M., Ahola, T., Forrester, N., Gould, E. A., Lafitte, D., Ferron, F., Lescar, J., Gorbalenya, A. E., de Lamballerie, X. and Canard, B. (2009). The crystal structures of Chikungunya and Venezuelan equine encephalitis virus nsP3 macro domains define a conserved adenosine binding pocket. *J. Virol.* **83**, 6534-6545.
- Martelli, G. P. and Jelkmann, W. (1998). *Foveavirus*, a new plant virus genus. *Arch Virol.* **143**, 1245-1249.
- Martínez-Salas, E., Pacheco, A., Serrano, P. and Fernandez, N. (2008). New insights into internal ribosome entry site elements relevant for viral gene expression. *J. Virol.* **89**, 611-626.
- Matthews, R. E. F (1982). Classification and nomenclature of viruses. Fourth report of the International Committee on Taxonomy of Viruses. *Intervirology.* **17**, 1-160.
- Matsuda, D., Bauer, L., Tinnes and, K. and Dreher, T. W. (2004). Expression of the two nested overlapping reading frames of turnip yellow mosaic virus RNA is enhanced by a 5' and 3' viral sequences. *J. Virol.* **78**, 9325-9335.
- Mayor, S. and Pagano, R. E. (2007). Pathways of clathrin-independent endocytosis. *Nat Rev Mol Cell Biol.* **8**, 603-612.
- McCartney, A. W., Greenwood, J. S., Fabian, M. R., White, K. A. and Mullen, R. T. (2005). Localization of the tomato bushy stunt virus replication protein p33 reveals a peroxisome-to-endoplasmic reticulum sorting pathway. *Plant Cell.* **17**, 3513-3531.
- Mendu, V., Chiu, M., Barajas, D., Li, Z. and Nagy, P. D. (2010). Cpr1 cyclophilin and Ess1 parvulin prolyl isomerases interact with the tombusvirus replication protein and inhibit viral replication in yeast model host. *Virol.* **406**, 342-351.
- Meulewaeter, F., Danthinne, X., VanMontagu, M., Comelissen, M. (1998a). 5' and 3' sequence of satellite tobacco necrosis virus RNA promoting translation in tobacco. *Plant Journal.* **14**, 169-176.

- Meulewaeter, F., VanMontagu, M., Comelissen, M. (1998b). Features of the autonomous function of the translation enhancer domain of satellite tobacco necrosis virus. *RNA*. **4**, 1347-1356.
- Miller, W. A., Waterhouse, P. M. and Gerlach, W. L. (1988). Sequence and organization of barley yellow dwarf virus genomic RNA. *Nucleic Acids Research*. **16**, 6097-6111.
- Mo, X. H., Qin, X. Y., Wu, J., Yang, C., Wu, J. Y., Duan, Y. Q., Li, T. F., Chen, H. R. (2003). Complete nucleotide sequence and genome organisation of a Chinese isolate of tobacco bushy top virus. *Arch. Virol.* **148**, 389-397.
- Molnár, A., Havelda, Z., Dalmay, T., Szutorisz, H. and Burgyán, J. (1997). Complete nucleotide sequence of tobacco necrosis virus strain D_H and genes required for RNA replication and virus movement. *J. Gen. Virol.* **78**, 1235-1239.
- Monkewich, S., Lin, H. X., Fabian, M. R., Xu, W., Na, H., Ray, D., Chernysheva, O.A., Nagy, P. D., White, K.A. (2005). The p92 polymerase coding region contains an internal RNA element required at an early step in tombusvirus genome replication. *J. Virol.* **79**, 4848-4858.
- Moore, N. F. (1991). The *Nudarelia* β family of insect viruses. In: Kursak, E., Mecel, D. (ed). *Viruses of Invertebrates*. New York. pp. 277-285.
- Morch, M. D, Boyer, J. C. and Haenni, A. L. (1988). Overlapping open reading frames revealed by complete nucleotide sequencing of turnip yellow mosaic virus genomic RNA. *Nucleic Acids Research*. **16**, 6157-6173.
- 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.

- Murant, A. F., Goold, R. A., Robert, I. M., Cathro, J. (1969). Carrot mottle-a persistent aphid-borne virus with unusual properties and particles. *J. Gen. Virol.* **4**, 329-341.
- Murphy, J. F., D'arcy, C. J. and Clark, J. M. (1998). Barley yellow dwarf virus RNA has a 5' terminal genome-linked protein. *J. Gen. Virol.* **70**, 2253-2256.
- Nagy, P. D. and Simon, A. E. (1998). *In vitro* characterization of late steps of RNA recombination in turnip crinkle virus. II. The role of the priming stem and flanking sequences. *Virol.* **249**, 393-405.
- 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.
- Nagy, P. D. and Pogany, J. (2008). Multiple roles of viral replication proteins in plant RNA virus replication. *Methods Mol. Biol.* **451**, 55-68.
- Nagy, P. D. and Pogany, J. (2012). The dependence of viral RNA replication on co-opted host factors. *Nature reviews microbiology.* **10**, 137-149.
- Napthine, S., Yek, C., Powell, M. L., Brown, T. D., Brierley, I. (2012). Characterization of the stop codon readthrough signal of colorado tick fever virus segment 9 RNA. *RNA.* **18**, 241-252.
- Navarro, B., Rubino, L., Russo, M. (2004). Expression of the cymbidium ringspot virus 33 kDa protein in *saccharomyces cerevisiae* and molecular dissection of the peroxisomal targeting signal. *J. Virol.* **78**, 4744-4752.
- Nelson, R. S. and van Bel, A. J. E. (1998). The mystery of virus trafficking into, through and out of the vascular tissue. *Progr. Botany.* **59**, 476-533.
- Ng, K. K., Cherney, M. M., Vázquez, A. L., Machín, A., Alonso, J. M., Parra, F. and James, M. N. (2002). Crystal structures of active and inactive conformations of a caliviral RNA-dependent RNA polymerase. *J. Biol. Chem.* **277**, 1381-1387.

- Nguyen Ba, A. N., Pogoutse, A., Provart, N. and Moses, A. M. (2009). NLStradamus: a simple Hidden Markov Model for nuclear localization signal prediction. *BMC Bioinformatics*. **10**, 202.
- Novoa, R. R., Calderita, G., Arranz, R., Fontana, J., Granzow, H and Risco, C. (2005). Virus factories: associations of cell organelles for viral replication and morphogenesis. *Biol.Cell*. **97**, 147-172.
- Nutter, R. C., Scheets, K. Panganiban, L. C. and Lomme, S. A. (1989). The complete nucleotide sequence of the maize chlorotic mottle virus genome. *Nucleic Acids Res.* **17**, 3163-3177.
- Okada, N., Noguchi, S., Kasai, H., Shindo-Okada, N., Ohgi, T., Goto, T. and Nishimura, S. (1979). Novel mechanism of post-transcriptional modification of tRNA. Insertion of bases of Q precursors into tRNA by a specific tRNA transglycosylase reaction. *J. Biol. Chem.* **254**, 3067-3073.
- Ollis, D. L., Brick, P., Hamlin, R., Xuong, N. G. and Steitz, T. A. (1985). Structure of the large fragment of Escherichia coli DNA polymerase I complexed with dTMP. *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* beta virus, a $T = 4$ insect virus. *J. Struct. Biol.* **105**, 111-122.
- O'Reilly, E. K. and Kao, C. C. (1998). Analysis of RNA-dependent RNA polymerase structure and function as guided by known polymerase structures and computer predictions of secondary structure. *Virology*. **252**, 287-303.
- Osman, T. A. and Buck, K. M. (1996). Complete replication *in vitro* of tobacco mosaic virus RNA by a template dependent, membrane bound RNA polymerase. *J. Virol.* **70**, 6227-6234.

- Oster, S. K., Wu, B. and White, K. A. (1998). Uncoupled expression of p33 and p92 permits amplification of tomato bushy stunt virus RNAs. *J. Virol.* **72**, 5845-5851.
- Panavas, T. and Nagy, P. D. (2003). The RNA replication enhancer element of tombusviruses contains two interchangeable hairpins that are functional during plus strand synthesis. *J. Virol.* **77**, 258-269.
- Panavas, T., Hawkins, C. M., Panaviene, Z. and Nagy, P. D. (2005). 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., Panavas, T., Nagy, P. D. (2005). Role of an internal and two 3' terminal RNA elements in assembly of tombusvirus replicase. *J. Virol.* **79**, 10608-10618.
- Panaviene, Z., Panavas, T., Serva, S. and Nagy, P. D. (2004). Purification of the cucumber necrosis virus replicase from yeast cells: Role of co-expressed 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.
- Pathak, K. B., Sasvari, Z. and Nagy, P. D. (2008). The host Pex19p plays a role in peroxisomal localization of tombusvirus replication proteins. *Virol.* **379**, 294-305.
- Petersen, T. N., Brunak, S., von Heijne, G. and Nielsen, H. (2011). SignalP 4.0: discriminating signal peptides from transmembrane regions. *Nat Methods.* **29**, 785-786.
- Pelham, H. R. (1978). Leaky UAG termination codon in tobacco mosaic virus RNA. *Nature.* **272**, 469-471.

- Pestova, T. V., Shatsky, I. N. and Hellen, C. U. T. (1996). Functional dissection of eukaryotic initiation factor 4F: the 4A subunit and the central domain of the 4G subunit are sufficient to mediate internal entry of 43S preinitiation complexes. *Molecular Cellular Biology*. **16**, 6870-6878.
- 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**, 3867-3874.
- Pogany, J. and Nagy, P. D. (2008). Authentic replication and recombination of tomato bushy stunt virus RNA in a cell-free extract from yeast. *J. Virol.* **82**, 5967-5980.
- Pogany, J., Stork, J., Li, Z., Nagy, P. D. (2008). *In vitro* assembly of the tomato bushy stunt virus replicase requires the host heat shock protein 70. *Proc. Natl. Acad. Sci. U.S.A.* **105**, 19956-19961.
- Pogany, J., White, K. A. and Nagy, P. D. (2005). Specific binding of tombusvirus replication protein p33 to an internal replication element in the viral RNA is essential for replication. *J. Virol.* **79**, 4859-4869.
- Pringle, F. M., Gordon, K. H. J., Hanzlik, T. N., Kalmakoff, J., Scotti, P. D. and Ward, V. (1999). A novel capsid expression strategy for *Thosea asigna* virus (*Tetraviridae*). *J. Gen. Virol.* **80**, 1855-1863.
- 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.
- Qu, F. and Morris, T. J. (2002). Efficient infection of *Nicotiana benthamiana* by tomato bushy stunt virus is facilitated by the coat protein and maintained by p19 through suppression of gene silencing. *Mol Plant Microbe Interact.* **15**, 193-202.

- Quaggio, R. B., Ferro, J. A., Monteiro, P. B. and Reinach, F. C. (1993). Cloning and expression of chicken skeletal muscle troponin I in *Escherichia coli*: the role of rare codons on the expression level. *Protein Sci.* **2**, 1053-1056.
- Raferty, L. A., Bermingham Jr, J. R. and Yarus, M. (1986). Mutation in the D arm enables a suppressor with a CUA anticodon to read both amber and ochre codons in *Escherichia coli*. *J. Mol. Biol.* **190**, 513-517.
- 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.* **77**, 9244-9258.
- Rajendran, K. S. and Nagy, P. D. (2004). Interaction between the replicase proteins of tomato bushy stunt virus *in vitro* and *in vivo*. *Virol.* **326**, 250-261.
- 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.
- Ray, D., and White, K. A. (2003). An internally located RNA hairpin enhances replication of tomato bushy stunt virus RNAs. *J. Virol.* **77**, 245-257.
- Reade, R., Kakani, K., Rochon, D. (2010). A highly basic KGKKGK sequence in the RNA-binding domain of the cucumber necrosis virus coat protein is associated with encapsidation of full-length CNV RNA during infection. *Virology.* **403**, 181-8.
- Reddy, D. V. R., Murant, A. F., Raschke, J. H., Mayo, M. A. and Ansa, O. A. (1985). Properties and partial purification of infective material from plants containing groundnut rosette virus. *Ann. Biol.* **107**, 65-78.
- Reinganum, C., Robertson, J. S. and Tinsley, T. W. (1978). A new group of RNA viruses from insects. *J. Gen. Virol.* **40**, 195-202.

- Rico, P. and Hernández, C. (2004). Complete nucleotide sequence and genome organization of pelargonium flower break virus. *Arch. Virol.* **149**, 641-651.
- Rochon, D., Lommel, S., Martelli, G.P., Rubino, L. and Russo, M. (2011). *Tombusviridae*. In: King, A. M. Q., Adams, M. J., Carstens, E. B. and Lefkowitz, E. J. (eds). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, Oxford, pp. 1111-1138.
- Rochon, D. M. and Tremaine, J. H. (1989). Complete nucleotide sequence of the cucumber necrosis virus genome. *Virology*. **169**, 251-259.
- Rost, B. and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584 - 599.
- Rost, B. Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins*. **19**, 55-72.
- Rozanov, M. N., Koonin, E. V. and Gorbalenya, A. E. (1992). Conservation of the putative methyltransferase domain: a hallmark of the 'sindbis-like' supergroup of positive strand RNA viruses. *J. Gen. Virol.* **73**, 2129-2134.
- Rubino, L., Russo, M., Martelli, G. P. (1995). Sequence analysis of pothos latent virus genomic RNA. *J. Gen. Virol.* **76**, 2835-2839.
- Rubino, L., Weber-Lotfi, F., Dietrich, A., Stussi-Garaud, C. and Russo, M. (2001). The open reading frame 1-encoded ('36K') protein of carnation Italian ringspot virus localizes to mitochondria. *J. Gen. Virol.* **82**, 29-34.
- Rubino, L., Pantaleo, V., Navarro, B. and Russo, M. (2004). Expression of tombusvirus open reading frames 1 and 2 is sufficient for the replication of defective interfering, but not satellite, RNA. *J. Gen. Virol.* **85**, 3115-3122.
- Russell, R. (2008). RNA misfolding and the action of chaperones. *Front. Biosci.* **13**, 1-20.

- Ryabov, E. V., Oparka, K. J., Santa Cruz, S., Robinson, D. J. and Taliansky, M. E. (1998). Intracellular location of two ground rosette umbravirus proteins delivered by PVX and TMV vectors. *Viol.* **242**, 303-313.
- Ryabov, E. V., Fraser, G., Mayo, M. A., Barker, H. and Taliansky, M. (2001). Umbravirus gene expression helps potato leafroll virus to invade mesophyll tissues and to be transmitted mechanically between plants. *Viol.* **286**, 363-372.
- Ryabov, E. V., Robinson, D. J. and Taliansky, M. E. (1999). A plant virus-encoded protein facilitates long-distance movement of heterologous viral RNA. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 1212-1217.
- Ryabov, E. V., Taliansky, M. E., Robinson, D. J., Waterhouse, P. M., Murrant, A. F., de Zoeten, G. A., Falk, B.W., Vetter, H. J. and Gibbs, M. J. (2011). *Umbraviridae*. In: King, A. M. Q., Adams, M. J., Carstens, E. B. and Lefkowitz, E. J. (eds). *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses*. Elsevier Academic Press, Oxford, pp. 1191-1195.
- Ryabov, V. E., Kim, H. S. and Taliansky, M. (2004). Identification of a nuclear localization signal and nuclear export signal of the umbraviral long distance RNA movement protein. *J. Gen. Virol.* **85**, 1329-33.
- Sabanadzovic, S., Ghanem-Sabanadzovic, N. A. and Gorbalenya, A. E. (2009). Permutation of the active site of putative RNA dependant RNA polymerase in a newly identified species of plant alpha-like virus. *Viol.* **394**, 1-7.
- Scholthof, H. B., Scholthof, K-B. G., Kikkert, M. and Jackson, A. O. (1995a). Tomato bushy stunt virus spread is regulated by two nested genes that function in cell to cell movement and host dependent systemic invasion. *Virology*. **213**, 425-438.
- Scholthof, K-B. G., Scholthof, H. B. and Jackson, A. O. (1995b). The tomato bushy stunt virus replicase proteins are coordinately expressed and membrane associated. *Virology*. **208**, 365-369.

- Shapka, N., Stork, J. and Nagy, P. D. (2005). Phosphorylation of the p33 replication protein of cucumber necrosis tomosvirus adjacent to the RNA binding site affects viral RNA replication. *Viol.* **343**, 65-78.
- Sharma, M., Sasvari, Z. and Nagy, P. D. (2010). Inhibition of sterol biosynthesis reduces tomosvirus replication in yeast and plants. *J. Viol.* **84**, 2270-2281.
- Short, J. R. and Dorrington, R. A. (2012). Membrane targeting of an alpha-like tetrairus replicase is directed by a region within the RNA-dependent RNA polymerase domain. *J. Gen Viol.* **93**, 1076-1716.
- Short, J. R., Knox, C. and Dorrington, R. A. (2010). Subcellular localization and live-cell imaging of the *Helicoverpa armigera* stunt virus replicase in mammalian and *Spodoptera frugiperda* cells. *J. Gen. Viol.* **91**, 1514-1523.
- Short, J. R., Nakayinga, R., Hughes, G. E., Walter, C. T. and Dorrington, R. A. (2013). Providence virus (family: *Carmotetraviridae*) replicates vRNA in association with the Golgi apparatus and secretory vesicles. *J. Gen. Viol.* **94**, 1073-1078.
- Skuzeski, J. M., Nicholas, 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.
- Smith, D. and Yarus, M. (1989). Transfer RNA structure and coding specificity. I. Evidence that a D-arm mutation reduces tRNA dissociation from the ribosome. *J. Mol. Biol.* **206**, 489-501.
- Somogyi, P., Jenner, A. J. Brierley, I. and Inglis, S. C. (1993). Ribosomal pausing during translation of an RNA pseudoknot. *Molecular and Cellular Biology.* **13**, 6931-6940.
- Sousa, R., Chung, Y. J., Rose, J. P. and Wang, B. C. (1993). Crystal structure of bacteriophage T7 RNA polymerase at 3.3 Å resolution. *Nature.* **364**, 595-599.

- Speir, J. A. and Johnson, J. E. (2008). Tetraviruses. In: Mahy, B. W. J. and Van Regenmortel, M. H. V. (eds). *Encyclopaedia of Virology*, 3rd edition. (Elsevier). pp 27-37.
- Speir, J. A., Taylor, D. J., Natarajan, P., Pringle, F. M., Ball, L. A. and Johnson, J. E. (2010). Evolution in action: N and C termini of subunits in related $T = 4$ viruses exchange roles as molecular switches. *Structure*. **18**, 700-709.
- Sprinzi, M., Horn, C., Brown, M., Ioudovitch, A. and Steinberg, S. (1998). Compilation of tRNA sequences and sequences of tRNA genes. *Nucleic Acids Res.* **26**, 148-153.
- Stork, J., Kovalev, N., Sasvari, Z., Nagy, P. D. (2011). RNA chaperone activity of the tombusviral p33 replication proteins facilitates initiation of RNA synthesis by the viral RdRp *in vitro*. *J. Virol.* **409**, 338-347.
- Studier, F. W. and Moffatt, B. A. (1986). Use of bacteriophage T7 RNA polymerase to direct selective high-level expression of cloned genes. *J. Mol. Biol.* **189**, 113-130.
- Suter, B., Altwegg, M., Choffat, Y. and Kubli, E. (1986). The nucleotide sequence of two homogeneous *Drosophila melanogaster* tRNA^{Try} isoacceptors: application of a rapid tRNA anticodon sequencing method using S1 nuclease. *Arch. Biochem. Biophys.* **247**, 233-237.
- Taliansky, M., Roberts, I. M., Kalinina, N., Ryabov, E. V., Raj, S. K., Robinson, D. J. and Oparka, K. J. (2003). An umbraviral protein, involved in long distance RNA movement, binds viral RNA and forms unique, protective ribonucleoprotein complexes. *J. Virol.* **77**, 3031-3040.
- Taliansky, M. E. and Robinson, D. J. (2003). Molecular biology of umbraviruses; phantom warriors. *J. Gen. Virol.* **84**, 1951-1960.

- Taliansky, M. E., Robinson, D. J. and Murant, A. F. (1996). Complete nucleotide sequence and organisation of the RNA genome of groundnut rosette umbravirus. *J. Gen. Virol.* **77**, 2335-2345.
- Taliansky, M. E., Robinson, D. J. and Murant, A. F. (2000). Groundnut rosette disease virus complex: biology and molecular biology. *Adv. Virus. Res.* **55**, 357-400.
- Tao, Y., Farsetta, D. L., Nibert, M. L. and Harrison, S. C. (2002). RNA synthesis in a cage: structural studies of reovirus polymerase $\lambda 3$. *Cell.* **111**, 733-745.
- Taylor, D. J., Krishna, N. K., Canady, M. A., Schneemann, A. and Johnson, J. E. (2002). Large-scale, pH-dependant, 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. (2006). Preliminary x-ray characterization of authentic providence virus and attempts to express its coat protein gene in recombinant baculovirus. *Arch Virol.* **151**, 155-165.
- Thiébauld, O., Pooggin, M. M. and Ryabova, L. A. (2007). Alternative translation strategies in plant viruses. *Plant viruses.* **1**, 1-20.
- Tompa, P. and Csermely, P. (2004). The role of structural disorder in the function of RNA and protein chaperons. *FASEB Journal.* **11**, 1169-1175.
- Tripconey, D. (1970). Studies on a non-occluded virus of the pine tree emperor moth. *J. Invertebr. Pathol.* **15**, 268-275.
- Turina, M., Maruoka, M., Monis, J., Jackson, A. O. Scholthof, K. B. (1998). Nucleotide sequence and infectivity of a full-length cDNA clone of panicum mosaic virus. *Virol.* **241**, 141-155.
- Urban, C. and Beier, H. (1995). Cysteine tRNAs of plant origin as novel UGA suppressors. *Nucleic Acids Res.* **23**, 4591-4597.

- Valle, R. P., Morch, M. D. and Haenni, A. L. (1987). Novel amber suppressor tRNAs of mammalian origin. *EMBO. J.* **6**, 3049-3055.
- Van der Heijden, M. W. and Bol, J. F. (2002). Composition of alphavirus-like replication complexes: involvement of virus and host encoded protein. *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.
- Walter, C. T. (2008). Establishment of experimental systems for studying the replication biology of providence virus. Rhodes University PhD thesis.
- Walter, C. T., Pringle, F. M., Nakayinga, R., De Felipe, P., Ryan, M. D., Ball, L. A. and Dorrington, R. A. (2010). Genome organisation and translation products of Providence virus: insight into a unique tetravirus. *J. Gen. Virol.* **91**, 2826-2835.
- Walter, C. T., Tomasicchio, M., Hodgson, V., Hendry, D. A., Hill, M. P. and Dorrington, R. A. (2008). Characterization of a succession of small insect viruses in a wild South African population of *Nudaurelia cytherea capensis*. *S. Afr. J. Sci.* **104**, 147-152.
- Wang, R.Y. and Nagy, P. D. (2008). Tomato bushy stunt virus co-opts the RNA binding function of a host metabolic enzyme for viral genomic RNA synthesis. *Cell Host Microbe.* **3**, 178-187.
- Wang, H. L., O'Rear, J. and Stollar, V. (1996). Mutagenesis of the Sindbis virus nsP1 protein: effects on methyltransferase activity and viral infectivity. *J. Gen. Virol.* **215**, 527-531.

- Wang, S., Browning, K. S. and Miller, W. A. (1997). A viral sequence in the 3' untranslated region mimics a 5' cap in facilitating translation of uncapped mRNA. *EMBO J.* **16**, 4107-4116.
- Weber-Lotfi, F., Dietrich, A., Russo, M. Rubino, L. (2002). Mitochondrial targeting and membrane anchoring of a viral replicase in plant and cells. *J. Virol.* **76**, 10485-10496.
- Weiner, A. M., Deininger, P. L. and Efstradiatis, A. (1986). Nonviral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. *Annu. Rev. Biochem.* **55**, 631-661.
- Weiss, W.A. and Friedberg, E. C (1986). Normal yeast tRNAs^{Gln} (CAG) can suppress amber codons encoded by an essential gene. *J. Mol. Biol.* **192**, 725-735.
- White, K. A. and Morris, T. J. (1999). Defective and defective interfering RNAs of monopartite plus strand RNA plant viruses. *Curr. Top. Microbiol. Immunol.* **239**, 1-17.
- Wills, N. M., Gesteland, R. F. and Atkins, J. F. (1994). Pseudoknot-dependent readthrough of retroviral *gag* termination codons: importance of sequences in the spacer and loop 2. *EMBO. J.* **13**, 4137-4144.
- Wimmer, E., Hellen, C. U. T. and Cao, X. (1993). Genetics of Poliovirus. *Ann. Rev. Genet.* **27**, 353-436.
- Wu, B. and White, K. A. (1998). Formation and amplification of a novel tombusvirus defective RNA which lacks the 5' nontranslated region of the viral genome. *J. Virol.* **72**, 9897-9905.
- Wu, B., Vanti, W. B., White, K. A. (2001). An RNA domain within the 5' untranslated region of the tomato bushy stunt virus genome modulates viral RNA replication. *J. Mol. Biol.* **305**, 741-756.

- Xiong, Z. and Lommel, S. A. (1989). The complete nucleotide sequence and genome organization of red clover necrotic mosaic virus RNA-1. *Virol.* **171**, 543-554.
- Xiong, Z., Kim, K. H., Kendal, T. L. and Lommel, S. A. (1993). Synthesis of the putative red clover necrotic mosaic virus RNA polymerase by ribosomal frame shifting *in vitro*. *Virol.* **193**, 213-221.
- Yang, X., Smidansky, E. D., Maksimchuk, K. R., Lum, D., Welch, J. L., Anorl, J. J., Cameron, C. E., Boehr, D. D. (2012). Motif D of viral RNA-dependent RNA polymerases determines efficiency and fidelity of nucleotide addition. *Structure.* **20**, 1519-1527.
- 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.
- Yoshinaka, Y., Katoh, I., Copeland, T. D. and Oroszlan, S. (1985). Murine leukemia virus protease is encoded by the *gag-pol* gene and is synthesized through suppression of an amber termination codon. *Proc. Natl. Acad. Sci. U.S.A.* **82**, 1618-1622.
- Zaccomer, B., Haenni, A. L. and Macaya, G. (1995). The remarkable variety of plant RNA virus genomes. *J. Gen. Virol.* **76**, 231-247.
- Zeddami, J-L., Gordon, K. H. J., Lauber, C., Felipe Alves, C. A., Luke, B. T., Hanzlik, T. N., Ward, V. K. and Gorbalenya, A. E. (2010). *Euprosteria elaeasa* virus genome sequence and evolution of the *Tetraviridae* family: Emergence of bipartite genomes and conservation of the VPg signal with the dsRNA *Birnaviridae* family. *Virol.* **397**, 145-154.
- Zerfass, K. and Beier, H. (1992a). Pseudouridine in the anticodon GΨA of plant cytoplasmic tRNA^{Try} is required for UAG and UAA suppression in the TMV specific context. *Nucleic Acids Res.* **20**, 5911-5918.

Zerfass, K. and Beier, H. (1992b). The leaky UGA termination codon of tobacco rattle virus RNA is suppressed by tobacco chloroplast and cytoplasmic tRNAs^{Trp} with CmCA anticodon. *EMBO. J.* **11**, 4167-4173.

Zhang, L., Zitter, T. A. and Palukaitis, P. (1991). Helper virus-dependent replication, nucleotide sequence and genome organization of the satellite virus of maize white line mosaic virus. *Virology*. **180**, 467-473.

Zuniga, S., Sola, I., Cruz, J. L. and Enjuanes, L. (2009). Role of RNA chaperones in virus replication. *Virus Res.* **139**, 253-266.

APPENDIX A

CONSTRUCTION OF RECOMBINANT PLASMIDS

Mutation of the readthrough stop codon to several codons was done by performing site directed mutagenesis using *Pfu* DNA polymerase. Briefly, the PCR reaction (50 *ul*) mixture consisted of 39.84 *ul* water, 5 *ul* of 10 x *Pfu* DNA polymerase buffer with MgSO₄, 1 *ul* dNTPs mix (0.2 mM final concentration), 0.625 *ul* DMSO (1.25 % final concentration), 1.5 *ul* each of the primers (0.3 *uM* final concentration), 0.74 *ul* of template DNA (50 ng) and 0.42 *ul* (1.25 U) *Pfu* DNA polymerase (Promega). The reaction mixture was amplified for 27 cycles consisting of an initial denaturing step at 95 °C for 2min (1 cycle), denaturing step at 95 °C for 1min, annealing step at 52 °C for 30 s, extension step at 72 °C for 11 min and a final extension step at 72 °C for 5 min (1 cycle). Characteristics of plasmids used in this study are summarized in Table A.1 below and diagrammatic representation of their construction is shown in detail below.

Table A. 1: Plasmids used in this study

Plasmid	Vector Backbone	Antibiotic resistance	Host	Description	Chapter	Reference
pT7-7	N/A	Amp	<i>E. coli</i>	<i>E. coli</i> expression vector	2	GE Healthcare
pEGFP-C1	N/A	Kan	HeLa	EGFP-fusion expression vector	3	ClonTech
pFastBac TM Dual vector	N/A	Amp/Gent	<i>E. coli</i>	Recombinant shuttle vector	2	Invitrogen
pUC18	N/A	Amp	<i>E. coli</i>	Cloning vector	2	Promega
pRN3	pCW41	Amp	<i>E. coli</i>	Encodes p104(CTT)	2 and 3	Constructed in this study
pRN11	pRN3	Amp	<i>E. coli</i>	6xHis-p104 ₍₄₃₇₋₉₁₅₎ cloning vector	2	Constructed in this study
pRN12	pT7-7	Amp	<i>E. coli</i>	6xHis-p104 ₍₄₃₇₋₉₁₅₎ expression vector	2	Constructed in this study
pRN18	pFastBac TM Dual vector	Amp, Gent	<i>E. coli</i>	Recombinant shuttle vector for generating 6xHis-p104 ₍₄₃₇₋₉₁₅₎	2	Constructed in this study
pRN39	pCW41	Amp	Sf9	p104 coding sequence	3	Constructed in this study
pCW41	pUC18	Amp	<i>E. coli</i>	p104 coding sequence	2 and 3	Walter, 2008
pRN54	pEGFP-C1	Kan	HeLa	Encodes p104 ₍₁₋₈₄₇₎	3	Constructed in this study
pRN55	pEGFP-C1	Kan	HeLa	Encodes p104 ₍₁₋₈₄₇₎ (CTT) downstream of the CMV promoter	3	Constructed in this study

Appendix A

						study
pCW56	N/A	Amp/Kan	Sf9	EGFP-fusion expression vector	3	Walter, 2008
pRN56	pRN54	Kan	HeLa	Encodes p104 ₍₈₄₈₋₉₁₅₎	3	Constructed in this study
pCW57	N/A	Amp	Sf9	Encodes EGFP-p104 ₍₁₋₃₅₉₎	3 and 4	Walter, 2008
pRN57	pRN55	Kan	HeLa	Encodes p104 ₍₈₄₈₋₉₁₅₎ (CTT) downstream of the CMV promoter	3	Constructed in this study
pCW58	N/A	Amp	Sf9	Encodes p104 ₍₁₋₃₅₉₎ or p40	4	Walter, 2008
pRN58	pRN56	Amp	Sf9	Encodes EGFP-p104 downstream of the <i>hr:AcMNPV</i> promoter	3	Constructed in this study
pRN59	pCW57	Amp	Sf9	Encodes EGFP-p104(CTT) downstream of the <i>hr:AcMNPV</i> promoter	3	Constructed in this study

Key to Table A.1: Antibiotic resistance-Amp: ampicillin, Kan: kanamycin, Gent: gentamicin. “Chapter”-chapter in which respective plasmid is first mentioned.

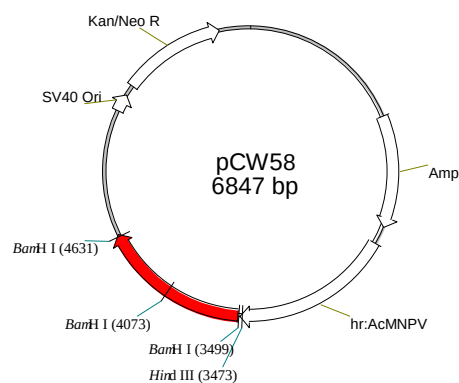


Figure A. 1: Graphic representation of the expression plasmid pCW58. The plasmid encodes the β -lactamase gene (AmpR), hybrid insect promoter (hr/AcMNPV Promoter), Neomycin phosphotransferase (Kan/Neo R), Simian Virus 40 origin of replication (SV40 ori).

APPENDIX B

ALTERNATIVE TRANSMEMBRANE TOPOLOGY MODELS OF

PrV REPLICASE

Two possible models for transmembrane topology have been considered based on the number of transmembrane helices predicted by TMpred (four) and Kyte and Doolittle (three) programmes. If all the four transmembrane helices predicted by TMpred are taken into consideration, then the predicted topology of p104 would traverse the hypothetical lipid membrane four times resulting in the positioning of the RdRp motifs I, II, III, IV (A), part of motif V (B), RBR 1, RBR 2 within the lumen space of the membrane organelle (Figure B.1). The rest of the motifs including VI (C), IV (A₁), VII (E), putative PrV ID and readthrough stop codon and regions would be in the cytoplasm (Figure B.1). According to this model, two of the ten putative antigenic sites would be located in the lumen while the reminder including the sequences 248-331 and 889-902 used to generate anti-p40₍₂₄₈₋₃₃₁₎ antiserum and anti-p104₍₈₈₉₋₉₀₂₎ peptide antibody would reside in the cytoplasm. This model has been eliminated because it does not allow for the RdRp to fold properly.

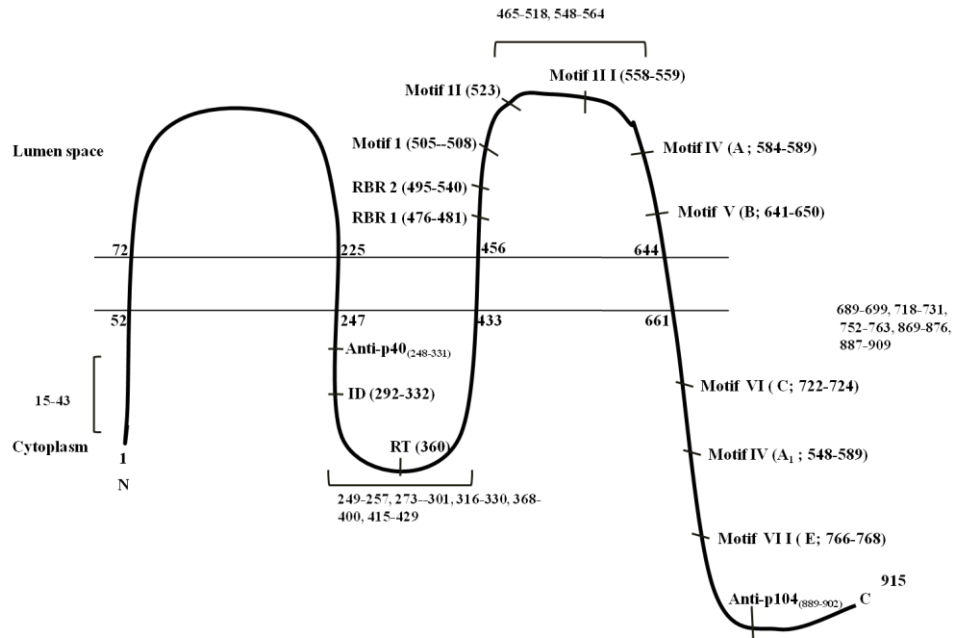


Figure B. 1: Model of the potential topology of PrV p104 with four transmembrane helices.

The p104 sequence traverses the hypothetical lipid bilayer indicated as an open space four times at positions 52 - 72 (predicted by Tmpred), 225 - 247, 433 - 456 and 647 - 661 (predicted by Tmpred and Kyte-Doolittle). The amino acid stretch of each transmembrane helice is annotated at either side of the lipid bilayer. The RdRp motifs I, II, III, IV (A), V (B), VI (C), IV (A₁), VII (E) are labeled on the protein, each with its amino acid position relative to the p104 amino acid sequence. The putative interaction domain is denoted as ID, RNA binding regions are denoted as RBR 1 and RBR 2. The readthrough stop codon is labeled as RT. Anti-p40₍₂₄₈₋₃₃₁₎ and Anti-p104₍₈₈₉₋₉₀₂₎ refer to regions used to generate antiserum and peptide antibody. Positions of the putative antigenic sites are indicated on the sides, below and above the p104 sequence.

An alternative model based upon the presence of three transmembrane helices predicted by Tmpred and Kyte and Doolittle programmes would ensure that the p104 traverses the hypothetical lipid membrane three times resulting in the positioning of the RdRp motifs VI (C), IV (A), VII (E), putative ID and the readthrough stop codon within the lumen space of the membrane organelle (Figure B.2). The remaining RdRp motifs I, II, III, IV (A₁), part of motif V (B) as well as putative RBR 1 and RBR 2 would reside in the cytoplasm (Figure B.2). According to this model, seven of ten putative antigenic sites including those used to generate anti-p40₍₂₄₈₋₃₃₁₎ antiserum and anti-p104₍₈₈₉₋₉₀₂₎ peptide

antibody would be located within the lumen (Figure B.2) and which is not likely. In addition, the position of the catalytic subdomain in the lumen is not likely and therefore this model would not allow proper folding of the RdRp.

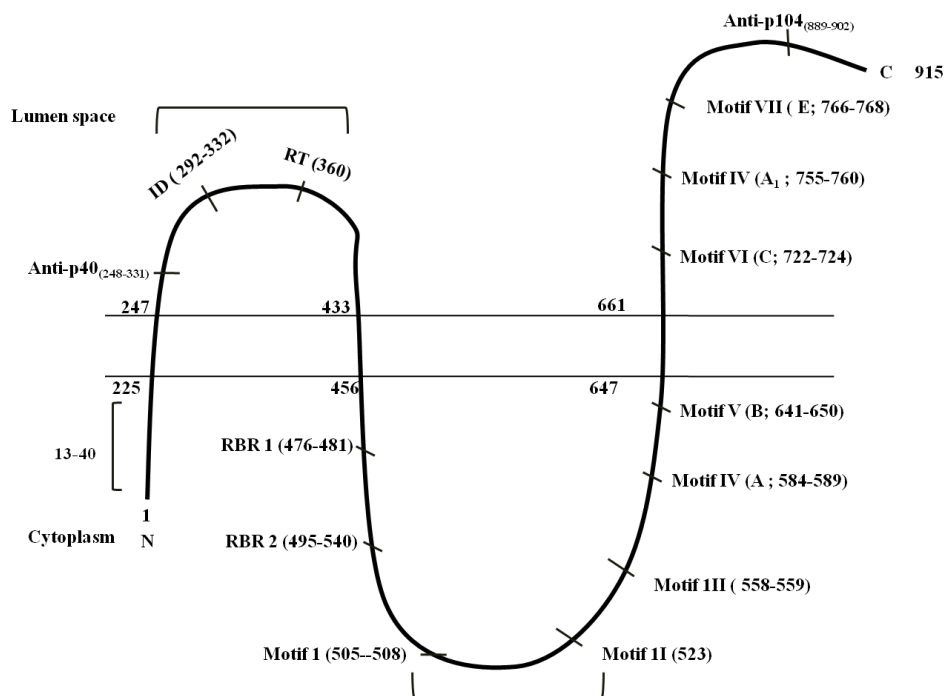


Figure B. 2: Model of the potential topology of p104 with three transmembrane helices.

The p104 sequence traverses the hypothetical lipid bilayer indicated as an open space three times at positions 225 - 247, 433 - 456 and 647 - 661 as predicted by TMpred and Kyte-Doolittle programmes. The amino acid stretch of each transmembrane helice is annotated at either side of the lipid bilayer and represented to scale. The RdRp motifs I, II, III, IV (A), V (B), VI (C), IV (A₁), VII (E) are labeled on the protein, each with its amino acid position relative to the p104 amino acid sequence. The putative interaction domain is denoted as ID, RNA binding regions are denoted as RBR 1 and RBR 2. The readthrough stop codon is labeled as RT. Anti-p40₍₂₄₈₋₃₃₁₎ and Anti-p104₍₈₈₉₋₉₀₂₎ refer to regions used to generate antiserum and peptide antibody. Positions of the putative antigenic sites are indicated on the sides, below and above the p104 sequence.

APPENDIX C

Short Communication

Providence virus (family: *Carmotetraviridae*) replicates vRNA in association with the Golgi apparatus and secretory vesicles

James R. Short,[†] Ritah Nakayinga, Gareth E. Hughes, Cheryl T. Walter[‡] and Rosemary A. Dorrington

Correspondence
Rosemary A. Dorrington
r.dorrington@ru.ac.za

Department of Biochemistry, Microbiology and Biotechnology, Rhodes University, Grahamstown 6140, South Africa

Received 3 September 2012
Accepted 22 January 2013

Providence virus (PrV) is the sole member of the family *Carmotetraviridae* (formerly *Tetraviridae*) sharing the characteristic $T=4$ capsid architecture with other tetravirus families. Despite significant structural similarities, PrV differs from other tetraviruses in terms of genome organization, non-structural protein sequence and regulation of gene expression. In addition, it is the only tetravirus that infects tissue culture cells. Previous studies showed that in persistently infected *Helicoverpa zea* MG8 cells, the PrV replicase associates with detergent-resistant membranes in punctate cytosolic structures, which is similar to the distribution of an alpha-like tetravirus replicase (*Helicoverpa armigera* stunt virus). Here, we demonstrate that the site of PrV vRNA replication coincides with the presence of PrV p40/p104 proteins in infected cells and that these replication proteins associate with the Golgi apparatus and secretory vesicles in transfected cells.

Positive sense (+) ssRNA viruses target their replication complexes to diverse locations, including the mitochondria (nodaviruses) (Garzon *et al.*, 1990; Miller *et al.*, 2001), endocytic pathway (togaviruses) (Froshauer *et al.*, 1988; Magliano *et al.*, 1998; Matthews *et al.*, 2009), endoplasmic reticulum (ER) (poliovirus and tomato ringspot virus) (Bienz *et al.*, 1987; Han and Sanfaçon, 2003; Rust *et al.*, 2001) and Golgi apparatus (poliovirus late in infection) (Bolten *et al.*, 1998). The mechanisms employed by (+) ssRNA viruses to target their vRNA replication complexes to particular cellular organelles have been well characterized, but the factors that drive the selection of these organelles as replication sites have yet to be elucidated.

Providence virus (PrV), the sole member of the newly created *Carmotetraviridae* tetravirus family, was isolated from a persistently infected *Helicoverpa zea* MG8 midgut tissue culture cell line and is the only tetravirus to replicate in tissue culture (Pringle *et al.*, 2003; ICTV, 2011). All tetraviruses have non-enveloped particles that display the characteristic $T=4$ icosahedral architecture and comprise 240 copies of a single capsid protein precursor (VCAP) that undergoes autoproteolytic cleavage to form the large (β) and

small (γ) subunits during capsid maturation (Dorrington and Short, 2010; Banerjee *et al.*, 2010). Tetraviruses were recently reclassified into three families, *Alphatetraviridae*, *Permutotetraviridae* and *Carmotetraviridae*, according to the sequence of their viral replicases which may be alpha-virus-like, picornavirus-like or carmovirus-like, respectively (ICTV, 2011).

PrV has a single genomic RNA of 6.4 kb encoding three ORFs, including the putative viral replicase (p104) followed by the VCAP ORF, which is translated from a subgenomic RNA of 2.4 kb (Fig. 1a). At the 5' end, upstream of and overlapping with the replicase is a third ORF that is unique to PrV, encoding a non-structural protein of unknown function with a predicted MR of 130 kDa (Walter *et al.*, 2010). Unlike other tetraviruses, the PrV replicase is most closely related to those of tombus- and umbraviruses, all of which are (+) ssRNA plant viruses. The activity of a type 1 read-through stop codon within the PrV replicase ORF results in the translation of a truncated 40 kDa protein (p40) as well as the full-length 104 kDa replicase protein (p104) at a ratio of 10:1 (p40:p104), a translation strategy also employed by tombusviruses (Beier and Grimm, 2001; Walter *et al.*, 2010).

Our interest is in studying the replication biology of alphatetraviruses, focusing on *Helicoverpa armigera* stunt virus (HaSV), but the ability of PrV to replicate in tissue culture has provided an experimental system for comparative studies on the establishment of tetravirus replication

[†]Present address: Department of Cell and Molecular Biology, Centre for Integrative Molecular Biosciences, Scripps Research Institute, La Jolla, CA 92037, USA

[‡]Present address: School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds LS2 9JT, UK

Supplementary material is available with the online version of this paper.

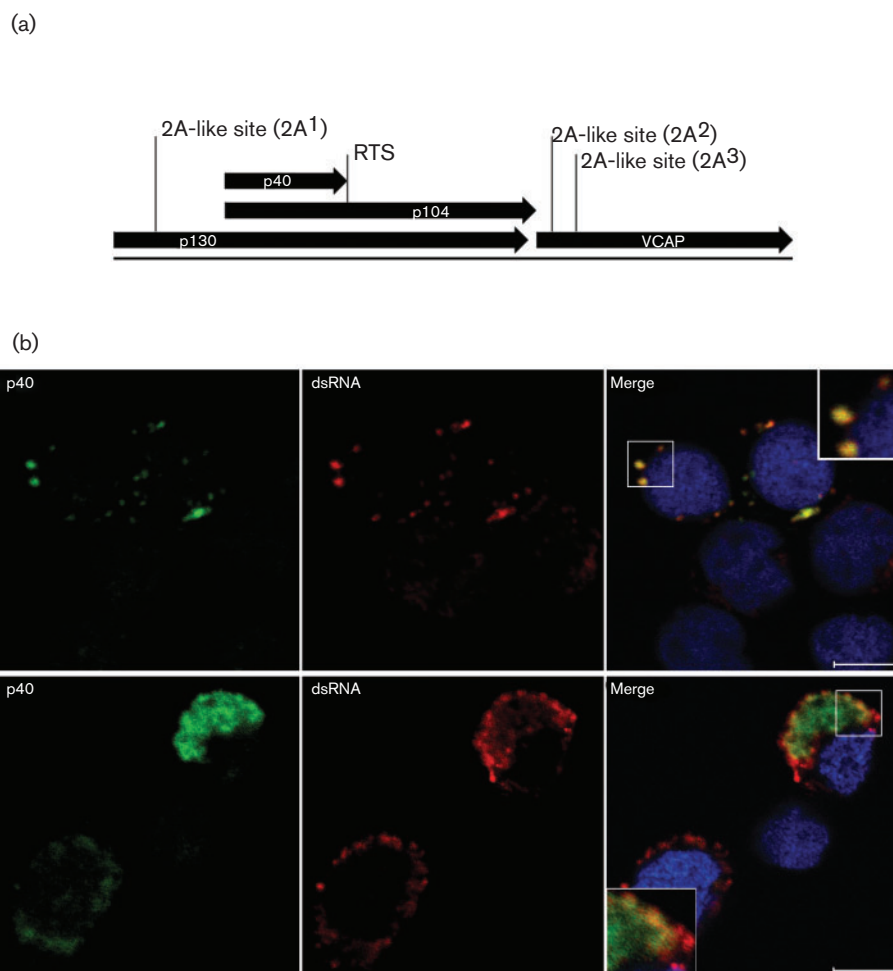


Fig. 1. PrV genome organization and co-localization of PrV p40/p104 with dsRNA in persistently infected *H. zea* MG8 cells. (a) The ORFs of the viral replicase (p40/p104), capsid protein precursor (VCAP) and a non-structural protein of unknown function (p130). Known 2A-like processing sites are indicated, as is the position of the read-through stop (RTS) codon that results in the generation of both p40 and p104 replication proteins from the p104 ORF. VCAP is translated from a subgenomic RNA (not shown). (b) Selected confocal images of infected MG8 cells stained by immunofluorescence using antibodies directed against PrV p40 (and p104) and dsRNA. Green staining (left) shows p40/p104, while the red (centre) shows dsRNA and the merged image (right) also includes nuclei stained with DAPI (blue). Inset image is magnified $\times 2$ over the area in the white box. All images represent 1 μ m optical slices. Bars, 10 μ m.

complexes *in vivo*. Previously, we conducted an investigation of the subcellular localization of HaSV replicase, showing that replicase–EGFP fusions were associated with detergent-resistant membranes (DRMs) and localized to punctate structures in the cytoplasm of transfected *Spodoptera frugiperda* cells (Short *et al.*, 2010; Short and Dorrington 2012). A preliminary study revealed a similar subcellular localization and strong association with DRMs for the PrV replicase in virus-infected cells raising the possibility of a common site of replication for both HaSV and PrV (Short and Dorrington, 2012).

We have examined more closely the subcellular localization of PrV replication proteins in relation to dsRNA in PrV-infected *H. zea* cells. To determine whether the

putative PrV replication proteins (p40 and p104) co-localize with replicating vRNA (dsRNA), persistently infected MG8 cells were prepared for confocal microscopy as previously described in Short *et al.* (2010), using rabbit anti-p40 antiserum that would detect both p40 and p104 and mouse monoclonal anti-dsRNA antibodies (clone K1; English & Scientific Consulting). The distribution of both p40/p104 and dsRNA was observed to range from co-localizing punctate cytosolic structures (Fig. 1b, top) to larger structures with stronger signal intensities spread over larger areas of the cytoplasm (Fig. 1b, bottom). The small punctate structures could represent earlier stages of infection, while the larger structures appear later in the infectious cycle. Interestingly, in cells with larger p40/p104-containing structures, the strongest dsRNA signal was

often concentrated around the periphery of the p40/p104 signal, possibly in virus particles. Taken together, these findings strongly suggest that p40 and p104 are involved in the formation of replication complexes and the replication of vRNA, as has been previously implied by bioinformatic analysis (Walter *et al.*, 2010).

In order to better understand the dynamic processes involved in the different distributions observed, more than 2000 cells were imaged and analysed for the presence and distribution of p40/p104 and dsRNA. Of these, 246 cells (~ 11 %) scored positive for p40/p104 expression and/or the presence of dsRNA as detected by immunofluorescence. The overlap and relative mean intensity of the signals were then analysed for each infected cell. Most of the cells (~71.5 %) displayed roughly equivalent relative fluorescence intensities and distribution for p40/p104 and dsRNA. The remainder of the cells exhibited a higher intensity and a greater distribution of p40/p104 (~5 %) or dsRNA (~20 %), while a small proportion (~3 %) were positive for only one of the immunofluorescence stains.

The mechanism by which the apparently persistent infection of MG8 cells by PrV is maintained is as yet unknown. The most likely possibilities are either: (i) the MG8 cells are only mildly susceptible to PrV infection and PrV multiplies within the population through cell division or cell-to-cell movement; or (ii) PrV infection of MG8 cells is inefficient. Consequently, the images of infected and uninfected cells were also analysed for signs of cell division (DNA metaphase or later) and cell death (chromosomal condensation and cell granularity). Of the cells imaged, 13 uninfected cells (0.66 %) showed signs of cell division. While no dividing cells were observed among the infected cells, a significantly larger population of cells would need to be scored for a statistically significant conclusion to be drawn regarding the movement of PrV between cells via cell division. However, an increased rate of chromosomal condensation and vesiculation in infected cells (41 % vs 5 % in uninfected cells) suggests that PrV replication ultimately results in cell death, which therefore implies a lytic life cycle. This increase in granularity could, however, be due to the formation of spherules in organelle membranes or a stimulation of membrane production as opposed to apoptosis or necrosis and further investigation would be necessary to determine if that is the case. Additional evidence of a lytic PrV infection lies in the observation of cells that were positive for dsRNA or p40/p104 only, indicating that replication progresses in infected cells and is not simply continuous at low levels. Finally, our preliminary attempts at superinfection of MG8 cells with purified PrV results in an increase in cell death, albeit with no significant increase in total PrV production (C. Adams and R. A. Dorrington, personal communication). Taken together, these data suggest that the persistence of PrV infection in the MG8 cell line is probably due to inefficiencies at establishing PrV infection, due to a combination of poor infectivity of progeny virions, a high

level of immunomodulation within infected insect host cells and low levels of PrV virus production.

Direct demonstration of the co-localization of the vRNA replication complexes with cellular organelles was not possible due to a lack of suitable antibodies directed against organelle markers in MG8 cells. As a result two constructs for the expression of EGFP-tagged PrV replication proteins from an hr/ACMNPV hybrid promoter in *S. frugiperda* Sf9 cells were generated. The constructs contained the entire p104 ORF fused to the C terminus of EGFP and generated EGFP-p104 (wt) (EGFP-p40 and EGFP-p104 expression) or EGFP-p104 (Leu) (producing only EGFP-p104 with the read-through 'UAG' stop codon mutated to 'CUU' encoding a leucine residue). Expression of only p104 from the p104 (Leu) ORF was confirmed using coupled transcription and translation reactions in insect cell extracts (data not shown).

First, we assessed the effect of the EGFP-tag on the distribution of p40 and p104 in the absence of other PrV proteins. *S. frugiperda* Sf9 cells expressing each of the EGFP-tagged fusion proteins were fixed with para-formaldehyde and examined by confocal fluorescence microscopy (Fig. 2). EGFP-associated fluorescence in cells expressing EGFP-tagged p104 or p40/p104 produced a small number (usually fewer than 10) of spherical structures distributed throughout the cytoplasm that typically appeared to be between 1.5 µm and 250 nm in diameter, reminiscent of the p40/p104 distribution in infected cells (compare Fig. 1b, top, to Fig. 2).

The subcellular location of the punctate structures observed for the EGFP-tagged PrV replication proteins was compared to that of fluorescent protein-tagged organelle marker proteins (Table S1, available in JGV Online). Comparison of the EGFP-tagged PrV replication proteins with co-expressed actin-YFP (Fig. 3a) revealed small amounts of actin accumulation at the same structures populated with EGFP-p104 (Leu) (Fig. 3a, top). Although no such accumulation was observed for EGFP-p104 (wt) (Fig. 3a,

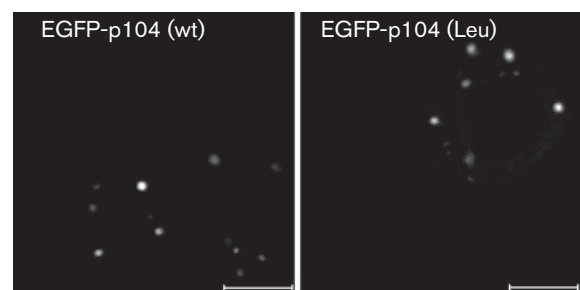


Fig. 2. Subcellular distribution of EGFP-tagged PrV p40 and p104. Confocal fluorescence images of *S. frugiperda* Sf9 cells expressing p104 [EGFP-p104 (Leu)], or both p40 and p104 [EGFP-p104 (wt)]. All images represent 1 µm optical slices. Bars, 10 µm.

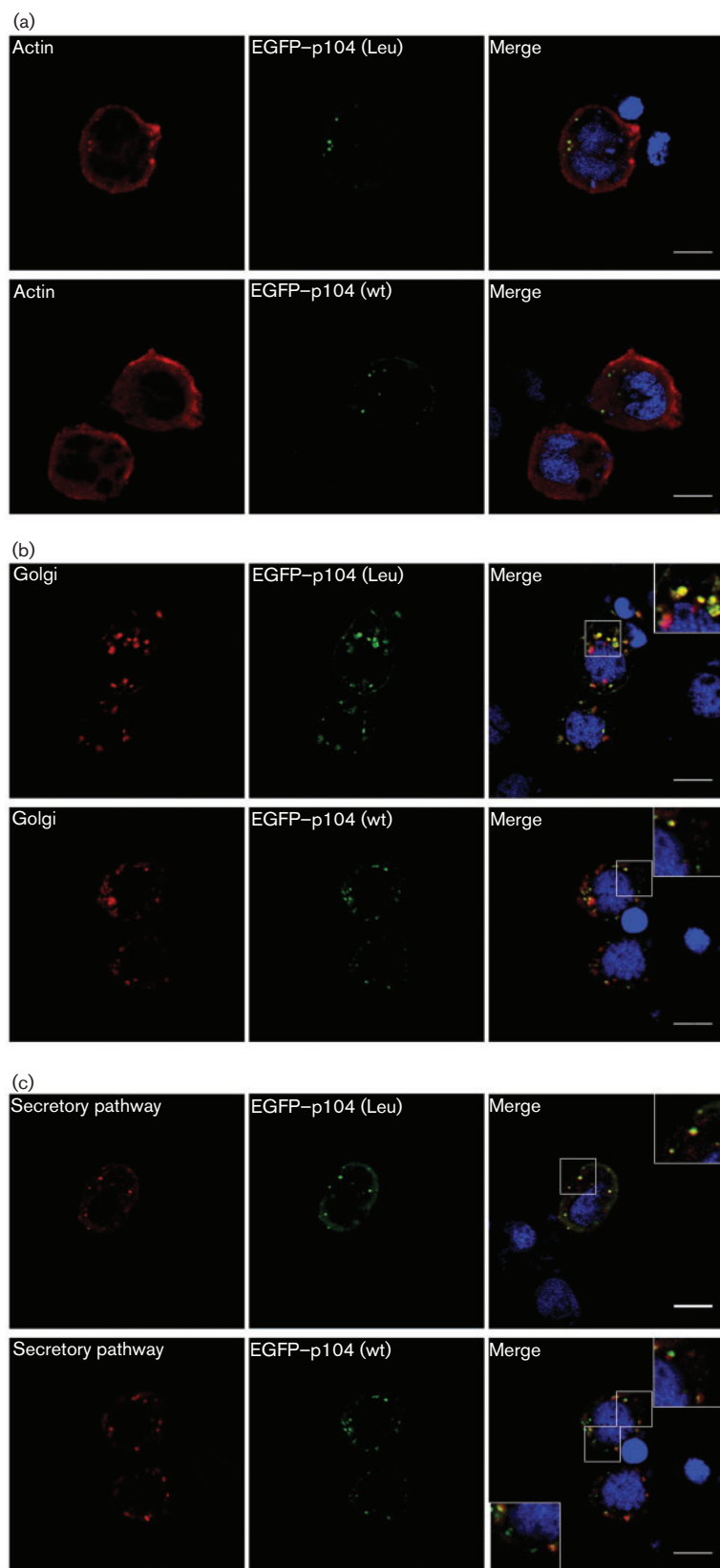


Fig. 3. Co-localization of EGFP-tagged p40 and p104 in *S. frugiperda* Sf9 cells co-transfected with fluorescent organelle marker proteins. Confocal fluorescence images of cells expressing EGFP-tagged PrV replication proteins and actin-YFP (a) or Golgi-targeted YFP (b) and/or CFP targeted to secretory vesicles (c). Insets represent a $\times 2$ magnification of the white box. All images represent 1 μm thick optical slices. Bars, 10 μm .

bottom row), these data suggested that actin might be involved in the formation of PrV replication complexes. To assess the role of actin in the initial formation of the

punctate structures, cells expressing EGFP-p104 (wt) were treated with 5 or 10 $\mu\text{g ml}^{-1}$ cytochalasin D for 15 h. Control cells expressing actin-YFP showed a collapse of the

actin cytoskeletal network at both concentrations, while the cellular distribution of EGFP-p40/p104 remained unchanged (data not shown). Therefore, if actin is involved in the replication of vRNA, it is likely to be involved in the condensation of membranes containing the viral replicase later in infection (Fig. 1b, bottom) and not in the initial formation of the punctate structures.

No overlap was observed between EGFP-tagged p40/p104 or p104 alone and the ER, peroxisomes or lysosomes (Fig. S1, data not shown). Since the lysosome-targeted RFP did not resolve the lysosomes satisfactorily in any of our experiments, this result was confirmed using LysoTracker (Fig. S2). In contrast, significant overlap between fluorescent proteins targeted to the Golgi apparatus (Fig. 3b) and secretory vesicles (Fig. 3c) and all of the EGFP-tagged p40/p104 variants was observed in more than 75 % of the co-transfected cells. The somewhat altered distribution of the PrV proteins in the presence of the Golgi-directed YFP suggests that some of the overlap observed in this instance may be due to the overloading of common protein transport machinery, resulting in the stalling of transport to the secretory pathway, which is more likely to be the ultimate destination for the replicase.

The subcellular trafficking of the PrV replicase proteins to the secretory pathway is unexpected. This is because the distribution and membrane association of p40/104 is similar to that of the HaSV replicase, which has been suggested to associate with membranes derived from an alternate endocytic pathway due to a partial overlap with the late endosomal marker, CD63, in transfected mammalian cells (Short *et al.*, 2010). Both replicases have been shown to associate strongly with cellular DRMs and the distribution of p40/104 and the HaSV replicase in insect cells appears virtually indistinguishable (Short and Dorrington, 2012). However, the data obtained with the HaSV replicase demonstrated only partial overlap with late endosomes, which was thought to be the result of endocytic organelle associations based upon an hypothesis that virus replication may be associated with acidic subcellular compartments (Tomasicchio *et al.*, 2007). This partial association with late endosomes could also be the result of the fusion of secretory vesicles containing the replicase with the late endosomes as part of normal vesicular trafficking. Further studies to confirm the targeting of the HaSV replicase to the secretory pathway are currently under way.

In conclusion, this study provides the first insight into the *in vivo* formation of tetravirus replication complexes and demonstrates that PrV replicates vRNA in association with secretory pathway vesicles. The data also shed light on the nature of the PrV infectious life cycle in *H. zea* MG8 cells, providing evidence for a low level lytic infection rather than a persistent infection. Our preliminary data indicate that the HaSV replicase also localizes to the secretory pathway (J.R. Short and R.A. Dorrington, unpublished results). The similarities between the distribution of the

PrV and HaSV replicases suggest that, although there are significant differences in the sequence and expression strategies of these viral replicases, there may be important biological and structural determinants that require replication to occur at the same location.

References

- Banerjee, M., Speir, J. A. & Johnson, J. E. (2010). Structural comparison of insect RNA viruses. In *Insect Virology*, pp. 327–346. Edited by S. Asgari & K. Johnson. Norfolk, UK: Caister Academic Press.
- Beier, H. & Grimm, M. (2001). Misreading of termination codons in eukaryotes by natural nonsense suppressor tRNAs. *Nucleic Acids Res* **29**, 4767–4782.
- Bienz, K., Egger, D. & Pasamontes, L. (1987). Association of polioviral proteins of the P2 genomic region with the viral replication complex and virus-induced membrane synthesis as visualised by electron microscopic immunocytochemistry and autoradiography. *Virol* **160**, 220–226.
- Bolten, R., Egger, D., Gosert, R., Schaub, G., Landmann, L. & Bienz, K. (1998). Intracellular localization of poliovirus plus- and minus-strand RNA visualized by strand-specific fluorescent *in situ* hybridization. *J Virol* **72**, 8578–8585.
- Dorrington, R. A. & Short, J. R. (2010). The tetraviruses. In *Insect Virology*, pp. 283–306. Edited by S. Asgari & K. Johnson. Norfolk, UK: Caister Academic Press.
- Froshauer, S., Kartenbeck, J. & Helenius, A. (1988). Alphavirus RNA replicase is located on the cytoplasmic surface of endosomes and lysosomes. *J Cell Biol* **107**, 2075–2086.
- Garzon, S., Strykowski, H. & Charpentier, G. (1990). Implication of mitochondria in the replication of Nodamura virus in larvae of the Lepidoptera, *Galleria mellonella* (L.) and in suckling mice. *Arch Virol* **113**, 165–176.
- Han, S. & Sanfaçon, H. (2003). Tomato ringspot virus proteins containing the nucleoside triphosphate binding domain are trans-membrane proteins that associate with the endoplasmic reticulum and cofractionate with replication complexes. *J Virol* **77**, 523–534.
- ICTV (2011). Release. <http://www.ictvonline.org/virusTaxonomy.asp?version=2011>
- Magliano, D., Marshall, J. A., Bowden, D. S., Vardaxis, N., Meanger, J. & Lee, J.-Y. (1998). Rubella virus replication complexes are virus-modified lysosomes. *Virology* **240**, 57–63.
- Mathews, J. D., Tzeng, W.-P. & Frey, T. K. (2009). Determinants of subcellular localization of the rubella virus nonstructural replicase proteins. *Virology* **390**, 315–323.
- Miller, D. J., Schwartz, M. D. & Ahlquist, P. (2001). Flock house virus RNA replicates on outer mitochondrial membranes in *Drosophila* cells. *J Virol* **75**, 11664–11676.
- Pringle, F. M., Johnson, K. N., Goodman, C. L., McIntosh, A. H. & Ball, L. A. (2003). Providence virus: a new member of the *Tetraviridae* that infects cultured insect cells. *Virology* **306**, 359–370.
- Rust, R. C., Landmann, L., Gosert, R., Tang, B. L., Hong, W., Hauri, H.-P., Egger, D. & Bienz, K. (2001). Cellular COPII proteins are involved in production of the vesicles that form the poliovirus replication complex. *J Virol* **75**, 9808–9818.
- Short, J. R. & Dorrington, R. A. (2012). Membrane targeting of an alpha-like tetravirus replicase is directed by a region within the RNA-dependent RNA polymerase domain. *J Gen Virol* **93**, 1706–1716.

Short, J. R., Knox, C. & Dorrington, R. A. (2010). Subcellular localization and live-cell imaging of the *Helicoverpa armigera* stunt virus replicase in mammalian and *Spodoptera frugiperda* cells. *J Gen Virol* **91**, 1514–1523.

Tomasicchio, M., Venter, P. A., Gordon, K. H. J., Hanzlik, T. N. & Dorrington, R. A. (2007). Induction of apoptosis in *Saccharomyces*

cerevisiae results in the spontaneous maturation of tetravirus procapsids *in vivo*. *J Gen Virol* **88**, 1576–1582.

Walter, C. T., Pringle, F. M., Nakayinga, R., de Felipe, P., Ryan, M. D., Ball, L. A. & Dorrington, R. A. (2010). Genome organization and translation products of Providence virus: insight into a unique tetravirus. *J Gen Virol* **91**, 2826–2835.

Genome organization and translation products of Providence virus: insight into a unique tetravirus

Cheryl T. Walter,^{1††} Fiona M. Pringle,^{2†§} Ritah Nakayinga,¹ Pablo de Felipe,^{3||} Martin D. Ryan,³ L. Andrew Ball² and Rosemary A. Dorrington¹

Correspondence

Rosemary A. Dorrington
r.dorrington@ru.ac.za

¹Dept of Biochemistry, Microbiology and Biotechnology, Rhodes University, Grahamstown 6140, South Africa

²Dept of Microbiology, University of Alabama at Birmingham, AL 35294, USA

³Centre for Biomolecular Sciences, University of St Andrews, North Haugh, St Andrews KY16 9ST, Scotland, UK

Providence virus (PrV) is a member of the family *Tetraviridae*, a family of small, positive-sense, ssRNA viruses that exclusively infect lepidopteran insects. PrV is the only known tetravirus that replicates in tissue culture. We have analysed the genome and characterized the viral translation products, showing that PrV has a monopartite genome encoding three ORFs: (i) p130, unique to PrV and of unknown function; (ii) p104, which contains a read-through stop signal, producing an N-terminal product of 40 kDa (p40) and (iii) the capsid protein precursor (p81). There are three 2A-like processing sequences: one at the N terminus of p130 (PrV-2A₁) and two more (PrV-2A₂ and PrV-2A₃) at the N terminus of p81. Metabolic radiolabelling identified viral translation products corresponding to all three ORFs in persistently infected cells and showed that the read-through stop in p104 and PrV-2A₃ in p81 are functional *in vivo* and these results were confirmed by *in vitro* translation experiments. The RNA-dependent RNA polymerase domain of the PrV replicase is phylogenetically most closely related to members of the families *Tombusviridae* and *Umbraviridae* rather than to members of the family *Tetraviridae*. The unique genome organization, translational control systems and phylogenetic relationship with the replicases of (+ve) plant viruses lead us to propose that PrV represents a novel family of small insect RNA viruses, distinct from current members of the family *Tetraviridae*.

Received 23 May 2010
Accepted 10 August 2010

INTRODUCTION

The members of the family *Tetraviridae* are small, non-enveloped insect viruses that encapsidate one or two single-stranded (ss) positive-sense (+ve) genomic RNAs (Hanzlik *et al.*, 2005). The members of the family *Tetraviridae* have an unusually narrow host range often restricted to only one or a few closely related lepidopteran species (Moore *et al.*, 1985; Gordon & Hanzlik, 1998; Christian *et al.*, 2001; Pringle *et al.*, 2003). Tetraviruses replicate in the midgut tissues of their hosts (Moore *et al.*,

1985; Bawden *et al.*, 1999; Brooks *et al.*, 2002) and sloughing of infected cells results in the stunting of infected larvae (Moore, 1991; Christian *et al.*, 2001; Brooks *et al.*, 2002).

Tetraviral genomic RNAs are encapsidated in an icosahedral capsid of approximately 40 nm in diameter, with *T*=4 symmetry (Olson *et al.*, 1990). The virus particles assemble as procapsids composed of 240 identical subunits that undergo maturation via assembly dependent autoproteolytic cleavage of the capsid protein precursor at its C terminus, leading to the production of the β and γ peptides (Agrawal & Johnson, 1992; Hanzlik *et al.*, 1995; Gordon *et al.*, 1999). This cleavage of the capsid protein precursor results in stabilization of the virus particle, which is proposed to be important for infectivity (Munshi *et al.*, 1996; Helgstrand *et al.*, 2004).

The members of the family *Tetraviridae* are classified into two genera according to their genome organization. Omegatetravirus particles encapsidate two genomic RNAs (RNA1 and RNA2) encoding the viral replicase and capsid protein precursor (VCAP), respectively. Betatetraviruses

†These authors contributed equally to this paper.

‡Present address: Institute of Molecular and Cellular Biology, University of Leeds, Leeds LS2 9JT, UK.

§Present address: Baldwins Intellectual Property, PO Box 852, Wellington 6140, New Zealand.

||Present address: Spanish Medicines Agency (AEMPS), Parque Empresarial Las Mercedes, Campezo 1 - Edificio 8, 28022 Madrid, Spain.

The GenBank/EMBL/DBJ accession number for the sequence reported in this paper is GU991616.

have a single (monopartite) genomic RNA that encodes both replicase and VCAP often on overlapping reading frames, with VCAP translated from a subgenomic (sg) RNA (Hanzlik *et al.*, 2005). This sgRNA may be encapsidated along with the genomic strand (Gordon *et al.*, 1999; Pringle *et al.*, 1999, 2003). Omegatetraviral RNAs are capped at their 5' ends and are not polyadenylated, encoding instead a tRNA-like structure at their 3' ends (Gordon *et al.*, 1995; Hanzlik *et al.*, 1995; Du Plessis *et al.*, 2005; Yi *et al.*, 2005). Also betatetraviral genomic RNAs are not polyadenylated at their 3' ends, nor are they blocked at the 5' end (King & Moore, 1984; Gordon *et al.*, 1999; Pringle *et al.*, 2003). The tRNA-like structure has thus far only been identified at the 3' end of the type species of this genus, *Nudaurelia capensis* β virus (N β V). Putative pseudoknots are present at the 3' ends of *Thosea asigna* virus (TaV) and *Euprosterna elaeasa* virus (EeV) (Gordon *et al.*, 1999; Zeddarn *et al.*, 2010).

Research on the members of the family *Tetraviridae* has focused on the molecular characterization of their genomes and fundamental studies on capsid structure and assembly. In contrast, little is known about the biology of this family because of their narrow host range and the lack of tissue culture cell lines that support tetravirus replication (Moore, 1991; Bawden *et al.*, 1999). Providence virus (PrV), which was isolated from a persistently infected *Helicoverpa zea* midgut cell line (MG8), is the only tetravirus known to replicate in cell culture. PrV does not appear to be very infectious, but high titres of the virus can infect *H. zea* fat body (FB33) cells and the virus also replicates in an embryonic *Spodoptera exigua* (Se-1) cell line (Pringle *et al.*, 2003). As with the other tetraviruses, PrV particles are assembled from a single capsid protein precursor (p68), which is autoproteolytically cleaved during maturation to form the β and γ peptides of 60 and 7.4 kDa, respectively. PrV particles encapsidate a genomic RNA of 6.4 kb as well as a sgRNA of 2.5 kb that encodes the viral capsid protein precursor, which has resulted in the classification of PrV as a betatetravirus (Pringle *et al.*, 2003; Hanzlik *et al.*, 2005). Analysis of a cDNA sequence derived from the sgRNA showed the presence of an ORF encoding a protein with predicted M_r of 81 kDa, encoding a 2A-like processing site at the N terminus of the VCAP-coding sequence. Processing at this 2A site would result in the production of two translation products of 13 (p13) and 68 (p68) kDa, the larger of which encodes the capsid protein precursor (Pringle *et al.*, 2003). An analogous 2A-like site is present in the VCAP ORFs of TaV and EeV (Pringle *et al.*, 1999; Zeddarn *et al.*, 2010), but is absent in N β V (Gordon *et al.*, 1999). Exactly how p13 functions in the betatetraviral life cycle is unknown, but it is thought to stabilize TaV virus-like particles (Pringle *et al.*, 2001).

The ability of PrV to replicate in tissue culture cells presents a unique opportunity for studying the replication biology of a member of the family *Tetraviridae* and so we have set about characterizing the PrV genome organization and the expression of viral translation products *in vitro* and *in vivo*. We report here that this virus has a novel genome

arrangement and unique co-translational processing strategies for regulating the expression of the viral gene products. We also show that PrV is phylogenetically more related to members of the families *Tombusviridae* and *Umbraviridae* than to members of the family *Tetraviridae*, leading to us to propose that this virus could be reclassified as the first member of a novel family of insect +ve ssRNA viruses.

RESULTS

PrV has a unique genome organization

PrV has a monocistronic genome comprising 6155 nt that is neither polyadenylated at its 3' end, nor capped at the 5' end. The first 8 nt of the PrV genomic RNA (vRNA) and sgRNAs are identical (GGUAAUGA), a feature not shared with other members of the genus *Betatetraviridae*. The PrV vRNA does not appear to terminate in a tRNA-like structure, as is the case for the omegatetraviruses and N β V (Hanzlik *et al.*, 1995; Gordon *et al.*, 1995, 1999; Du Plessis *et al.*, 2005; Yi *et al.* 2005), nor is there evidence of the presence of a pseudoknot as has been found for EeV and TaV (Zeddarn *et al.*, 2010).

The PrV genome encodes three major ORFs (Fig. 1). Nearest to the 5' end of the genome is the largest ORF (nt 45–3707) encoding a translation product with a theoretical molecular mass of 130 kDa (p130). The translational start of p130 could also be at a second downstream AUG codon (nt 354–356), which like the first AUG, is in a good context for translation (ACAAAUGG and GGAGAUGA, respectively) compared to the consensus insect Kozak, CA(A/G)(C/G)AUG(G/A) (Cavener, 1987; Cavener & Ray, 1991). Translation from the second AUG (M103) would result in a protein of approximately 115 kDa. A putative 2A-like processing sequence (GDVESNPG↓P) is located at the N terminus of p130, the activity of which would produce two translation products of 17 and 113 kDa, respectively. BLASTP analyses failed to identify any homologous proteins in the available sequence databases and thus the function of these translation products is unknown.

Positioned between nt 1027 and 3775 is the replicase ORF producing a protein of 104 kDa (p104) in a +1 frame relative to p130. The p104 ORF has a Kozak sequence of UUGUAUGU, which is not in a particularly good context for initiation of translation, compared with the optimal invertebrate Kozak consensus sequence. An in-frame UAG stop codon is present in this ORF at nt position 2104, which would produce a truncated protein with a predicted molecular mass of 40 kDa (p40). The sequence and context of the stop codon (UAGCAACUA) corresponds to the consensus sequence for a read-through stop codon (UAGCARYYA) (Skuzeski *et al.*, 1991) that is most similar to the group 1 read-through stop signals (UAGCAAUU/CA), described by Harrell *et al.* (2002). More than 80 % of known read-through stop signals are clustered in group 1,

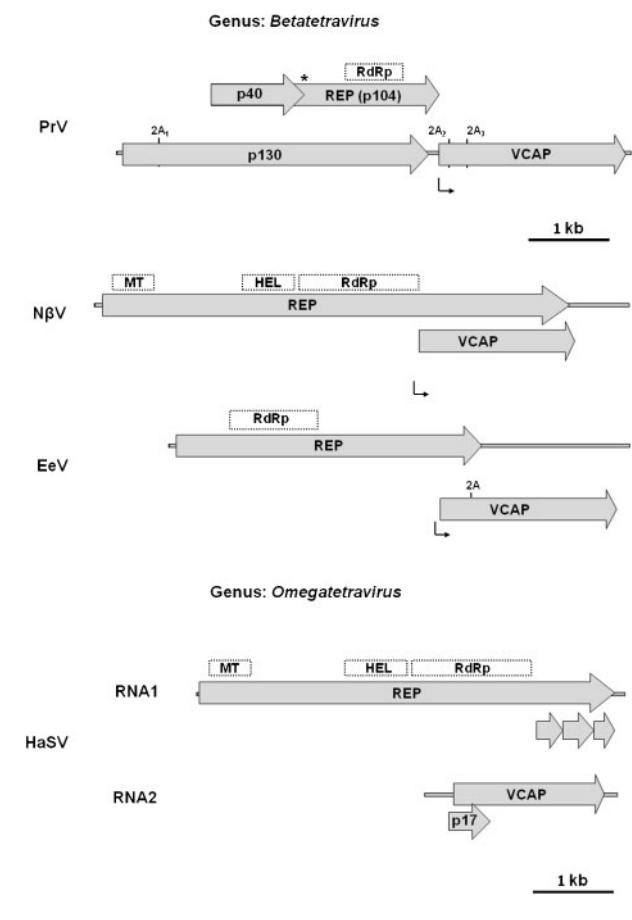


Fig. 1. Comparison of the genome organization of PrV with other members of the family *Tetraviridae*. The three PrV ORFs are indicated as p130, the viral replicase (REP) and the capsid protein precursor (VCAP). The positions of the three PrV 2A-like processing sites (2A₁, 2A₂ and 2A₃) are shown within the p130 and VCAP ORFs as well as the read-through stop codon (*) within the REP ORF that results in the translation of p40. The position of a 2A-like cleavage site within the N-terminal of the EeV VCAP is also indicated. The positions of RdRp as well as methyltransferase (MT) and helicase (HEL) domains are shown within the coding sequences of the tetraviral replicases (REP). The arrow indicates the start of the betatetraviral sgRNA. Virus sequence accession numbers: PrV, GU991616; *Nudaurelia capensis* β virus (NβV), AF102884; *Thosea asigna* virus (TaV), AF282930 and AF062037; *Helicoverpa armigera* stunt virus (HaSV), U18246 and L37299.

derived from (+ve) ssRNA plant viruses belonging to the families *Tobamovirus*, *Benyivirus* and *Pomovirus* (Harrell *et al.*, 2002). The divergent C nucleotide at the start of the second codon following the stop codon is unique to PrV within the group 1 sequences. There is little sequence conservation with stop signals found in other insect and animal viruses (data not shown), suggesting that this read-through stop signal is of group 1 plant virus origin.

The capsid precursor ORF begins just 4 nt after the UGA stop codon of p104 and is 2265 nt in length (nt 3779–6044),

with a Kozak sequence of GAUCAUGC, encoding a protein of 81 kDa (p81). There are two 2A-like processing sites positioned at the N terminus of p81, PrV-2A₂ (GDIESNPG↓P) and PrV-2A₃ (GDIESNPG↓P), which, if active, would produce translation products of 7, 8 and 68 kDa. Previous research by Pringle *et al.* (2003) has shown that the virus particle is composed of two proteins derived from maturation-dependent autoproteolytic cleavage of p68 to form the β and γ peptides of 60 and 7.4 kDa, respectively. This suggests that at least PrV-2A₃ is functional *in vivo*. Neither p7 nor p8 was detected in purified virus particles (Pringle *et al.*, 2003).

Phylogenetic relationship between PrV and other tetraviruses

Phylogenetic analysis of tetravirus RNA-dependent RNA polymerase (RdRp) sequences shows that they cluster in three distinct groups (Fig. 2), each belonging to a different superfamily of RNA viral replicases (Koonin, 1991). The alpha-like RdRp's of NβV and HaSV (and *Dendrolimus punctatus* tetravirus, DpTV) form a distinct cluster, while those of TaV and EeV are found in a second group, related to the birnaviruses, infectious pancreatic necrosis virus (IPNV) and infectious bursal disease virus (IBDV), both of which are dsRNA viruses. PrV is most closely related to the tombusvirus, *Pelargonium* chlorotic ring pattern virus (PCRPV), a (+ve) ssRNA plant virus and not to either of the other two groups of tetraviruses (Fig. 3). A BLASTP analysis using the PrV replicase sequence revealed significant similarity (up to 26% identity) with the replicase proteins of members of the (+ve) ssRNA virus families,

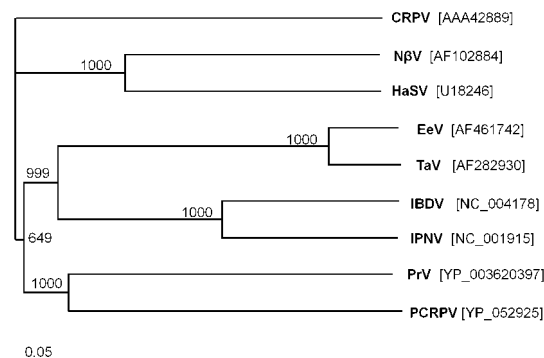


Fig. 2. Phylogenetic relationship between the RdRp domains of PrV and other members of the families *Tetraviridae*, *Birnaviridae* and *Tombusviridae*. The PrV RdRp sequence was aligned with those of the other tetraviruses, two birnaviruses, IPNV and IBDV, related to EeV and TaV, the tombusvirus PCRPV, related to PrV, and the genetic outlier CrPV. Bootstrap confidence values were calculated from 1000 bootstrap trials and are indicated at the node of each branch. The distance scale is shown in the bottom left-hand corner of the phylogram, generated in TreeView. GenBank accession numbers are indicated next to the sequence name in parentheses.

Umbraviridae and *Tombusviridae* and not with any of the other tetravirus replicases. Tombus- and umbraviruses also utilize a read-through stop signal as a means to regulating RdRp expression yet interestingly, the plant viruses that are most closely related to PrV contain group 3 read-through stop signals with the exception of carrot mottle mimic virus which contains a group 4 signal. The consensus for group 3 and group 4 read-through signals, whose members are comprised mainly of tombus- and umbraviruses, are defined as UAGGGGUGU/C and UAGGGAGGC, respectively (Harrell *et al.*, 2002). Bioinformatic analysis of the PrV replicase sequence shows the presence of an RdRp located in the C-terminal region of the protein, downstream of the read-through stop sequence (Fig. 3). As in the tombusvirus and umbraviruses, the PrV RdRp contains a canonical A-B-C palm subdomain located between amino acid residues 583 and 724 and four other conserved sequences identified by Koonin (1991), all of which are also conserved in tombus- and umbraviral replicases (Fig. 3). Interestingly, there appears to be a second potential A motif downstream of the GDD box (C motif) and preceding the conserved VII or E motif (not present in the tombus- and umbraviral replicases), raising the possibility of a permuted (C-A-B) palm subdomain similar to those identified in EeV and TaV (Gorbalenya *et al.*, 2002). However, a poorly conserved B motif further downstream and lack of sequence conservation of the second potential A motif suggest that the presence of a second, permuted, palm subdomain is unlikely. The absence of methyltransferase and helicase domains together with the A-B-C arrangement of the palm subdomain suggests that the PrV replicase may belong within the carmo-like supergroup. This is supported by the observation that the PrV replicase is most closely related to members of the families *Umbraviridae* and *Tombusviridae*, which are classified within the carmo-like virus supergroup (Fig. 3).

Analysis of viral translation products

We used the persistently PrV-infected *H. zea* MG8 cell line for *in vivo* metabolic labelling of viral translation products because the infection of the *H. zea* FB33 and *S. exigua* 1 cell lines with PrV was not high enough to detect labelled proteins. MG8 cells were treated with actinomycin D to inhibit cellular transcription, the viral translation products labelled with [³⁵S] methionine and then analysed by SDS-PAGE followed by autoradiography. Four dominant protein species were detected with relative molecular masses of approximately 130, 63, 60 and 40 kDa of which the 130 kDa band was the most prominent. In addition, fainter bands of approximately 104 and 68 kDa were also detected (Fig. 4a). In the absence of an uninfected MG8 cell line, Western blot analysis was used to determine the origin of these proteins, using rabbit polyclonal antibodies raised against the N-terminal translation product of the PrV replicase (anti-p40) and PrV virus particles (anti-VCAP). Proteins corresponding to 40 and 104 kDa were detected by the anti-p40 antiserum in protein extracts derived from

MG8 cells, but not in uninfected FB33 cells (Fig. 4b, compare lanes 1 and 2), indicating that the bands of similar size detected by *in vivo* metabolic labelling (Fig. 4a) were probably the translation products of the replicase ORF. As was observed in metabolic labelling experiments, the 40 kDa protein appeared to be present at higher concentrations than the 104 kDa protein (Fig. 4a, b), suggesting co-translational regulation of the expression of the two replicase translation products relative to each other. A Western blot using anti-VCAP antiserum detected proteins of 68 and 60 kDa, corresponding in size to the capsid protein precursor (p68) and mature capsid protein β (p60) as well as two degradation products of approximately 40 and 30 kDa (Fig. 4c, lane 3). The 8 kDa gamma peptide was not detected using two different sources of polyclonal anti-antisera raised against wild-type virus particles, consistent with the observations of Pringle *et al.* (2003). These proteins were not present in uninfected FB33 cells (Fig. 4c, lane 4). We were unable to confirm the identity of the prominent band of approximately 130 kDa detected by *in vivo* metabolic labelling of MG8 cells (Fig. 4a) because of a lack of suitable antibodies but, based upon its size, the assumption is that this protein is translated from the large ORF at the 5' end of the PrV genome.

Coupled *in vitro* transcription-translation reactions were used to further characterize translation of p104. Two proteins of approximately 40 and 104 kDa were detected when rabbit reticulocyte extracts were programmed with the plasmid, pSTA5, encoding the native p104 ORF (Fig. 5a, lane 3). In contrast to PrV-infected cells, p104 was the major translation product with very little p40 produced. When extracts were programmed with pSTA4, which encodes the N-terminal half of p104 encoding p40 with the native read-through stop replaced with a legitimate (UAA) stop codon, only the 40 kDa protein was detected (Fig. 5a, lane 2). Neither of these two proteins was present in unprogrammed control reactions (Fig. 5a, lane 1). A protein of 40 kDa was also translated in wheatgerm extracts programmed with pSTA4 (Fig. 5a, lane 4) and both p40 and p104 were translated from pSTA5 (Fig. 5a, lane 5). However, p40 was the major translation product in the wheatgerm extracts, with substantially lower levels of p104. These levels of expression are consistent with those observed in *in vivo* labelling experiments (Fig. 4a).

Finally, we conducted coupled *in vitro* transcription-translation experiments to characterize the translation products of the p81 ORF. Rabbit reticulocyte and wheat germ lysates were programmed with constructs expressing either the full-length p81 ORF (pSTA8) or a truncated p81 ORF encoding the virus protein precursor, p68 (pSTA7). Two proteins, of 68 and 15 kDa, corresponding in size to the capsid protein precursor and the product of 2A-like processing by PrV-2A₃, were detected in both rabbit reticulate and wheatgerm lysates programmed with pSTA8 (Fig. 5b, lanes 3 and 7). Only the 68 kDa protein was present in lysates programmed with pSTA7 (Fig. 5b, lanes 2 and 6). Western blot analysis using anti-VCAP

Tetravirus	PRV	WEWLLGRITPTKXQLKWVIRPGVCDPADPPVREKARIIEPELMTQCQDYPK--TOMLSI	403
	PCRPFV	RVVLCFCMPVQEAQFAKXGCLMQTCGVDTKVFERGEHHWVKEAHEASYIP--PRLFSI	278
	MNSP	RTIDNLCGLPDWKAFLINXGCEELAGFCTSVHRGTHPDMTEYPPQERPVK--TKLFCL	307
	MCMV	NWVRHLGGDDPALFQELKXGCEELWGVQSRRTARHPKLRSHFQDTRHK--TRRVFR	477
	OCST	VAGENLFSRIAQWNP-GKXGCPVVVTGGCVISRAPDCAQLRVKRLGVT-KNRTCMRV	252
Tombusvirus	GRV	WFQPTLCDPGYQGRGTFKLVGEVYARSQVETKTRPKILSPQIKVLAARPRPVKVSXNV	327
	CMV	HTHPG-----FLLGDVIARDGVCTKTPVILSPSLHIRFSARPRPVKVSQI	70
	OPMV	WFQPTLVDPYSLGLRIFSWGCVRSS-CGVGLDGR-NIITPAIRVIRARPRAVEVSYQI	161
	TBTV	WAQG-----FLLGDVVARRGVETKVKTKIH-PKIRVLRAARPRPVKVSQI	47
	PEMV	VADPDRRPSLWSAGGFLVGAFLPRLGDTSVRFNDLHPSIEVIRAKARPRPTQMSYQI	343
* : : *			
Tetravirus	PRV	QFPEFYSLWAHEDCSHNQLRAVKGRVLGKVPV--TPVATSMLLPVMALVASSLPST	460
	PCRPFV	SGLGPTIDWGVHNCYVNLRLRGIMRVYFVER---KGKLEPCPPPAEGEFARLFKKYG	333
	MNSP	GGVGTSVQVFNHNSMRNLRLRGIMRVYFVENE---KKELEPAPKPLDGVFG-NLAWFR	362
	MCMV	AGLGNLYEFGVHNSAVNLERGLMRVYFVKDD---KGLVSCPEPISGIFWKLKGRF	533
	OCST	SGVSPNIQIIIPFNNDITTLERAIKERVFFVKNLDKGSPTKFPVSPRPAPGFFAQLSNTL	312
Tombusvirus	GRV	DVLRPSADYGVHNSLNLVRCVNERVYFTNNQ---KLPLAPAGAYQIDCAALKQFR	384
	CMV	DAVRPSADFGVHNSLNLVRCVNERVYFTDNRG---SLPVRPTPGFRDLT-VDIKHF	126
	OPMV	DVVRPCADFGVHNSLNLVRCVNERVYFTDHRK---SSPTKPLPGAFDEICVSLKSF	218
	TBTV	DVVRPCADFGVHNSLNLVRCVNERVYFTDHRK---KEPRRPSAGSFDKIDISEIKAF	104
	PEMV	DVVRPLGDFGVHNSLNLVRCVNERVYFTDNAR---TEPLKP---KVPFSSRELKTF	397
: : * : *			
Tetravirus	PRV	-----APLSLEECISEFPQNRNRRTALSEYLSRGITPK-DATVKAIVRMERLQ	509
	PCRPFV	HRVVGLCGSHSKISREEFVGSYQGRKRTIYNAMTSLLDKPLELR-DSFLSHVRCERLN	392
	MNSP	RKLHNIVGTHSCVSPGQFLDYTGRRRTIYEGAVKSLEGLGVQR-DAYLKVTVKAEKIN	421
	MCMV	NSIVHHVGRHRFPVRETFLAYTGPKRITMEKAVNSLYEMFVSVD-DAKLKVTVKAEKIN	592
	OCST	GLLVFPLPSTAPMSHQFVDSYTPSRKRKVYQALIEDISCHGLNLTDSKVKVYVVEKTD	372
Tombusvirus	GRV	-----VTPFLDDVMSYKGSQVRVYKQAVDSLGMRPLSKS-DARVSTIRKAEVN	434
	CMV	-----INAWMEDVNSYTGROARVEQAMLSIMERPLERR-DARVSTIRKAEKIN	176
	OPMV	-----VTPWIMEQVDSYAGSQNTRYQNALESINTQPLTRS-DARVSTIRKAEKIN	268
	TBTV	-----VQFWLEEVVDSYTGSRVRYGQAVESLAVTFLSRN-DARVSTIRKAEKIN	154
	PEMV	-----VTPWIMDRVVESYTGSRTRMANARDSILSNFLSPK-DARVSTIRKAEKIN	447
: : * : *			
Tetravirus	PRV	IIILRHG-DPRLTQSGRTVTHLLFSTMTKPLTHALMHVSDPETNLPMAKCRNLDQRAALL	568
	PCRPFV	FSKRPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	452
	MNSP	TTKRDPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	481
	MCMV	LTKNADFPVRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	652
	OCST	HTSKADFPVRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	432
Tombusvirus	GRV	FRKRPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	494
	CMV	FSKRPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	236
	OPMV	FTAKRPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	328
	TBTV	FTAKRPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	214
	PEMV	FTAKRPDPAPRVHQPDPRENVELGCLYKREHSAIRALDKLWGGPTVMKGTIVQEIIGKII	507
* : : * : *			
Tetravirus	PRV	QRMWRLYERFVSISFDSFDMHYOVPLSRVLEISQHVIT-----CPLLDQMCQNLL	621
	PCRPFV	ESKWNEYRFPVAIGFDMRSRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	509
	MNSP	RDANDSFVSPVAIGFDMRSRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	538
	MCMV	ENARSEFTDEVAIGFDMRSRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	709
	OCST	QHKWNNENKFPVAIGFDMRSRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	492
Tombusvirus	GRV	AKKWMEANFPVCVGLDASRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	550
	CMV	RKKWDSEFNFPVCVGLDASRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	292
	OPMV	AKKWMEANFPVCVGLDASRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	384
	TBTV	AKKWMEANFPVCVGLDASRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	270
	PEMV	AKKWMEANFPVCVGLDASRDOHVSVDALKEFKKIKRVFR---GDQNLARLDQWQLENR	563
: : * : *			

antiserum confirmed the presence of p68, but failed to detect the 15 kDa peptide (Fig. 5b, lane 5)

DISCUSSION

PrV is unique amongst the members of the family *Tetraviridae* in that it was first discovered in a persistently infected *H. zea* midgut cell line and it is currently the only tetravirus able to replicate in tissue culture. Our analysis of the viral genome shows that PrV has a monopartite genome (hence its classification as a betatetravirus), with both the vRNA and sgRNA starting with the same eight nucleotide sequence. These conserved sequences may have a function in directing replication of viral RNAs since both vRNA and sgRNAs are replicated *in vivo* (Pringle *et al.*, 2003). As in other betatetraviruses, the viral genome

encodes the viral replicase, p104, followed by the virus capsid protein precursor gene, p81. However, unlike other tetraviruses, the PrV genome encodes a third ORF (p130) positioned 5' to and almost entirely overlapping the replicase ORF. *In vivo* radio-labelling of viral translation products indicated the presence of proteins corresponding in size to p130, but we were unable to confirm the identity of this protein due to the lack of suitable antibodies. The absence of any homology to proteins in the electronic databases suggests that p130 may have a unique function in the virus life cycle and it is tempting to speculate on a role in maintaining the persistent infection in midgut cells.

Our data has shed light on the potential for co-translational regulation of the expression of PrV proteins. The viral replicase ORF encodes a functional group 1 read-through stop signal that results in the differential expres-

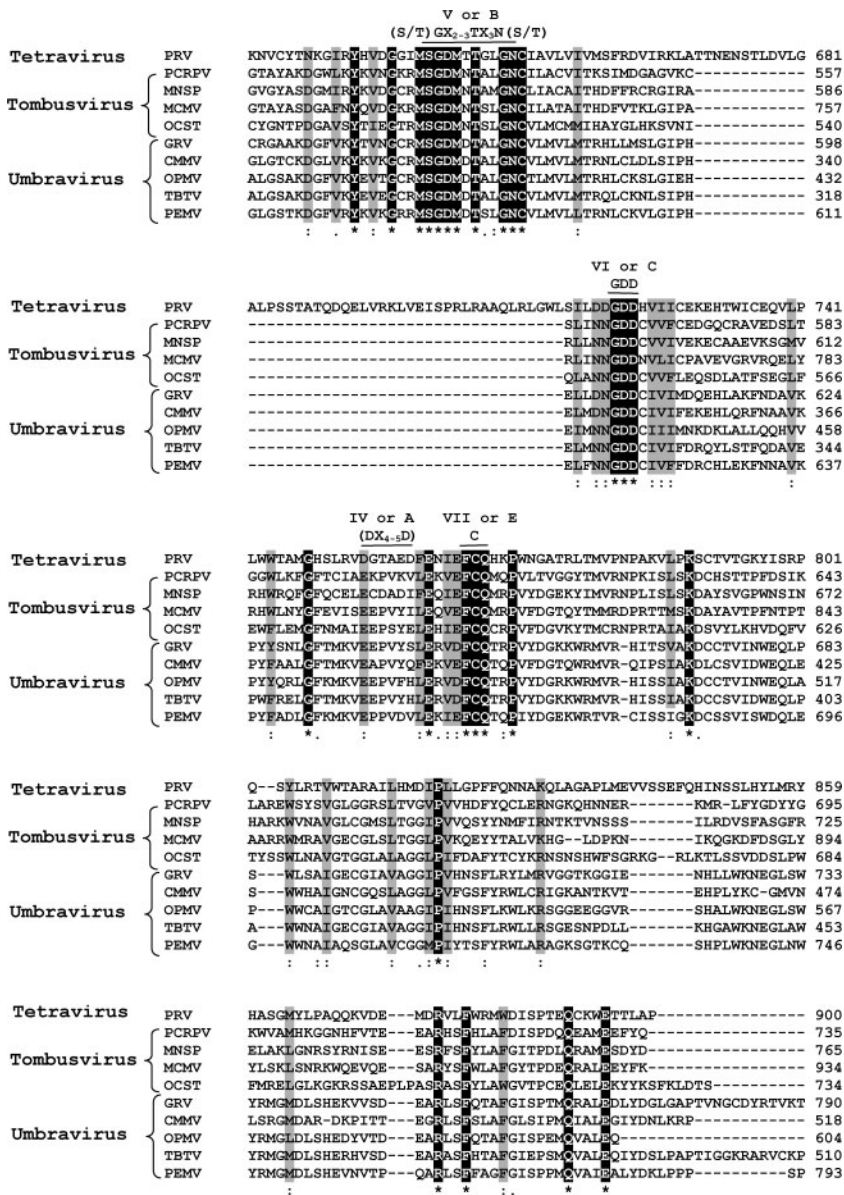


Fig. 3. Multiple sequence alignment of the RdRp domain of PrV p104 with the nine most similar tombusvirus and umbravirus replicases. Amino acid residues which are conserved in all the virus sequences are shaded in black and indicated by an asterisk (*), while semi-conserved amino acids (grey shading) are indicated by a colon (:). (▼) indicates the position of the read-through stop in the PrV replicase (p104). The positions of seven conserved domains are indicated above the sequence with consensus sequence according to Kamer & Argos (1984), Poch *et al.* (1989) and Koonin (1991) indicated.

sion of two translation products, where there appears to be significantly higher levels of p40 than the full-length p104, which encodes the RdRp domain. p40 was also the major translation product in coupled *in vitro* transcription-translation reactions using wheatgerm lysates programmed with plasmids expressing wild-type p104. This result is consistent with the observation that the PrV stop signal sequence is similar to group 1 read-through stop signals, all of which are found in plant viruses. Interestingly, the activity of the PrV read-through stop signal appears to be suppressed in mammalian expression systems, because p104 was the major translation product observed in rabbit reticulocyte lysates. The PrV ORFs encode three 2A-like processing sequences that have been shown to be functional *in vitro* (Luke *et al.*, 2008). Coupled *in vitro* transcription-translation experiments show that PrV-2A₃ is functional, since a 15 kDa peptide in addition to the capsid

protein precursor (p68) was detected in both rabbit reticulocyte and wheatgerm lysates. A protein corresponding in size to this peptide was not observed in metabolically labelled MG8 cells. The data imply that PrV-2A₂ is not functional in this specific polyprotein context or was less active than PrV-2A₃. We are at present unable to confirm this result *in vivo* since our anti-VCAP antiserum does not cross-react with the 15 kDa peptide. Likewise, the activity of PrV-2A₁ *in vivo* remains to be confirmed pending the availability of suitable antibodies.

The current members of the family *Tetraviridae* have been classified according to their characteristic *T*=4 capsid morphology with the two genera representing viruses with a monopartite (*Betatetravirus*) or bipartite (*Omegatetravirus*) genome organization. The recent elucidation of the PrV crystal structure has revealed structural conservation between

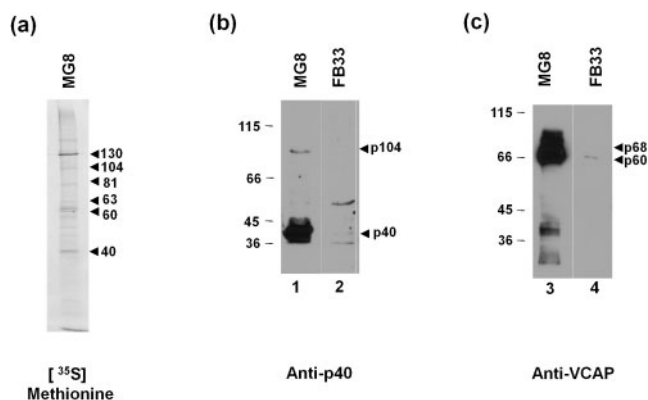


Fig. 4. *In vivo* metabolic labelling of PrV translation products. (a) Autoradiograph of SDS-PAGE used for analysis of [³⁵S]-labelled methionine viral translation products in *H. zea* MG8 cells. A total of 10 µg of a crude protein extract derived from MG8 cells treated with actinomycin D and then cultured in medium containing [³⁵S]-labelled methionine. The relative molecular mass of each protein species was determined by linear regression, using the marker proteins (visualized by Coomassie staining of the SDS-PAGE gels prior to autoradiography). (b) Western blot analysis of crude protein extracts derived from *H. zea* MG8 and FB33 cells using anti-p40 antiserum to detect the viral replicase or (c) using anti-VCAP antiserum to detect the viral capsid proteins. Relative molecular mass is indicated in kDa on the right.

the jellyroll folds and autoproteolytic cleavage sites of PrV and NoV. Unexpectedly, the N and C termini, which comprise the molecular switches that determine the $T=4$ architecture of these two virus particles, are completely different (Speir *et al.*, 2010). The PrV molecular switch most closely resembles that of the $T=3$ nodavirus particles, which are proposed to have given rise to the primordial tetravirus particle. The data have led Speir *et al.* (2010) to hypothesize that the PrV capsid is more closely related to an ancestral $T=4$ capsid from which the NoV particle evolved. This hypothesis is supported by the results of the phylogenetic analysis that shows that the three omegatetravirus capsid protein precursors (including NoV), may have evolved more recently, and share a common ancestor with PrV.

Phylogenetic analysis of tetravirus RdRp sequences has shown that they cluster into three distinct groups, each belonging to a different superfamily of viral RNA replicases (Koonin, 1991; van der Heijden & Bol, 2002). The alpha-like RdRp's of NβV and HaSV (and DpTV) form a distinct cluster. All three of these tetraviruses encode a tRNA-like structure at the 3' end of their vRNAs. Gorbalenya *et al.* (2002) and Zeddard *et al.* (2010) have shown that TaV and EeV form a group distantly related to the birnaviruses, IPNV and IBDV, which are picorna-like viruses. Unlike the tetraviruses with alpha-like replicases, TaV and EeV encode pseudoknots and not tRNA-like structures at the 3' ends of their vRNA. PrV represents a third group, which is more closely related to the tombusvirus, PCRPV, and other

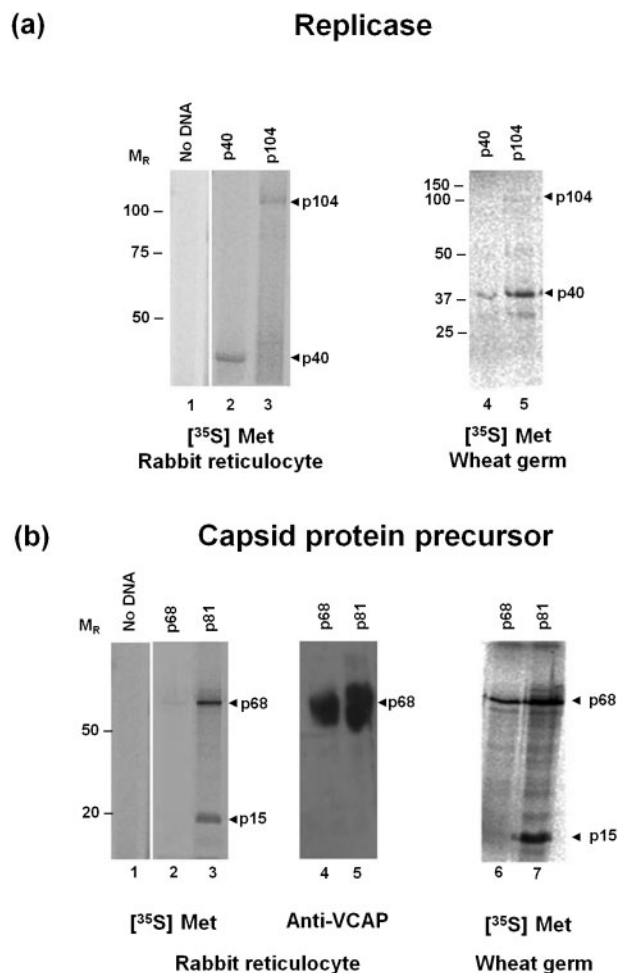


Fig. 5. Coupled *in vitro* transcription-translation products of constructs expressing the replicase or VCAP ORFs. (a) Coupled *in vitro* transcription-translation products of constructs expressing p40/p104. Lanes 1–3, autoradiograph of SDS-PAGE used to resolve [³⁵S]-labelled methionine translation products in rabbit reticulocyte lysates: lane 1, no input DNA; lane 2, expression construct pSTA4 (p40) and lane 3, pSTA5 (p104). Lanes 4 and 5, autoradiograph of SDS-PAGE used to resolve [³⁵S]-labelled methionine translation products in wheatgerm lysates: lane 4, expression construct pSTA4 (p40) and lane 5, pSTA5 (p104). Relative molecular mass is indicated in kDa on the right. (b) Lanes 1–3, autoradiograph of SDS-PAGE used to resolve [³⁵S]-labelled methionine translation products in rabbit reticulocyte lysates. Lane 1, no input DNA; lane 2, expression construct pSTA8 (p68) and lane 3, pSTA7 (p81). Lanes 4 and 5, Western blot analysis of *in vitro* translation products produced in rabbit reticulocyte extracts, using anti-VCAP antiserum. Lane 4, pSTA8 and lane 5, pSTA7. Lanes 6 and 7, autoradiograph of SDS-PAGE used to resolve [³⁵S]-labelled methionine translation products in wheatgerm lysates. Lane 6, pSTA8 (p68) and lane 7, pSTA7 (p81). The positions of molecular mass markers (in kDa) are indicated on the left and viral translation products on the right.

carmo-like viruses than to any of the tetraviruses. The PrV replicase is unique among the members of the family *Tetraviridae* in that it encodes a functional read-through

stop signal that results in two translation products from the same ORF. PrV is also distinct from all other tetraviruses with respect to genome organization and the presence of a third large ORF (p130) as well as the absence of 3' tRNA-like structures or pseudoknots.

Taken together, these observations lead us to propose that the family *Tetraviridae* could be reclassified into three separate virus families: (i) a prototypic family, comprising existing tetraviruses with alpha-like replicases (HaSV, DpTV, NβV and most likely NωV), including two genera, the monopartite betatetraviruses and the bipartite omegatetraviruses; (ii) a second family proposed by Zeddum *et al.* (2010), comprising the tetraviruses with permuted picorna-like replicases (TaV and EeV) and (iii) a third family with PrV as the sole member representing tetraviruses with carmo-like viral replicases.

METHODS

Analysis of the PrV cDNA sequence. The cDNA sequence of PrV was obtained by sequencing overlapping cDNA clones and using 5'-RACE and/or primer extension to determine the 5' ends of the genomic and sgRNAs. Junction RT-PCR (as described in Albarrino *et al.*, 2001) was used to determine the sequence of the 3' ends because attempts to obtain the sequence from the (+ve) RNA (3'-RACE) or negative strand using 5'-RACE were unsuccessful. The consensus cDNA sequence has been deposited in GenBank (accession no. GU991616). Full-length cDNA clones of vRNA and sgRNA were assembled from assorted RT-PCR, 5'-RACE and 3'-RACE clones. The presence of the ORFs within the cDNA sequence was determined using the GCG software package. CLUSTAL W (Thompson *et al.*, 1994) was used to align the amino acid sequences of the replicase proteins of PrV versus nine tombusvirus replicase sequences, the replicase sequences of the tetraviruses NβV, HaSV, TaV, EeV, the birnaviruses IPNV and IBDV as well as a genetic outlier, cricket paralysis virus (CrPV). CLUSTAL_X (Larkin *et al.*, 2007) was used to analyse the resultant alignment and predict bootstrap confidence values of the alignment using the neighbour-joining algorithm. A total of 1000 bootstrap trials were conducted on each submitted alignment file and 111 used as a random number generator seed. The graphical representation of the phylogeny was created using TreeView (Page, 1996). The GenBank accession numbers of the tombusvirus sequences used for the multiple sequence alignment were as follows: PCRPV, AAT69549.1; melon necrotic spot virus (MNSV), AAB02430.1; maize chlorotic mottle virus (MCMV), NP_619718.1; oat chlorotic stunt virus (OCSV), NP_619751.1; groundnut rosette virus (GRV), NP_619714.1; carrot mottle mimic virus (CMMV), AAB81639.1; opium poppies mosaic virus (OPMV), ABV66257.1; tobacco bushy top virus (TBTv), NP733848.1; pea enation mosaic virus-2 (PEMV), NP_620846.2.

Tissue culture cell lines. *H. zea* fat body (BCIRL-HZ-FB33 and referred to here as FB33) cells (Kariuki *et al.*, 2000) were grown in TC100 medium containing 1% penicillin and streptomycin as well as 5% heat-inactivated FCS. *H. zea* midgut cells (BCIRL-HZ-MG8 and referred to here as MG8) were grown in ExCell 420 medium containing 0.5% penicillin and streptomycin without FCS (Pringle *et al.*, 2003).

Polyclonal antibodies. Virus was extracted from MG8 cells as described by Pringle *et al.* (2003) and a total of 300 µg purified PrV was used to generate rabbit polyclonal antiserum against the virions (anti-VCAP antiserum). A predicted antigenic peptide corresponding to amino acid residues 248–331 of the p104-coding sequence

immediately upstream of the read-through stop codon was used to raise anti-p104 rabbit polyclonal antibodies. The peptide (p104_{248–331}) was fused at its N terminus to glutathione S-transferase (GST), overexpressed in *Escherichia coli* BL21(DE3) cells and the resulting 37 kDa fusion protein purified using the affinity GSTrap system (Amersham Biosciences). A total of 500 µg purified fusion protein was used to generate rabbit polyclonal antibodies that were able to detect both p104 and p40 *in vivo*.

In vivo radiolabelling of PrV translation products. *H. zea* MG8 cells were allowed to grow to 90% confluence in ExCell 420 medium containing 0.5% penicillin and streptomycin in a T25 flask. The culture medium was removed and replaced with ExCell 420 medium minus methionine and containing 0.5% penicillin, 0.5% streptomycin and 20 mg actinomycin D ml⁻¹. The cells were incubated at 28 °C for 30 min after which the culture medium was removed and the cells rinsed twice with 1 ml methionine-free culture medium. Three millilitres of methionine-free medium containing actinomycin D (20 mg ml⁻¹) together with 5.55 MBq of [³⁵S]-labelled methionine (Amersham Biosciences) were added to the T25 flask and the mixture incubated at 28 °C for 24 h. Whole cells were dislodged from the flasks, pelleted by centrifugation (5000 g for 2 min), resuspended in PBS buffer (pH 7.4) containing 10 mM PMSF and lysed by passing through a 22-gauge syringe needle 20 times. A total of 10 µg cell-free extract was analysed by SDS-PAGE, the gel was fixed and stained with Coomassie brilliant blue, dried under vacuum at 80 °C and exposed to hyperfilm (GE Healthcare) for an appropriate duration of time to visualize radiolabelled proteins.

In vitro translation and radiolabel incorporation of expressed PrV ORFs. The expression plasmids used for analysing PrV translation products (Table 1) were constructed by cloning PCR amplification products between the *NcoI* and *NotI* restriction sites in the pBiEX-3 vector (Novagen). Expression vector pSTA4 contains the p40-coding sequence with a legitimate stop codon (UAA) introduced to replace the read-through stop signal (UAG) found in the native PrV sequence. The p40 ORF was amplified using primers CW04 (5'-TGCGAGCCTTTAAGTTTACGC-3') and CW06 (5'-TATATAGCG-GCCGCTTACTTTGTCGGGATTGTTCTTCC-3'). Construct pSTA5 was generated by amplifying the entire p104 ORF including the read-through stop signal, using primers CW06 (5'-TATATAGCGG-CCGCTCAATACCCTGACGGCCAGTCAATATC-3') and CW05. Plasmid pSTA7 was constructed by amplifying the capsid precursor-coding sequence (p81) using the primers CW09 (5'-CAG-AATTTACCAGTACCCAATG-3') and CW10 (5'-TATATAGCGGC-

Table 1. Expression constructs used for coupled *in vitro* transcription–translation studies

Plasmid	PrV genome sequence	PrV ORF	Predicted translation products (kDa)
pSTA4	1027–2013	p104 _(1–259) *	40
pSTA5	1027–3374	p104	40 + 104
pSTA7	3779–6044	p81	81 or 68 + 13† or 68 + 7 + 8‡
pSTA8	4148–6044	p81 _(122–754)	68

*Introduced UAA and not the native UAG stop codon.

†Dependent upon the activity of the PrV-2A₃ processing site.

‡Dependent upon the activity of the PrV-2A₂ and PrV-2A₃ processing sites.

CGCTTACTTGGCTCGTCTAGCGCGG-3'). Finally, pSTA8 was generated by amplifying a truncated VCAP ORF (p68) starting immediately downstream of the PrV-2A₃ processing site and introducing an AUG start codon at nt 4148 using the primers CW11 (5'-CCCCAGATGACCCGCGAGAATAAAG-3') and CW10. Coupled *in vitro* transcription-translation reactions were performed using a rabbit reticulocyte lysate TnT T7 system or wheatgerm extract (Promega) according to the manufacturer's instructions. The reactions (in a total volume of 25 µl) were incubated at 30 °C for 90 min, 5 µl each radiolabelled translation reaction was analysed by SDS-PAGE and the labelled proteins visualized by autoradiography.

ACKNOWLEDGEMENTS

This research was supported by a grant from the South African National Research Foundation (NRF) (GUN: 65512). Cheryl Walter and Ritah Nakayinga were supported by scholarships from the South African NRF. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to any of the above-mentioned donors.

REFERENCES

- Agrawal, D. K. & Johnson, J. E. (1992). Sequence and analysis of the capsid protein of *Nudaurelia capensis* omega virus, an insect virus with $T=4$ icosahedral symmetry. *Virology* **190**, 806–814.
- Albariño, C. G., Price, B. D., Eckerle, L. D. & Ball, L. A. (2001). Characterization and template properties of RNA dimers generated during flock house virus RNA replication. *Virology* **289**, 269–282.
- Bawden, A. L., Gordon, K. H. J. & Hanzlik, T. N. (1999). The specificity of *Helicoverpa armigera* stunt virus infectivity. *J Invertebr Pathol* **74**, 156–163.
- Brooks, E. M., Gordon, K. H. J., Dorrian, S. J., Hines, E. R. & Hanzlik, T. N. (2002). Infection of its lepidopteran host by the *Helicoverpa armigera* stunt virus (Tetraviridae). *J Invertebr Pathol* **80**, 97–111.
- Cavener, D. R. (1987). Comparison of the consensus sequence flanking translational start sites in *Drosophila* and vertebrates. *Nucleic Acids Res* **15**, 1353–1361.
- Cavener, D. R. & Ray, S. C. (1991). Eukaryotic start and stop translation sites. *Nucleic Acids Res* **19**, 3185–3192.
- Christian, P. D., Dorian, S. J., Gordon, K. H. J. & Hanzlik, T. N. (2001). Pathology and properties of the tetravirus *Helicoverpa armigera* stunt virus. *Biol Control* **20**, 65–75.
- Du Plessis, L., Hendry, D. A., Dorrington, R. A., Hanzlik, T. N., Johnson, J. E. & Appel, M. (2005). Revised RNA2 sequence of the tetravirus, *Nudaurelia capensis* ω virus (N ω V). *Arch Virol* **150**, 2397–2402.
- Gorbalenya, A. E., Pringle, F. M., Zeddam, Z.-L., Luck, B. T., Cameron, C. E., Kalkmakoff, J., Hanzlik, T. N., Gordon, K. H. & 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. & Hanzlik, T. N. (1998). The Tetraviridae. In *The Insect Viruses*, pp. 269–299. Edited by L. K. Miller & L. A. Ball. New York: Plenum.
- Gordon, K. H. J., Johnson, K. N. & 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. *Virology* **208**, 84–98.
- Gordon, K. H. J., Williams, M. R., Hendry, D. A. & Hanzlik, T. N. (1999). Sequence of the genomic RNA of *Nudaurelia capensis* β virus (Tetraviridae) defines a novel genome organization. *Virology* **258**, 42–53.
- Hanzlik, T. N., Dorrian, S. J., Johnson, K. N., Brooks, E. M. & Gordon, K. H. J. (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., Gordon, K. H. J., Gorbalenya, A. E., Hendry, D. A., Pringle, F. M., Ward, V. K. & Zeddam, J.-L. (2005). Family Tetraviridae. In *Virus Taxonomy: Eighth Report of the International Committee on Taxonomy of Viruses*, pp. 877–883. Edited by C. M. Hanzlik, M. A. Mayo, J. Maniloff, U. Desselberger & L. A. Ball. San Diego, London: Elsevier Academic Press.
- Harrell, L., Melcher, U. & Atkins, F. J. (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. & Liljas, L. (2004). The refined structure of *Nudaurelia capensis* ω virus reveals control elements for a $T=4$ capsid maturation. *Virology* **318**, 192–203.
- Kamer, G. & Argos, P. (1984). Primary structural comparison of RNA dependent polymerases from plant, animal and bacterial virus. *Nucleic Acids Res* **12**, 7269–7282.
- Kariuki, C. W., McIntosh, A. H. & 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.
- King, L. A. & Moore, N. F. (1984). The RNAs of two viruses of the *Nudaurelia* β family share little homology and have no terminal poly(A) tracts. *FEMS Microbiol Lett* **26**, 41–43.
- Koonin, E. V. (1991). The phylogeny of RNA-dependent RNA polymerases of positive strand RNA viruses. *J Gen Virol* **72**, 2197–2206.
- Larkin, M. A., Blackshields, G., Brown, N. P., Chenna, R., McGettigan, P. A., McWilliam, H., Valentin, F., Wallace, I. M., Wilm, A. & other authors (2007). CLUSTAL W and CLUSTAL_X version 2.0. *Bioinformatics* **23**, 2947–2948.
- Luke, G. A., de Felipe, P., Lukashev, A., Kallioinen, S. E., Bruno, E. A. & Ryan, M. D. (2008). Occurrence, function and evolutionary origins of '2A-like' sequences in virus genomes. *J Gen Virol* **89**, 1036–1042.
- Moore, N. F. (1991). The *Nudaurelia* β family of insect viruses. In *Viruses of Invertebrates*, pp. 277–285. Edited by E. Kursak. New York: Marcel Dekker.
- Moore, N. F., Reavy, B. & King, L. A. (1985). General characteristics, gene organization and expression of small RNA viruses of insects. *J Gen Virol* **66**, 647–659.
- Munshi, S., Liljas, L., Cavarelli, J., Bomu, W., McKinney, B., Reddy, V. & 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.
- Olson, N. H., Baker, T. S., Johnson, J. E. & 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.
- Page, R. D. M. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Comput Appl Biosci* **12**, 357–358.
- Poch, O., Sauvaget, I., Delarue, M. & Tordo, N. (1989). Identification of four conserved motifs among the RNA-dependent polymerase encoding elements. *EMBO J* **8**, 3867–3874.
- Pringle, F. M., Gordon, K. H. J., Hanzlik, T. N., Kalkmakoff, J., Scottii, P. D. & Ward, V. (1999). A novel capsid expression strategy for *Thosea asigna* virus (Tetraviridae). *J Gen Virol* **80**, 1855–1863.

- Pringle, F. M., Kalmakoff, J. & 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. & Ball, L. A. (2003). Providence virus: a new member of the *Tetraviridae* that infects cultured insect cells. *Virology* **306**, 359–370.
- Skuzeski, J. M., Nichols, L. M., Gesteland, R. F. & 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.
- Speir, J. A., Taylor, D. J., Natarajana, P., Pringle, F. M., Ball, L. A. & Johnson, J. E. (2010). Evolution in action: N and C termini of subunits in related *T=4* viruses exchange roles as molecular switches. *Structure* **18**, 700–709.
- Thompson, J. D., Higgins, D. G. & 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.
- van der Heijden, M. W. & Bol, J. F. (2002). Composition of alphavirus-like replication complexes: involvement of virus and host encoded proteins. *Arch Virol* **147**, 875–898.
- Yi, F., Zhang, J., Yu, H., Liu, C., Wang, J. & 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.
- Zeddari, J.-L., Gordon, K. H. J., Lauber, C., Felipe Alves, C. A., Luke, B. T., Hanzlik, T. N., Ward, V. K. & Gorbalenya, A. E. (2010). *Euprosterna elaeasa* virus genome sequence and evolution of the *Tetraviridae* family: emergence of bipartite genomes and conservation of the VPg signal with the dsRNA *Birnaviridae* family. *Virology* **397**, 145–154.