Development of an experimental system to investigate the interaction between the *Helicoverpa armigera stunt virus* capsid protein and viral RNA

Andrew James Mascré Nel

Development of an experimental system to investigate the interaction between the *Helicoverpa armigera stunt virus* capsid protein and viral RNA

Thesis submitted in fulfilment of the requirements for the degree of

Master of Science

at

Rhodes University

by

Andrew James Mascré Nel

October 2004

Abstract

Tetraviruses are entomopathogenic viruses that propagate solely in lepidopteran hosts. Viruses of this group possess non-enveloped 38- to 40-nm capsids arranged in *T* = 4 surface symmetry. The viral genome consists of one or two single stranded positive sense RNA strands, which define the two genera of this family, the monopartite betatetraviruses and the bipartite omegatetraviruses. Two extensively studied members of the tetraviruses are the omegatetraviruses, *Helicoverpa armigera stunt virus* (HaSV) and the closely related *Nudaurelia capensis ω virus* (NωV). The larger genomic strand of HaSV (RNA1) encodes the viral replicase, while the other (RNA2) encodes the 71-kDa capsid precursor protein (p71). The pro-capsid is assembled from 240 copies of p71, which undergo a maturation auto-catalytic cleavage into the 64-kDa (p64) capsid protein and a 7-kDa peptide (p7) forming the capsid shell. The mechanism for the recognition and packaging of the viral genome is poorly understood for these viruses.

The principle objective of the research described in this study was to develop *in vitro* and *in vivo* experimental systems to investigate interactions between the N terminal domain of HaSV p71 and viral RNAs. More specifically, the two positively charged clusters of predominantly arginine residues that are conserved amongst tetraviruses and the structurally analologous nodaviruses capsid protomers' N terminal domains were investigated.

An *in vitro* RNA-protein "pull down" system was developed using the rapid protein purification technique of the IMPACT^{TM}-CN system. The coding sequence of the N terminal domain of p71 was fused to that of a chitin binding affinity tag (intein). This fusion protein was used as protein bait for the viral RNA. It was proposed that if RNA interacted with the fusion protein, it would be pulled down by the mass of affinity matrix and be precipitated and fluoresce when analysed by agarose gel electrophoresis using ethidium bromide. Despite optimisation of the *in vitro* assay, results were affected by the interaction between the intein-tag and nucleic acids, the state of the expressed fusion protein (in particular self-cleavage) and the excessive fluorescence present on the gels.

The *ADH2-GAPDH* yeast expression system was used to investigate the *in vivo* assembly of p71 containing deletions of either one or both clusters within N terminal domain. It was found that all p71 mutants were expressed with the exception of the mutant containing a deletion of the second cluster. The reasons for this still require further investigation. The expressed p71 mutants were not processed into p64 and were degraded *in vivo*. In addition, an experimental attempt to purify assembled p71 mutant VLPs was unsuccessful. The assembly defect of p71 mutants emphasised the significance of the clusters, which are possibly required for interaction with viral RNAs for efficient VLP assembly.

The results of this study suggest that an alternative tag or *in vitro* RNA-protein interaction assay be used. In addition, further experiments are required to investigate whether the co-expression of full length viral RNAs are required to rescue the *in vivo* assembly defect of p71 mutants into VLPs.

Contents

List of Figures

Chapter 1

Chapter 2

Chapter 3

List of Tables

Chapter 1

Chapter 2

Chapter 3

Abbreviations

General

Bp	Base pair
BSA	Bovine serum albumin
cryoEM	Cryoelectron microscopy
ddH_2O	Double-distilled water
DEPC	Diethyl Pyrocarbonate
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetate
Hepes	N-(2-Hydroxyethyl)piperazine-N'-(2-ethanesulfonic acid) hemisodium salt
IMPACT	Intein Mediated Purification with an Affinity Chitin-binding Tag
IPTG	Isopropyl-β-D-thiogalactoside
NEB	New England Biolabs
Nt	Nucleotide
OD_{600nm}	Optical Density at 600nm
ORF	Open Reading Frame
P_{AG}	*Saccharomyces cerevisiae ADH2-GAPDH* promoter
PAGE	Polyacrylamide Gel Electrophoresis
PCR	Polymerase Chain Reaction
P_{GAL}	*Saccharomyces cerevisiae GAL* promoter
PMSF	Phenylmethyl-sulfonyl fluoride
RNA	Ribonucleic Acid
RNase	Ribonuclease
SAP	Shrimp Alkaline Phosphatase
SDM	Site Directed Mutagenesis
SDS	Sodium Dodecyl Sulfate (Sodium lauryl sulfate)
SDS-PAGE	Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis
SMM	Synthetic Minimal Medium
ss +RNA	Single stranded, positive sense RNA
TEMED	N,N,N′,N′-Tetramethylethylenediamine
Tris	Tris(hydroxymethyl)aminomethane
tRNA	transfer RNA
VLPs	Virus-like particles

Viruses

BBV	*Black beetle virus*
BMV	*Brome mosaic virus*
BoV	*Boolarra virus*
EeV	*Euprosterna elaeasa virus*
FHV	*Flock House virus*
GGNNV	*Greasy grouper nervous necrosis virus*
HaSV	*Helicoverpa armigera stunt virus*
NβV	*Nudaurelia capensis beta virus*
NoV	*Nodamura virus*
NωV	*Nudaurelia capensis omega virus*
PaV	*Pariacoto virus*
PrV	*Providence virus*
SeMV	*Sesbania mosaic virus*
SJNNV	*Striped jack nervous necrosis virus*
TaV	*Thosea asigna virus*

Acknowledgments

Professor Rosemary Dorrington (project supervisor): I would like to thank Rosie for her inspiration, her no-nonsense approach, her time and patience. She has also encouraged and supported me through the course of this study. I acknowledge that I still have a lot of work to do in my training as a scientist, I look forward to working with her in the future.

Dr Caroline Knox (project co-supervisor): I would like to thank Caroline for her endless discussions and advice that have helped me to get through this project. In addition, I would like to thank her for her time and valuable input in the writing of this thesis.

Lab 417 & Rhodes University Microbiology, Biochemistry and Biotechnology Department: Firstly, I must thank Val Hodgson for being the manager of the lab and making sure everything is running smoothly also to the past members (Arno Venter, Fergus Manford, Carol Hartley, Anu Idicula, Sally Ann-Clark, and Kekeletso Mosisili) and the present members (Gwyneth Matcher, Mesbah Jiwaji, Petra Gentz, Cheryl Walter, Michele Tomasicchio) and the honours students. Thank you all for your enthusiasm, never-ending arguments, "extra"-smooth running of the lab., assistance, interesting gossip, advice, etc. In addition, to the many members of the Department, thank you for your assistant and support during my years at the department.

My Family: I would like to thank my parents, Owen & Michèle Nel for their love, support, advice, and patience during the course of my masters. In addition, I would like to acknowledge Nicole, my sister for her enthusiasm and being my closest friend, and, my grampa, François Mascré for his many interests and to my late gran, for her love. Thanks too, to the rest of the family for their interest and encouragement.

In addition, I must thank my friends Tanya & Nigel Mullins and their daughter Sofia, Marie Blanche Ting aka "Covy", Simisha Pather, Tracey Platt, Anu Idicula for listening to my woes and being my friends and also to my digsmates of 5 South Street, the twins Charles and Richard Antrobus, Brendan Robus and Amy Heydenrych - A big thank you for making digs life fun!

For funding of this research, I gratefully acknowledge the financial support received from Rhodes University in the form of a Post-graduate scholarship.

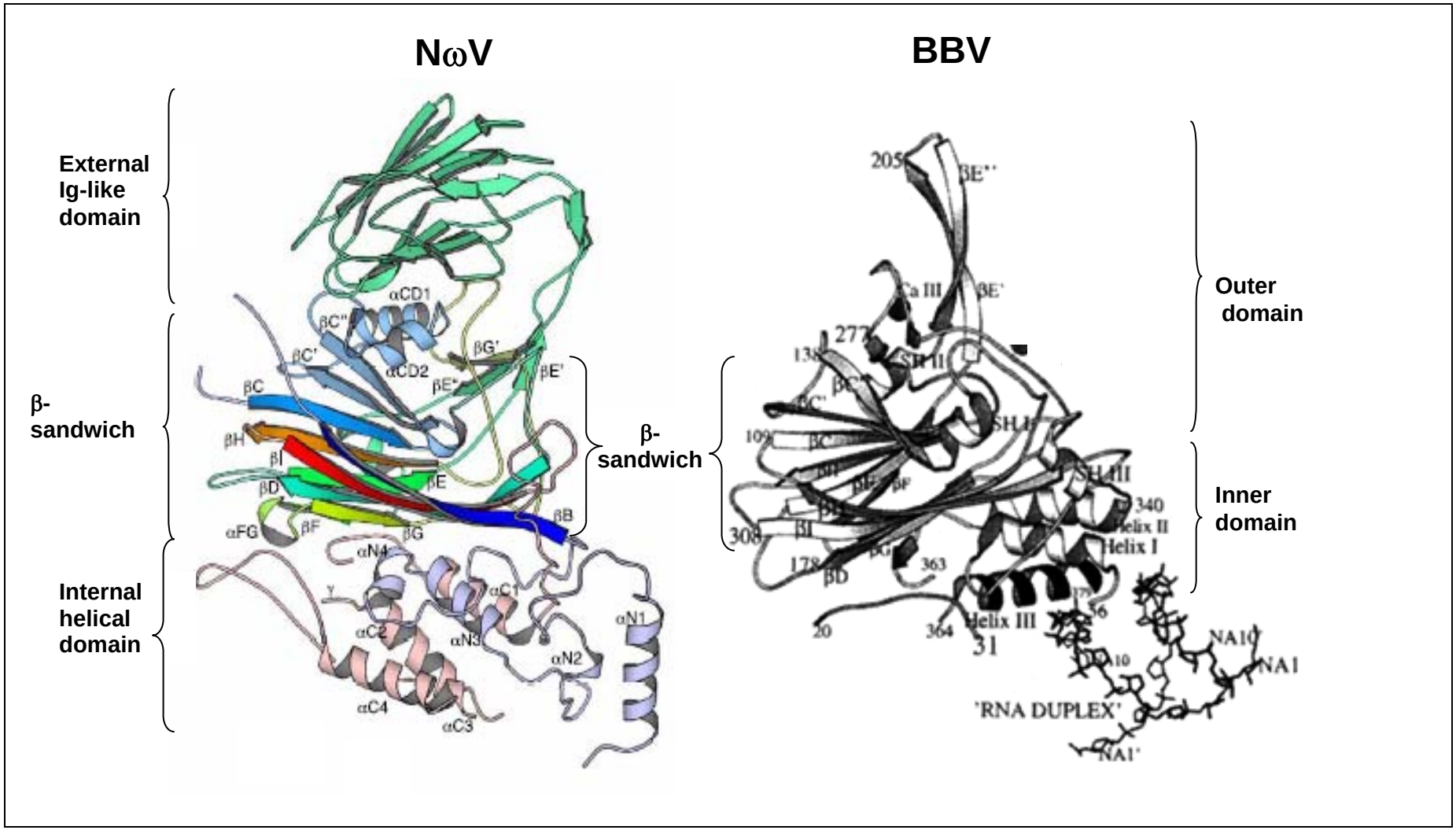


Figure 1.5 Subunit tertiary structures of NωV and BBV in C conformations. The respective structural domains are outlined and the similar β-sandwich for both virus subunits is outlined in the middle. NωV taken from Helgstrand *et al*., (2004) and BBV taken from Johnson and Reddy (1998).

Chapter 1: Literature review

1.1 Introduction

Tetra- and nodaviruses are non enveloped, single-stranded positive sense RNA (ss +RNA) viruses, which assemble their capsid shells from numerous copies of a single coat protein (Johnson and Reddy, 1998). These capsid shells are icosahedral and respectively display *T*= 4 and *T*= 3 quasi-symmetry. Much interest has been shown in these viruses, since their simple capsid structure provides a model for studying similar icosahedral complex viruses that assemble their capsids from multiple copies of different coat proteins (Schneemann *et al*., 1998).

Assembly of noda- and tetraviruses involve at least two types of macromolecular interactions: Protein protein interactions, which are required for the formation of the viral capsid, and protein RNA interactions, which are required for recognition and packaging of the viral genome (Johnson and Reddy, 1998). Protein protein interactions are well understood for both virus families, since most information on their capsid structure has been resolved to near atomic resolution. The capsid structure has been resolved for the tetravirus, *Nudaurelia capensis ω virus* (NωV) (Helgstrand *et al*., 2004), and four nodaviruses, *Black beetle virus* (BBV) (Hosur *et al.,* 1987; Wery *et al*., 1994), *Flock House virus* (FHV) (Fisher and Johnson, 1993), *Nodamura virus* (NoV) (Zlotnick *et al*., 1994) and the *Pariacoto virus* (PaV) (Tang *et al*., 2001). Comparisons of features of the capsid structure determined for NωV with those of the nodaviruses, FHV and BBV, revealed that these two families are structurally analogous due to the presence of highly conserved regions within their tertiary structures (Munshi *et al*., 1996).

Despite the structural similarities between tetra- and nodaviruses, studies on most aspects of tetravirus biology have progressed far more slowly than for nodaviruses. This is largely due to the lack of suitable cell culture systems that support the propagation of these tetraviruses. The lack of suitable experimental systems is possibly attributed to the exceptionally narrow host range of the tetraviruses (Bawden *et al*., 1999). However, the *Providence virus* (PrV) has recently been shown to replicate in cultured cells (Pringle *et al.,* 2003). As a result of this, there has only been a limited amount of data relating the structure of tetraviral capsids to their

biology. In contrast, there is more than a decade of research on nodaviruses, which replicate in a wide variety of cell culture systems. This has resulted in a considerable understanding of the principles relating nodavirus structure to biological function (for reviews, refer to Johnson, 1996, Ball and Johnson, 1998, Schneemann *et al*., 1998, and Johnson and Reddy, 1998).

Since the packaging of the viral genome during the encapsidation process of nodaviruses has been studied, this literature review will focus on these nodaviruses. A review of tetraviruses with comparisons to the nodaviruses is also considered, focusing mainly on the section of the tetraviral capsid protein that is proposed to interact with packaged RNA.

1.2 The *Tetraviridae*

1.2.1 History and taxonomy

Tetraviruses first attracted attention in South Africa when massive epizootics occurred in larval populations of the pine tree emperor moth, *Nudaurelia cytherea capensis.* Diseased cadavers were found to be infected with at least five structurally distinct RNA viruses which were termed *Nudaurelia* γ, δ, α, ε and β viruses in order of increasing prevalence (Hendry *et al*., 1968; Juckes, 1970). Of these viruses, *Nudaurelia β virus* (NβV) was extensively researched biophysically, and this virus type formed the new virus group, the *Nudaurelia* β-like viruses (Juckes, 1979; Matthews, 1982). NβV possessed a ss +RNA genome, encased in a non enveloped icosahedral capsid with a diameter of 35 nm (Struthers and Hendry, 1974). In addition, further studies on NβV capsid revealed that it had an icosahedral shape possessing *T*= 4 quasi-symmetry (Finch *et al*., 1974).

The *Nudaurelia* β-like virus group had to be re-classified later due to the discovery of NωV by Hendry *et al*. (1985). NωV has similar biophysical properties with NβV, but NωV is serologically distinct and contains a bipartite genome. The *T*= 4 capsid symmetry of these viruses formed the basis for the family, *Tetraviridae* (from the Greek *tettares*, four) (Francki *et al*., 1991). NωV became the type virus for a second genus, ω -like genus, while NβV was the virus type for the β-like genus (Murphy *et*

al., 1995). Later, the two genera were renamed and virus members were classified into the monopartite *Betatetraviridae* and the bipartite *Omegatetraviridae* (van Regenmortel *et al*., 2000) according to their genome structure. Of the tetraviruses, the most intensively studied members are the omegatetraviruses, NωV and *Helicoverpa armigera stunt virus* (HaSV) (Hanzlik *et al*., 1993; Hanzlik and Gordon, 1997).

Tetraviruses are the only recognized insect virus family that have an exceptionally narrow host range, propagating solely in closely related species of leptidopteran insects, namely butterflies and moths (Hanzlik and Gordon, 1997). These viruses are reported to display a tropism for the midgut cells of their hosts (Reinganum, 1991; Hanzlik and Gordon 1997). In agriculture, tetraviruses are seen to have attractive prospects as biopesticides for the control of caterpillar pests of worldwide importance (Bawden *et al*., 1999). This is due to their narrow host range and ability to act rapidly and selectively on their lepidopteran hosts.

1.2.2 Tetraviral genomes

Currently, the complete genome sequences of three betatetraviruses, namely NβV (Gordon *et al*., 1999) , *Thosea asigna virus* (TaV) (Pringle *et al*, 1999; Gorbalenya *et al*., 2002) and *Euprosterna elaeasa virus* (EeV) (Gorbalenya *et al*., 2002) and one omegatetravirus, HaSV (Gordon *et al*., 1995; Hanzlik *et al*., 1995), are known. In addition, the sequence of RNA2 of the only other omegatetravirus, NωV (Agrawal and Johnson, 1992) and the capsid-encoding sequence of a further betatetravirus, *Providence virus* (PrV) have been determined (Pringle *et al*., 2003).

The organisation of the 6.6-kb RNA genome of the betatetravirus, NβV, consists of a 5' proximal ORF that encodes the viral replicase (215-kDa) and a 3' proximal ORF that encodes the capsid precursor protein (70-kDa), which is expressed from a 1.8-kb subgenomic RNA (Gordon *et al*., 1999) (Figure 1.1, panel A). This subgenomic RNA was found to be encapsidated along with the genomic RNA, by NβV capsids (Gordon *et al*., 1999). The capsid precursor protein α is cleaved at the carboxy-terminus between Asn536 and Gly537 to produce β- (62 kDa) and γ- (8 kDa) proteins that make up the capsid shell. However, according to the sequence of the capsid precursor protein of two other betatetraviruses, TaV and EeV revealed that these

viruses undergo an additional cleavage at the amino-terminus (Pringle *et al.,* 1999; Gorbalenya *et al*., 2002). In the case of TaV, two of these proteins are the capsid proteins (58.3-kDa and 6.8-kDa), but the third N terminal protein (17-kDa) could not be detected in the capsid shell (Pringle *et al*., 1999).

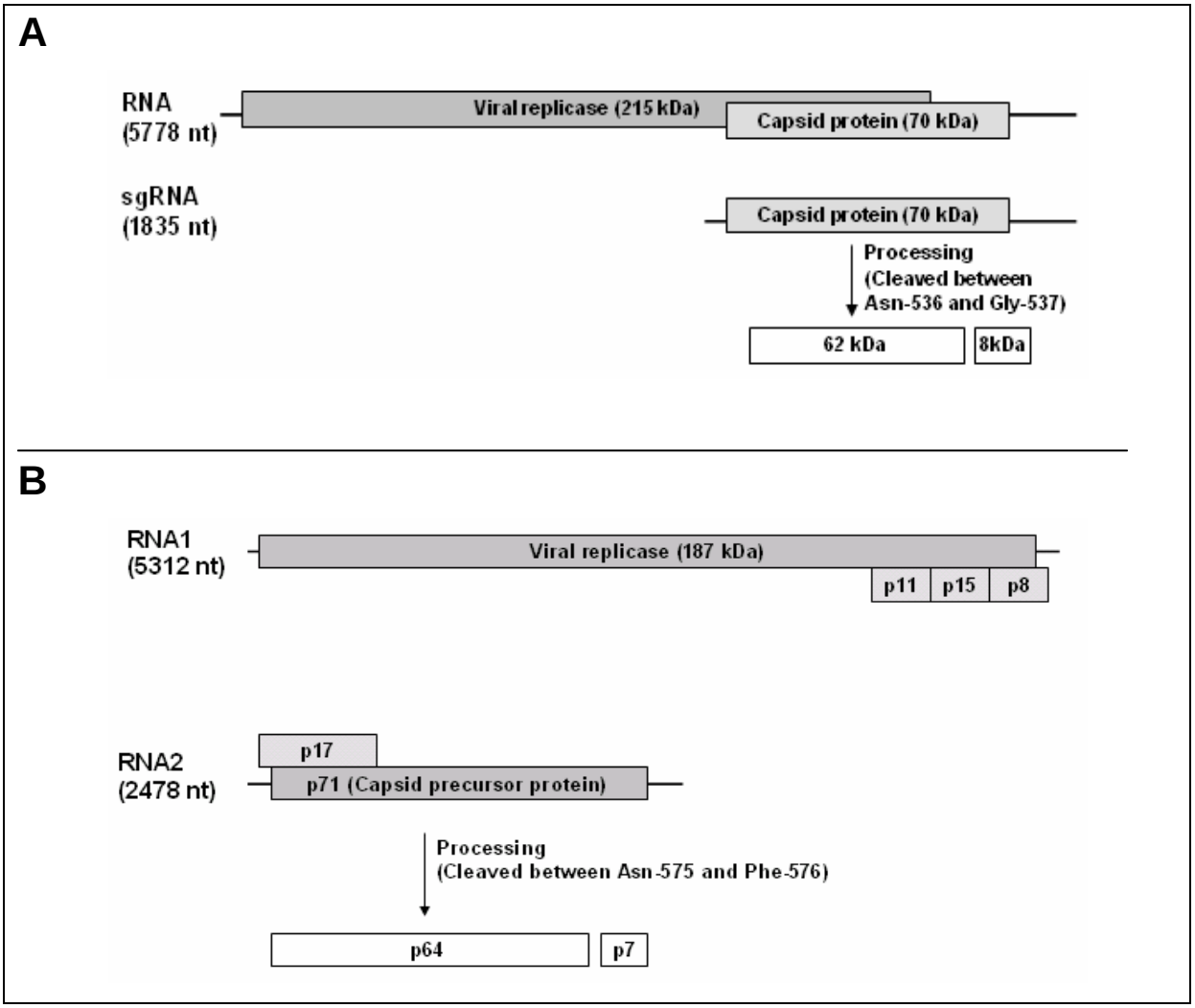


Figure 1.1 Genome organisation of a betatetravirus and an omegatetravirus. Panel A represents the genome organisation of a betatetravirus, NβV. The monopartite genome of NβV 5778 nt and its derived subgenomic RNA (sgRNA) are represented with labelled ORFs. Panel B represents the genome organisation of an omegatetravirus, HaSV. HaSV RNA1 and RNA2 are also represented with labelled ORFs. Panel A and panel B are adapted from Gorbalenya *et al*. (2002) and Gordon and Hanzlik (1998), respectively.

The larger RNA1 of HaSV (5312 nt) carries a large ORF that encodes the viral replicase (187-kDa), and three smaller ORFs at its 3' end (Gordon *et al*., 1995) (Figure 1.1, panel B). The presence, or function, of protein(s) expressed by the latter

ORFs has not yet been established. The smaller RNA2 (2478 nt) has two overlapping ORFs, the first encoding a 17-kDa protein (p17) and the second encoding the 71-kDa capsid precursor protein α (p71) (Hanzlik *et al*., 1995) (Figure 1.1, panel B). Two possible functions have been attributed to p17. This p17 may have a regulatory function in the replication of the two genomic RNAs, or it may have a function analogous to that of movement proteins encoded by plant viruses. p71 is cleaved between Asn575 and Phe576 into the β-(64 kDa) and γ-(7 kDa) proteins designated p64 and p7, respectively. These proteins make up the capsid shell of the mature virion (Hanzlik *et al*., 1995).

1.2.3 RNA secondary structure

The 5' termini of HaSV, NβV RNAs and NωV RNA2 contain a cap structure and the 3' termini have distinctive tRNA-like structures (Hanzlik *et al.,* 1995; Gordon *et al*., 1995, 1999). The 3' termini for HaSV RNAs and NβV RNA can be folded into a tRNA-like secondary structure with a valine anti-codon that shows direct sequence homology to eukaryotic $tRNA^{VAL}$ (Gordon *et al*., 1995, 1999) (Figure 1.2).

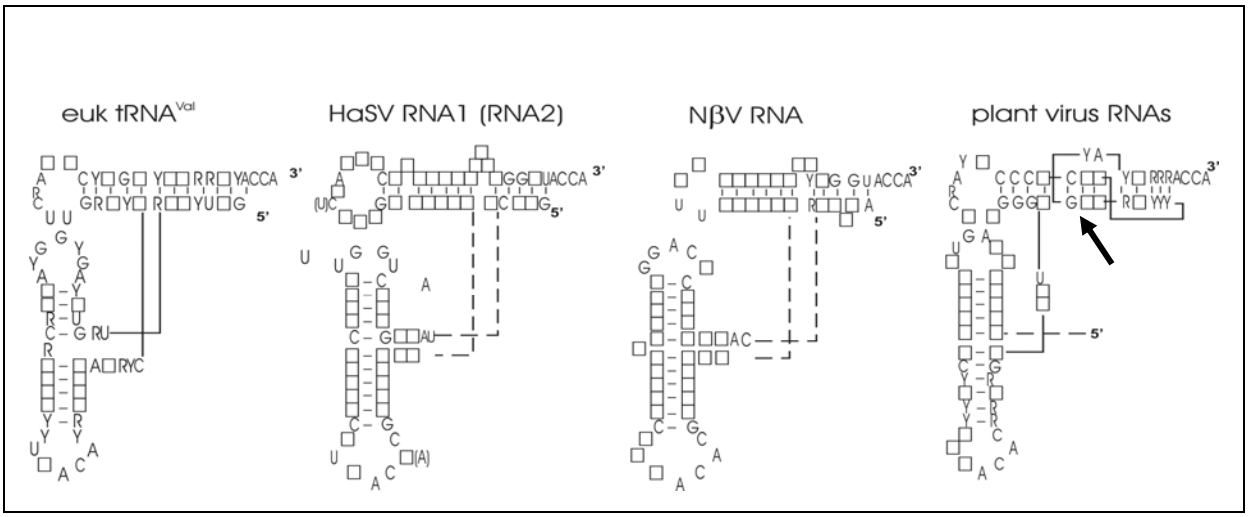


Figure 1.2 Comparison of tRNA-like structures on NβV and HaSV RNAs with eukaryotic $tRNA^{VAL}$ and plant viral $tRNA^{VAL}$-like structures. Only conserved residues are specified for the tRNA, with others being represented by open squares. In the case of viral structures, only residues corresponding to the conserved ones in the tRNA are shown. The plant virus structure shows the pseudoknot in the amino-acyl stem (indicated with an arrow). Dashed lines in HaSV structure represent single stranded regions and a hairpin structure corresponding to the variable loop. HaSV residues shown in parentheses refer to RNA2. Figure taken from Gordon and Hanzlik (1998).

The presence of these 3' tRNA-like structures of tetraviral RNAs are the first known for animal viruses, since these have been previously described only for plant viruses (Gordon and Hanzlik, 1998). However, the tetraviral 3' tRNA-like structure differs from plants viruses as it lacks the pseudoknot in the amino-acyl stem (Gordon *et al*., 1999) (Figure 1.2). It is suggested that these structures may play, either a similar role to that of 3' poly-A tails, which protect cellular mRNAs from degradation (Hanzlik and Gordon, 1997), or a role in viral replication as in plant RNA viruses (Gordon *et al*., 1999).

Both RNA strands of HaSV are capable of forming prominent hairpins located just downstream from the 5' cap site, which contains the conserved hexamer sequence, GGUAAA, in the loop (Gordon *et al*., 1995; Hanzlik *et al*., 1995). It is suggested that both the hairpin loops and 3' tRNA-like structures may be involved in RNA replication (Gordon and Hanzlik, 1998).

1.3 The *Nodaviridae*

1.3.1 A brief background

Apart from omegatetraviruses, members of the nodavirus family are the only other spherical ss +RNA animal viruses with bipartite genomes (Johnson and Reddy, 1998). This family is subdivided into the *Alphanodavirus* and *Betanodavirus* genera, due to a low RNA2 sequence similarity between members of the respective genera (van Regenmortel *et al*., 2000). Alphanodaviruses primarily infect insects, but the type-species of the genus, NoV, seems to be unique in its ability to infect and kill invertebrates (mosquitoes) and vertebrates (mice) (Hendry, 1991). Most of the molecular biology was unravelled using members of this alphanodavirus group, notably BBV and FHV, both of which can grow to high titers in cultured *Drosophila* cells (Schneemann *et al*., 1998). Betanodaviruses are commonly referred to as the piscine nodaviruses that infect more than 20 species of fish throughout Asia, Europe and Australia (Lu and Lin, 2003). The sequences of two genomic RNAs of the *Striped jack nervous necrosis virus* (SJNNV, type member) (Mori *et al*., 1992), and the *Greasy grouper nervous necrosis virus* (GGNNV) (Tan *et al*., 2001) have been documented. This review will focus on the molecular and structural biology of the alphanodaviruses.

BBV and FHV are similar in their molecular properties, including the amino acid sequence of virally encoded proteins (Kaesberg *et al*., 1990). Most research has been focused on FHV, since it is more cytolytic for *Drosophila* cells than BBV (Schneemann *et al*., 1998). On the other hand, PaV is a distantly related virus of the alphanodavirus genera, since its capsid protein shares 40% sequence identity with other nodaviruses, and its viral replicase shares 26% identity with FHV (Zeddam *et al*., 1999). Despite the low sequence homology, the crystal structure of PaV determined by Tang *et al*. (2001) is virtually super-imposable upon the FHV capsid and provides more detail of the interaction between the capsid shell and genomic RNA. This will be described later in section 1.6.4.

1.3.2 Genome organisation of FHV

The large ORF of FHV RNA1 encodes the viral replicase (protein A, 112kDa) (Friesen and Rueckert, 1981; Gallagher *et al.,* 1983), while a second and third ORF, both much smaller than the first, are present on a subgenomic RNA (Figure 1.3).

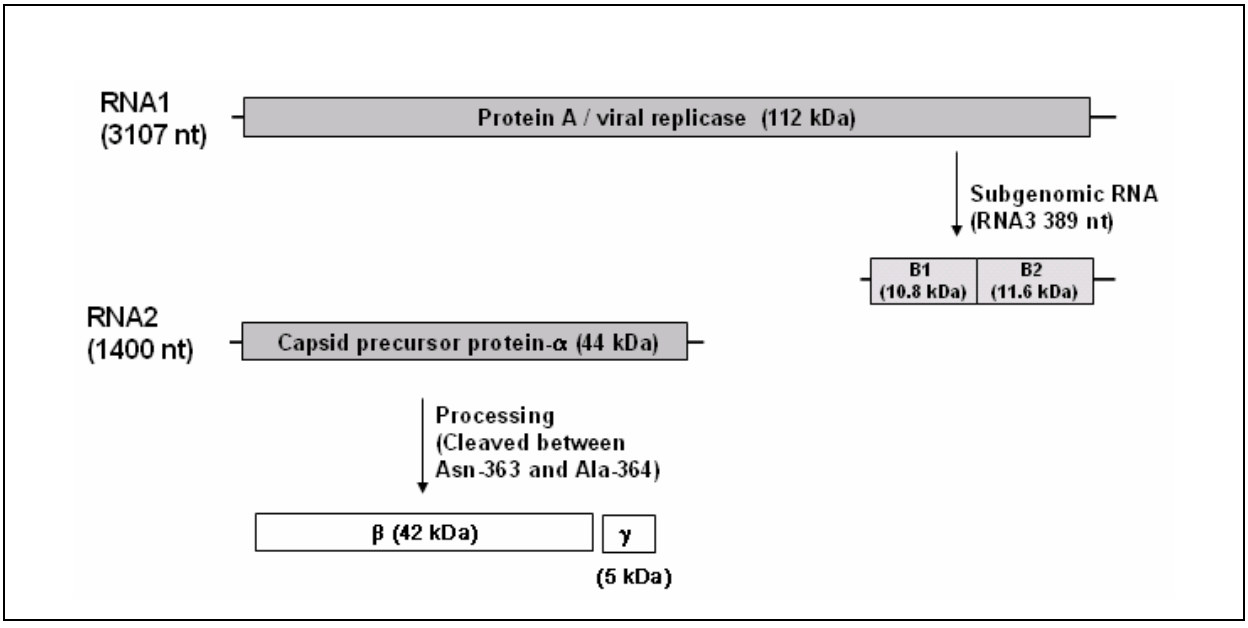


Figure 1.3 Genome organisation of FHV RNA1 and RNA2. Both genomic RNAs are represented with ORFs found on the RNAs (Schneemann *et al*. 1998).

This subgenomic RNA, RNA3 (389 bp) derived from the 3' end of RNA1, encodes two proteins B1 (10.8kDa) and B2 (11.6kDa) (Ball, 1995; Dasmahaptra *et al*., 1985; Gallagher and Rueckert, 1988). The function of protein B1 still remains unclear (Eckerle and Ball, 2002), and protein B2, particularly for NoV, has been shown to

enhance viral RNA accumulation within host cells (Li *et al*., 2002; KL Johnson, *et al*., 2004). Protein B2 suppresses or circumvents an innate antiviral response such as RNA interference. RNA2 encodes the capsid precursor protein (47 kDa) (Friesen and Rueckert, 1981) (Figure 1.3). As in tetraviruses, the capsid precursor protein is auto-catalytically cleaved between Asn363 and Ala364 residues, to produce β (42-kDa) and γ (5-kDa), the two capsid proteins, in the mature virion (Hosur *et al.,* 1987).

1.3.3 RNA secondary structure

FHV RNA1 and RNA2, as well as the subgenomic RNA, are capped at the 5' end and are non polyadenylated at their 3' ends (Moore and Tinsley, 1982; Dasmahapatra *et al*., 1985). The 3' proximal end of FHV RNA2 consists of two hairpin loops that have been proposed to serve as replication signals (Ball and Li, 1993; Kaesberg *et al.,* 1993).

1.4 Capsid structure of noda- and tetraviruses

Most viruses possess icosahedral capsids because they form the largest internal cavity sufficient to package the viral genome (Johnson, 1996). The simplest icosahedral capsid is composed of 60 identical subunits adequate to package a viral genome. The formation of larger shells is explained by the quasi-equivalence theory proposed by Caspar and Klug (1962). This theory refers to the axes that relate the different subunits within the quasi-equivalent shells. The theory states that the number of identical subunits organised in an icosahedral lattice is limited to 60*T*, and that *T* is equal to 1, 3, 4, 7 etc. For *T* values > 1 the subunits are chemically identical, but are structurally quasi-equivalent due to *T* different environments. The formation of quasi-equivalent particles requires a molecular switching mechanism for the formation of flat and bent contacts between subunits (Johnson, 1996). This molecular switch is usually a protein segment, a segment of viral genome, or both, that acts as a wedge in the formation of flat contacts. This wedge is absent for the formation of bent contacts (Johnson, 1996).

Nodaviruses assemble their *T*= 3 capsid from 180 subunits that assume one of the three quasi-equivalent conformations denoted A, B and C (Figure 1.4: Panel A)

(Johnson, 1996). These particles have the geometry of the rhombic triacontahedron, and can be viewed as constructed from 60 copies of ABC trimers. The bent contacts with a dihedral angle of 144^{o} occur along the quasi-twofold axes (open ellipses) that relate the icosahedral fivefold (solid pentagon) and the quasi-six fold axes (solid triangle) (Figure 1.4: Panel B). The flat contacts occur along the icosahedral 2-fold axis (solid ellipses) that relates two quasi six-fold axes (Figure 1.4: Panel C) (Johnson, 1996).

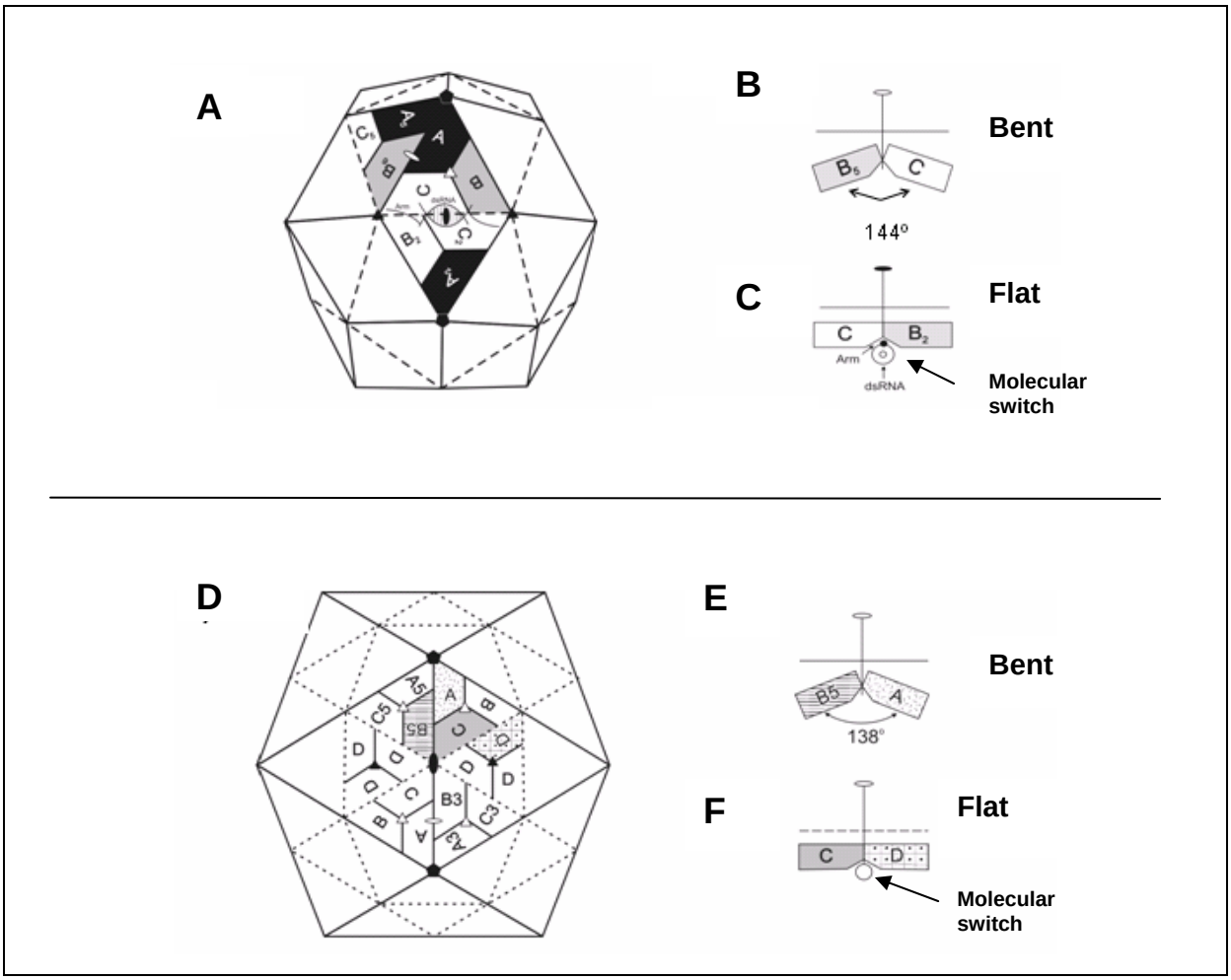


Figure 1.4 Comparison of *T*= 4 and *T*= 3 capsid shells. Panel A and B represents capsid shells arranged with *T*= 3 (nodaviruses) and *T*= 4 (tetraviruses) quasi-symmetry, respectively. Panel B, and E represent bent contacts and panel C and F represents flat contacts found within *T*= 3 and *T*= 4 capsid shells, respectively. Solid ellipses, triangles, and pentagons, indicate icosahedral 2-fold, 3-fold and 5-fold axes, respectively; while empty ellipses and triangles indicate quasi-2-fold and quasi-3-fold axes, respectively. The subunits marked A, B, C, and D correspond to the same gene product located in different conformations. The flat and bent contacts required for the formation of *T*= 3 and *T*= 4 quasi-symmetrical particles are indicated in Panel B, C, E and F. All panels taken from Johnson and Reddy (1998).

The tetravirus *T*= 4 capsid, which has the geometry of a standard icosahedron, assembles from 240 copies of the capsid precursor protein. The latter assumes one of the four quasi-equivalent conformations denoted A, B, C, and D (Olson *et al*., 1990). The shell can be viewed as constructed from two different types of trimers, 60 copies of ABC and 20 copies of DDD trimers (Helgstrand *et al*., 2004) (Figure 1.4: Panel D).

Unlike nodaviruses, the *T*= 4 quasi-symmetry has two unique quasi-twofold axes (open ellipses): one relating two quasi-six fold axes (related by two copies of C, B and D subunits) and the other relating an icosahedral fivefold (solid pentagon) and a quasi-six fold axis. In the case of flat and bent contacts, the bent contact that has a dihedral angle of 144^{o}, occurs along the quasi-twofold axis relating the icosahedral fivefold and quasi-six fold axes (Johnson and Reddy, 1998) (Figure 1.4: Panel E). The flat contacts that have a dihedral angle of 180^{o}, occur along the quasi-twofold axes relating two quasi-six fold axes (Figure 1.4: Panel F) (Johnson and Reddy, 1998).

1.5 Similarities between the capsid subunit tertiary structure of tetra- and nodaviruses

Early structural studies of capsids suggested that NωV and nodaviruses are structurally related (Johnson *et al*., 1994; Agrawal and Johnson, 1995). This was based on the similar cleavage mechanism of the coat proteins that releases the γ-peptide and the structure of the viral capsids. When the NωV structure was resolved to near atomic resolution, Munshi *et al.* (1996) noted that the subunit tertiary structure was very similar to the known structures of the nodaviruses BBV (Hosur *et al.,* 1987; Wery *et al*., 1994) and FHV (Fisher and Johnson, 1993) despite little sequence similarity between the capsid precursor proteins. The sequence identity of above mentioned capsid precursor proteins is about 12% (Munshi *et al*., 1996; Helgstrand *et al*., 2004). The following discussion will be based on Munshi *et al*, (1996), but the subunit tertiary structure of NωV that has recently been refined by Helgstrand *et al*. (2004) will be used for the comparison.

The tertiary structure of NωV subunit in C conformation is visualised as three protein domains: the internal helical domain, β-sandwich domain, and the external / Ig-like domain (Munshi *et al*., 1996) (Figure 1.5). The nodavirus BBV (also in C conformation), subunit is subdivided into the inner and outer domain (Schneemann *et al*., 1998) (Figure 1.5). The outer domain of BBV subunit corresponds to the β-sandwich domain and the external domain of the NωV subunit (Johnson and Reddy, 1998). The inner domain of BBV and the internal helical domain of NωV, which are in close proximity to the encapsidated RNA, will be described later in detail in Section 1.6.4.

1.5.1 β-sandwich and external domain

The β-sandwich with jelly-roll topology forms the shell of the capsids, and this structure is very similar for both NωV and BBV (Figure 1.5), despite a lack of sequence homology (Munshi *et al*., 1996). According to the subunit tertiary structure for both NωV and BBV, the β-sandwich is composed of eight anti-parallel strands labelled βB-I. Of the eight anti-parallel strands, two sets of four strands form two β-sheets that form the walls of the β-barrel (Munshi *et al*., 1996).

Based on structural superimposition calculations, the eight strands of the β-sandwich domain (βB-I), including the βC', βC'' strands, superimpose well for both viruses (Helgstrand *et al*., 2004). However, the major difference between nodaviruses and NωV occurs at the insertion between strands E and F (Munshi *et al*., 1996).

Nodaviruses have a 34 residue extended loop at this location that interacts with neighbouring subunits at the quasi-threefold axes (Figure 1.4, panel A: open triangle) to form the outermost protrusion of the particle. In contrast to nodaviruses, NωV external domain is composed of 133 residues, which fold into a structure resembling that, observed in immunoglobulins (Munshi *et al*,. 1996). This domain of NωV interacts with related subunits at either the quasi-threefold or the icosahedral threefold axes (Figure 1.4, panel B: solid triangle and empty triangle, respectively) giving NωV its characteristic appearance in electron micrographs (refer to Figure 1.6).

Table 1.1 Amino acid homologies among capsid protein domains of three tetraviruses NβV, NωV, and HaSV. Adapted from Gordon *et al*., (1999).

Region	NβV Residues	Percentage identity to NωV	Percentage identity to HaSV[a]
N terminal	**1-39**	**29**	**33 (40)**
β-barrel (N terminal half)	**40-257**	**28**	**27 (81)**
Ig-like (central)	**258-407**	**21**	**22 (37)**
β-barrel (C terminal half)	**408-536**	**25**	**28 (85)**
γ-peptide	**537-612**	**28**	**29 (53)**

[a] Figures in brackets indicate the sequence homology (identity) between HaSV and NωV for the relevant domain

Based on the alignment of the coat proteins of, HaSV with NωV, the region representing the β-sandwich domain, has 80% sequence identity (Hanzlik *et al*., 1995). However, for the betatetravirus, NβV, there is a low sequence identity (20-30%) with that of NωV (Gordon *et al*., 1999) (Table 1.1), thus suggesting a divergence in evolution between betatetraviruses and omegatetraviruses. For the region representing the Ig-like domain there is a low level of conservation (37%) between HaSV and NωV, despite a high level of sequence conservation within the rest of the capsid precursor proteins (Munshi *et al*., 1996) (Table 1.1).

Although, nodaviruses do not posses an Ig-like domain, their 34 residue extended loop has been suggested to be implicated in host cell recognition. This is because a single mutation within this region of the nodavirus, FHV, greatly reduced its infectivity of cultured *Drosophila melanogaster* cells (Dasgupta *et al*., 1994).

1.5.2 Internal helical domain

The internal helical domain is composed of the N and the C terminal ends of the capsid protein that proceed from the β-barrel domain for both NωV and BBV (Munshi *et al*., 1996) (Figure 1.5). The positions of the N and C termini are nearly identical for both structures. This region consists predominantly of α-helices.

For NωV, both N and C termini have four helices labelled αN1–αN4 and αC1–αC4, respectively (Helgstrand *et al*., 2004) (Figure 1.5). These helices from both termini of

one subunit, interact closely with each other and extensively with the internal helical domains of neighbouring subunits within the capsid shell. The αC2–C4 helices of the C terminus end form the cleaved γ-peptide. Of the polypeptides that form the N terminus end, residues 1-40 are disordered for all the subunit conformations. These disordered 40 residues are proposed to project toward the interior of the virus particle where they may interact with the packaged RNA (Helgstrand *et al*., 2004).

According to the subunit tertiary structure of BBV, residues 20 – 31 and helix I, form the N terminal polypeptide (Hosur *et al*., 1987) (Figure 1.5). The C terminal polypeptide is composed of helices II and III. The latter helix represents the cleaved γ-peptide. The N terminal visible residues 20-31 are visible for this subunit C, but structurally disordered for subunits A and B (Hosur *et al*., 1987). Another unique feature of BBV subunit C, is its close proximity to ten nucleotides of the encapsidated RNA that conforms to icosahedral symmetry (Fisher and Johnson, 1993). Structural superimposition calculations for the internal helical domain of NωV revealed that αN3, αC1 and the region close to the cleavage, superimpose with similar regions in BBV (Helgstrand *et al*., 2004).

1.6 Functions of the internal helical domain

The internal helical domain has four roles within NωV and nodaviral capsids: 1) It controls the maturation of the particle (Taylor *et al.,* 2002; Canady *et al*., 2000, 2001; Schneemann *et al*., 1992); 2) It is required for the establishment of flat and bent contacts required for the formation of quasi-symmetrical particles (Johnson, 1996; Helgstrand *et al*., 2004); 3) It forms the helical bundle that plays a role in membrane translocation (Cheng *et al*., 1994; Munshi *et al*, 1996; Helgstrand *et al.,* 2004); 4) It lies in close proximity to the packaged viral RNA within the capsid (Tang *et al*., 2001; Helgstrand *et al*., 2004). These roles of the internal helical domain will be described briefly in the current order with comparisons between NωV and nodaviruses. The fourth role will be expanded in order to discuss the protein RNA interactions, which might be required for the packaging of the viral genome by the capsid shell.

1.6.1 The γ-peptide

Agrawal and Johnson (1995) first reported the assembly of NωV VLPs by the baculovirus-mediated expression of the gene encoding NωV capsid precursor protein in *Spodoptera frugiperda* (Sf9) cells. These NωV VLPs were found to be structurally indistinguishable from the wild type. In addition, they specifically encapsidated viral mRNA that encodes the capsid precursor protein. The latter discovery will be discussed later in Section 1.10. Canady *et al* (2000) performed cryo-EM analysis on these NωV VLPs, and revealed that there were two distinct capsid structures purified at pH 7 and pH 5, respectively, termed the pro-capsid and the mature capsid (Figure 1.6). The pro-capsid purified at pH 7, was approximately 16% larger than the capsid with a diameter of 450 Å (Figure 1.6: procapsid). In addition, these pro-capsids in which the 240 subunits are all in similar conformation were observed to be round and porous.

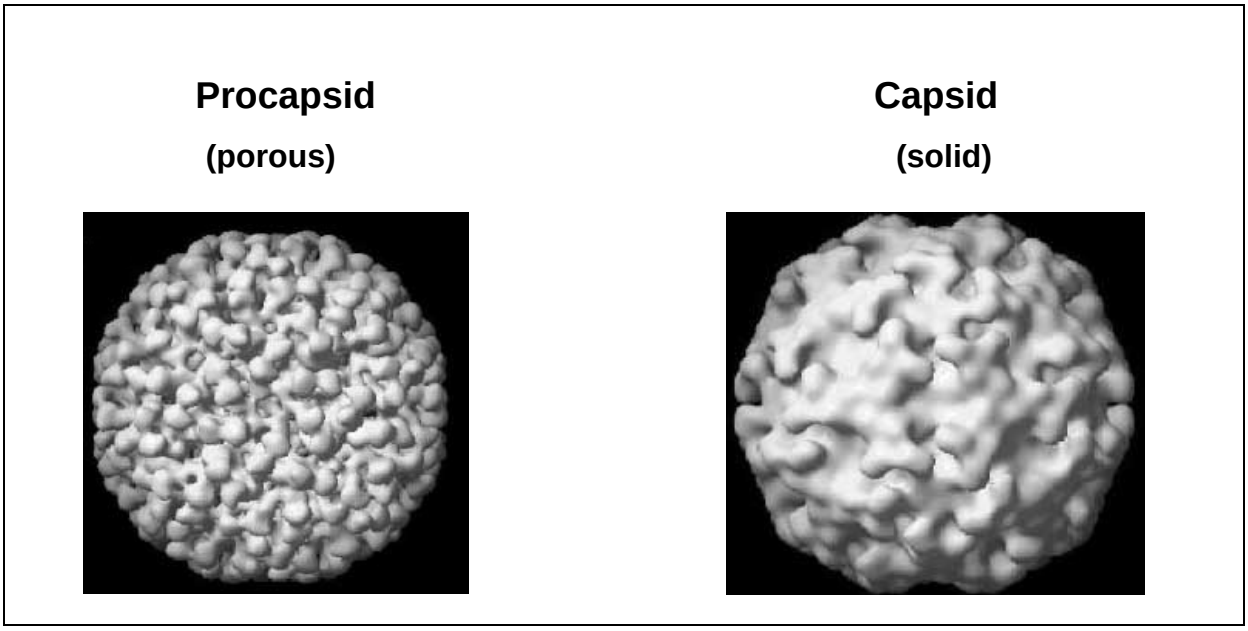


Figure 1.6 3D-surface shaded structure of NωV capsid shell. During assembly, the protein protomers form a porous pro-capsid (left). After assembly it differentiates into a solid mature capsid (right) due to the autocatalytic cleavage of the carboxy-terminus of the capsid pre-cursor protein α within the internal helical domain. Taken from Canady *et al*. (2000).

Exposure of the purified pro-capsid to acidic conditions (pH 5), resulted in at least 15% of the subunits undergoing autocatalytic cleavage at the carboxy-terminus of, protein α, releasing the γ-peptide (Canady *et al.,* 2000, 2001). Mature capsids are

solid and polyhedral in shape, with a diameter of 410 Å (Figure 1.6: capsid) (Canady *et al*., 2000). During the pro-capsid to capsid transition there is a re-organisation of the internal helical domain that causes the subunits to undergo conformational change, adopting one of the four possible conformations as described for the *T*= 4 quasi-symmetry (Canady *et al*., 2000). These conformational changes of the subunits, lead to the formation of flat and bent contacts required for the formation of a *T*= 4 quasi-symmetrical particle.

Based on studies that revealed that the pro-capsid to capsid transition was reversible until more than 15% of the subunits were cleaved (Canady *et al*., 2001), Taylor *et al*. (2002) performed studies on NωV VLPs. VLPs were assembled from capsid precursor protein containing a mutant that abolished the auto-catalytic cleavage site. These NωV VLPs were capable of reversible transition between the pro-capsid and capsid conformations along with fluctuations in pH. This indicated that the auto-catalytic cleavage that releases the γ-peptide is required for the fixed transition of the capsid conformation (Taylor *et al*., 2002).

Nodaviruses have a similar maturation cleavage, but unlike NωV, maturation occurs automatically after viral assembly and there is no reversible intermediate (Johnson and Reddy, 1998). The cleavage event is critical in the life cycle of authentic FHV, because cleavage-defective FHV mutants were shown to be un-infectious (Schneemann *et al*., 1992).

1.6.2 The molecular switch

For nodaviruses, the molecular switch that regulates the flat contacts is composed of the N terminal peptide (residues 20-31) and the RNA. This RNA is associated with the two C subunits along the icosahedral twofold axis (Fisher and Johnson, 1993). The RNA duplex together with the two peptides, referred to as the peptide ARM, serve as a wedge. This fills the groove formed by the CC_2 interface establishing the flat contact (Figure 1.7: Panel A). In contrast, at the quasi-twofold joints, relating A and B_6 subunits, there is no structural RNA presence and residues 20 - 30 are disordered (Figure 1.7: Panel B). As a consequence, the A – B_6 contact is bent, with a dihedral angle of 144° (Schneemann *et al*., 1998). Unlike BBV, the crystal

structure of PaV has continuity of the N terminal polypeptide. The latter is derived from the A-subunits surrounding the icosahedral fivefold axis and corresponds to residues 20-31 of BBV (Tang *et al*., 2001).

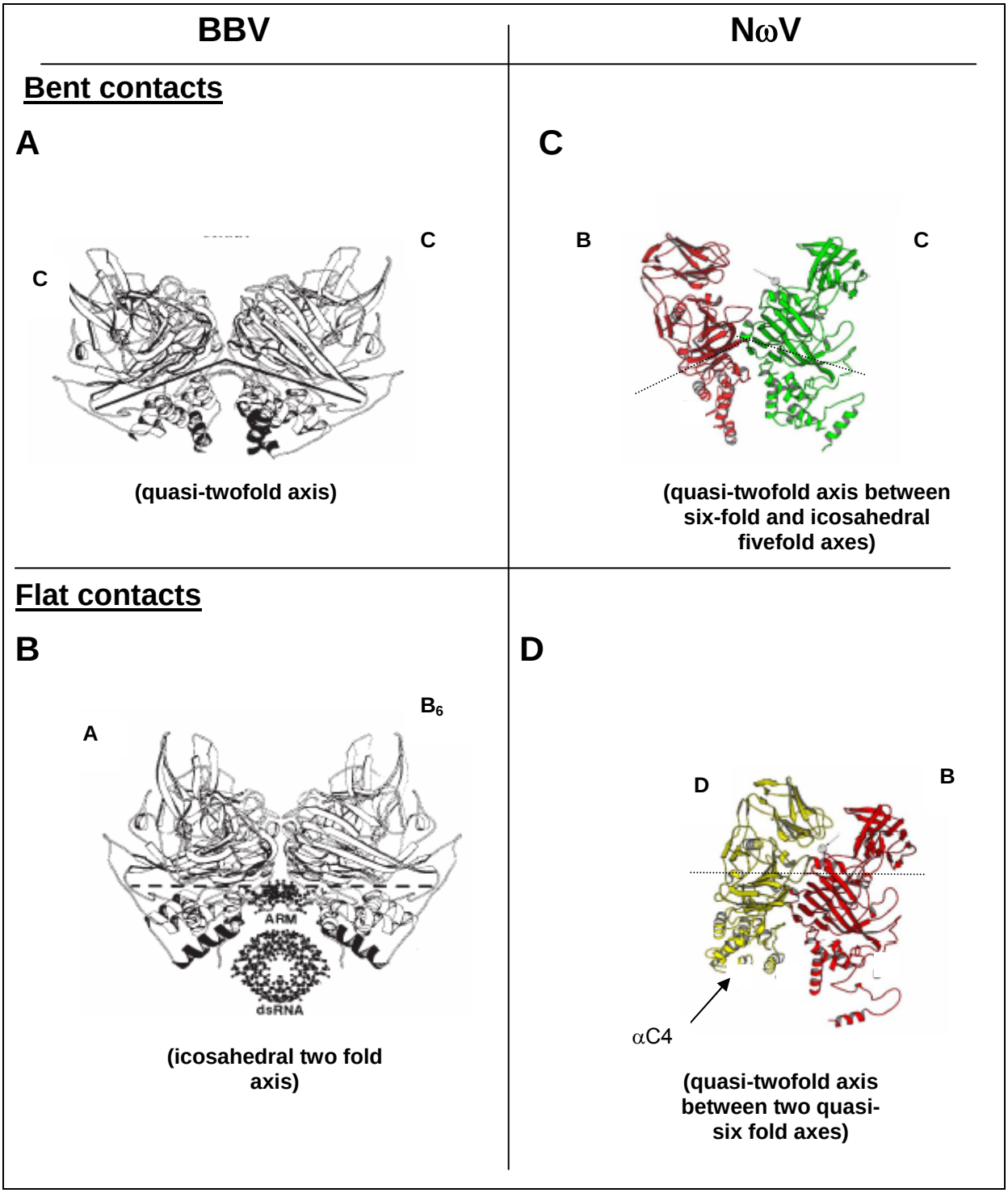


Figure 1.7 Role of the internal helical domain in the establishment of flat and bent contacts required for the formation of *T*= 3 and *T*= 4 quasi-symmetrical capsid of BBV and NωV, respectively. The compositions of the molecular switch are indicated below subunits related by flat contacts. Panel A & B taken from Johnson (1996) and Panel C & D taken from Helgstrand *et al.* (2004).

The molecular switch for NωV is composed of the αC4 helix, which is ordered for C and D subunits (Helgstrand *et al*., 2004) (Figure 1.7: Panel D). The flat contact is due to the insertion of αC4, that forms the wedge at the CB_2 interface along the quasi-twofold axis relating two quasi-six fold axes (Figure 1.7: Panel D). This wedge is absent from the B_6-C interface along the quasi-twofold axis relating the quasi-six fold axis and the icosahedral fivefold axis (Figure 1.7: Panel C).

1.6.3 Membrane translocation of RNA

Munshi *et al*. (1996) found that the arrangement of the cleaved γ-peptides of the internal helical domain of five A-subunits related by the icosahedral fivefold axis (Figure 1.4, panel A & B: solid pentagon) was very similar for both NωV and BBV. These γ-peptides form a helical or pentameric bundle (Figure 1.8: Panel A & B).

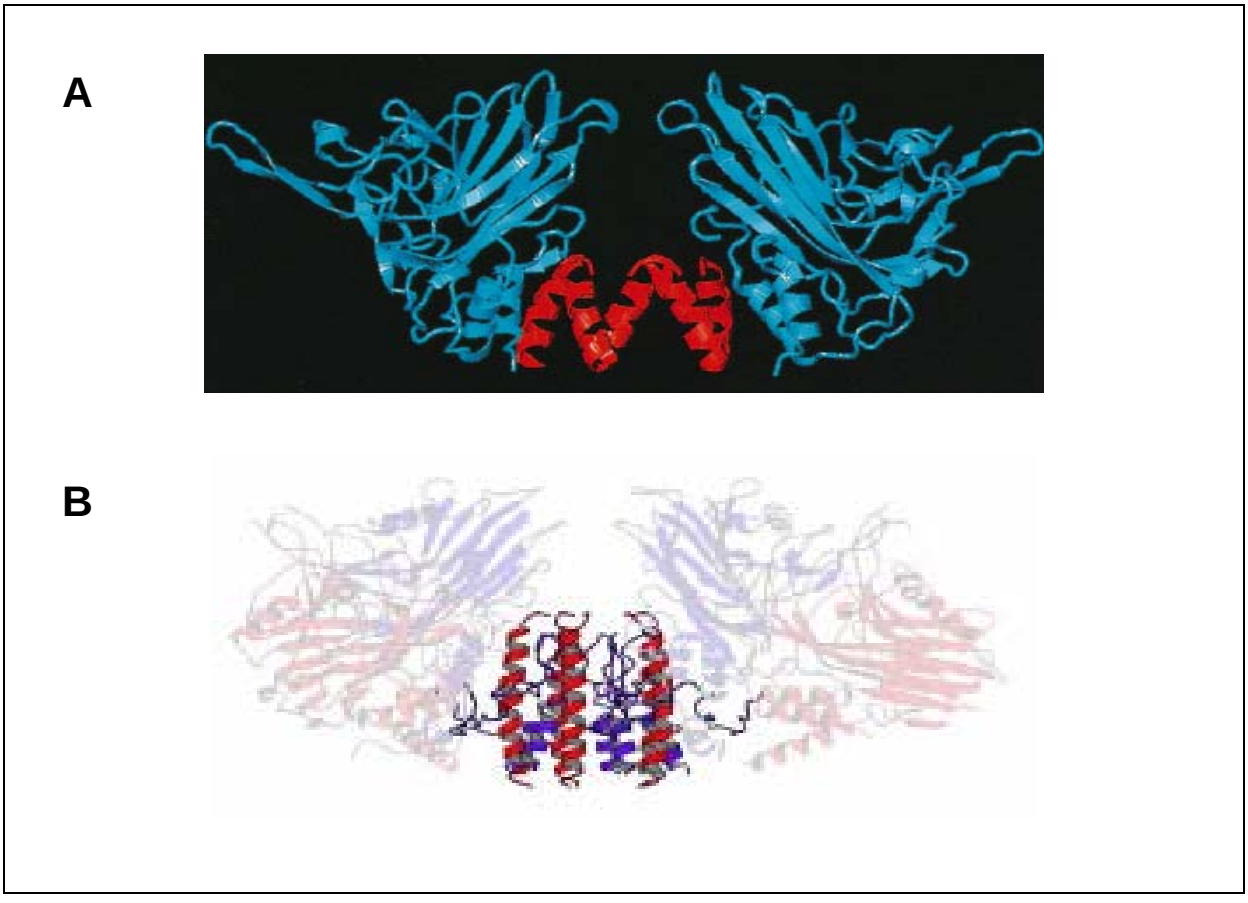


Figure 1.8 A view perpendicular to the icosahedral fivefold axis in BBV (Panel A) and NωV (Panel B). For both BBV and NωV the residues in red correspond to the ordered cleaved γ-peptide of subunits A (red residues) surrounding the five-fold axes. Only two of the five subunits are shown for clarity, although all five helices are represented. Unlike BBV, NωV γ-peptides are intercalated with αN1 helices (dark blue) of the N terminus from B subunits. Panel A take from Munshi *et al.* (1996), and panel B taken from Helgstrand *et al*. (2004).

The helical bundle of FHV has been proposed to function in membrane translocation of RNA (Cheng *et al*., 1994). Part of the helical bundle faces the interior of the capsid and is associated with the genomic RNA. The latter observation was based on two studies where small portions of genomic RNA were released from intact FHV particles at the icosahedral fivefold axes, when the γ-peptides were removed by either treating FHV particles with SDS, or exposing them to heat followed by low pH (Gallagher and Rueckert, 1988; Cheng *et al*., 1994). The inside of the helical bundle is lined with polar residues, while the outside of the bundle is mainly hydrophobic allowing insertion into, or association with, the cell membrane (Fisher and Johnson, 1993). In addition, the surface of the FHV capsid has a hydrophobic patch at the fivefold axes that would allow the particle to make contact with the cell membrane. It is hypothesised that upon interaction, the residues surrounding the fivefold axis which are hydrophobic, would allow the bundle to ''slide out'' following a structural change translocating the genomic RNA into the cytoplasm (Fisher and Johnson, 1993) (Figure 1.9). It has been demonstrated that a region of the γ-peptide of FHV is membrane active and readily forms pores that release dyes from liposomes (Bong *et al*., 1999; Janshoff *et al*., 1999).

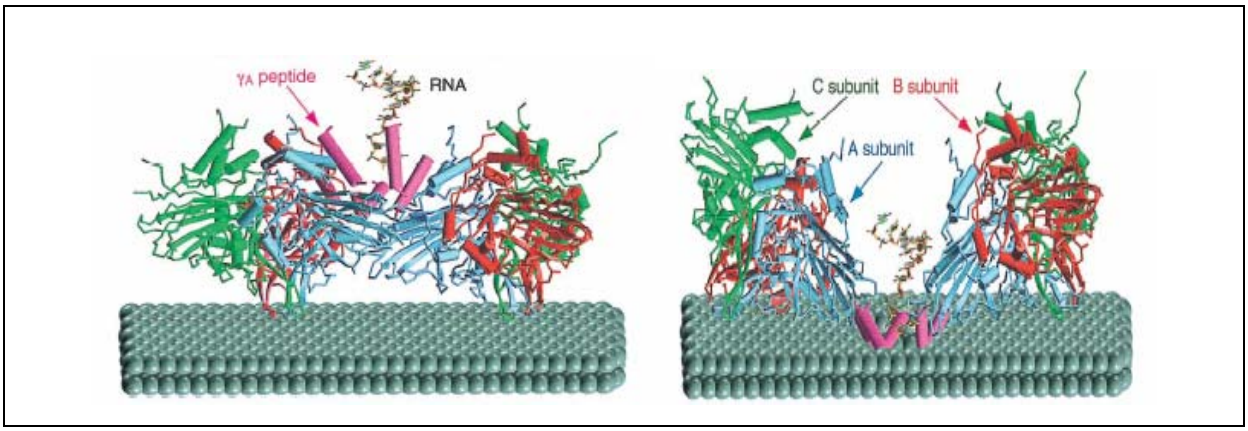


Figure 1.9 Hypothetical entry pathway for FHV, role of the fivefold axes in membrane translocation of RNA. Left figure: Initial interaction of the icosahedral fivefold axis exterior of the capsid section with a bilayer membrane (grey balls). Right figure: Upon interaction, subunits rotate toward the membrane, allowing the γ-A peptides to be inserted into the membrane and forming a channel releasing the viral RNA into the host cell. Taken from Baker *et al.* (1999).

According to the re-defined crystal structure of NωV (Helgstrand *et al.,* 2004), there are some differences in the helical bundle of NωV compared with BBV and FHV.

NωV γ-peptides are significantly longer (Figure 1.8: Panel B) (approximately 30 residues longer) than the nodaviral γ-peptides (Helgstrand *et al*., 2004; Munshi *et al*., 1996). Unlike nodaviruses, the five γ-peptides of the NωV helical bundle are intercalated by N terminal helices (αN1) contributed by five neighbouring B chains (Figure 1.8: Panel B). This results in a 10 helical bundle that is covalently attached to the rest of the capsid protein through the αN1 helices. With these covalent interactions, it is not certain how the NωV helical bundle would be externalised due to the limitations of conformation changes (Helgstrand *et al*., 2004). Evidence that there could be interaction between the C terminal ends of the helical bundle and RNA comes from earlier cryo-EM studies on NβV. This was as a result of a high density of partially or highly ordered RNA close to icosahedral fivefold axes (Olson *et al*., 1990).

1.6.4 Interaction of the internal helical domains and genomic RNAs

Nodaviruses

According to the computational 3-D re-construction of the crystal structure of PaV capsid, the N terminal polypeptide (Figure 1.10: coloured yellow) is observed to interact extensively with a 25 nt RNA duplex along the icosahedral two-fold axes (Tang *et al*., 2001).

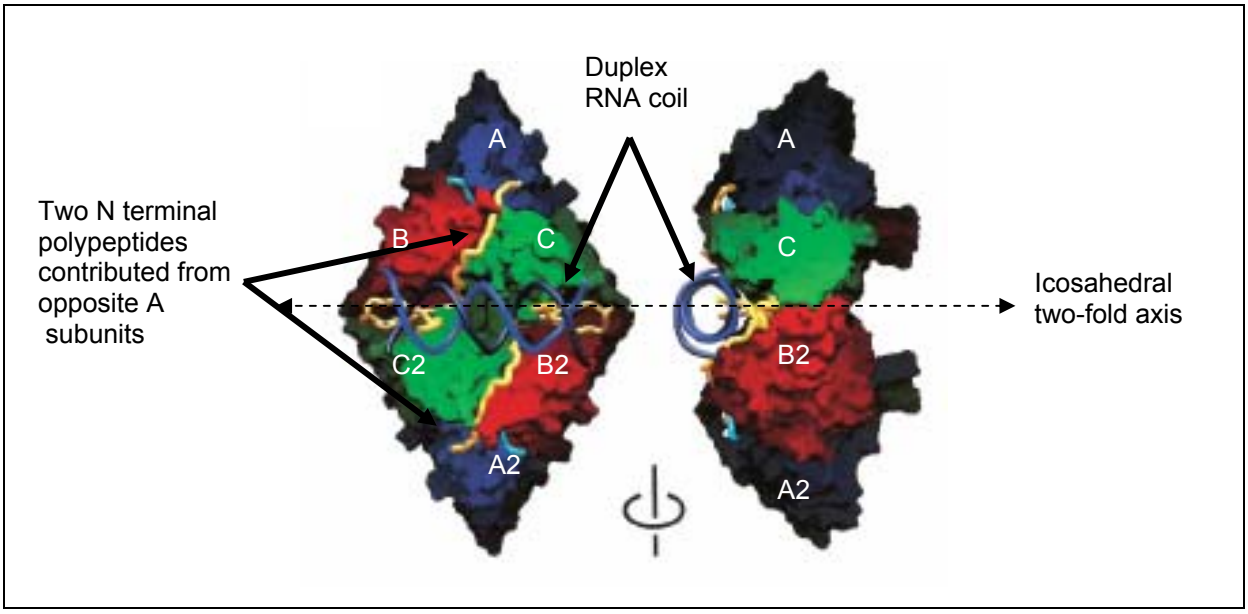


Figure 1.10 3D-computational reconstruction of the region interior of the capsid icosahedral two-fold axis of PaV. Left image, interior of the capsid shows extensive interaction of the N terminal polypeptide derived from the A subunits with RNA duplex. Right image, view rotated 90° relative to the left panel. The RNA duplex is shown as blue coils. The extended N terminal polypeptide derived from A-subunits is coloured yellow. Figure taken from Tang *et al.* (2001).

Along this axis, the packaged RNA forms a hetero-duplex formed by two single stranded RNA molecules (Figure 1.10: coloured blue). It is not known how the two single stranded RNA molecules form a hetero-duplex. It is suggested that the capsid protein plays a role in this conformation of RNA (Tang *et al*., 2001). Two N terminal polypeptides observed along the icosahedral two-fold axes are derived from two opposite A-subunits (A and A_2) (Figure 1.10) around the five-fold axes (A-A_5 and A_2-A) (Tang *et al*., 2001). The portion of the N terminal polypeptide that forms the peptide arm of the wedge (the molecular switch that determines the flat contact within the *T*= 3 lattice) lies along the CC_2 axis. From this wedge section of the N terminal polypeptide, the remaining section of the polypeptide protrudes toward the RNA duplex, and inserts into the major groove of the RNA-duplex (Tang *et al*., 2001) (Figure 1.10).

At low resolution, the 3D-figure re-construction of the side chains of the two PaV N terminal polypeptides contributes 24 basic residues available for interaction with RNA. Of the 12 basic residues, which are clustered around residues 11 and 40 of the capsid protein, positive charge appears to neutralize the abundant negative charges of the phosphate groups of the RNA backbone (Figure 1.11).

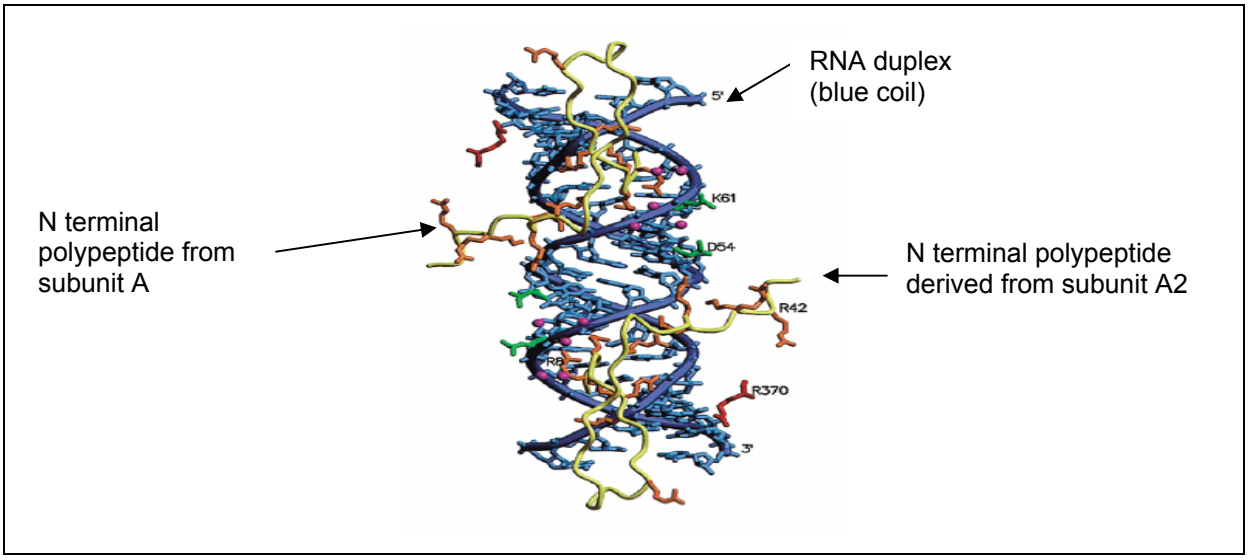


Figure 1.11 3D-computational Figure re-construction of PaV N terminal polypeptide side chains that interact with the RNA duplex along the icosahedral two-fold axis. The extended N terminus of subunit A interacts with the duplex RNA along the icosahedral two-fold axes (the N termini of the A subunits are in yellow and the RNA duplex is in blue). Basic residues from subunits A, B and C are colored orange, red and green, respectively. Figure taken from Tang *et al.,* (2001).

The side chains of Arg8, Arg9, Lys11, Arg13, Lys14, Arg37, Lys38 and Lys41 (coloured orange) are in direct contact with the phosphate groups of the RNA duplex. The side chains of Arg8, Arg9, and Lys11, in particular, are located in the major groove of the duplex where they form hydrogen bonds with bases in the RNA (Tang *et al*., 2001).

The 30 RNA duplexes that were observed along the 30 icosahedral two-fold axes of PaV, form an RNA dodecahedral cage (Figure 1.12: Panel A & B) and this cage accounts for ~35% of PaV genomic RNA (1500 of 4322 nucleotides). The remaining ~65% RNA did not adhere to the symmetrical pattern of the capsid, and was therefore disordered. This disordered RNA connected to the ordered RNA duplexes via the 3 way junctions (Tang *et al*., 2001) (Figure 1.12: Panel B). This dodecahedral RNA cage has recently been described for FHV as well, but only 16% of RNA was observed, which is approximately 10 nt opposed to 25 nt for PaV of nucleotides along the icosahedral twofold axis (Tihova *et al*, 2004).

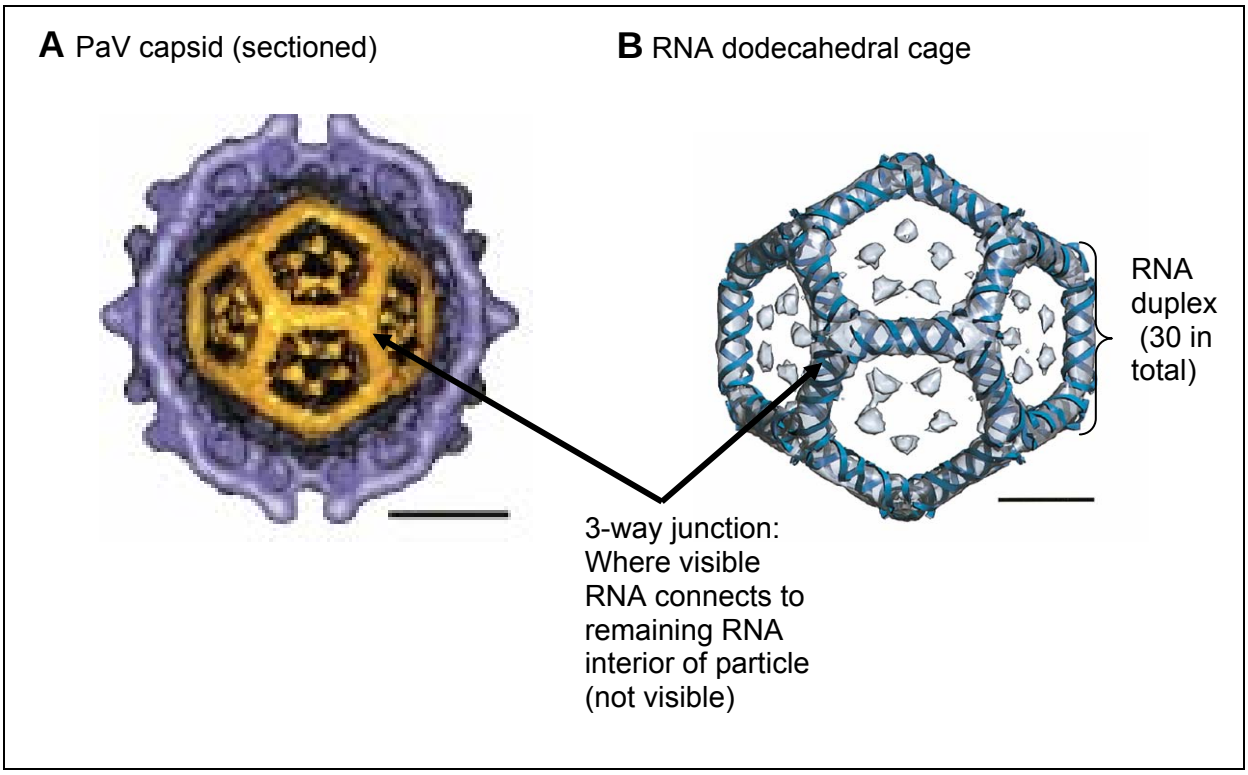


Figure 1.12 PaV packages its RNA that is visualised as a dodecahedral cage structure. Panel A, sectioned view showing the packaged RNA genome (gold) within the capsid shell of PaV Bar = 100 Å. Panel B represents the dodecahedral cage structure of the packaged RNA (blue ribbons) Bar = 50 Å. Both panels taken from Tang *et al*., (2001).

It cannot be determined with which segment of the PaV bipartite genome the N terminal polypeptide interacts (Tang *et al*., 2001). This is because Cryo-EM analysis is an average of the symmetrical pattern. In this case the 25 nt RNA observed along the icosahedral twofold axis, is an average of the 30 RNA-duplexes along the 30 icosahedral two-fold axes. In other words, the phosphate backbone and ribose sugar molecules of the RNA strand can be visualised, since these are uniform. In contrast, the nucleosides cannot be visualised, as there are four types of nucleosides present in RNA.

Tetraviruses

Unlike nodaviruses, the crystal structure of NωV does not appear to contain ordered RNA components within its capsid shell (Munshi *et al.*, 1996; Helgstrand *et al*., 2004). Despite the lack of ordered RNA, the primary amino acid sequence of NωV reveals that the proximal ends of the C and N terminus are abundant in positively charged amino acids (Agrawal and Johnson, 1995). The C terminus, particularly the γ-peptide has a net charge of +8 over the last twenty residues, and together with the net +11 charge of the first 40 residues of the N terminus, contributes to the positively charged character of the internal helical domain (Agrawal and Johnson, 1995). This combined positive charge is sufficient to neutralise phosphate groups on over 50% of the viral RNA genome (Agrawal and Johnson, 1995). In addition, further evidence for interactions between the internal helical domain and the encapsidated RNA is based on a comparison between the amino acid sequences of the NωV and HaSV capsid precursor proteins, which reveals the presence of two low sequence homology domains at N and C termini of these proteins (Gordon and Hanzlik, 1998). This may reflect a prerequisite for different interactions with the genomes of these two viruses.

Taking the redefined crystal structure (Helgstrand *et al*., 2004) into account, the NωV γ-peptide is disordered for the A and B subunits, and ordered for the C and D subunits. For the eight arginine residues in the region of 628-640 of the ordered γ-peptides, only two are exposed to the interior of the capsid and are probably available for interaction with encapsidated RNA. Of the remaining 6 arginine residues, two are involved in Protein protein interactions, and the other four are not exposed to the interior. All eight arginine residues for the disordered γ-peptide are suggested to be available for interaction with encapsidated RNA. For the N terminus,

residues 1-40 of the N terminus are disordered for all quasi-equivalent subunits. This segment contains 10 arginine and 2 lysine residues, which contribute to the positive charged nature of this protein segment as observed in nodaviruses coat proteins, and were proposed to be associated with the encapsidated RNA (Helgstrand *et al*., 2004).

1.7 RNA encapsidation by nodaviruses

The first direct evidence that FHV particles co-packaged its bipartite genome into a single virion came from heat treatment studies of FHV particles (Krishna and Schneemann, 1999). Heat treatment caused FHV RNA1 and RNA2 to form a stable complex that migrated as a single species on agarose gel electrophoresis.

RNA encapsidation has been investigated for FHV particles that were assembled from capsid precursor protein containing deletions within the N or C terminus region (Schneemann and Marshall, 1998; Marshall and Schneemann, 2001). These investigations were carried out by transfection of *Drosophila melanogaster* cells, which support FHV replication, with full length FHV RNA1 and RNA2 containing mutations in either the N or C terminal coding region.

Residues 382-407 of the γ-peptide were crucial, since FHV particles lacking these residues encapsidate primarily cellular RNAs (Schneemann and Marshall, 1998). Further mutational analysis revealed that amino acid residues, particularly phenylalanines at positions 402, 405, and 407, were critically important for the specific encapsidation of FHV RNAs. There was, however, no evidence of contacts between the γ-peptides and genomic RNA in the crystal structures of FHV, BBV and PaV (Fisher and Johnson, 1993; Hosur *et al*., 1987; Wery *et al*., 1994; Tang *et al*., 2001). Referring to studies on the PaV crystal structure, it has been suggested that the ordered γ-peptides of the A-subunits may play a role in RNA packaging during assembly, but not in the mature virus state (Tang *et al*., 2001). The disordered γ-peptide of PaV B and C subunits are proposed to project toward the interior of the capsid where they interact with the disordered 65% remaining RNA.

RNA-encapsidation studies on mutant FHV particles assembled from the capsid precursor protein with 2-31 residues deleted from the N terminal end, revealed that this terminus was required for the specific encapsidation of RNA2 (Marshall and Schneemann, 2001). These Δ31 FHV particles encapsidated less RNA2, normal amounts of RNA1 and co-encapsidated large amount of cellular RNA, probably compensating for the loss of RNA2 encapsidation. Marshall and Schneemann (2001) proposed that the 2-31 residues deletion removed a crucial basic cluster of five arginine residues. This proposal was based on the fact that the side chains of arginine residues are involved in many RNA-protein complexes (Tan and Frankel, 1995). Further experiments proved that these arginine residues were crucial for RNA2 encapsidation, since site directed mutagenesis of all five arginine with alanine residues reduced encapsidation of RNA2 (Marshall and Schneemann, 2001). However, the Δ31 FHV particles had the most reduced RNA2 encapsidation, which suggested that there are additional residues besides the cluster of five arginine residues within the N terminus 2-31 region, which may be involved in RNA2 encapsidation (Marshall and Schneemann, 2001).

Comparison of positions of basic residues within the N termini of PaV and FHV, revealed that arginine residues at position 8 and 9 are equivalent to FHV, but residues at position 11 and 14 are lysine instead of arginine (Marshall and Schneemann, 2001). As described earlier, residues Arg8, Arg9 and Lys11 of PaV are located in the major groove of the duplex where they form hydrogen bonds with the bases in the RNA (Tang *et al*., 2001). It has been suggested that the N terminus of FHV is likely to form analogous interactions with its packaged genome (Marshall and Schneemann, 2001).

In summary, the role of the C terminus is fundamentally different from that of the N terminus in nodaviruses. Instead of selective packaging of a particular genomic segment as demonstrated for the N terminus for RNA2, the C terminal mutants preferentially package cellular RNAs and only small amounts of viral RNAs (Schneemann and Marshall, 1998; Marshall and Schneemann, 2001).

Schneemann *et al*. (1993) first reported the assembly of ordered structures representing FHV VLPs. This was done by the baculovirus-mediated expression of FHV RNA2 in the *Spodoptera frugiperda* (Sf9) cell line. These VLPs were found to

be structurally indistinguishable from the wild type, but they encapsidated predominantly cellular RNAs (Schneemann *et al*., 1993). In contrast, FHV VLPs encapsidated both viral RNA1 and RNA2 in the presence of a biologically active viral replicase (Krishna *et al*., 2003). The viral replicase encoded by RNA1 was found to amplify its own message, to synthesize the subgenomic RNA3, and also to replicate RNA2, which led to the assembly of infectious VLPs. It has been suggested that the specific encapsidation of FHV RNAs was dependent on viral replication, where high concentrations of viral RNAs are required near the vicinity of FHV capsid assembly (Krishna *et al*., 2003).

Recently Tihova *et al*. (2004) analysed the high-resolution X-ray structure of the dodecahedral RNA cage of four types of FHV particles that differed in RNA and protein content. These four types of FHV particles were the FHV native particles, Δ31 FHV particles, FHV VLPs and the Δ31 FHV VLPs. With the exception of Δ31 FHV VLP, the RNA contents have been described earlier (Figure 1.13: Panel A). The Δ31 FHV VLPs RNA content consisted of primarily cellular RNA (Figure 1.13: Panel A), and possessed multiple types of assembled structures, which were probably influenced by the capsid protein N terminal deletion (Dong *et al*., 1998). Although the majority of these VLPs have a similar shape and dimension as the wild type, these were subsequently selected for the dodecahedral RNA cage study (Tihova *et al*., 2004).

The RNA dodecahedral cages were similar for all four types of FHV particles, irrespective of the differing RNA contents (Tihova *et al*., 2004) (Figure 1.13: Panel B). In terms of similar dodecahedral RNA cages, these FHV particles still possess RNA duplexes along the icosahedral twofold axes that meet each other at the 3-way junctions (Figure 1.13: Panel B). Most importantly, the capsid protein N terminal polypeptide (which is known from the PaV crystal structure to interact with duplex RNA (Tang *et al*., 2001)), was not required for the positioning of duplex RNA along the icosahedral twofold axes (Tihova *et al*., 2004).

Tihova *et al*., (2004) proposed that the C terminus played an important role not only in the initial selection of the viral RNAs for packaging, but was also required for the positioning of RNA duplex along the icosahedral twofold axes. In addition, during

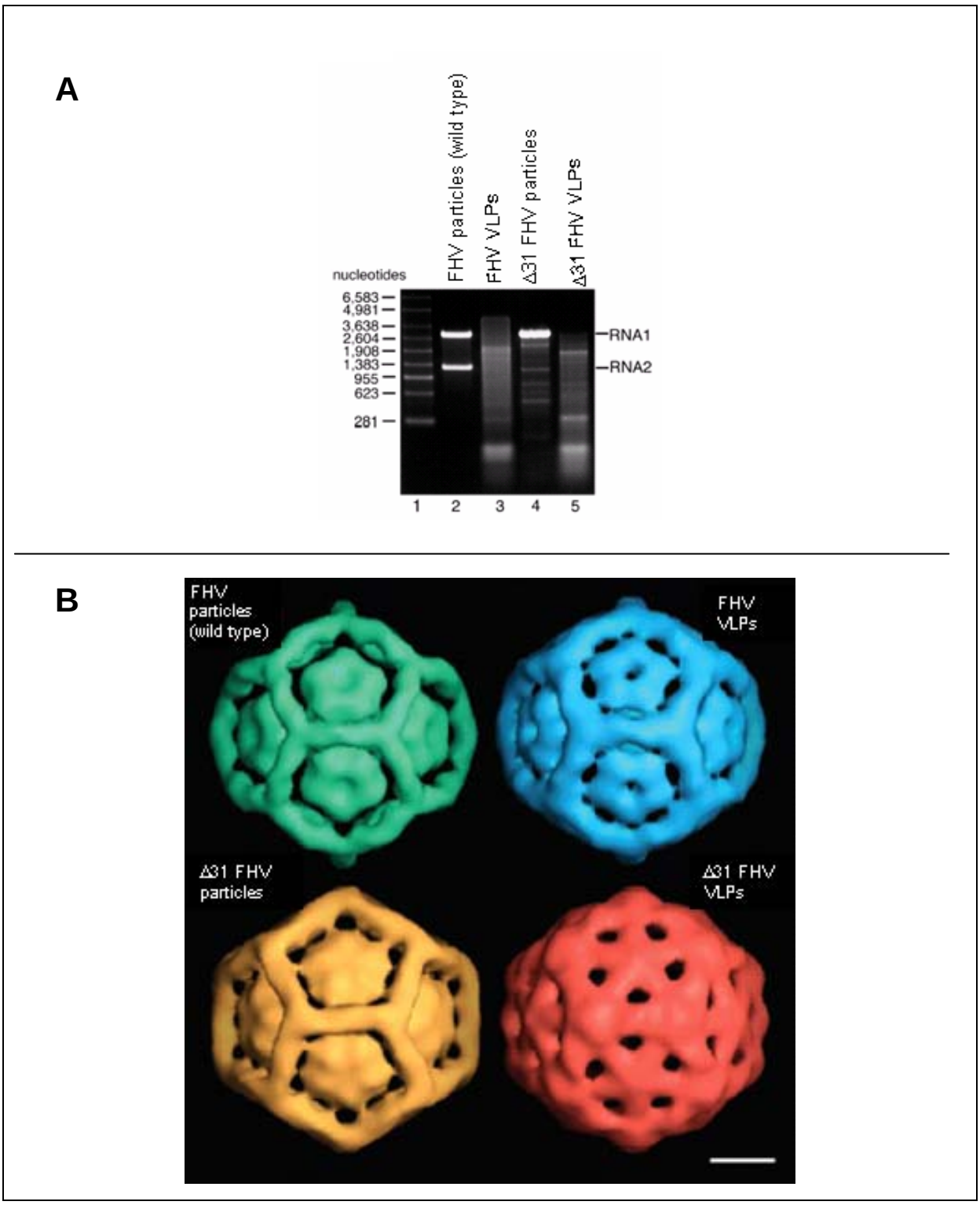


Figure 1.13 RNA content and arrangement of four different types of FHV particles. Panel A represents gel electrophoretic analysis of RNA extracted from the four types of FHV particles. Lanes: 1, RNA molecular size markers; 2, RNA from FHV particles; 3, RNA from FHV VLPs; 4, RNA from Δ31 FHV particles; 5, RNA from Δ31 FHV VLPs. Panel B represents 3D surface-shaded reconstructions of FHV particles (green), Δ31 FHV particles (yellow), FHV VLPs (blue), and Δ31 FHV VLPs (red) reveal nearly identical dodecahedral RNA cages. Bar = 100 A° (Tihova *et al*., 2004).

virion assembly the highly basic flexible N terminus was proposed to project toward the interior of the capsid in order to organise the central core of RNA.

Most importantly, FHV VLPs containing cellular RNAs have been found to display less stability in solution (Bothner *et al*., 1998, 1999). In addition, no "empty" particles, devoid of RNA, have been reported for Nodaviruses. "Empty particles" have been reported in other virus systems such as the *Norwalk* virus (Prasad *et al*., 1999). These observations further implied that viral RNA not only carries the genetic information, it contributes to particle integrity and structural mobility (Bothner *et al*., 1998, 1999).

1.8 RNA encapsidation by tetraviruses

Although, there is a lack of visible RNA within the NωV crystal structure, this does not imply that these capsids could assemble into normal virus particles in the absence of RNA. This is because the highly charged inner surface of these particles would almost certainly require some form of poly-anion to allow capsid formation (Munshi *et al*., 1996). In contrast to FHV VLPs that encapsidated primarily cellular RNAs, NωV VLPs have been shown to encapsidate mRNA encoding the capsid precursor protein (Agrawal and Johnson, 1995). In addition, unlike FHV VLPs that encapsidated viral FHV RNA1 and RNA2 under replication conditions, Gordon *et al*. (2001) showed that infectious HaSV VLPs were assembled in plants despite no evidence of RNA replication. This might indicate that the capsid protein encoding region of RNA2 may contain a recognition signal for the capsid protein to confer selectivity. In addition, encapsidation of viral RNAs by HaSV VLPs is independent of RNA replication. It is still not known what section of the capsid protein confers selectivity for the viral mRNA.

1.9 Project proposal

The mechanism for the recognition and packaging of the viral genome is poorly understood for tetraviruses. The coat protein N terminus of NωV can be seen to be

analogous in two respects with those of nodaviruses. Namely, the N termini are highly positively charged, due to the presence of arginine and or lysine residues that are clustered proximally to this end of the protein (Agrawal and Johnson, 1995; Marshall and Schneemann, 2001); and in addition, it has been demonstrated that these basic residues are found to contact packaged RNA as depicted in the crystal structure of PaV, and are required for the specific encapsidation of RNA2 by FHV particles (Tang *et al*., 2001; Marshall and Schneemann, 2001). Secondly, the region of the NωV N terminal polypeptide (residues 1-48), which is disordered in that it does not adhere to the symmetrical pattern of the capsid, has been proposed to project towards the interior of the capsid shell to organize the encapsidated RNA (Helgstrand *et al*., 2004). This disordered region is seen to be analogous to that of nodavirus FHV, where residues 2-31 have been proposed to have a similar function (Tihova *et al*., 2004).

Despite the lack of visible RNA and its contact with the capsid protein within the crystal structure of NωV, its VLPs have been shown to specifically encapsidate RNA2 (the capsid precursor protein mRNA) (Agrawal and Johnson, 1995). It is still not known what region of the capsid protein confers this specificity. Since FHV N terminus is required for the specific interaction of RNA2 (Marshall and Schneemann, 2001), the analogous NωV N terminus may have a similar role.

The principle objective of this project was to develop experimental systems to investigate the role of the N terminal domain of the tetravirus capsid protein in RNA encapsidation. Since cDNAs of both genomic RNAs for the tetravirus, *Helicoverpa armigera stunt virus* are available, this virus was used.

The objectives for the research described in this thesis are as follows:

- Development of an *in vitro* assay in order to investigate the interaction between the N terminal domain of HaSV capsid protein and genomic RNA.
- Expression of HaSV capsid precursor protein *in vivo,* to examine the effect of N terminal deletions on the assembly of VLPs

Chapter 2: Development of an *in vitro* system

2.1 Introduction

The lack of *in vivo* experimental systems has limited our understanding of tetravirus biology, and in particular, the mechanisms by which these viruses assemble and encapsidate genomic RNA in their host cells. In the laboratory, we have developed non-host expression systems to study capsid assembly (Venter, 2001; Tomasicchio, unpublished results). However, the heterologous expression of the gene, *P71*, encoding HaSV capsid precursor protein in yeast, *Saccharomyces cerevisiae*, produces very low yields of assembled ordered structures, representing HaSV VLPs (Venter, 2001). Venter showed that more than 95% of expressed p71 formed insoluble aggregates, which were rapidly lost, probably through proteolytic degradation. The formation of insoluble aggregates would obviously make the study of the interaction between the capsid protein and viral RNA *in vivo*, impractical.

In order to circumvent the problem of p71 insoluble aggregates *in vivo*, it was necessary to design an experimental system whereby the N terminal domain of p71 could be isolated to detect interaction between the protein and viral RNA. This region of p71 was chosen due to its position within the interior of the virus particle where its positively charged residues might be interacting with the encapsidated RNA. Thus, an RNA-protein "pull down" assay was developed to investigate this interaction. This was instead of electrophoretic gel mobility shift assays (EMSA) and NorthWestern blot experiments which are two common approaches used to investigate RNA-protein interaction in viruses (Brantley and Hunt, 1993; Choi and Rao, 2000; Xu *et al*., 2002). The assay developed does not require radiolabelled probes for the detection of any RNA-complex, since this complex is disrupted by the addition of SDS. The released RNA would be visualised by separation on an agarose gel containing ethidium bromide.

The IMPACT^{TM}-CN system (New England Biolabs) was used to construct the protein component of the RNA-protein "pull down" assay (Figure 2.1). IMPACT which stands for "intein mediated purification with an affinity chitin-binding tag" is a novel protein purification system, which utilises the inducible self-cleavage activity of a protein-splicing element (named intein) to separate the target protein from the affinity tag. Unlike other protein purification systems, the target protein is purified in a single chromatographic step without the use of a protease. The self-cleavage of the intein

peptide is induced in the presence of thiols such as dithiothreitol, β-mercaptoethanol, or cysteine. This releases the target protein from the chitin-bound intein peptide resulting in a single column purification of the target protein (Figure 2.1).

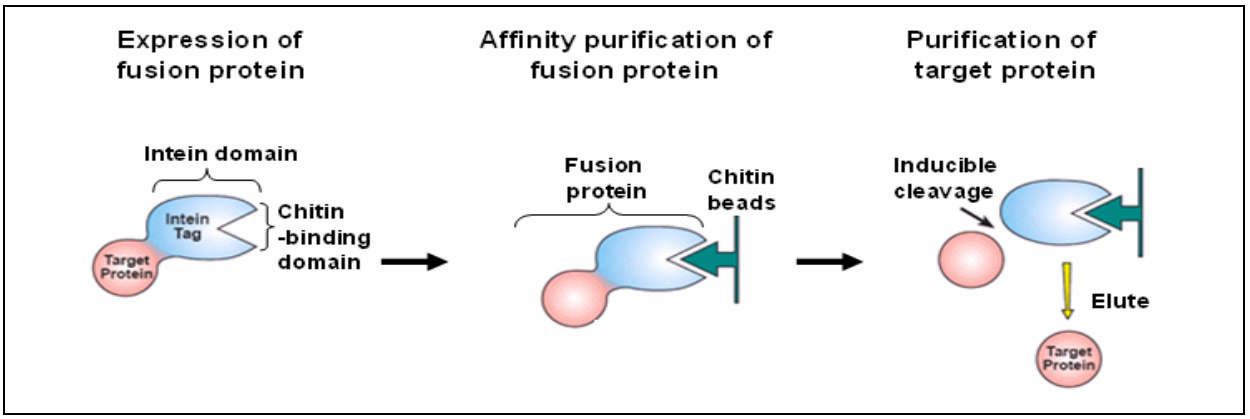


Figure 2.1 Schematic flow diagram of the IMPACT^{TM}-CN purification of target protein. Taken from IMPACT^{TM}-CN manual (NEB).

This chapter describes the development of an RNA-protein "pull down" assay (hereafter referred to as "binding assay") to investigate the interaction between the N terminal domain of p71 and *in vitro* synthesized HaSV genomic RNAs.

2.2 Materials and Methods

2.2.1 Construction of a fusion vector for the expression of the N terminal domain of HaSV p71

General recombinant DNA techniques were carried out as described in Appendix 1. PCR amplification of the coding sequence for the N terminal domain of HaSV p71 (hereafter referred to as HN) was carried out using primers AN4 and AN5 (Table 2.1), which correspond to nucleotides 1-26 and 345-363 of *P71*, respectively.

These primers would allow the isolation of 363 bp from *P71*, which is the coding sequence of HN (Figure 2.2). The yeast expression vector, pAV3, was used as the template, since it contains the entire HaSV *P71* coding sequence constructed by Venter (2001). Panel A of Figure 2.3 summarises the cloning methodology for the construction of the vector containing the coding sequences HN isolated from pAV3

and cloned in frame with that of the intein peptide within the expression vector, pTYB1 (NEB). Of the two primers, AN4 anneals at the 5' ATG start codon of the HaSV *P71* ORF and introduces an upstream *Nde* I restriction site, while AN5 anneals downstream introducing *Xho* I and *Sap* I restriction sites. These restriction sites were required for later cloning applications. PCR amplification was performed with the Expand High Fidelity PCR system (Roche) using thermal cycling programme 2 (Appendix 4).

Table 2.1 Primers AN4 and AN5 used in the PCR isolation of the coding sequence representing the N terminal domain of HaSV *P71*.

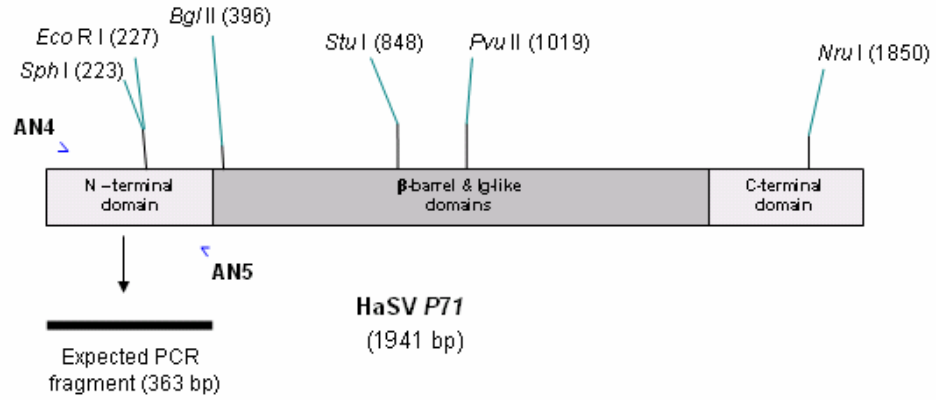
Primer	Anneals[1]	Sequence[2]
AN4	1-26	*Nde* I 5'-cAT ATG GGA GAT GCT GGA GTG GCG TCA CA–3'
AN5	345-363	*Sap* I *Xho* I 5'–ggt ggt tgc tct tcc gga ctc GAG GCC GTC AGG GAT CTT–3'

[1] Corresponds to nucleotides of *P71*

[2] Small letters indicate a heterologous sequence that was added to introduce restriction sites for downstream cloning applications. Primer AN4 introduces a recognition sequence for the *Nde* I restriction digest at the 5' end and primer AN5 introduces the recognition sequences *Sap* I and *Xho* I restriction digests at the 3' end of the PCR transcript.

The PCR product was gel purified using the Nucleospin Extract 2 in 1 (Macherey-Nagel), and inserted into pGEM-T Easy vector (Promega) (Appendix 2) according to the manufacturer's instructions. The ligation reaction was transformed into competent *E. coli* DH5α cells, which were plated onto Luria agar plates containing 100 μg/mL Ampicillin, 100 mM IPTG and 50 mg/mL X-Gal for blue/white screening. For the confirmation of the correct PCR product within pGEM-T Easy, the insert was sequenced using the pUCF sequencing primer (Appendix 3), which anneals to the 5' end of the *lacZ* gene within pGEM-T Easy vector. The resultant construct was named pAN1.

A



B

AN4

```
     M   G   D   A   G   V   A   S   Q   R   P   H   N   R   R   G   T
1    ATG GGA GAT GCT GGA GTG GCG TCA CAG CGA CCT CAC AAC CGT CGC GGA AC
     TAC CCT CTA CGA CCT CAC CGC AGT GTC GCT GGA GTG TTG GCA GCG CCT TG

       R   N   V   R   V   S   A   N   T   V   T   V   N   G   R   R
51   C CGT AAC GTT CGG GTC AGC GCC AAC ACC GTC ACC GTC AAT GGT AGA AGA A
     G GCA TTG CAA GCC CAG TCG CGG TTG TGG CAG TGG CAG TTA CCA TCT TCT T

     N   Q   R   R   R   T   G   R   Q   V   S   P   P   D   N   F   T
101  AC CAA CGG CGT CGG ACC GGA AGG CAA GTT TCT CCC CCT GAC AAT TTC ACC
     TG GTT GCC GCA GCC TGG CCT TCC GTT CAA AGA GGG GGA CTG TTA AAG TGG

     A   A   A   Q   D   L   A   Q   S   L   D   A   N   T   V   T   F
151  GCT GCT GCA CAA GAC CTC GCG CAA AGC CTT GAC GCC AAC ACC GTC ACT TT
     CGA CGA CGT GTT CTG GAG CGC GTT TCG GAA CTG CGG TTG TGG CAG TGA AA
                                  Sph I       Eco RI
       P   A   N   I   S   S   M   P   E   F   R   N   W   A   K   G
201  C CCC GCT AAC ATC TCT AGC ATG CCC GAA TTC CGG AAT TGG GCC AAG GGA A
     G GGG CGA TTG TAG AGA TCG TAC GGG CTT AAG GCC TTA ACC CGG TTC CCT T

     K   I   D   L   D   S   D   S   I   G   W   Y   F   K   Y   L   D
251  AG ATC GAC CTC GAC TCC GAT TCC ATC GGC TGG TAC TTC AAG TAC CTT GAC
     TC TAG CTG GAG CTG AGG CTA AGG TAG CCG ACC ATG AAG TTC ATG GAA CTG

     P   A   G   A   T   E   S   A   R   A   V   G   E   Y   S   K   I
301  CCA GCG GGT GCT ACA GAG TCT GCG CGC GCC GTC GGC GAG TAC TCG AAG AT
     GGT CGC CCA CGA TGT CTC AGA CGC GCG CGG CAG CCG CTC ATG AGC TTC TA

         P   D   G   L
351  C CCT GAC GGC CTC
     G GGA CTG CCG GAG
```

AN5

Figure 2.2 Schematic representation of *P71* with primers AN4 and AN5 designed to isolate HN coding sequence (panel A). Protein structural domains found on p71 are illustrated on *P71*. Panel B represents the coding sequence of HN. The amino acid sequence of HN is indicated in bold above the coding sequence and the position of primers AN4 and AN5 are indicated with double underlining. Restriction sites authentic to the coding sequence are underlined.

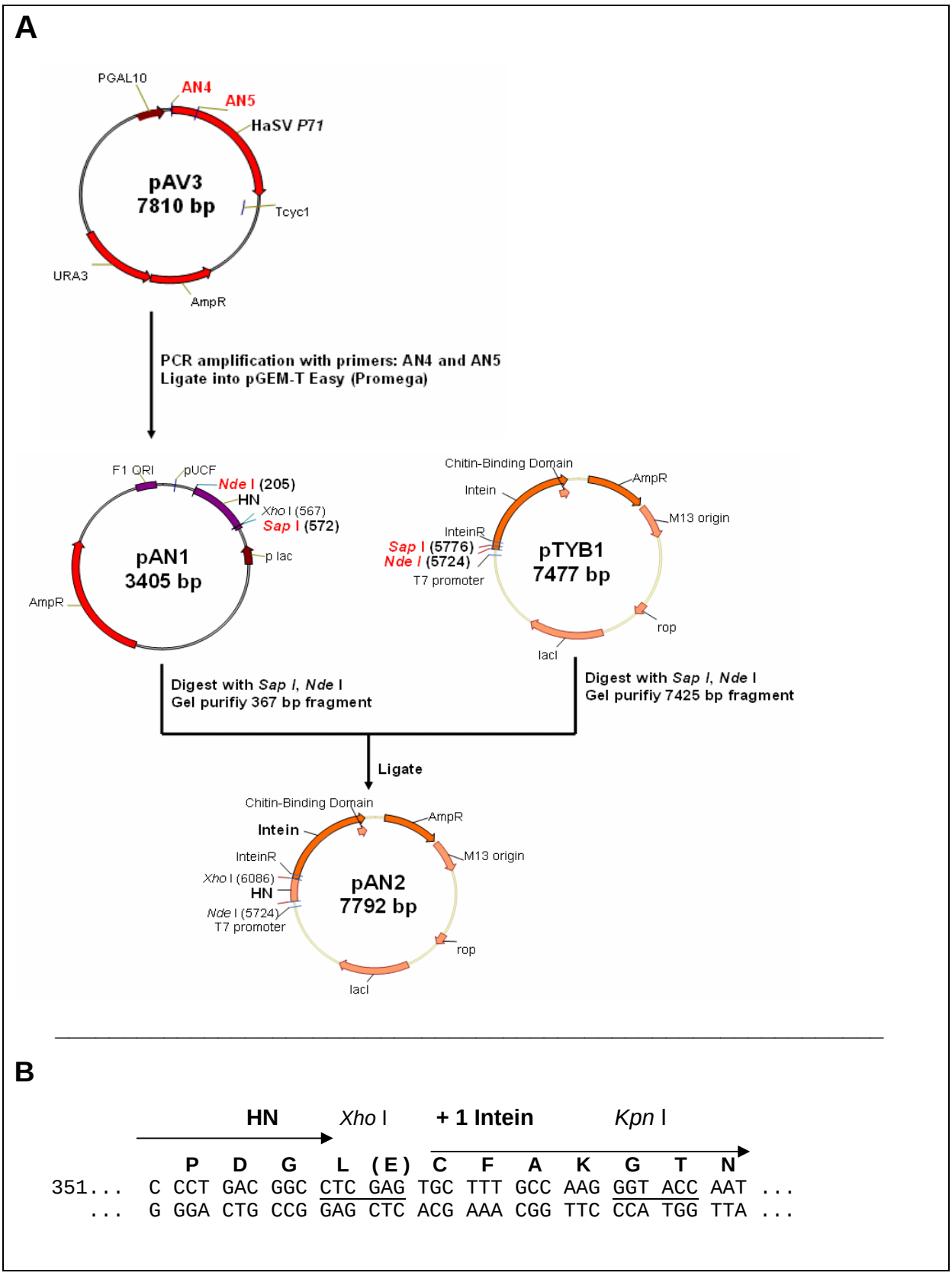


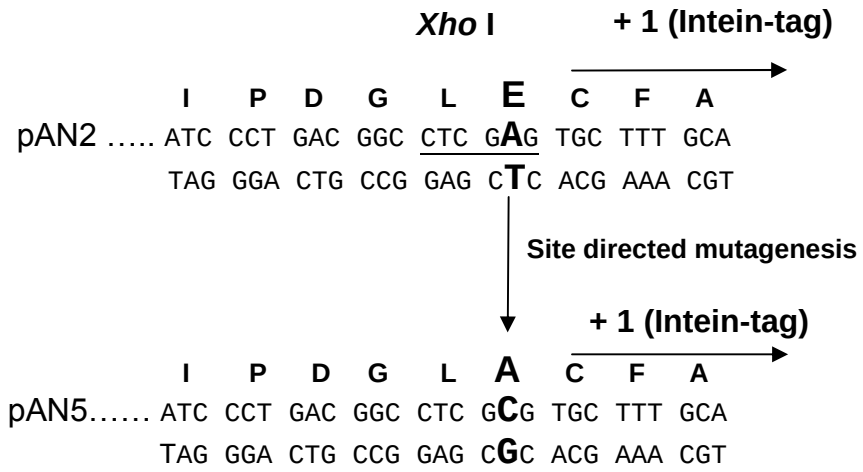
Figure 2.3 Flow diagram summarising the construction of fusion plasmid pAN2 (panel A). Panel B represents a section of the coding sequence of HN fused to the amino terminus of the intein peptide within the construct, pAN2. The amino acid sequence is indicated in bold above the coding sequence. The residue glutamate (E), in parenthesis, is a heterologous addition to the carboxy-end of the HN protein introduced by the primer AN4. Restriction sites within the nucleotide sequence are underlined.

To construct the recombinant vector containing the fused coding sequences of HN in frame with that of the intein-tag, pAN1 and vector pTYB1 (NEB) were digested with *Sap* I and *Nde* I. pTYB1 is a *T7* promoter-driven system expression vector that contains the coding sequence of the intein-tag with an upstream multiple cloning site. This site is for the C-terminal fusion of the target protein coding sequence in frame with that of the intein peptide (Appendix 2). The resultant 7462 and 363 bp transcripts of pTYB1 and pAN1 respectively were gel-purified and ligated using *T4* DNA ligase (Promega). The selected construct, pAN2, was sequenced with Intein Reverse primer (NEB) to ensure that the isolated HaSV N terminal sequence was in frame with the intein-tag, in order to produce the full-length fusion protein (Figure 2.3, panel B). The selected construct was sequenced and labeled pAN2.

Table 2.2 Sequence of primers AN6 and AN7 used to introduce the mutation Δ122 Glu→Ala.

Primer	Sequence[1]
AN6	5' – CGA TGA TCC CTG ACG GCC TCG cGT GCT TTG CCA AGG G – 3'
AN7	5' – CCC TTG GCA AAG CAC gCG AGG CCG TCA GGG ATC TTC G – 3'

Δ120Glu→Ala mutagenesis[2]



[1] Primers AN6 and AN7 used to introduce the single base pair (small letters) change within pAN2.

[2] Coding sequence before and after Site directed mutagenesis. Elimination of the *Xho* I restriction site distinguishes the mutant from the parental. Amino acid sequence is in bold above coding sequence. Large font indicates the single amino acid change.

Due to an occurring instability of the expressed fusion protein from pAN2 *in vivo*, a second version of the fusion was constructed. The QuikChange® Site-Directed Mutagenesis Kit (StrataGene) protocol was used to introduce a single amino acid change Δ122Glu→Ala within pAN2. Two primers (AN6 and AN7, corresponding to nucleotides 344-381 of the fusion protein, were designed to introduce the mutation within the nucleotide sequence (Table 2.2) by PCR using thermal cycling programme 5 (Appendix 4). Elimination of the *Xho* I recognition site distinguished the new construct (pAN5) from the parental construct, pAN2.

2.2.2 Construction of vector for the *in vitro T7* transcription of HaSV RNA1

Figure 2.4 represent the construction of the construct, pAN6, which is used for the synthesis of RNA1 transcripts.

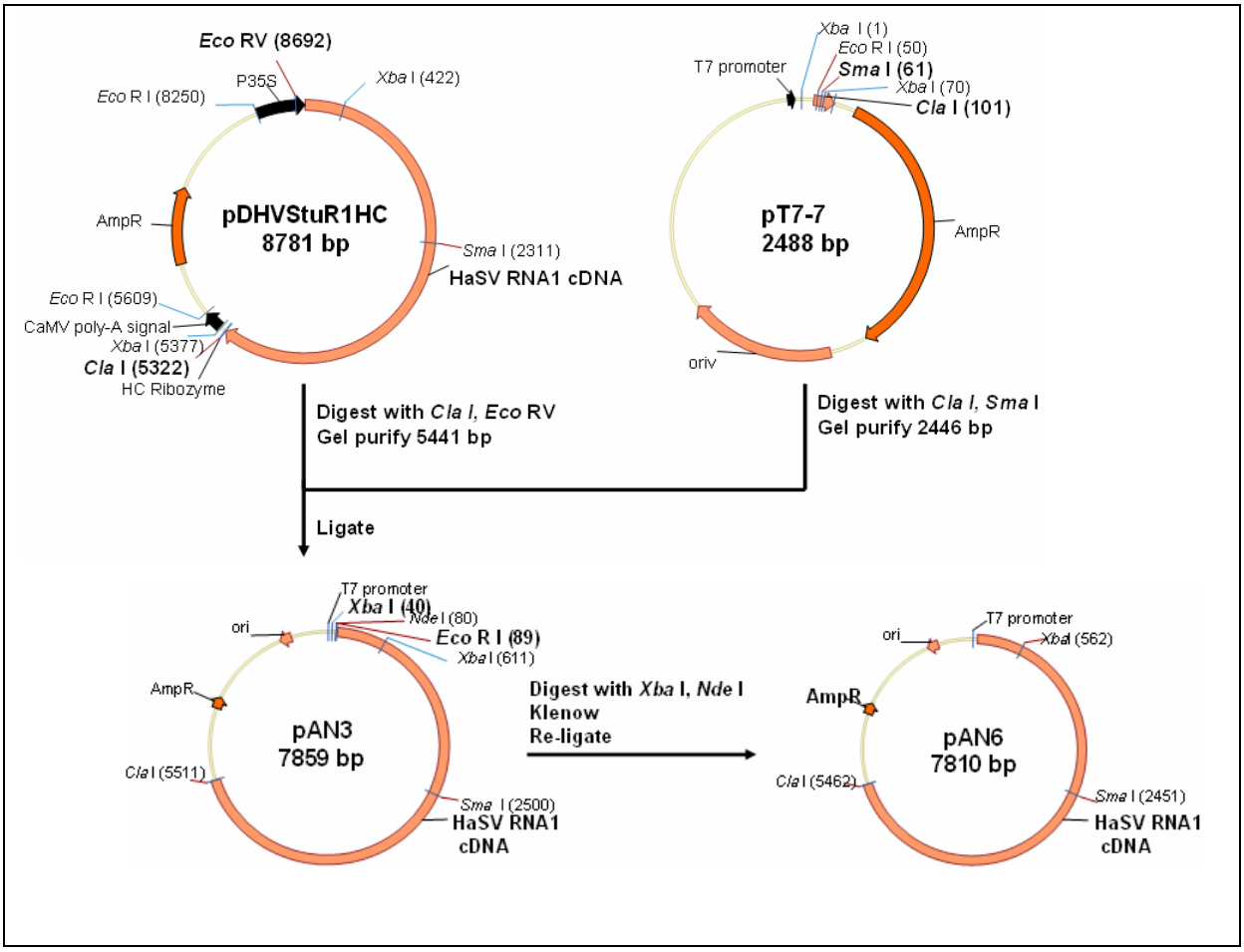


Figure 2.4 Flow diagram describing of the construction of pAN6 used for the synthesis of RNA1 transcripts.

The plasmid pDHVStuR1HC (Gordon *et al*., 2001), a plant expression vector, was used as a source of HaSV RNA1 cDNA (Figure 2.4). p*T7*-7 and pDHVStuR1HC were digested with *Sma* I and *Eco* RV, respectively and then both with *Cla* I. The resulting 5441 bp transcript was gel purified. Further digestion using *Xba* I and *Eco* RI deleted 45 bp from pAN3 in order to reduce heterologous sequence between the transcriptional start of the T7 promoter and the cDNA encoding HaSV RNA1. Klenow DNA polymerase I was used to blunt end the 3' overhang generated by *Xba* I, and to fill the *Eco* RI 5'-overhang. The treated plasmid was gel purified, re-ligated in the presence of *T4* ligase and transformed into competent *E. coli* DH5α cells. The loss of the *Nde* I between *Xba* I and *Eco* RI sites distinguished the desired construct from that of the parental, pAN3. The selected construct was named, pAN6.

2.2.3 Heterologous expression in *E. coli* BL21 (DE3)

The *E. coli* strain BL21 (DE3) (Studier and Moffat, 1986) was used for the heterologous expression of recombinant proteins (Appendix 1). Overnight, confluent starter cultures of transformed cells were used to inoculate Luria broth at an OD_{600nm} of 0.1. The cultures were incubated at 37^{o}C with shaking to midlog phase (OD_{600} ~0.5-0.8), then induced with IPTG at a final concentration of 1 mM and harvested after 4 hours. For whole cell protein analyses, 1 OD_{600} unit of cells was harvested by centrifugation at 12 000 $\times$ g for 1 minute, resuspended in 250 μL of 1 $\times$ SDS-PAGE sample buffer (62.6 mM Tris-HCl (pH 6.8, 25°C); 2% (w/v) SDS; 10% glycerol; 0.015% (w/v) bromophenol blue) and boiled for 5 minutes. Samples were stored at -20^{o}C prior to SDS-PAGE analysis.

2.2.4 Preparation of clarified extract for affinity purification

Transformed *E. coli* BL21 (DE3) cells were inoculated into 1 L of Luria broth and these cultures were induced as described in Section 2.2.3. Four hours after induction, the cells were harvested and resuspended in 50 mL Tris-buffer (20 mM Tris-HCl (pH 7.4, 25^{o}C); 500mM NaCl). 2 mL aliquots of the resuspended pellet were centrifuged at 12 000 $\times$g for 2 minutes. The buffer was discarded and the pellets stored at -20^{o}C until use. When required, the pellets were thawed on ice and resuspended in 2 mL of cell lysis buffer (20 mM Tris-HCl (pH 7.4, 25^{o}C); 500mM NaCl; 0.1% Triton X-100). To obtain total protein extracts, the samples were

sonicated twice for 30 seconds, with 30 seconds incubation on ice between each burst. The total protein extracts were pooled and centrifuged at 20 000 ×g for 30 minutes, and the supernatant (clarified extract) either used immediately or stored at -80°C with 20% glycerol. For confirmation of protein solubility, 40 μL of total cell extract was mixed with 20 μL of 3 × SDS sample buffer (187.5 mM Tris-HCl (pH 6.8, 25°C); 6% (w/v) SDS; 30% glycerol; 0.03% (w/v) bromophenol blue) and 30 μL was used for SDS-PAGE and Western analysis.

2.2.5 Affinity purification of chitin beads bound-protein

In order to purify the fusion protein from the clarified extract, approximately 1 mL of pre-washed chitin beads was added to 25 mL of clarified extract and incubated with agitation for 1 hr. This was followed by centrifugation at 5000 × g for 2 minutes to collect the chitin beads. The beads were rinsed with approximately 40 bed volumes of wash buffer (20 mM Tris HCl, pH 7.4; 500 mM NaCl) and dried by aspiration using a small bore Prot/Elec Tip (BioRad). To the dried chitin beads, an equal volume of wash buffer was then added to give a 50:50 beads/buffer suspension. This suspension was stored at 4°C. To estimate the amount of protein bound to the chitin beads, 5 μL of protein bound beads were analysed by SDS-PAGE with known concentrations (0.2, 1, 2, 5 and 10 μg) of bovine albumin serum (BSA). Protein samples were mixed with an equal volume of 1 × SDS-PAGE sample buffer (62.6 mM Tris-HCl (pH 6.8, 25°C); 2% (w/v) SDS; 10% glycerol; 0.015% (w/v) bromophenol blue), boiled for 5 minutes and loaded onto a SDS-PAGE gel for separation.

2.2.6 SDS-PAGE and Western analysis

SDS-PAGE and Western analysis were carried out according to Laemmli (1970) and Towbin *et al*., (1979), respectively. The miniVE Electrophoresis and Electrotransfer Unit (Amersham Biosciences) was used for SDS-PAGE and transfer of protein samples onto Hybond-C+ membranes (Amersham) (Appendix 5). Gels were run in duplicate and either used for Western analysis, or stained for 2 hours or O/N with Commassie Brilliant Blue followed by destaining (Appendix 5). For Western analysis the Anti-Chitin Binding Domain polyclonal antibodies (NEB) were used as the primary antibodies.

2.2.7 *In vitro* transcription of HaSV RNA1 and RNA2 cDNAs

The plasmid pL3.16 constructed by Dr Lana du Plessis (Rhodes University) (Appendix 2), contains the full-length cDNA of HaSV RNA2 under the control of the *T7* promoter sequence. pL3.16 and pAN6 were digested with *Cla* I to linearize the plasmid. The digested plasmids were purified using the DNA Clean and Concentrator Kit (Zymo Research). The *T7* transcription kit (MBI Fermentas) or Riboprobe® *in vitro T7* Transcription Systems (Promega) were used for the synthesis of RNA transcripts according to the manufacturer's instructions. However, the transcription reactions were scaled up to 100 μL. 10 μL aliquots of transcription reaction were placed into fresh RNAse free PCR tubes, and stored at -80^{o}C until use. Since the yield of RNA transcript is dependant on the amount of DNA template used, the scale up of the transcription reaction to 100 μL was expected to generate between 10-20 μg of RNA transcript (100-200 ng/μL) from 2 μg of DNA template. The RNA transcripts were visualised by separation on a 1% agarose RNA denaturing gel (Appendix 6).

2.2.8 Standard binding assay

20 μL of protein-bound beads were placed into an eppendorf tube and washed with 100 bed volumes of RNAse free buffer (20 mM HEPES pH 7.4) to eliminate NaCl from the beads:buffer homogenate and provide an RNAse free environment. The beads were dried by aspiration in the manner described in section 2.2.5. To the aspirated beads, 100 μL of 5 × RNA binding buffer (100 mM HEPES pH7.4; 25 mM NaCl; 5 mM $MgCl_2$; 1 mM DTT; 25% Glycerol) was added and the volume was brought up to the total volume of 500 μl with RNAse free H_2O. 5 μL (500 – 1000 ng) of RNA transcript was added and the mixture allowed to incubate with agitation at room temperature for 25 minutes. After incubation, the beads were collected by centrifugation at 3000 × g for 3 minutes and washed with approximately 100 bed volume of 1 × RNA binding buffer. The beads were again dried by aspiration and 30 μL of 1/20 diluted gel loading buffer (50% glycerol; 1mM Na_2EDTA; 0.4% bromophenol blue) containing 1% SDS added. The mixture was allowed to incubate for 10 minutes to ensure complete disruption of bound entities by SDS. The samples were separated by electrophoresis on a 1% non-denaturing agarose gel containing 0.5 μg/mL ethidium bromide and using Tris-acetate-NA_2EDTA buffer as the running buffer.

2.3 Results

2.3.1 Outline of the binding assay

The protein component of the binding assay is schematically illustrated in Figure 2.5, panel A. The fusion protein is composed of the 55-kDa affinity tag (hereafter the "intein peptide") fused to the potential RNA-binding domain, the HN peptide. The latter peptide represents the N terminal domain of HaSV p71.

The approach used in developing the binding assay is summarised in Figure 2.5, panel B. For the preparation of the protein component of the binding assay, the fusion protein was purified attached to the chitin beads (Figure 2.5, panel B, section I-II), because these were required for the "pull down" part of the binding assay. The purified protein bound beads was stored in an equal volume of buffer at 4^{o}C (Figure 2.5, panel B, section II). This eliminated variability in protein concentration between each binding assay. For SDS-PAGE analysis of the protein bound to the chitin beads, a solution of 1% SDS was added to disrupt the affinity interaction between the chitin beads and the fusion protein (Figure 2.5, panel B, section IV). This disruption was necessary as the large diameters (50-100 μm) of the beads would have an effect on SDS-PAGE.

To perform the binding assay, a certain amount of RNAses-free protein bound beads was incubated for 30 min at 25^{o}C in a buffer solution containing synthesized RNA transcripts (Figure 2.5, panel B, section V). The excess or unbound RNA was removed by rinsing the protein bound beads several times followed by brief centrifugation to settle the beads (Figure 2.5, panel B, section VI). To confirm interaction between the fusion protein and nucleic acid, RNA sample buffer containing SDS was added to the dried beads, thereby precipitating any bound RNA (Figure 2.5, panel B, section VI). This precipitate was analysed by gel electrophoresis on an agarose gel containing ethidium bromide.

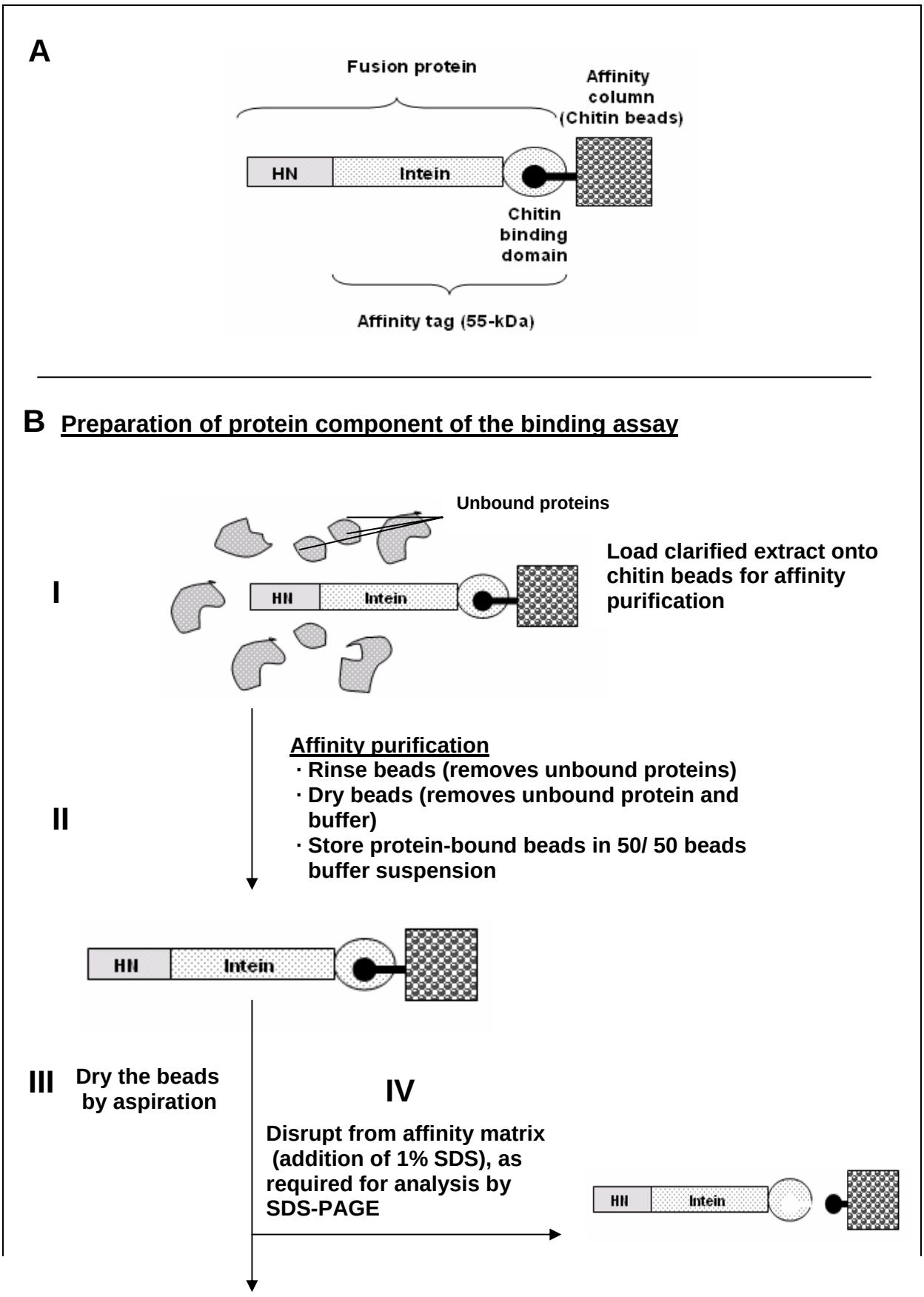


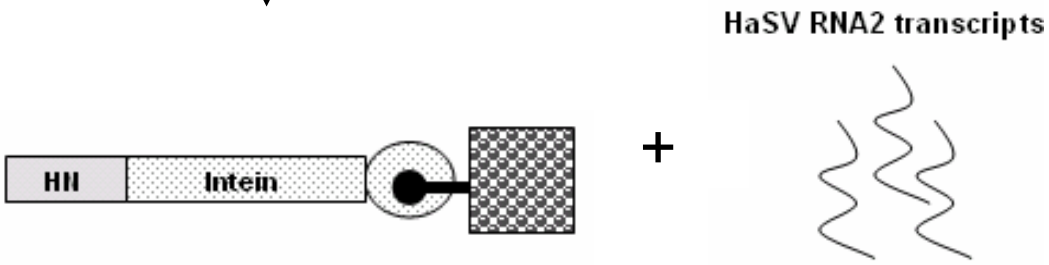
Figure 2.5 Schematic flow diagram of the binding assay. Panel A, represents the HN peptide fused to the 55-kDa intein peptide bound to the chitin beads. Panel B, outlines the preparation of the fusion protein bound to the chitin beads and the binding assay itself.

Binding assay

V

Rinse beads with RNAse free RNA binding buffer
Dry beads (Aspiration)

Add RNAse free RNA binding buffer
Add RNA transcripts
Incubate at 25°C with agitation



VI

Wash beads, brief centrifugation to settle beads
Aspirate dry beads
Add RNA sample buffer containing 1% SDS

Analyse by agarose gel electrophoresis

Figure 2.5 Continued…

2.3.2 Identification of the coding sequence representing the N terminal domain of p71

An alignment of the amino acid sequences of p71 and NωV coat proteins constructed by Hanzlik *et al*. (1995) was used to determine the region of p71 representing the N terminal domain (Figure 2.6). It was determined that residues 1-121 of p71 represent this domain, since a high sequence similarity (81%) existed particularly for residues 48-274 between NωV and p71 coat proteins (Figure 2.6). The isolated coding sequence for residues 1-121 of p71 would encode for a peptide, labeled HN, with a calculated M_R of approximately 13-kDa.

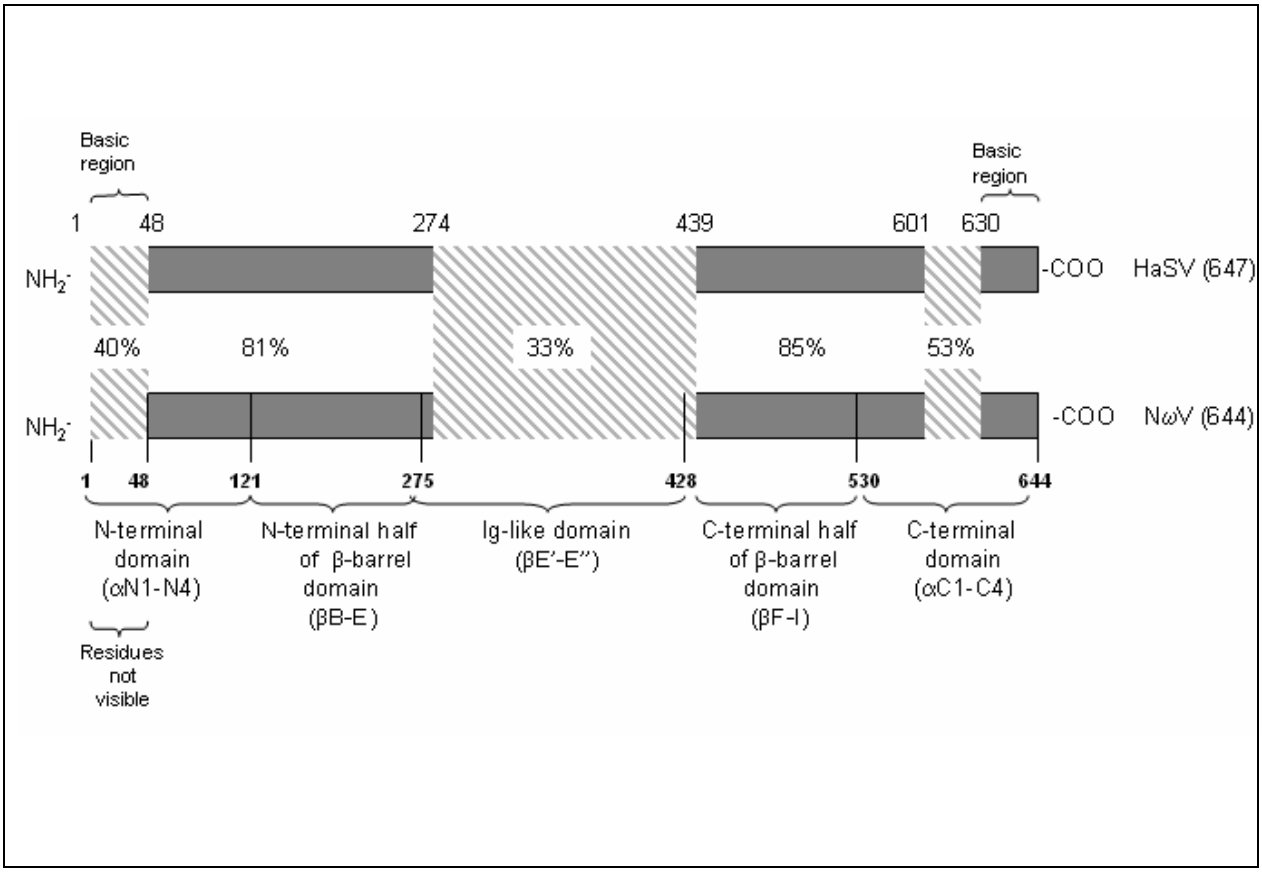


Figure 2.6 Alignment of the amino acid sequences of p71 and NωV coat proteins. Distinct regions that share greater and lesser similarities are indicated with percentage identities between the two coat proteins. For NωV coat protein, the distinct protein structural domains are outlined below with secondary structures indicated in parenthesis. The length of the viral coat proteins is indicated in parenthesis adjacent to the names of the viruses. The schematic alignment with percentage similarities is obtained from Hanzlik *et al*. (1995) and the structural domains and protein secondary structures are obtained from Helgstrand *et al*. (2004).

2.3.3 Preparation of protein component of RNA-binding assay

Expression of the HN-fusion protein in E. coli BL21 (DE3)

It was necessary to characterise the expression of the HN-fusion protein from the construct pAN2 in *E. coli* BL21 (DE3), since expression from pTYB vectors could be affected by several factors: (i) the *E. coli* strain; (ii) cell culture conditions (e.g., temperature, aeration, cell density, etc.); and (iii) protein expression induction conditions (temperature, duration, the IPTG concentration, etc). Since the pTYB vectors were not recommended for testing expression, the vector pMYB5, which carries the coding sequences of the *E. coli* 42-kDa maltose binding protein (MBP) fused in frame to that of the intein peptide was used as a positive control.

To determine whether the HN-fusion protein was expressed, total protein extracts from induced and uninduced *E. coli* BL21 (DE3) cells, transformed with pAN2, were analysed by SDS-PAGE. Despite expression of the 55-kDa intein peptide and the 97-kDa MPB fusion protein, there was no visible expression of the 68-kDa HN-fusion protein (Figure 2.7, compare lanes 2 and 3 with 1).

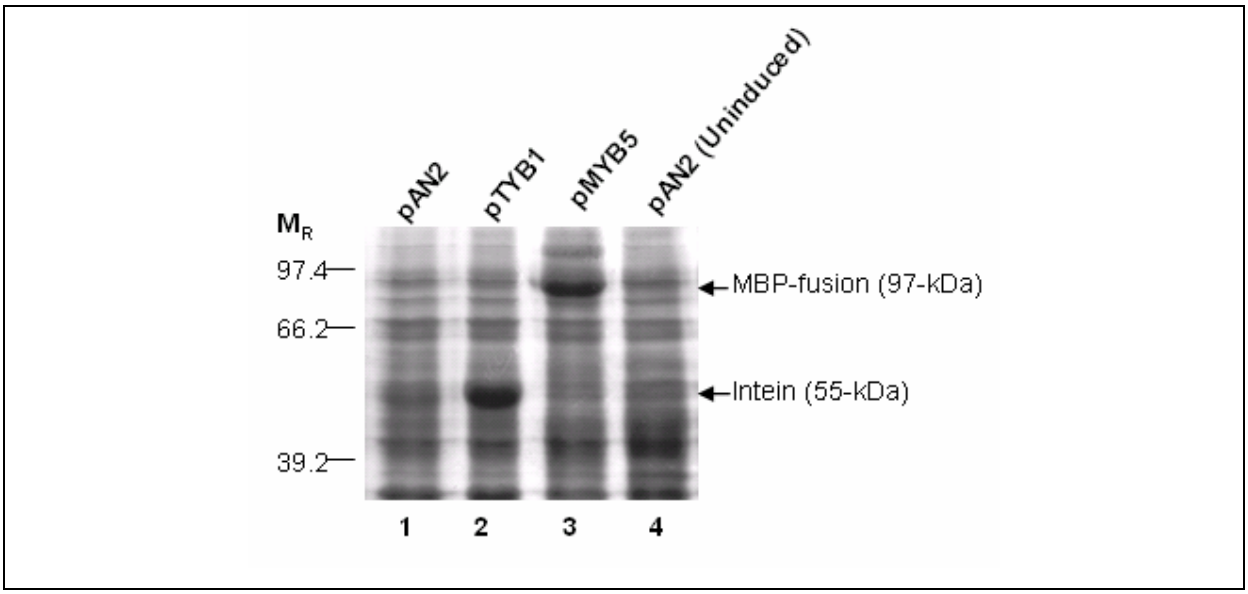


Figure 2.7 Expression of HN-fusion protein in BL21 cells. Commassie stained SDS-PAGE gel of cell extracts obtained from BL21 cells harbouring pAN2, pTYB1 or pMYB5 (lanes 1, 2, and 3, respectively). Lane 4 denotes cells harbouring pAN2 that were not induced with 1 mM IPTG. M_R denotes the sizes of the protein molecular weight markers (kDa). The position of 97-kDa MPB fusion (pMYB5) and the 55-kDa intein peptide (pTYB1) are indicated on the right hand side of the gel.

Inspection of the amino acid sequence of HN showed that the C-terminal residue adjacent to the N terminus of the intein peptide is glutamate (Table 2.2). A glutamate at this position could cause up to 50% *in vivo* cleavage by inducing the cleavage between the HN and the intein peptides, thus causing 50% loss of the expressed fusion protein. Since *in vivo* cleavage could be reduced by lowering the induction temperature, four different temperatures (15, 20, 30, and 37°C) were tested. The cells were harvested 4 and 16 hours after induction. The 16 hour induction period was necessary at the lower induction temperatures, so that sufficient *T7* RNA polymerase could accumulate within the cells.

Lowering the temperature had no effect on the expression of HN-fusion protein, as there was no visible presence of a 68-kDa band in induced cells compared with uninduced cells (Figure 2.8, compare lanes 1-8 with 11). A band co-migrating with the 55-kDa intein peptide was visible at all temperatures used to express the HN-fusion protein, indicating that *in vivo* cleavage still took place (Figure 2.8, compare lanes 1-8 with 9). The 13-kDa HN protein could not be observed on the SDS-PAGE gel due to its small size.

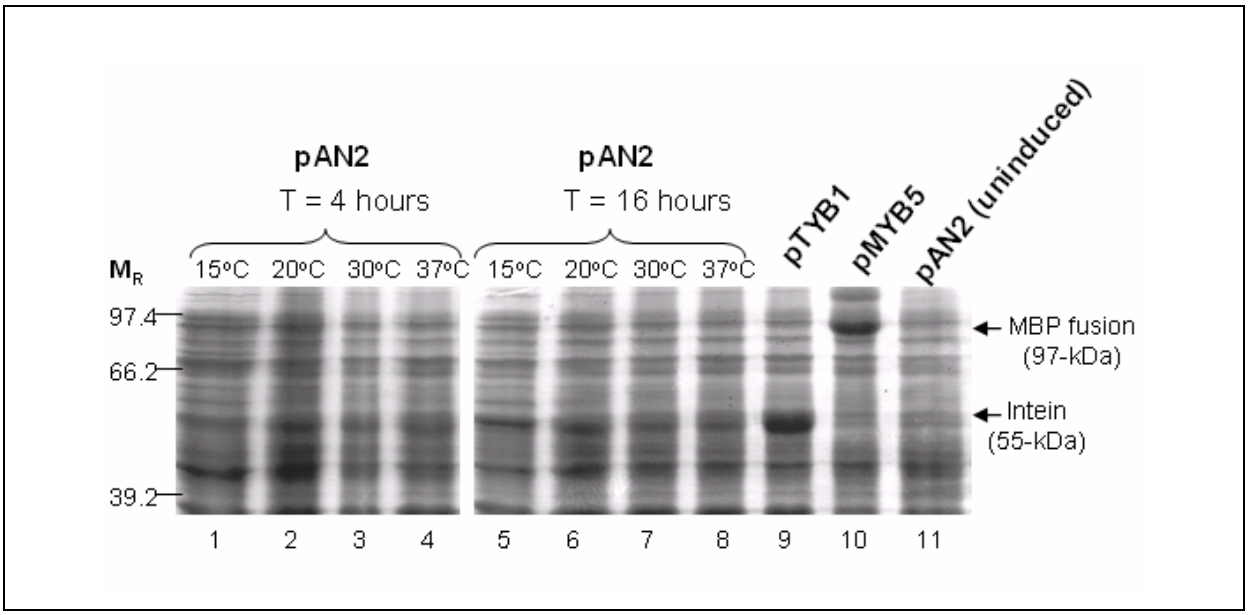


Figure 2.8 Effect of different induction temperatures on the expression of HN-fusion protein in *E. coli* BL21 cells. Commassie stained SDS-PAGE gel of cell extracts obtained from BL21 cells harbouring pAN2 that were harvested at T= 4 and 16 hours after induction at 15-37°C (lanes 1-4 and 5-8, respectively). Cells harbouring pTYB1, pMYB5 and pAN2 (uninduced) were harvested at 4 hours after induction that was carried out at 37°C (lanes 9, 10 and 11, respectively). Lane 11 denotes cells harbouring pAN2 that were not induced with 1 mM IPTG (lane 11). M_R denotes the sizes of the molecular weight markers (kDa). The position of 97-kDa MPB fusion and the 55-kDa intein peptide are indicated on the right hand side of the gel.

To address the problem of *in vivo* cleavage, site directed mutagenesis was used to substitute alanine for glutamate (Δ122Glu→Ala) as described in Section 2.2.1. Alanine was chosen, as it has not been found to cause *in vivo* cleavage of the fusion protein.

Expression of the 68-kDa HN-fusion protein from pAN5, showed a significant reduction of *in vivo* cleavage of the HN-fusion protein at 2 and 4 hours after induction compared with the expression of the fusion protein from the former construct, pAN2 (Figure 2.9, panel A, compare lanes 2 and 3 with 6 and 7). To confirm that the 68-kDa band observed was the HN-fusion protein (Figure 2.9 panel A, lanes 2 and 3), the proteins were electroblotted onto nitrocellulose membrane and probed with antibodies specific for the intein peptide. The bands that migrated at a M_R of 68-kDa (Figure 2.9, panel A, lanes 2 and 3) were detected confirming HN-fusion protein (Figure 2.9, panel B, lanes 2 and 3). The detected bands representing the 55-kDa intein peptide and the 97-kDa MBP-fusion protein, confirmed the specificity of the primary antibodies for the intein peptide (Figure 2.9, panel B, lanes 8 and 9).

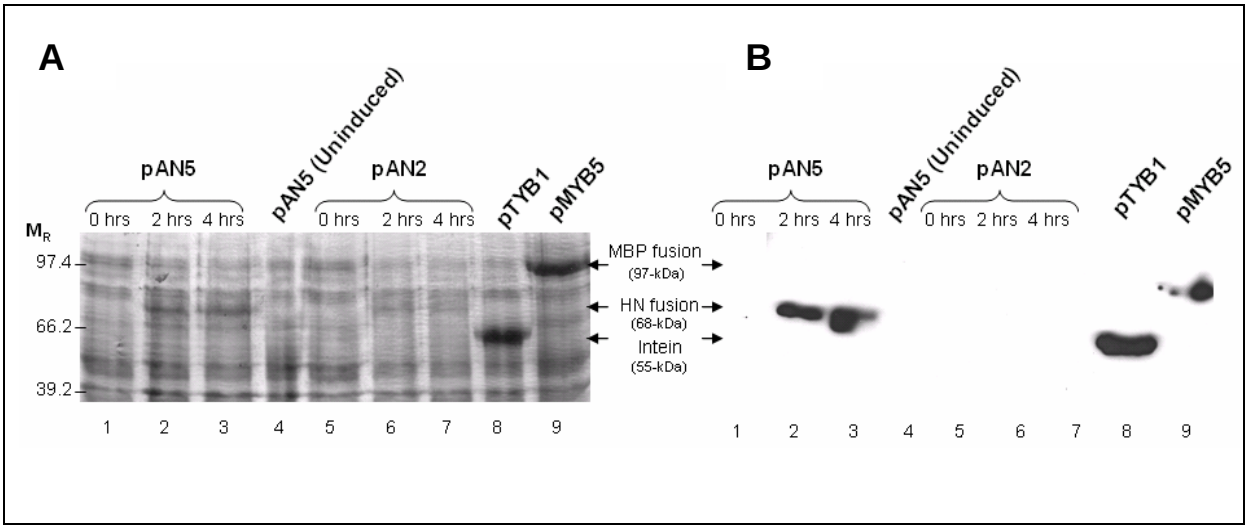


Figure 2.9 Effect on the expression of HN-fusion protein by a Δ122Glu→Ala within its coding sequence. Panel A, Commassie stained SDS-PAGE gel of cell extracts obtained from cells harbouring pAN2 and pAN5 harvested at 0-4 hours after induction (lanes 1-3 and 5-7, respectively). Cells harbouring pAN5 (uninduced), pTYB1 and pMYB5 were harvested at 4 hours after induction at 37°C (lanes 4, 8 and 9, respectively). Lane 4 denotes cells harbouring pAN5 that were not induced with 1 mM IPTG. Induction was carried out at 37°C for all cells. Panel B, Western analysis of protein extracts that were separated on a SDS-PAGE gel as depicted in Panel A, electroblotted onto a nitrocellulose membrane and probed with anti-chitin binding domain polyclonal antibodies. M_R denotes sizes of the molecular weight markers (kDa). The position of 97-kDa MPB fusion protein, the 68-kDa HN-fusion protein and the 55-kDa intein peptide are indicated in between the gels.

Optimisation of HN-fusion protein expression

An attempt was made to determine whether the yield of HN-fusion protein expressed would be affected by increasing the length of induction time. This was attempted, since the expression of the HN-fusion protein was low compared with that of 97-kDa MPB fusion protein (Figure 2.9, panel A, compare lanes 3 with 9). Cells expressing HN-fusion protein were harvested at 1 hour intervals up to a maximum of 6 hours. According to the analysis of total protein extracts, the position of the 68-kDa HN-fusion protein band could be noted 1 hour after induction (Figure 2.10, lane 2). This protein band was absent within uninduced cells (Figure 2.10, lane 7). The intensity of the HN-fusion protein band increased up to 4 hours after induction (Figure 2.10, lanes 1-4); thereafter, there was no significant increase (Figure 2.10, compare lanes 4 with 5 and 6). The intensity of the HN-fusion protein band was still significantly less than that of the 55-kDa intein peptide (Figure 2.10, compare lanes 4-6 with 8).

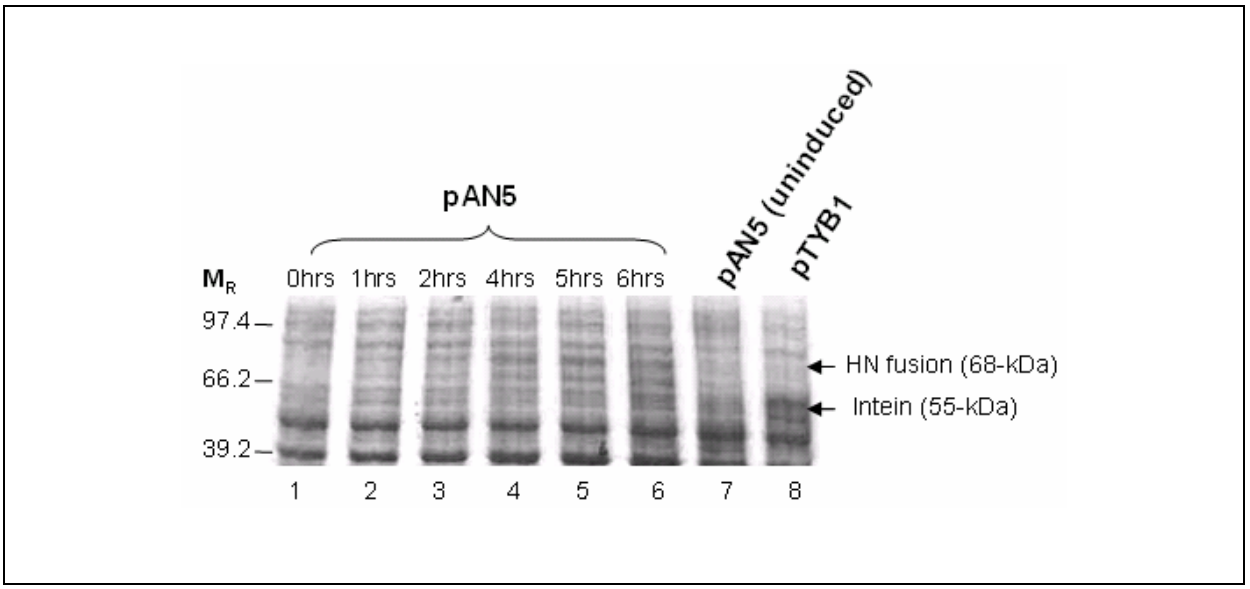


Figure 2.10 Effect of induction time on the expression of HN-fusion protein. Commassie stained SDS-PAGE of cell extracts from cells harbouring pAN5 harvested at 0-6 hours after induction (lanes 1-6, respectively). Cells harbouring pAN5 (uninduced) and pTYB1 were harvested at 4 hours after induction (lanes 7 and 8, respectively). Lane 7 denotes cells harbouring pAN5 that were not induced with 1 mM IPTG. Induction was carried out at 37°C for all cells. M_R denotes the sizes of the molecular weight markers (kDa). The position of 68-kDa HN-fusion protein and the 55-kDa intein peptide are indicated on the right hand side of the gel.

An alternative explanation for the low expression of the HN-fusion protein could be the presence of infrequently used *E. coli* arginine codons such as AGG and AGA within the coding sequence. Consecutive infrequently used arginine codons can terminate transcription, or cause frameshift during translation (Zhang *et al*., 1991; Rosenburg *et al*., 1993).

Inspection of the HN-coding sequence revealed the presence of three infrequently used arginine codons. Two of these codons were consecutive within the sequence, which might have been the cause of low levels of HN-fusion protein expression. A helper plasmid containing the *argU* gene encoding tRNAs that recognised rare arginine codons was co-expressed with the construct, pAN5, as an attempt to increase HN-fusion protein expression. There was no increase in the yield of protein expressed (data not shown).

Affinity purification of chitin bound HN-fusion protein

SDS-PAGE analysis was used to monitor the affinity purification of the HN-fusion protein from clarified cell extracts of cells expressing the protein. Protein samples were collected at each step of the purification procedure for this analysis.

No comparison could be made between the clarified extracts of cells expressing the HN-fusion protein before and after incubation with the chitin beads, as there was no visible band denoting the 68-kDa protein (Figure 2.11, panel A, lanes 1 and 2). Despite this, there was purified protein that migrated at a M_R of 68-kDa (Figure 2.11, panel A, lane 6). This M_R corresponded to the estimated size of the HN-fusion protein that was disrupted from the chitin beads by the addition of SDS.

As anticipated for the first gel, there was a significant reduction of the 55-kDa intein peptide band intensity between the clarified extracts of cells expressing the peptide before and after incubation with the chitin beads (Figure 2.11, panel B, compare lanes 1 with 2). This showed that the binding of the intein peptide to the chitin beads was efficient. This could be confirmed by the presence of purified peptide that was disrupted from the chitin beads by the addition of SDS (Figure 2.11, panel B, lane 6). Analysis of samples of the wash buffer used to rinse the chitin beads after incubation

showed that the removal of the excess unbound protein was successful (Figure 2.11, panels A and B, lanes 3-5).

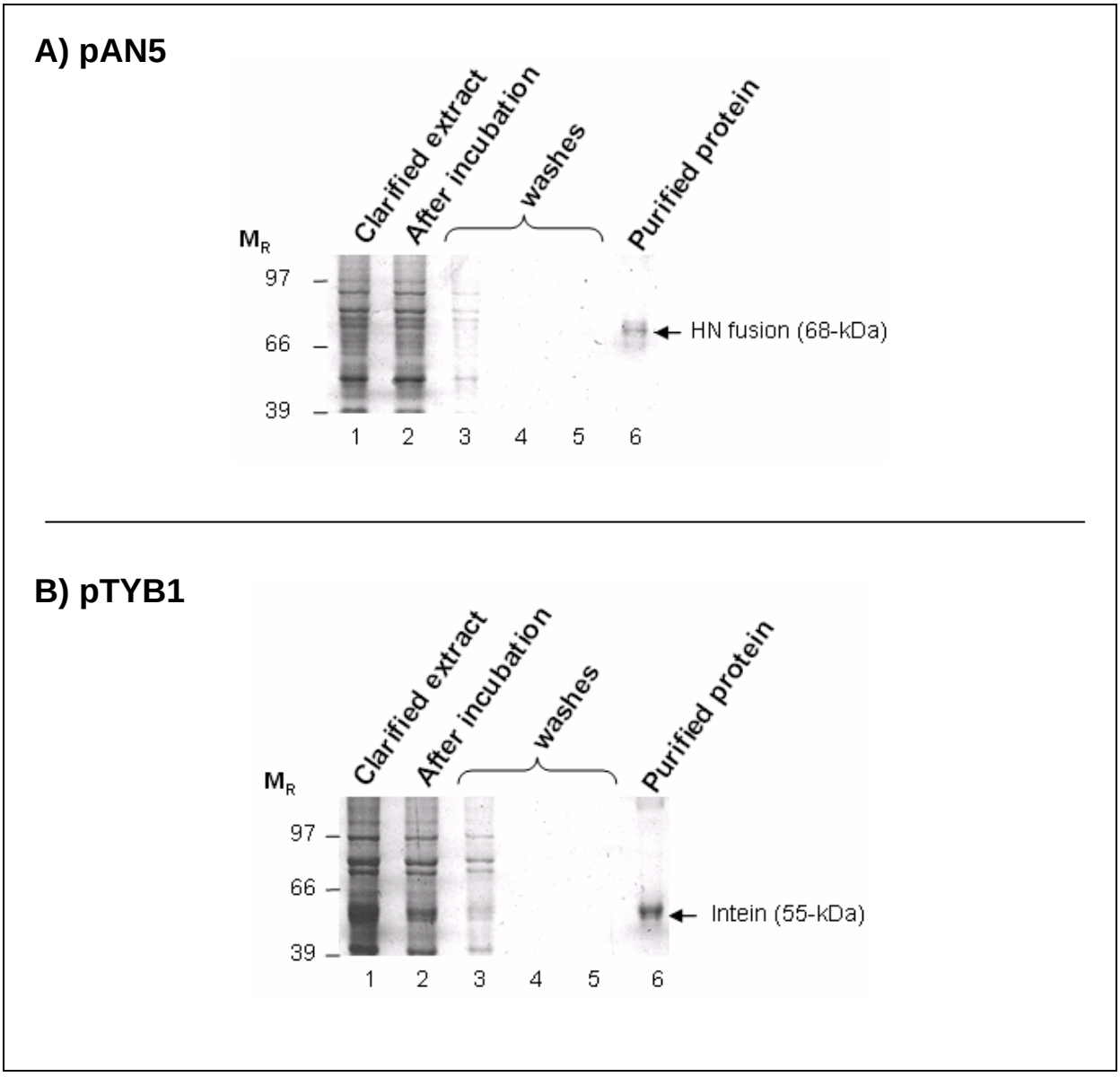


Figure 2.11 Affinity purification of the HN-fusion protein and intein peptide from BL21 cells clarified extract. Panel A, represents a Commassie stained SDS-PAGE gel of protein samples obtained from different steps during the purification procedure of the HN-fusion protein and panel B, the intein peptide from clarified extract of BL21 cells. For both panel A and B: lanes 1, represent clarified extract of cells harbouring pAN5 and pTYB1; lane 2, clarified extract after incubation with chitin beads; lane 3-5 several washes used to remove unbound protein; lane 6, purified protein disrupted from the chitin beads (5 μL of protein-bound beads were used) using a solution of 1% SDS, respectively. M_R denotes the sizes of the molecular weight markers (kDa). The position of the purified 68-kDa HN-fusion and the 55-kDa intein peptides are indicated on the right hand side of the gels.

Stability of protein-chitin bead complex

As described in section 2.2.5, a 50:50 beads/buffer suspension was prepared for the HN-fusion protein, which was stored at 4°C. This was done in order to eliminate variability in protein concentration between experiments, since the same batch could be used in each binding assay. It was important that the stability of the proteins bound to the chitin beads over long storage periods be tested, as the binding assay requires a stable protein component. To test this, two equal aliquots of protein bound beads were placed into separate eppendorfs, which were stored for 8 and 14 days. Figure 2.12 represents an SDS-PAGE analysis of protein bound beads obtained from the initial batch of beads/buffer suspension (day 1). Also represented in Figure 2.12 are the two aliquots that were stored for 8 and 14 days.

This SDS-PAGE analysis revealed that although the batch of the HN-fusion protein bound chitin beads was stored in column buffer, which contained no intein self-cleavage inducers (such as thiols), there was a clear increase in the presence of a 55-kDa band (Figure 2.12, lanes 2 and 3), which corresponds to the intein peptide (lanes 4-6). This suggested that the HN-fusion protein was undergoing self-cleavage over time, as there was a decrease in intensity of the 68-kDa band (Figure 2.12, compare lanes 2 and 3 with 1).

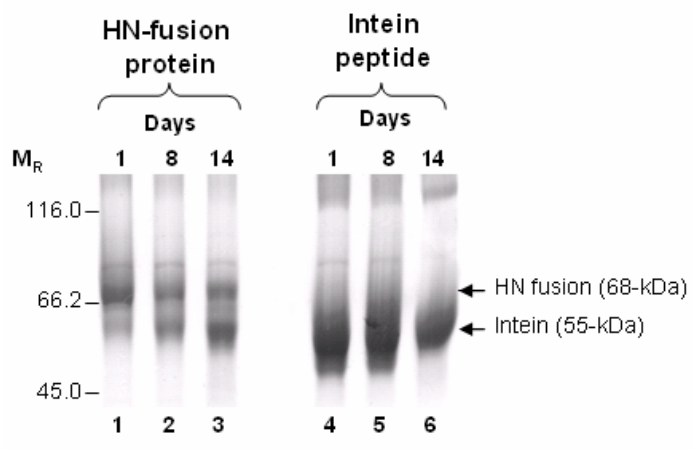


Figure 2.12 Stability of HN-fusion protein over time at 4°C. Commassie stained SDS-PAGE of purified HN-fusion protein, and intein bound to chitin beads were stored at 4°C for 1-14 days (lanes 1-3 and 4-6, respectively). The sizes of the molecular weight markers are indicated on the left hand side of the gel, while the position of 68-kDa HN-fusion protein and the purified 55-kDa intein peptide are indicated on the right hand side of the gel.

The 55-kDa intein peptide remained bound to the chitin beads over the same period of time (Figure 2.12, lanes 4-6). The presence of the protein band corresponding to the 55-kDa intein peptide within the fusion protein lanes, suggested that self-cleavage is the most likely cause of the HN-fusion protein instability.

The manufacturer's instructions of the IMPACT^{TM}-CN system suggest that no amino acud that induces *in vivo* cleavage be placed within three residues upstream of the target protein C-terminus adjacent to the intein. Inspection of HN amino acid sequence revealed that the fourth residue upstream of the peptide C-terminus is aspartate (Table 2.2, pAN5). Aspartate is known to cause 100% *in vivo* cleavage if positioned within three amino acids adjacent to the N terminus of the intein. The IMPACT^{TM}-CN *in vivo* cleavage data was based on the MPB protein fused to intein, with amino acid substitutions within three residues upstream of the C-terminus of MPB. The fourth position of aspartate within the C-terminus of HN protein might have been responsible for the breakdown of the fusion protein (Figure 2.12). Since sufficient recombinant HN-fusion protein was present until eight days, the construct, pAN5, was used to express the protein for the binding assays.

Saturation of chitin beads with protein and protein estimation

It could be possible that the low expression of HN-fusion protein in *E. coli* BL21 cells affected the yield of the protein purified. This suggestion was based on comparison of the amount of purified HN-fusion protein with that of the intein peptide (Figure 2.11, panel A and B, compare lanes 6; Figure 2.12 compare lanes 1 with 4).

An attempt to increase the amount of purified HN-fusion protein was to incubate 1 mL of chitin beads with 50 mL of clarified extract of cells expressing the HN-fusion protein. SDS-PAGE analysis was used to estimate the amount of purified protein by comparing it with known amounts of BSA protein.

As anticipated, the intensity of the 68-kDa HN-fusion protein band within the clarified cell extract of cells expressing the protein was lower than that of the 55-kDa intein peptide band (Figure 2.13, compare lanes 1 and 3). The amount of the purified HN-fusion protein and intein peptide (Figure 2.13, lane 2 and 4) could be estimated by comparing the intensity of these protein bands with those of known amounts of BSA

protein (Figure 2.13, lanes 5-9). It was calculated that per mL of chitin beads there is approximately 0.5 mg of HN-fusion protein and 2 mg of intein peptide. Only the topmost of the two protein bands observed for the HN-fusion protein was used for the calculation of protein amount (Figure 2.13, lane 2).

Although a fresh batch of HN-fusion protein beads/buffer suspension was used, there was evidence of protein degradation (Figure 2.13, lane 2) as discussed earlier. It was suggested that instability associated with the HN-fusion protein might have been occurring to a certain extent within the expressing cells. The breakdown of the HN-fusion protein within the expressing cells might be visible when larger volumes of cell extracts are used.

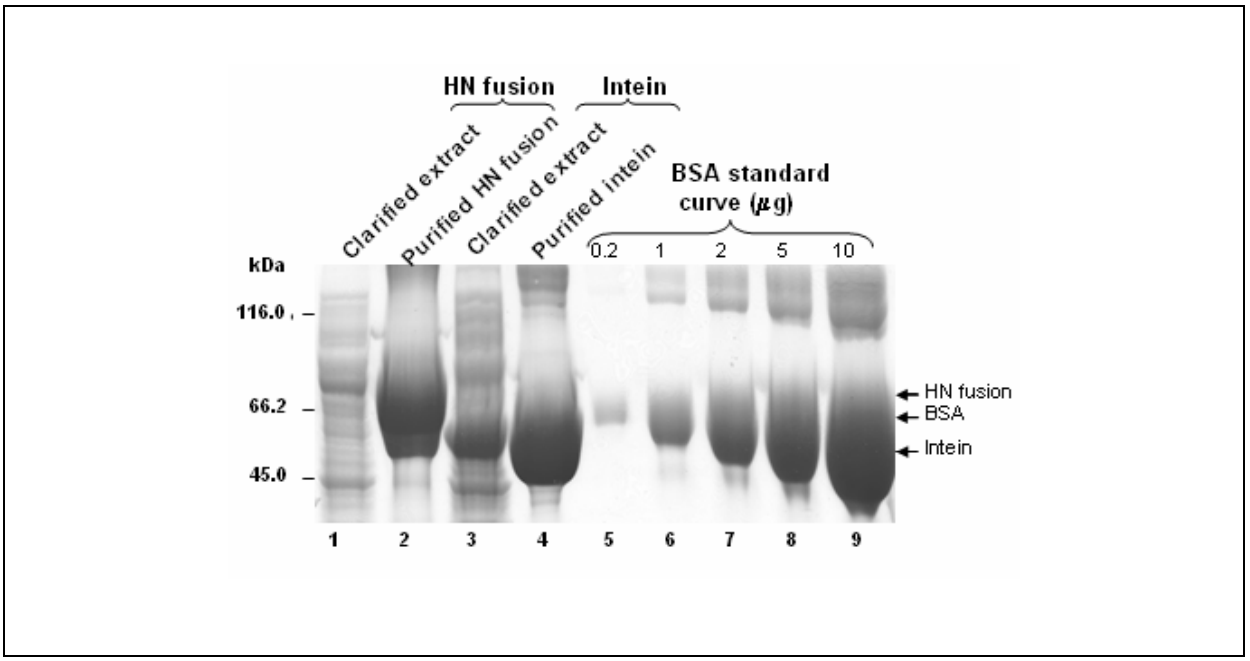


Figure 2.13 Estimation of HN-fusion protein bound to chitin beads. Lanes 2 and 4 denotes the Commassie stained SDS-PAGE gel of HN-fusion protein and intein peptide that was disrupted from the chitin beads using a solution of 1% SDS. In addition, known amounts of BSA (0.2-10 μg) were used to estimate the amount of protein bound to the chitin beads (lanes 5-9). Lanes 1 and 3 represent clarified extract prepared from crude extract of BL21 cells harbouring pAN5 and pTYB1, respectively. The sizes of the molecular weight markers are indicated on the left hand side of the gel, while the position of 68-kDa HN-fusion protein, 66.2-kDa BSA and the purified 55-kDa intein peptide are indicated on the right hand side of the gel.

Due to the low amount of HN-fusion protein purified from 1 mL of chitin beads, it was decided to use four times the volume of chitin beads originally used for the binding assays. This would ensure that the amount of HN-fusion protein being used is the

same as that used as the control (intein peptide). The reason for this decision was that if no RNA was pulled down by the intein peptide alone, but was pulled down by the HN-fusion protein, it was reasonable to assume that the HN portion of the HN-fusion protein would interact specifically with the RNA introduced.

2.3.4 Evaluation of RNA transcripts

In vitro synthesized RNA transcripts were analysed by denaturing gel electrophoresis in order to confirm the sizes. It was confirmed that *in vitro* synthesized RNA1 and RNA2 had the same M_R as calculated between the transcriptional start of the *T7* promoter and the *Cla* I restriction site of plasmids pAN6 and pL3.16, respectively (Figure 2.14, panel A and C). However, gel electrophoresis of the HaSV RNA2 transcription reaction revealed a further transcript that was calculated to be approximately 600 nt (Figure 2.14, panel A, *iv*R2), which may have resulted from premature transcription termination of the RNA. This RNA transcript was more apparent on the native gel (Figure 2.14, panel B).

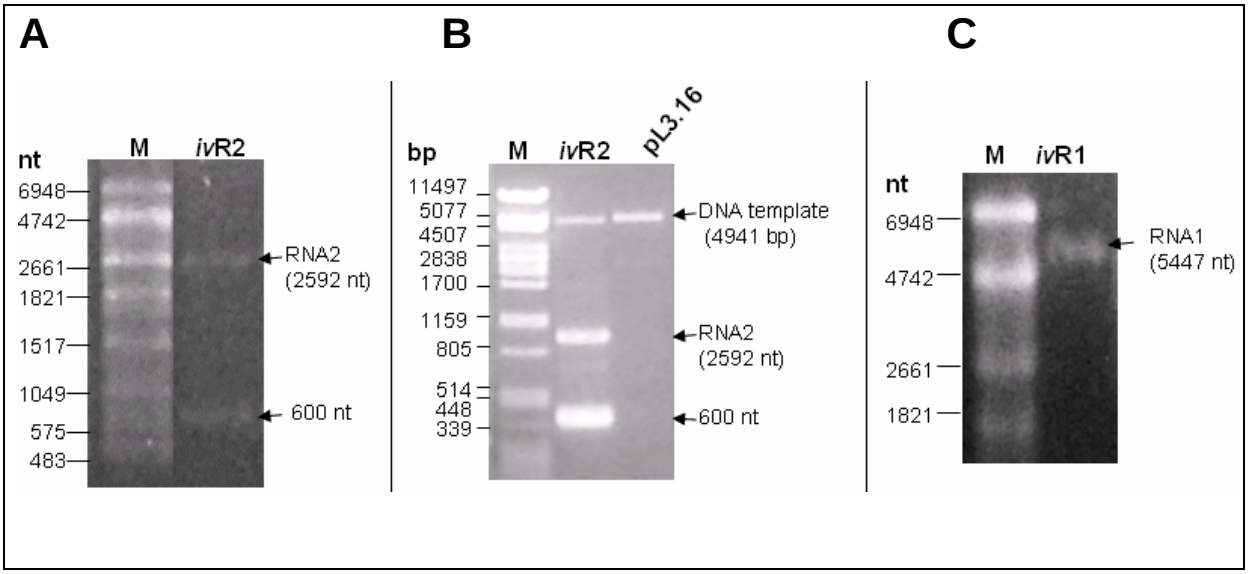


Figure 2.14 Analysis of *in vitro* synthesized HaSV RNA1 and RNA2 transcripts. Panel A, analysis of RNA transcripts by denaturing gel electrophoresis of *in vitro* synthesized HaSV RNA2 (*iv*R2). *iv*R2 was generated from the plasmid, pL3.16. Panel B, analysis of ivR2 on native RNA gel. Panel C, analysis of RNA by denaturing gel electrophoresis of *in vitro* synthesized HaSV RNA1 (*iv*R1). *iv*R1 was generated from the plasmid, pAN6. The sizes of the *Cla* I linearized DNA templates and RNA transcripts are indicated on the right hand side of the gels. The molecular weight marker (M) used for the RNA denaturing gels is the Roche RNA molecular weight marker 1 and for the native gels a *Pst* I digest of lambda-DNA.

2.3.5 Binding of HaSV genomic RNAs to HN-fusion protein *in vitro*

In the first binding experiment, RNA2 transcripts (2592 and 600 nt transcripts) were incubated for 25 min in binding buffer with HN-fusion protein bound to the chitin beads. The same was done with the intein peptide bound to chitin beads. The samples were then precipitated and after several washes, analysed by agarose gel electrophoresis. No RNA was pulled down in the sample containing chitin beads (Figure 2.15, lane 3), while the 600 nt transcript was pulled down by both the HN-fusion protein and the intein peptide (Figure 2.15, lanes 1 and 2). There did not appear to be any interaction between the HN-fusion protein or intein peptide with the 2592 nt transcript (Figure 2.15, lanes 1 and 2). The data suggested that the intein peptide was capable of interaction with nucleic acids, and that there was no pull down of 2592 nt transcript by the HN-fusion protein. Thus, further optimisation of the binding assay was necessary.

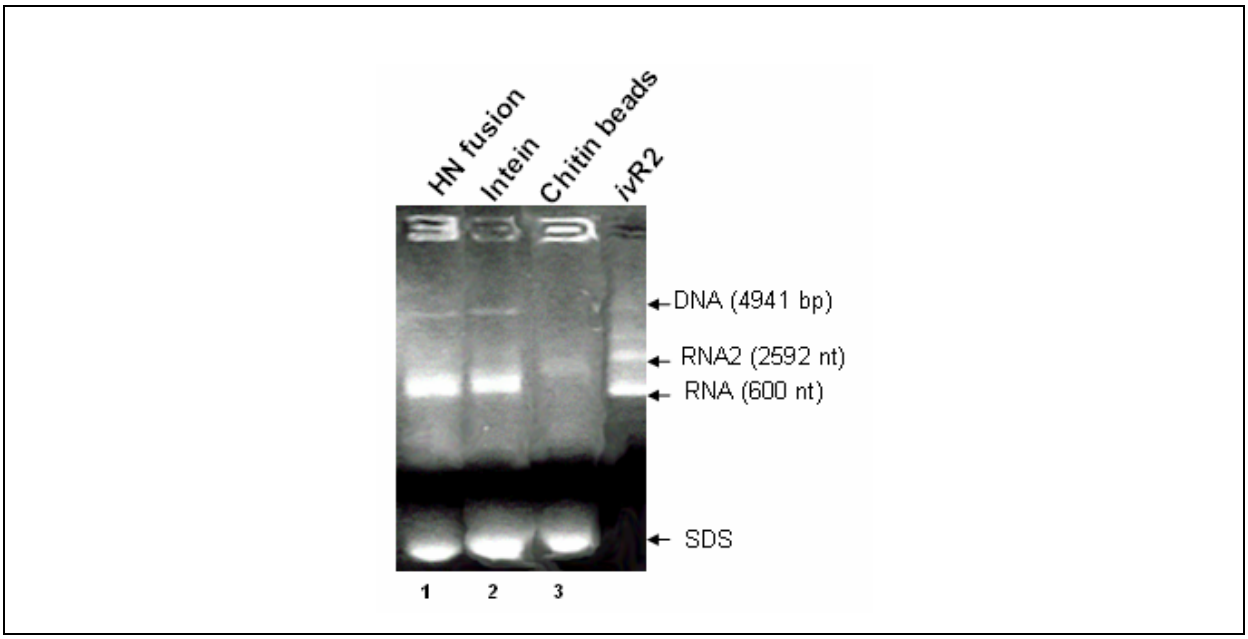


Figure 2.15 Initial binding assay. Agarose gel electrophoresis of precipitated RNA obtained from binding assays of HN-fusion protein (lane 1), intein peptide (lane 2) and chitin beads (lane 3) were incubated with 2592 and 600 nt RNA transcripts. Components of the RNA2 transcription reaction (*iv*R2) are indicated on the right hand side of the gel. SDS used to precipitate the nucleic acids is also indicated on the right hand side of the gel

Two additional bands were also present on the gel; a faint band masked by the 600 nt transcript (Figure 2.15, lane 3) and a lower band that ran further down the gel. This lower band was probably due to the presence of SDS, as it was absent in the

sample used to denote the components of the RNA transcription reaction (Figure 2.15, lane *iv*R2). SDS was not included in the RNA-loading buffer used for this sample. However, the identity of the bands not representing nucleic acids needed to be investigated.

Characterisation of bands other than nucleic acids

Agarose gel electrophoresis revealed that the band that migrated the furthest was the 1% SDS used to separate the HN-fusion protein from the chitin beads. This was revealed by running gel-loading buffer with and without the addition of 1% SDS (Figure 2.16, panel B, lanes 4 and 5). The more blurred band that migrated around the 600 nt transcript was a component of the RNA binding buffer (Figure 2.16, panel A, lanes 2-5). This was determined when the 2592 and 600 nt transcripts were incubated with and without the RNA binding buffer alone and analysed (Figure 2.16, panel A, compare lanes 1 with 2). The RNA binding buffer was not further analysed to determine which component was causing the fluorescence. These extra bands were not indicated on future gel electrophoresis of binding assays.

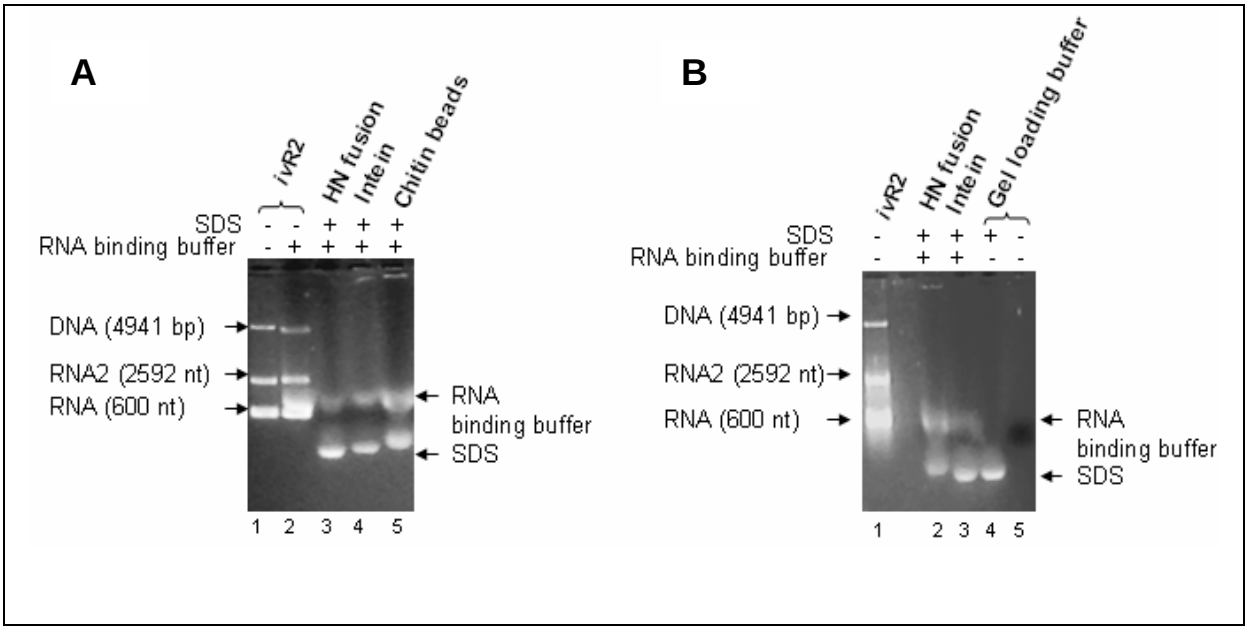


Figure 2.16 Characterisation of additional bands other than nucleic acids in binding assays. Agarose gel electrophoresis of RNA transcripts (*iv*R2) (panel A, lanes 1, 2, panel B, lane 1), HN-fusion protein (panel A, lane 3, panel B, lane 2), intein peptide (panel A, lane 4, panel B, lane 3), chitin beads (panel A, lane 5) and gel loading buffer (panel B, lanes 4 and 5) were mixed with either SDS or RNA binding buffer. For each sample indicated on top, either SDS or RNA binding buffer were added (+) or left out (-). The identified extra bands are indicated on the right hand side of the gel. Components of the RNA2 transcription reaction (*iv*R2) are indicated on the left hand side of the gel.

2.3.6 Optimisations of the binding assay

Effect of ionic strength

Since no interaction occurred between the HN-fusion protein and the 2592 nt transcript in the initial binding assay (Figure 2.15, lane 2), it was decided to optimise the binding assay starting with the ionic component of the binding buffer. The ionic strength of the binding buffer is one of the main factors that influences the binding between RNA and protein (Pierce Biotechnology Inc., 2004). The effect of varying ionic strength upon 2592 nt transcript binding by HN-fusion protein was therefore investigated. Increasing concentrations of NaCl (25, 50, 75, 100, 125, and 150 mM) in the RNA binding buffer were employed. This was an attempt to reduce the interactions observed between the intein peptide and nucleic acids (Figure 2.17).

Of the intein peptide and chitin beads used in the binding assays in order to eliminate false positive results, the intein peptide (Figure 2.17, lanes 6-9) was found to interact with nucleic acids. Its interaction was reduced at 150 mM NaCl (Figure 2.17, lane 10). The chitin beads (affinity matrix) were found not to interact with nucleic acids (Figure 2.17, lanes 11-15), and therefore omitted as a control in subsequent assays.

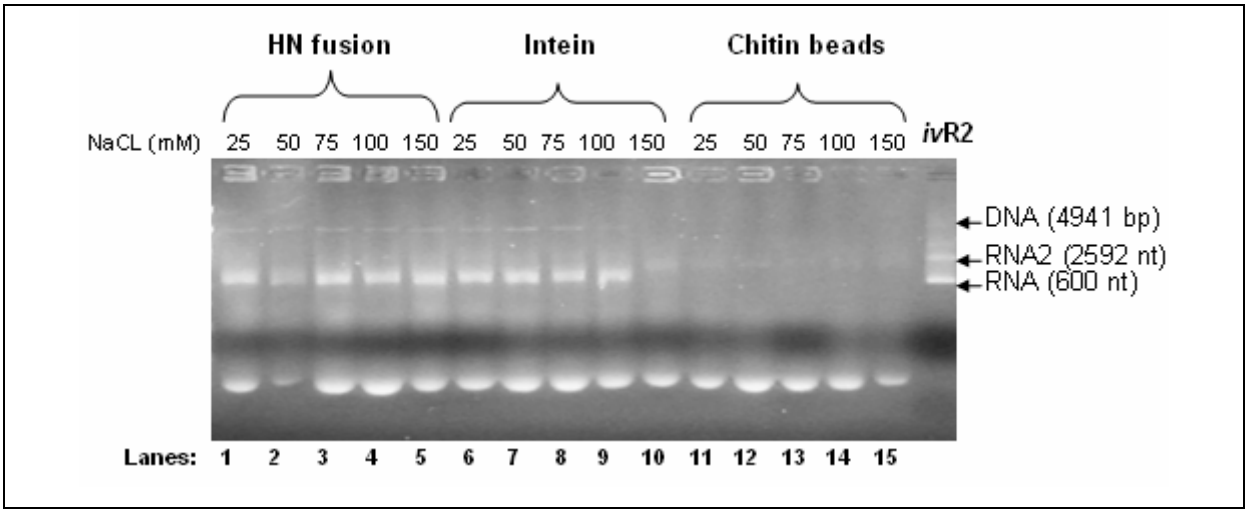


Figure 2.17 Effect of increased ionic strength on the binding assays. Agarose gel electrophoresis of precipitated RNA obtained from binding assays tested with increasing ionic strength (25-150 mM NaCl) of the binding buffer. Lanes 1-5, 6-10, and 11-12 represent the HN-fusion protein, intein peptide and the chitin beads that were incubated with 2592 and 600 nt RNA transcripts in different ionic strength buffers, respectively. Components of the RNA2 transcription reaction (*iv*R2) are indicated on the right hand side of the gel.

Although the intein peptide interaction with nucleic acid was reduced at 150 mM NaCl (Figure 2.17, lane 5), the interaction was retained for the HN-fusion protein. This suggested that the interaction of the intein peptide portion of the HN-fusion protein with nucleic acid could be reduced.

The binding assay had two major problems. Firstly, the poor yield of the 2592 nt transcript pulled down and secondly, the stickiness of the chitin beads (described in the next section). The binding assay which was carried out at 150 mM NaCl, showed a reduction in interaction between the intein peptide and nucleic acid (Figure 2.17, lane 10). It was therefore decided to use this concentration of NaCl in the RNA binding buffer for the next set of binding assays.

Effect of addition of a detergent

During the washing procedure of protein bound chitin beads, the chitin beads were found to clump together and stick to the sides of the eppendorf and pipette tips. These characteristics of the chitin beads were unsuitable for binding assays, as the clumping could have minimised the beads-protein exposure to RNA present in the mixing reaction. This problem would have caused inaccurate results, due to the loss of beads and therefore proteins. This can be observed in lane 2 (Figure 2.17), where less of the 600 nt transcript was present, compared with the yield of 600 nt in lanes 1 and 3-5 (Figure 2.17). It was thus decided to include a detergent to circumvent the problems associated with the chitin beads. Triton X-100 was added to the RNA binding buffer and it not only prevented the clumping and stickiness of the chitin beads, but also appeared to increase the amount of the 2592 nt transcript pulled down (Figure 2.18, compare lanes 3 and 4 with 1 and 2).

An explanation for the increased yield of particularly the 2592 nt transcript of the RNA2 transcripts (Figure 2.18, lanes 3 and 4), is that the detergent present might have prevented the loss of chitin beads thus increasing the exposure of the protein bound portion of the chitin beads to RNA transcripts. The effect of chitin bead loss is obvious from the observation of the differing amounts of 600 nt being pulled down in the binding assays where the detergent was not added (Figure 2.18, compare lanes 1 with 2).

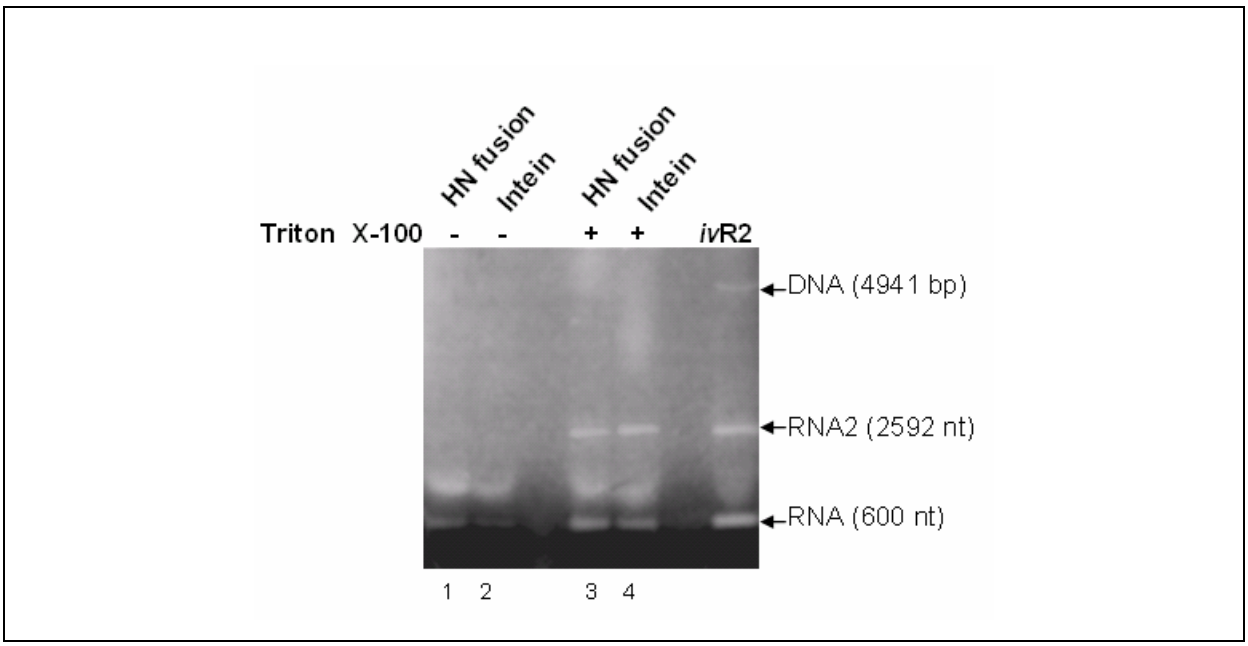


Figure 2.18 Effect of addition of Triton X-100 to reduce stickiness of the chitin beads and loss of the beads. Agarose gel electrophoresis of precipitated RNA obtained from binding assays experimented in the presence (lanes 3 and 4) and absence (1 and 2) of Triton X-100. Lanes 1 and 3, 2 and 4 represent HN-fusion protein and intein peptide that were incubated with 2592 and 600 nt RNA transcripts, respectively. The NaCl concentration of the RNA binding reaction is 150 mM. Components of the RNA2 transcription reaction (*iv*R2) are indicated on the right hand side of the gel.

Hereafter, Triton X-100 was included in the binding buffer for subsequent binding assays. Although, the RNA binding buffer contained 150 mM NaCl, the interaction between the intein peptide and nucleic acids was not reduced (Figure 2.17, lane 4), as it had been previously (Figure 2.17, lane 10). It was decided to include a non-specific competitor to reduce the interaction between the intein peptide and nucleic acid.

Addition of a non-specific competitor, ssDNA

In DNA-protein pull down assays, a non-specific competitor, such as polyA, is included in binding assays to minimise the non-specific binding of the DNA to the target protein (particularly the positively charged amino acid residues). This repetitive polymer provides an excess of non-specific sites to absorb proteins that will bind to any general negatively charged DNA sequence. To maximise the effectiveness of the competitor DNA it must be added to the reaction along with the protein prior to the addition of DNA (Pierce Biotechnology Inc., 2004). Single stranded DNA (ssDNA), prepared from salmon sperm DNA, was used as a non-

specific competitor, as it mimics single stranded RNA. Two protocols were used; firstly, ssDNA was incubated with the protein for 30 minutes, then after several washes and aspiration of the chitin beads, RNA was incubated with the pre-treated protein. In the second protocol, all components of the binding assay were added simultaneously.

Regardless of the protocols used, the addition of 2 μg of ssDNA inhibited the pull down of DNA template by the intein peptide (Figure 2.19, lanes 2 and 4). The DNA template was pulled down in the binding assay carried out without the non-specific competitor (Figure 2.19, lane 6). The addition of ssDNA significantly increased the amount of 2592 nt transcript pulled down for the HN-fusion protein (Figure 2.19, compare lanes 1 with 5). ssDNA did not reduce interaction between the intein peptide and RNA components of the RNA transcription reaction (Figure 2.19, compare lanes 2 and 4 with 6).

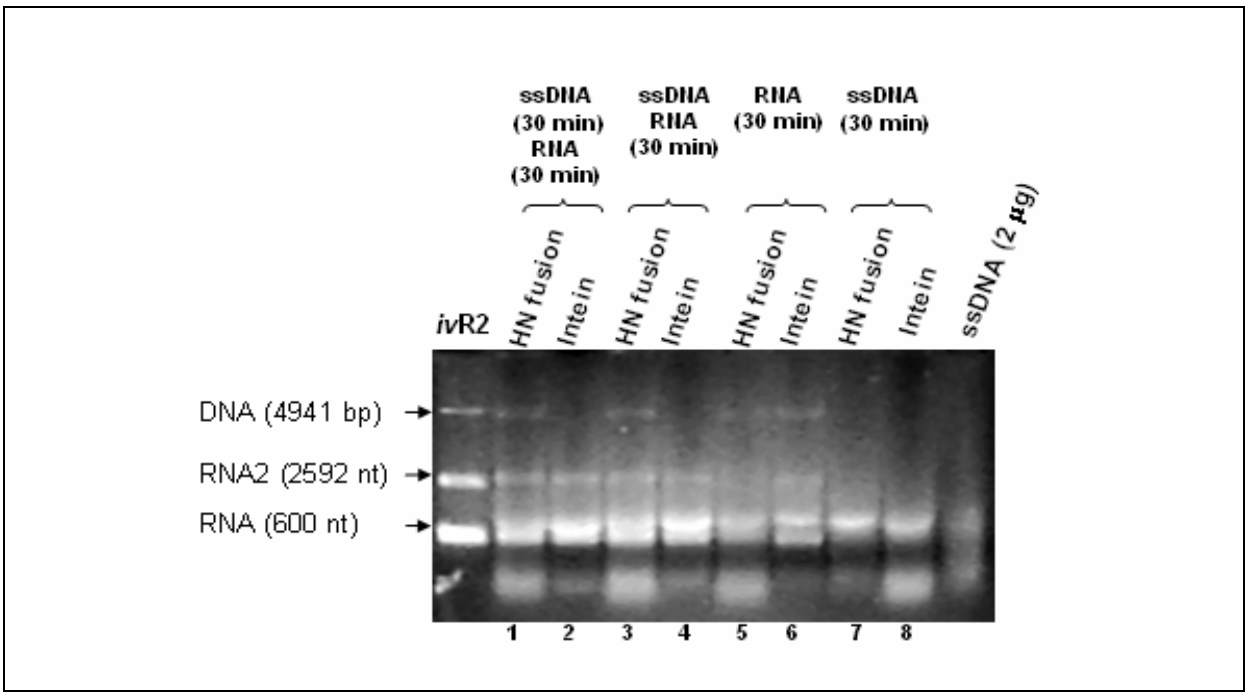


Figure 2.19 Effect of non-specific competitor on binding assays. Agarose gel electrophoresis of precipitated RNA obtained from binding assays performed in the presence (lanes 1-4) and absence (lanes 5 and 6) of 2 μg of ssDNA. HN-fusion and intein peptide bound to chitin beads were either incubated with ssDNA, washed, and then incubated with RNA transcripts (lanes 1 and 2, respectively), or incubated with both RNA and ssDNA (lanes 3 and 4, respectively) or incubated with RNA (lanes 5 and 6, respectively) or with ssDNA (lanes 7 and 8, respectively). The length of incubation with respective nucleic acid is indicated in parenthesis. Optimised RNA binding buffer contained 150 mM NaCl and 0.5% Triton X-100. Components of the RNA2 transcription reaction (*iv*R2) are indicated on the left hand side of the gel. The last lane consisted of 2 μg of ssDNA.

Effect of substitution of NaCl by $KC_2H_3O_2$

Due to the binding of the 2592 and 600 nt RNA2 transcripts to the intein peptide, it was decided to test whether the salt, $KC_2H_3O_2$, instead of NaCl, would reduce the interaction of the intein. According to the literature, different types of salts, such as KCl, $KC_2H_3O_2$, $NaC_2H_3O_2$, and K_2HPO_4, were used to enhance specificity in these assays (Pierce Biotechnology Inc., 2004).

The effect of $KC_2H_3O_2$ at a concentration of 100 mM on the binding assays was compared with 150 mM NaCl, using the optimised RNA binding buffer containing 0.5% Triton X-100 and 2 μg ssDNA. A change of salt from NaCl to $KC_2H_3O_2$ increased the amount nucleic acid being pulled down by the HN-fusion protein (Figure 2.20, compare lanes 1 with 3). This effect is more noticeable for the 600 nt transcripts within the HN-fusion protein binding assays (Figure 2.20, compare lanes 1 with 3). In addition, the interaction between the intein peptide and 2592 and 600 nt transcripts was reduced (Figure 2.20, compare lanes 4 with 2).

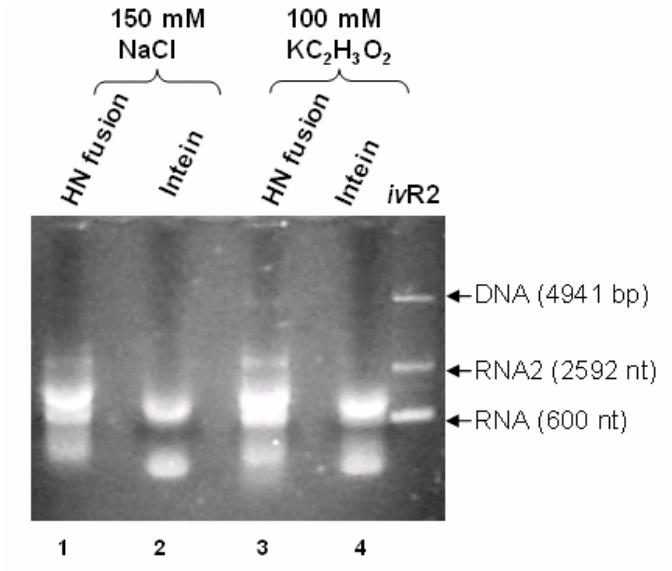


Figure 2.20 Effect of different ionic strength on the binding assays. Agarose gel electrophoresis of precipitated RNA obtained from binding assays tested with two different salts, 150 mM NaCl and 100 mM $KC_2H_3O_2$. HN-fusion protein and intein peptide bound to chitin beads were incubated with RNA in RNA binding buffer containing either 150 mM NaCl (lanes 1 and 2, respectively), or 100 mM $KC_2H_3O_2$ (lanes 3 and 4, respectively). Optimised RNA binding buffer contained 0.5% Triton X-100 and 2 μg of ssDNA. Components of the RNA2 transcription reaction (*iv*R2) are indicated on the right hand side of the gel.

In subsequent assays, $KC_2H_3O_2$ at a concentration of 100 mM was used instead of 150 mM NaCl. $KC_2H_3O_2$ was found to strengthen the interaction between the HN-fusion protein and RNA (particularly the 2592 nt transcript). A reason for this is not known.

2.3.7 Reproducibility of binding assay with new batch of protein

When it was attempted to reproduce the binding assay using a fresh batch of protein, the results were unexpectedly different, since the intein peptide that had previously not interacted with nucleic acid (Figure 2.20, lane 4), now did so (Figure 2.21 A, lane 2). The sole variable in these two experiments was the protein component. A constant batch of protein bound beads/buffer suspension was required. In other words, since a fresh batch of protein bound beads/buffer suspension was made for each set of binding assays (approximately 20 binding assays), the intein peptide portion of the HN-fusion protein had to be estimated so as to achieve the same amount of intein peptide as described in section 2.3.3.

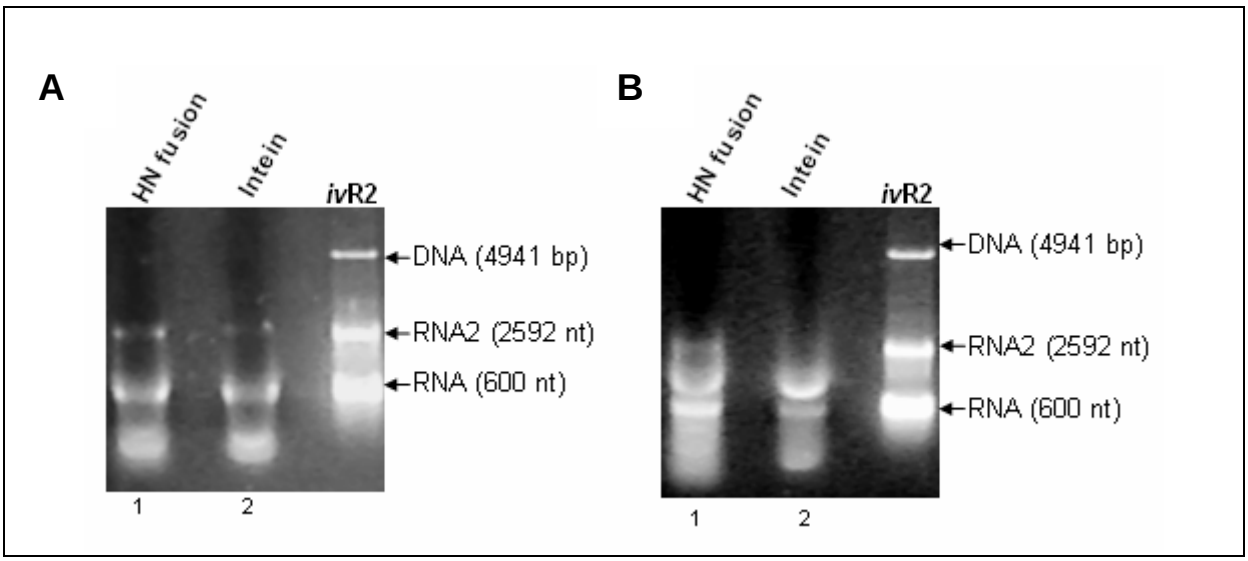


Figure 2.21 Reproducibility of binding assays with a fresh batch of beads/buffer suspension. Agarose gel electrophoresis of precipitated RNA obtained from binding assays tested with a new batch of protein bound to chitin beads (panel A), and its optimisation with increased of amount of ssDNA added (panel B). Lanes 1 and 2 represent the fresh batch of HN-fusion protein and intein peptide bound to chitin beads, respectively, that were incubated with RNA in the presence of 2 μg of ssDNA (panel A) and 4 μg ssDNA (panel B). Components of the RNA2 transcription reaction (*iv*R2) are indicated on the right hand side of the gels. Optimised RNA binding buffer includes 0.5% Triton X-100, 100mM $KC_2H_3O_2$.

It was decided that for each fresh batch of protein bound beads/buffer suspension a re-optimisation to reduce the interaction between the intein peptide and RNA would have to be made. To do this the amount of non-specific competitor ssDNA added to the binding assay was doubled from 2 μg to 4 μg (Figure 2.21, panel B). The increased amount of ssDNA reduced the interaction between the intein peptide and RNA2 (Figure 2.21, panel B, lane 2).

This method of optimisation, however, increased the background fluorescence within the lanes of the gel due to the larger amount of ssDNA fragments present. Having reduced the interaction between the intein peptide and nucleic acids, and determined that HN might be binding RNA2 (Figure 2.21, panel B, lane 1), experiments were attempted to determine whether the HN protein could bind RNA1.

2.3.8 Evaluation binding assay with HaSV RNA1 and non-viral RNA

As a control, HN-fusion protein was tested for its ability to bind non-viral RNAs, which were 2346 and 1065 nt in size. For this binding assay, the beads/buffer suspension used in the previous experiment (Figure 2.21, panel B) with the optimised RNA binding buffer containing 4 μg of ssDNA was used.

The reduction of the interaction between the intein peptide and RNA2 in the presence of 4 μg of ssDNA (Figure 2.22, panels A and B, lanes 2) was reproduced in a way similar to that previously observed (Figure 2.21, panel B, lane 2). The HN-fusion protein pulled down both RNA1 and non-viral RNA (Figure 2.22, panels A and B, lanes 1). However, the results were inconsistent and further experiments such as competition assays, where the affinity of the HN for certain RNA molecules, such as viral RNA over cellular RNAs, were not carried out.

A	B

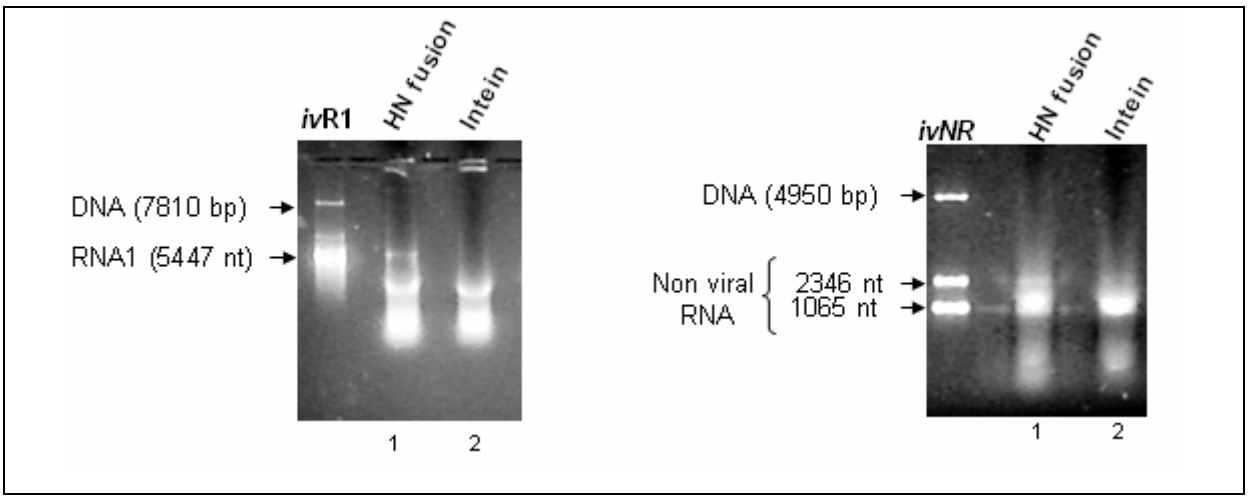


Figure 2.22 Evaluation of HaSV genomic RNA1 and non-viral RNA. Agarose gel electrophoresis of precipitated RNA obtained from binding assays of HN-fusion and intein peptide bound to chitin beads that were incubated with HaSV RNA1 (panel A, lanes 1 and 2, respectively) and non-viral RNA (panel B, lanes 1 and 2, respectively). Components of the RNA1 transcription reaction (*iv*R1) and non-viral RNA (*iv*NR) are indicated on the left hand side of the gels. Non-viral RNA was derived from the *in vitro T7* transcription of the pGEM® Express Positive Control Template (Promega). Optimised RNA binding buffer contains 0.5% Triton X-100, 100 mM $KC_2H_3O_2$, and 4 μg ssDNA.

2.4 Discussion

The objective of the set of experiments reported in this chapter was to develop an *in vitro* experimental system to investigate the interaction between the N terminal domain of HaSV p71 and viral genomic RNA.

Two aspects of the protein purification system, IMPACT^{TM}-CN, were adapted to develop the RNA-protein "pull down" assay. These were the rapid purification of fusion protein from cells expressing the protein and the utilisation of the chitin-binding domain of the affinity tag (intein peptide) for the "pull down" assays. It was proposed that if RNA interacted with the potential RNA-binding domain of the fusion protein, it would be pulled down by the mass of the chitin beads. To confirm this interaction, gel electrophoresis analysis of precipitated RNA on an agarose gel containing ethidium bromide was used.

Residues 1-121 of p71 were used as the potential RNA-binding domain of the fusion protein. This 121 residues peptide represented the N terminal domain of p71 (HN), which was outlined using a protein alignment between p71 and the structurally characterised NωV coat protein constructed by Hanzlik *et al*. (1995). This was possible since the N terminal domain of NωV coat protein, residues 48-121, shares 81% sequence similarity with the same region of p71 (Hanzlik *et al*., 1995) (Figure 2.6). Most importantly, residues 1-48 of this isolated peptide from p71 not only share low sequence similarity with NωV coat protein, but also consists of predominantly arginine residues. It is the high arginine content that has been proposed to be responsible for the potential RNA-binding property of this polypeptide (Helgstrand *et al*., 2004). In addition, this region was proposed to project towards the interior of the capsid shell, as observed in the crystal structure of the NωV capsid shell.

Referring to the set of SDS-PAGE analyses of protein samples obtained during the preparation of the protein component of the binding assay, the amount of the expressed 68-kDa HN-fusion protein in *E. coli* cells harbouring the construct, pAN5, was low compared with that of the expressed controls (Figure 2.9, panel A, compare lanes 3 with 8 and 9). This was attributed to the two consecutive rare *E. coli* arginine codons found within the coding sequence of HN. Consecutive rare arginine codons are known to reduce translation efficiency in *E. coli* (Rosenberg *et al*., 1993). Despite the two unsuccessful attempts to increase the amount of HN-fusion protein expressed (section 2.3.3), there was sufficient fusion protein presence for protein purification (discussed later).

As required for the “pull down” part of the binding assays, the HN-fusion protein was purified attached to the affinity matrix (chitin beads). Before proceeding with optimisation to increase the amount of purified HN-fusion protein, it was crucial to investigate whether the protein bound to the chitin beads was stable. Data showed that there was a considerable breakdown of 68-kDa HN-fusion protein beyond eight days of storage (Figure 2.12, lane 2 and 3) of the beads/buffer suspension prepared in order to eliminate variability in protein concentrations between each binding assay. This breakdown occurred between the HN and the intein peptides of the HN-fusion protein, as there was the presence of a 55-kDa band corresponding to the M_R of the intein peptide (Figure 2.12, compare lane 3 with 4-5). A possible reason for this breakdown could be attributed to the position of an unfavourable amino acid,

aspartate, positioned within the fourth position upstream of the C-terminus of HN protein adjacent to the intein peptide (Figure 2.2, pAN5). Aspartate at this position may have induced a substantial amount of self-cleavage within the HN-fusion protein during storage. Despite this instability, it was suggested that HN-fusion protein beads/buffer suspension should be used within 8 days as there was sufficient stable 69-kDa fusion protein present (Figure 2.12, compare lanes 1 with 2).

As expected, the amount of the purified HN-fusion protein obtained was low compared with the purified control (Figure 2.12, panels A and B, compare lanes 6). To compensate for this, double the volume of the clarified cell extract of cells expressing the HN-fusion protein was incubated with the chitin beads. Although the amount of purified HN-fusion protein was increased, the amount of the protein representing the stable 68-kDa HN-fusion protein was estimated to be four times less than that of the purified intein peptide (Figure 2.13, lanes 2 and 3, respectively). Therefore, for the binding assays, a ratio of 4:1 of HN-fusion protein and intein peptide amounts was used to ensure that equal amounts of protein were present.

Since the protein component of the binding assay included the intein peptide and the chitin beads, it was inevitably essential to test whether these two components interacted with nucleic acids. This proved to be true for the intein peptide, as it was found to pull down nucleic acids in the initial binding assay under standard binding conditions (Figure 2.15, lane 2). In addition, there was no pull down of the 2592 nt transcript of the two RNA2 transcripts (2592 and 600 nt) by the HN-fusion protein (Figure 2.15, lane 1). The data obtained from the initial binding assay implied that two optimisation objectives should be set. Firstly, the interaction between the nucleic acid and intein peptide needed to be reduced or eliminated and secondly, whether there was any interaction between the HN-fusion protein and the 2592 nt RNA transcript. The latter transcript represents full length HaSV RNA2.

The first attempt at optimisation involved several binding assays carried out at different concentrations of NaCl. Although results showed that at 150 mM NaCl the interaction between the intein peptide and nucleic acids was reduced (Figure 2.17, lane 10), the interaction between 2592 nt RNA transcript and HN-fusion protein, was non-existent (Figure 2.17, lanes 1-5). The yield of the latter was increased in a second attempt by the addition of a detergent, Triton X-100, to the RNA-binding

buffer (Figure 2.18, lanes 3 and 4), which was used to prevent chitin beads loss. This was to reduce the clumping together of the beads and their sticking to the sides of the eppendorfs, which affected the data obtained. Despite this optimisation, the interaction observed between the intein peptide and nucleic acid was not reduced at 150 mM NaCl as found previously (Figure 2.18, lane 4).

In a third attempt at optimisation, the addition of a non-specific competitor, ssDNA, to the RNA-binding buffer showed a small, but promising reduction of interaction, particularly between the intein peptide and DNA template of the RNA2 transcription reaction (Figure 2.19, lanes 2). However, a repetition of the same experiment significantly reduced intein peptide interaction with all nucleic acid components of the *in vitro* RNA2 transcription reaction (Figure 2.20, lane 2). This reduction of interaction was reproduced in the fourth attempt at optimisation, when the binding assay was carried out with $KC_2H_3O_2$ instead of NaCl as the RNA-binding buffer ionic component (Figure 2.20, lane 4). This substitution effectively increased the yield of both 600 and 2592 nt RNA2 transcripts being pulled down by the HN-fusion protein (Figure 2.20, lanes 3). Potassium acetate might have strengthened the interaction between the protein and nucleic acids. A reason for this is not known.

During the course of the binding assays, there were several irreproducible results obtained. For example, the interaction between the intein peptide and the nucleic acid carried out in lane 2 and 4 (Figure 2.19) differed with that in lane 2 (Figure 2.20), which were identical experiments. It was suggested that the only variable between these binding assays was the protein component. It was found that optimisations used to reduce the interaction between the intein peptide and nucleic acids were suited to only certain batches of HN-fusion protein bead/buffer suspensions (compare lane 4 (Figure 2.20) with lane 2 (Figure 2.21, panel A)). This emphasized the need for a stable HN-fusion protein for the binding assays as the instability associated with the current HN-fusion protein made protein amount estimations, difficult.

To circumvent this, a re-optimisation using 4 μg ssDNA in the RNA-binding buffer significantly reduced the non-specific interaction between the intein peptide and nucleic acid (Figure 2.21, panel B, lane 2). With this current optimisation, it was decided to evaluate the binding of HaSV RNA1 and non-viral RNAs to the HN-fusion protein. The latter RNAs were found to interact with the HN-fusion protein (Figure

2.22, panel A and B, lanes 1). However, the objective of this experimental attempt was to develop competition assays to investigate the specificity of HN protein by incubating the fusion protein with mixtures of viral RNAs (RNA1 and RNA2) as well as non-viral RNAs. Instead, with problems associated with the binding assays it was proving to be difficult to conclude whether the synthesized RNA interacted with the HN peptide or the intein peptide of the HN-fusion protein.

The major problems associated with the current RNA-protein pull down assay were, firstly, the inconsistent results (Figure 2.21, panel A and B) and secondly the interaction between the intein peptide and nucleic acids (Figure: 2.16). The inconsistent results may be related to the instability of the current HN-fusion protein due to *in vivo* cleavage occurring as described in Section 2.3.3. Protein estimation using the Commassie blue staining method was proven to be unreliable, since there was instability of the HN-fusion protein. Estimation of the intein peptide with that of the intein portion of the HN-fusion protein was necessary for optimisations in the reduction of interaction between the intein peptide and nucleic acids, more specifically RNA. A more reliable method of estimation is therefore required to ensure that the amounts of protein used do not vary between experiments.

Inconsistent results were not only due to instability. Whether RNA2 binds or not, depends on the state of the HN-fusion protein expressed. Problems could have arisen from the folding characteristics of the HN-fusion protein. In theory, the HN portion of the HN-fusion protein was expected to expose the HN portion to the RNA in the binding assay. However, it was not known how the HN-fusion protein folds. The 13-kDa HN protein has a highly basic section near the N terminus and it might have interacted with any protein region within the intein. An alternative affinity tag in the construction of a fusion protein is discussed later.

Other problems encountered were, the difficulty in visualising any bound RNA due to the presence of extra bands other than nucleic acids and the presence of a non-competitor, ssDNA. Alternative protein denaturants such as 6 M guanidium, 5-8 M urea and 0.3 M NaOH as suggested in the IMPACT^{TM}-CN system, should be tested to disrupt the protein-RNA complex as opposed to SDS. In order to eliminate the excessive fluorescence caused by ssDNA, bovine albumin serum (BSA) could be tested as an alternative non-competitor. BSA might also prevent the adsorption of

nucleic acids to the intein. BSA has been included in many protein-protein and in some DNA-protein pull down assays as a non-specific adsorbant.

The current RNA-protein pull down assay was not effective in answering the question whether the HN specifically binds viral RNAs. For a RNA-protein pull down assay the 6 × His-tag protein purification system (Promega) could be an attractive alternative option. Although the six histidine residues comprising the tag are positively charged, this protein purification system is used as a suggestion for an alternative system. It could be proposed that these 6 × His residues would be electrostatically attracted to the affinity matrix and not exposed to the binding solution. Since, the tag is only a six amino acid residue polypeptide, the resulting His-HN-fusion would be approximately 14-kDa and be less likely to exhibit abnormal folding properties, compared with the 68-kDa HN-fusion protein. This alternative could also produce a more stable fusion complex with no *in vivo* cleavage mishaps. However, it is recommended that before construction of such fusion proteins for the binding assays, it should be investigated whether the tag itself, or the affinity matrix, interacts with the nucleic acid of interest. In addition to the above, future RNA-protein pull down assays should involve investigating increasing concentrations of Mg^{2+}; as this is a common co-factor in the folding of RNA. This increase in the concentration of Mg^{2+} might optimise the folding of RNA. The co-factor might assist the formation of specific secondary and/or tertiary structures that in turn might confer specific encapsidation signals (JE Johnson *et al.*, 2004).

In the literature, various methods of investigating RNA-protein interaction are described (Pierce Biotechnology Inc., 2004). Electrophoretic mobility shift assay (EMSA), and NorthWestern blot experiments are two commonly used approaches to analyse RNA-protein interactions in RNA viruses to identify potential RNA binding domains (Xu *et al.,* 2002; Brantley and Hunt, 1993; Choi and Rao, 2000). The technique of EMSA is very similar to RNA-protein pull down assays. However, in the application of EMSA, the small size of the 13-kDa HN protein would probably cause no detectable shift if bound to the high molecular weight of HaSV RNAs. An alternative approach would be probing for the shift in protein rather than the RNA. Unbound protein would obviously run further on an electrophoretic gel. NorthWestern blots experiments might be the ideal approach since the 13-kDa HN

protein of interest is resolved by an SDS-PAGE gel and transferred onto a nitrocellulose membrane, which is probed with RNA for interaction.

This *in vitro* RNA-protein interaction assay was a first attempt in the investigation of the interaction between the capsid protein and viral RNAs for tetraviruses. However, an improved assay must be developed first to ensure that the HN protein does interact with viral RNA. With an improved RNA-protein pull down assay, mutations can be made within the protein to map the region that interacts with RNA.

Chapter 3: Expression of p71 mutants *in vivo*

3.1 Introduction

In the previous chapter, it was reported that the heterologous expression of the gene encoding HaSV capsid precursor protein, *P71*, in *S. cerevisiae* was an unsuitable system for *in vivo* investigations to study the interaction between the capsid protein and viral RNA. This was because 95% of the translated p71 formed insoluble aggregates and very low quantities of VLPs were assembled when *P71* was expressed using the yeast *GAL1* promoter (Venter, 2001). An estimated 0.1% of the capsid protein could be purified as soluble VLPs. These VLPs did not mature *in vivo* according to data obtained on their buoyancy density.

Until recently, the expression of *P71*, using the yeast hybrid promoter *ADH2-GAPDH* (alcohol dehydrogenase II-glyceraldehyde-3-phosphate dehydrogenase), increased the levels of soluble p71 and assembled VLPs (Tomasicchio, unpublished results). These VLPs matured *in vivo* as observed by the detection of p64 by Western blot. This detection was confirmed using electron microscopy of purified VLPs that were shown to be mature virus particles (Tomasicchio, unpublished results). Given these results, the P_{AG} expression system proved to be suitable for *in vivo* studies on the assembly of HaSV VLPs.

This 5'-*ADH2-GAPDH*-3' hybrid promoter (P_{AG}) consists of the 5'*ADH2* regulatory sequence fused to the 3' end of the *GAPDH* promoter, which contains the TATA box and the mRNA cap site (Cousens *et al*., 1987). The induction strategy requires depletion of glucose from the culture media (as in the late fermentation stages) for derepression of the hybrid promoter and active transcription of the heterologous gene. This promoter is advantageous due to its strong constitutive promoter, intrinsic strength, and elements incorporated for its regulation to control expression of the heterologous gene (Marino, 1991).

The objective of this chapter was to investigate whether deletions that removed the positively charged residues from the N terminal domain of p71 would affect its assembly into ordered like structures representing VLPs *in vivo*. It was suggested that such deletions within p71 should be done so as to draw comparisons with those performed on the nodavirus, FHV.

3.2 Materials and Methods

3.2.1 Construction and expression of p71 mutants

General recombinant DNA techniques were carried out as described in Appendix 1. Two plasmids, pMT8 and pAV17 (Figure 3.1) were used for the construction of yeast recombinant vectors containing deletions within *P71*.

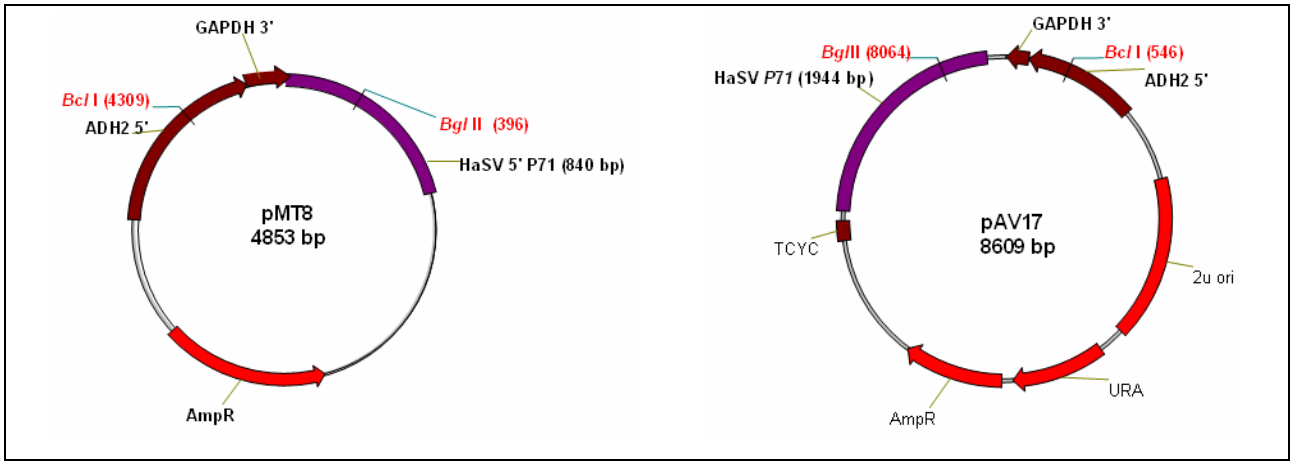


Figure 3.1 Schematic representation of pMT8 and pAV17, which were used for the construction of yeast recombinant vectors containing deletions within *P71*.

The plasmid pMT8, which contains 840 bp of the 5' HaSV *P71* fused to the 3' end of P_{AG} with an optimised yeast Kozak sequence, was used as a template to introduce deletions within the N terminal domain coding sequence of p71. pAV17 is a yeast shuttle vector that contains the full length of *P71* (1944 bp) under the control of the P_{AG} (constructed by Venter, 2001). To construct the yeast shuttle vector, the restriction digests *Bcl* I and *Bgl* II were used to subclone 940 bp from pMT8 into pAV17. Plasmid pMT9 that was constructed by Tomasicchio, M (Rhodes University, Appendix 2) using this cloning method, was used as a positive control for the expression studies. Studies on *P71* under the control of the P_{AG} expression system were based on this construct.

To generate deletions within the coding sequence of the N terminal domain of p71, two pairs of primers namely, AN10 & AN11 and AN12 & AN13 (Table 3.1) were designed to delete the coding sequence for residues 10-18 and 32-38 from *P71,* respectively.

Table 3.1 Sequences of primers used for construction of HaSV *P71* mutants.

Primers	Anneals[1]	Sequence[2]
AN10	4-24	5' – cag CTG TGA CGC CAC TCC AGC ATC -3'
AN11	52-69	5' – AAC GTT CGG GTC AGC GCC– 3'
AN12	72-93	5 '- ATT GAC GGT GAC GGT GTT GGC – 3'
AN13	114-135	5' – GGT ACC GGA AGG CAA GTT TCT CCC - 3'

[1] Corresponds to nucleotides of *P71*
[2] Lower case letters refer to the heterologous sequence added to primer AN10 that introduces a codon for residue leucine

For clarity, deletion of residues 10-18 and 32-38 will be referred to as ΔDI and ΔDII, respectively. AN11 was designed to add three nucleotides at position 25-27 of *P71*. This introduced a *Pvu* II recognition sequence for the identification of a DI deletion. This heterologous addition to *P71* coding sequence, encodes for a leucine residue at position 9 of p71. For the confirmation of a DII deletion, a recognition sequence for *Kpn* I was formed by the deletion of the 93-114 coding sequence. Figure 3.2 represents the cloning outcome of the deletions of DI and DII using the inverse PCR technique at the coding sequence level.

The Expand High Fidelity PCR system (Roche) was used to generate the deletions within the N terminal domain coding sequence of *P71.* The 5' ends of the primers were phosphorylated with *T4* polynucleotide Kinase, prior to inverse PCR. The thermal cycling programmes for the DI and DII deletions are programmes 6 and 7, respectively (Appendix 4). The PCR product was analysed by gel electrophoresis prior to treatment with *Dpn* I. *Dpn* I was used to eliminate the methylated parental template after the PCR. The PCR product was further treated with Klenow DNA polymerase I, to remove the 3' A overhangs that were generated by the *Taq* polymerase of the Expand High Fidelity PCR system. The treated PCR product was then purified and concentrated using the DNA Clean and Concentrator Kit (Zymo Research), and religated using *T4* DNA ligase before transformation into *E. coli* DH5α.

Pvu II and *Kpn* I restriction digests were used to select the constructs that contained either of the deleted sequences as described in Figure 3.2. The selected constructs were sequenced with sequencing primer MT4 (Appendix 3), which annealed to the region 100 bp upstream from the 3' end of P_{AG}. This primer would allow the sequencing of 25 bp upstream of *P71* 5' end and cover approximately 500 bp downstream. The constructs containing ΔDI and ΔDII within the p71 coding sequence were named pAN9 and pAN10, respectively (Figure 3.3). The process of inverse PCR deletion was repeated again using pAN10 as the template with primers AN10 and AN11. This created a p71 mutant with the coding sequence of both DI and DII deleted (ΔDI&DII). This construct was named pAN11 (Figure 3.3).

A schematic diagram describing the cloning methodology for the construction of yeast shuttle vectors expressing p71 mutants, is represented in Figure 3.4. Plasmids pAN9, pAN12, pAN11 and the pAV17, were transformed into *E. coli* JM 110 competent cells in order to generate an unmethylated *BcI* I restriction site, freeing the promoter region. *BcI* I and *BgI* II digests were performed and the DNA separated by gel electrophoresis. The 372, 375, 354 and 7418 bp fragments of digested pAN9, pAN10, pAN11 and pAV17, respectively, were gel purified. The 7418 bp gel purified fragment of pAV17 was pre-treated with Shrimp alkaline phosphatase (Roche) to prevent self-ligation. *T4* DNA ligase was used for ligation of 7418 bp fragment with one of each of the above fragments containing deletions within the coding sequence of p71.

The ligations were transformed into *E. coli* DH5α. Since *Bgl* II and *Bcl* I form compatible ends, the orientation of the insert was confirmed using *Eco* RI and *Eco* RV restriction digests. The resulting recombinant plasmids that had the complete *P71* gene containing the three types of deletions, DI, DII and DI & DII, under the control of P_{AG} were named pAN12, pAN13 and pAN14, respectively.

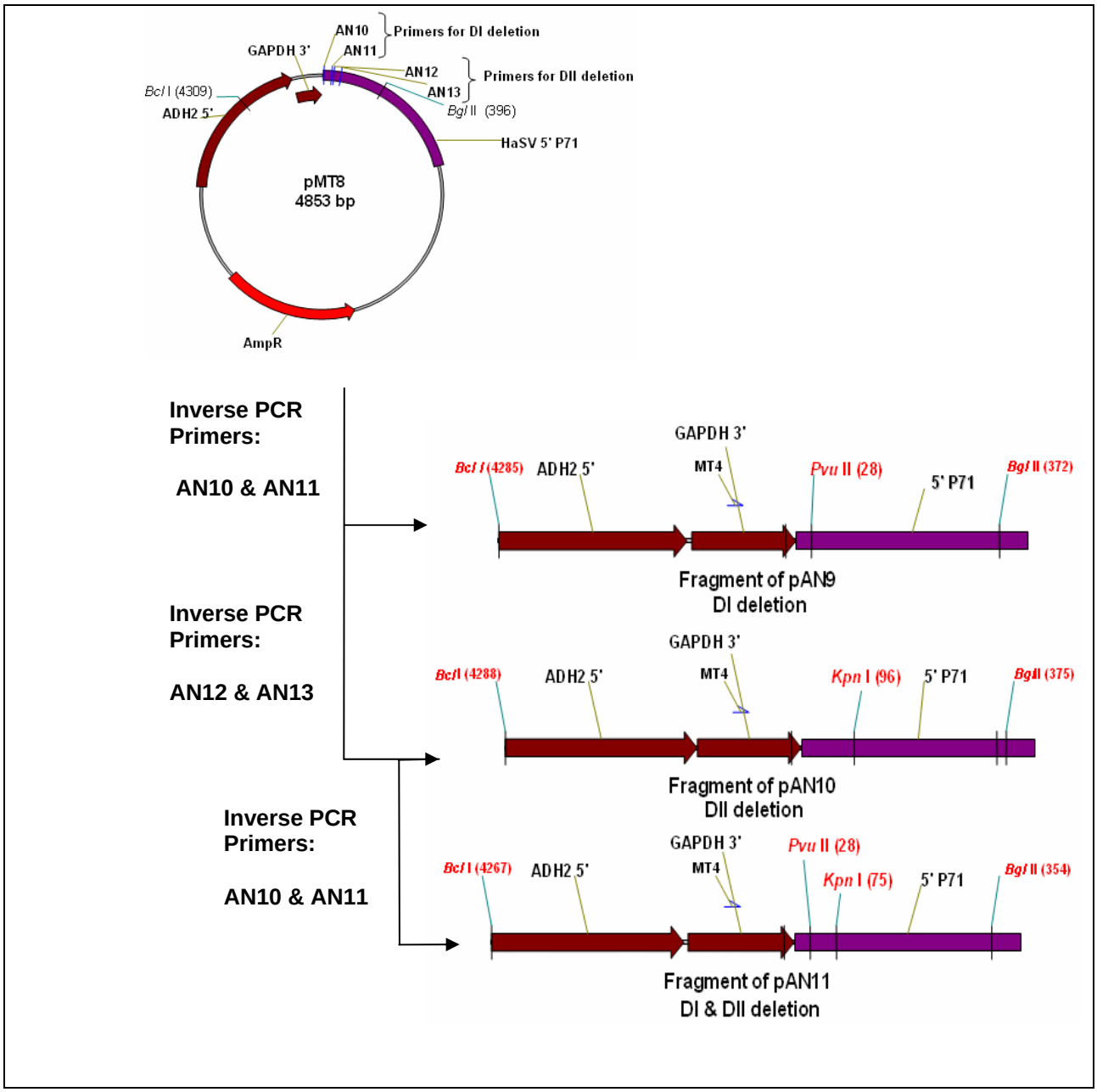


Figure 3.3 Schematic representation of the cloning methodology for the construction of deletions within *P71*. Restriction digests *Pvu* II and *Kpn* I indicated in red bold refer to confirmation of deletion of coding sequence for DI and DII, respectively. *Bcl* I and *Bgl* II restriction enzymes were used to subclone these fragments into yeast shuttle vector, pAV17.

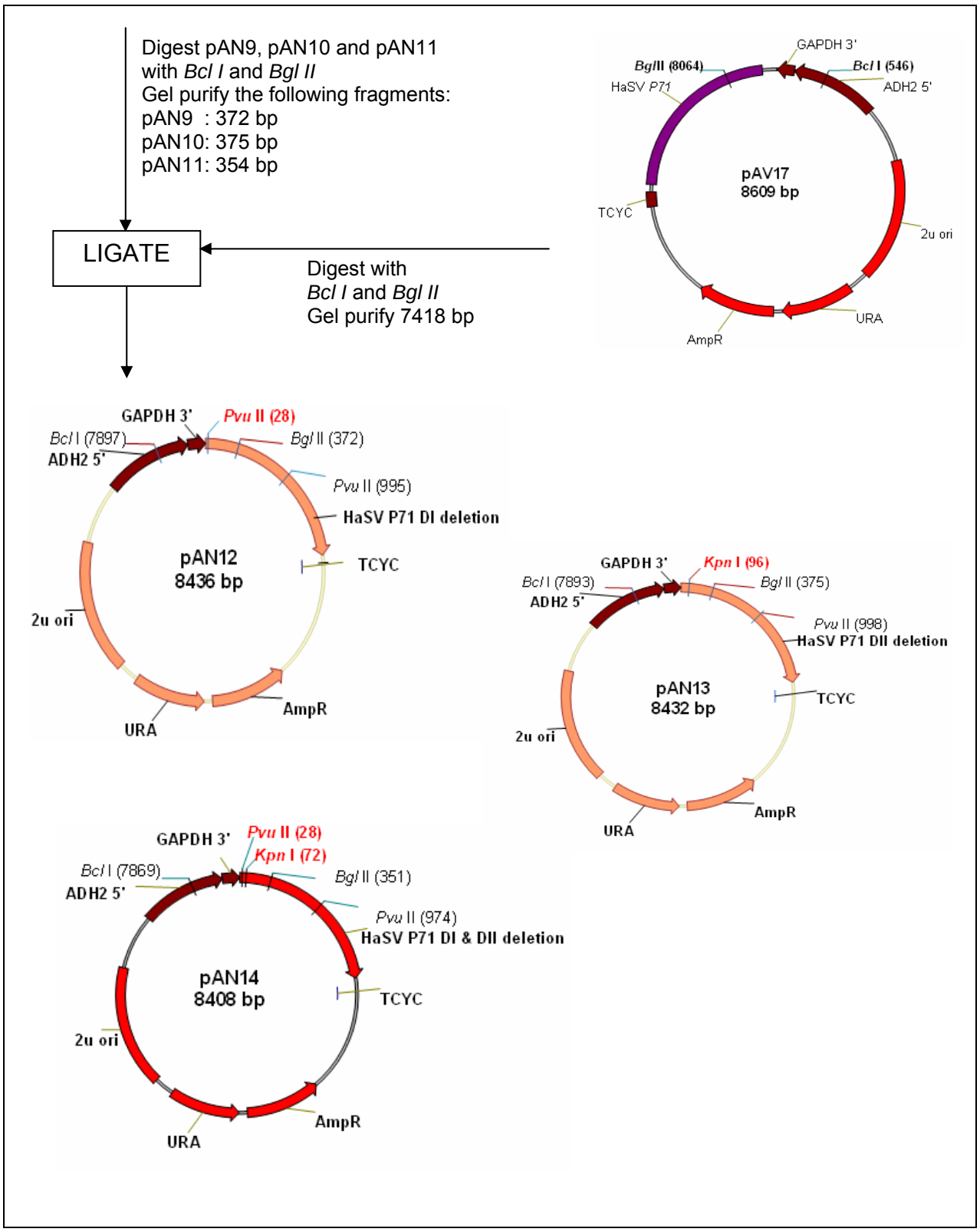


Figure 3.4 Schematic representation of the cloning methodology for the construction of yeast shuttle vectors used for the expression of p71 mutants.

3.2.2 Expression of *P71* in *S. cerevisiae*

The yeast strain *S. cerevisiae* INVSc1 (*Matα, trp1, his3, leu2, ura3*) (Invitrogen) was used for the heterologous expression of HaSV capsid precursor proteins. Competent yeast cells were prepared and transformed using Frozen-EZ Yeast Transformation II Kit (Zymo Research). Prior to transformation, the plasmids were checked with *Kpn* I and *Pvu* II digests.

Yeast cells were grown in supplemented minimal media (SMM) containing amino acids to maintain selection. Glucose was added as a carbon source (Kaiser *et al*., 1994; Appendix 8). Transformed yeast colonies were patched onto selective SMM agar plates and grown for two days at 30^{o}C. SMM was inoculated at an OD_{600nm} = 0.1 and the cultures grown for 32 hours at 30^{o}C.

At 16, 24 and 32 hours, $2OD_{600nm}$ units of cells were harvested, rinsed with $_{ddd}H_2O$, and stored at -80^{o}C until use. For crude yeast protein analyses, the pellets were resuspended in 30 μL EB buffer (50 mM Tris-HCL pH 7.4; 250 mM NaCL; 1 mM DTT; 1 mM PMSF; and 1 μM Pepstatin A). 0.1 g of acid-washed glass beads (Sigma) was added to each and the mixture vortexed for 30 sec periods, with 30 sec intervals on ice for a total homogenisation time of 10 min. To the vortexed mixture, 2.3 v/v (69 μL) of EB was added. For SDS-PAGE analyses of crude yeast protein extracts, 5 μL of 3 $\times$ SDS-PAGE sample buffer (187.5 mM Tris-HCl at pH 6.8); 6% (w/v) SDS; 30% glycerol; 0.03% (w/v) bromophenol blue; 1% β-mercaptoethanol) was added to 15 μL of protein sample and boiled for 5 min. In order to load equal amounts of crude yeast protein extracts, protein concentrations of each sample were estimated by Commassie stained SDS-PAGE analysis. After obtaining equal protein estimations, SDS-PAGE gels were run in duplicate and used for Western blot analysis and stained with Commassie blue. SDS-PAGE and Western blot were carried out as described in section 2.2.6. The presence of 71-kDa HaSV capsid precursor (p71) and the 64-kDa mature protein (p64) were detected using a rabbit polyclonal antibody raised against wild-type HaSV.

3.3 Results

3.3.1 Alignment of the N terminal domains of HaSV, NωV, FHV and PaV coat proteins

Alignment of the amino-termini of the capsid precursor protein of two omegatetraviruses, NωV, HaSV and two nodaviruses, FHV and PaV, revealed conservation in organisation of the positively charged residues. These residues, predominantly arginine, were organised into two clusters separated by a stretch of small, aliphatic amino acids (Table 3.2). The first and second basic clusters within the N terminal domain of p71 are labeled domain I (DI) for residues 10-18 and domain II (DII) for residues 32-38, respectively.

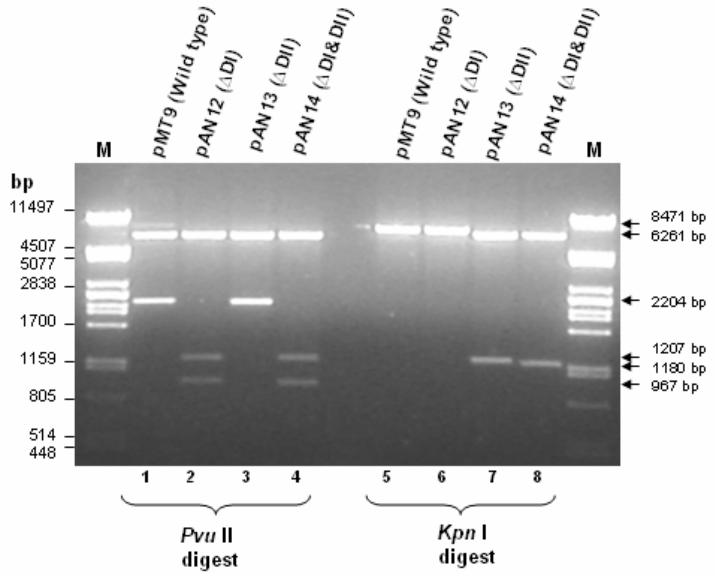
Whereas the large deletion of 2-31 residues investigated in FHV coat protein abolished the first basic cluster (Marshall and Schneemann, 2001), the deletions constructed within p71 involved one or both of the two clusters. These deletions allowed the rest of the N terminal domain of p71, including the stretch of aliphatic amino acids (residues 19-29) that separate the two basic clusters to remain intact.

The three types of deletions investigated such as the deletion of DI (residues 10-18), of DII (residues 32-38) and the double domain deletions of DI and DII (residues 10-18 and 32-38) are referred to as ΔDI, ΔDII and ΔDI&DII, respectively.

3.3.2 Expression of p71 mutants

To confirm whether detection or no detection of expressed p71 mutants was as a result of the deletions constructed within *P71*, the following experimental procedure was used. As described in section 3.2.2, the yeast recombinant vectors expressing p71 mutants, indicated in parenthesis, pAN12 (ΔDI), pAN13 (ΔDII), pAN14 (ΔDI& DII) and the wild type p71 (expressed from pMT9) were placed in triplicate tubes. One was used for transformation into yeast competent cells and the other two were digested with *Pvu* II and *Kpn* I. The digests confirmed the correct order of plasmids transformed.

The yeast vector pMT9 served as the wild type for the expression studies. The expression profile from pMT9, was used for comparison with the expression of p71 mutants. The restriction digest *Pvu* II, which represented ΔDI within the coding sequence of p71, cleaved pAN12 (Figure 3.5, lane 2) and pAN14 (Figure 3.5, lane 4) three times compared with the double cleavage of the pMT9 (Figure 3.5, lane 1). pAN13 was cleaved twice, since it does not have a DI deletion (Figure 3.5, lane 3). *Kpn* I, which represented ΔDII, cleaved plasmids pAN13 (Figure 3.5, lane 7) and pAN14 (Figure 3.5, lane 8) twice, compared with the single cleavage of pMT9 (Figure 3.5, lane 5) and pAN12 (Figure 3.5, lane 6).



Digest	*Pvu* II	*Kpn* I
Plasmid		
pMT9	6261 2204	8471
pAN12	6261 1207 967	8441
pAN13	6261 2177	8374 1180
pAN14	6261 1207 946	7264 1156

Figure 3.5 Confirmation of DI and DII deletions within the coding sequence of p71. Agarose gel electrophoresis of yeast recombinant vectors pMT9, pAN12, pAN13 and pAN14 that were digested with *Pvu* II restriction digest (lane 1-4) and *Kpn* I restriction digest (lane 5-8). The sizes of the expected digest fragments are represented in the table on the right. A *Pst* I digest of lambda DNA was used as a molecular weight marker (M), with the sizes of DNA fragments indicated on the left hand side of the gel.

Plasmids pAN12 (ΔDI), pAN13 (ΔDII), and pAN14 (ΔDI&DII) were transformed in competent yeast cells together with wild type construct (pMT9). Cells were inoculated into selective medium and incubated for 32 hours. To monitor expression levels, samples were collected after 16, 24 and 32 hours and analysed for the presence of p71

by Western blot analysis. To determine that differences in expression levels from the various constructs used were not due to the differences in the amount of protein loaded, the concentration of proteins within yeast crude extract was estimated by Commassie blue staining of a duplicated SDS-PAGE gel (Figure 3.6, panel A). It was estimated that the density of protein bands within the SDS-PAGE was equal for each protein extract of cells harvested at 16, 24 and 32 hours (Figure 3.6, panel A, compare lanes 1 with 3-4, 5 with 6-8 and 9 with 10-12).

Referring to the Western blots (Figure 3.6, panel B), mutant p71 containing ΔDI was expressed at similar levels to those of the wild type after 16 and 24 hours (Figure 3.6, panel B, compare lanes 1 and 7 with 2 and 8, respectively), but after 32 hours the amount of protein was significantly reduced (compare lanes 13 with 12, respectively). In contrast to the wild type, there was no sign of any protein migrating at 64-kDa in cells expressing p71 ΔDI (Figure 3.6, panel B, compare lanes 13 with 12, respectively). The position of p64 is represented in the lane containing 1 ng of wild-type HaSV virus particles (Figure 3.6, panel B, lanes 5, 6, 11). No expression of p71 ΔDII could be detected (Figure 3.6, panel B, lanes 3, 9 and 14).

The expression of p71 containing the double domain deletion, ΔDI&DII, had the same expression characteristics as the p71 ΔDI at 24 and 32 hours (Figure 3.6, panel B, compare lanes 10 and 15 with 8 and 13). There was no detection of expressed protein at 16 hours. At 32 hours, levels of p71 ΔDI&DII were significantly reduced compared with that of the wild type and no band representing p64 was observed (Figure 3.6, panel B, compare lanes 15 with 12).

Although there was no detection of p64 for p71 ΔDI and ΔDI&DII at 32 hours, an experiment was attempted to purify VLPs. To do this, cells from 3L cultures expressing p71 mutants as well as the wild type, under the same growth conditions as described in section 3.2.2, were harvested. Purified yeast extract was obtained as described in Tomasicchio (unpublished results), and VLPs were purified through a 10-40% sucrose gradient according to Taylor *et al*. (2002). Despite Western blot detection of p71 and p64, which confirmed the presence of immature and mature HaSV VLPs in cells

expressing the wild type, no detection of purified VLPs assembled from p71 mutants was obtained (data not shown).

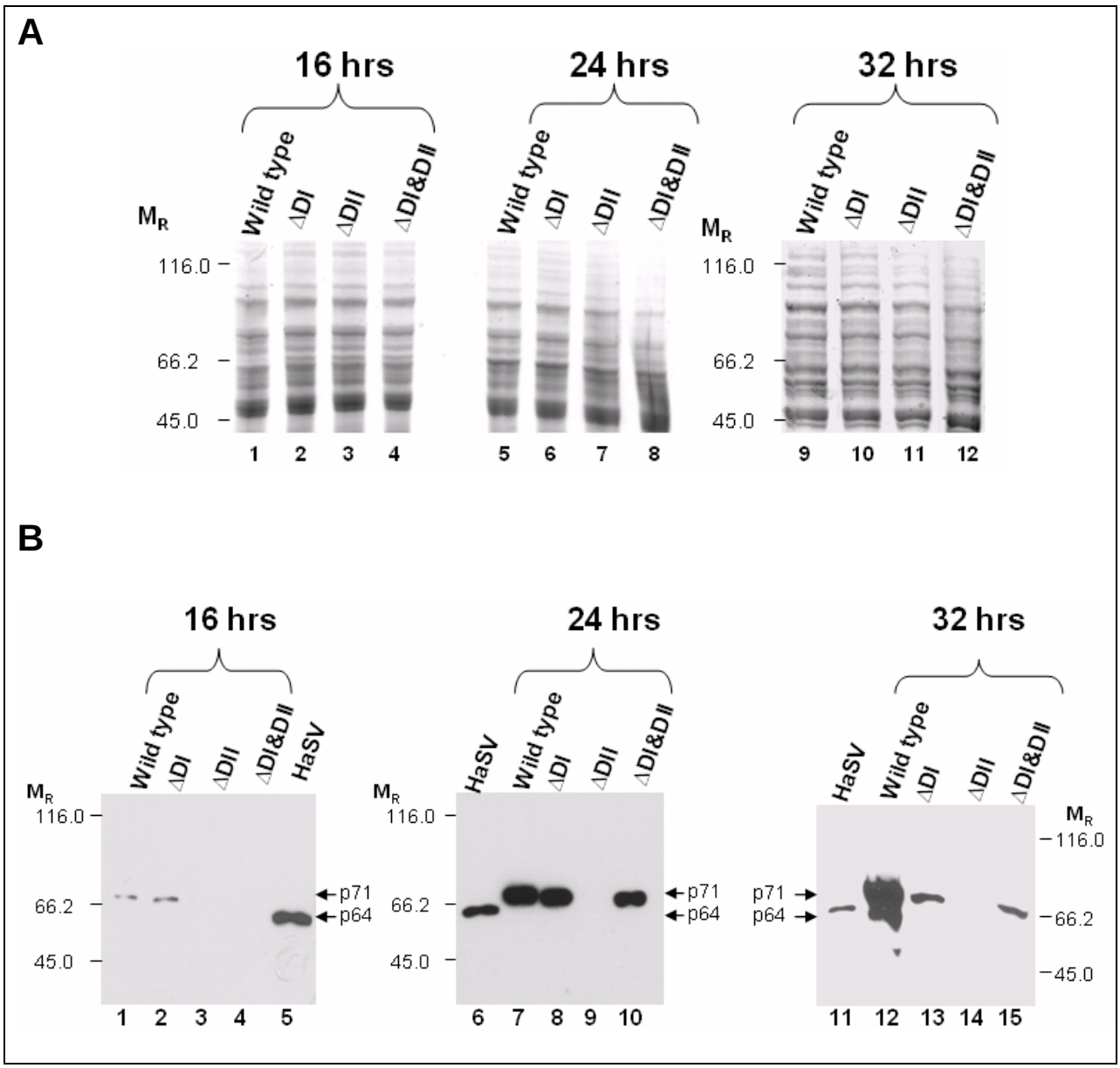


Figure 3.6 Expression of mutant p71 constructs. Panel A, Commassie stained SDS-PAGE of estimated protein concentrations obtained from yeast cells harbouring pMT9 (wild type), pAN12 (ΔDI), pAN13 (ΔDII) and pAN14 (ΔDI&DII) harvested at 16, 24 and 32 hours. Panel B, Western blot detection of p71 and p64 with polyclonal antibodies raised against the wild type HaSV. Lanes 5, 6 and 7 represent 1 ng of HaSV native virus (p64). Lanes 1-4, 7-10 and 12-15 represent yeast crude extracts from cells harbouring pMT9 (wild type), pAN12 (ΔDI), pAN13 (ΔDII) and pAN14 (ΔDI&DII) harvested at 16, 24 and 32 hours, respectively. The positions of p71 (71-kDa) and p64 (64-kDa) are indicated on the side of the gels.

3.4 Discussion

The main objective of the work described in this chapter was to investigate whether deletions within the N terminal domain of HaSV p71, would have an effect on its assembly into ordered structures representing VLPs.

Despite the 12% sequence homology that exists between nodaviruses and NωV coat proteins (Helgstrand *et al*., 2004) and the 40% sequence homology existing between residues 1-48 of the two omegatetraviruses (NωV and HaSV) coat proteins (Hanzlik *et al*., 1995), the organisation of the positively charged residues was found to be conserved within the amino termini region of the viral coat proteins (Table 3.2). It was revealed from the alignment that the conserved clusters residing within the N terminal domain of p71 are of some significance. This is because the first positively charged cluster of p71 (residues 10-18(DI)) overlapped the five arginine residues (R6,8,9,11, and 12) of FHV coat protein that were shown to be crucial for the specific encapsidation of RNA2 (Marshall and Schneemann, 2001). In addition, DI of p71 overlapped that of PaV (residues 4-18), which was observed in the major groove of the RNA duplex along the icosahedral twofold axes (Tang *et al*., 2001). In the case of the DII of p71 (residues 32-38), an overlap with that of PaV (residues 36-50) was observed to neutralise the backbone of the RNA duplex (Tang *et al*., 2001). Since DI and DII reside within residues 1-48 of NωV coat protein, which is disordered (Helgstrand *et al*., 2004), it was probable that the same region of p71 could project towards the interior of the capsid shell possibly organising the genomic RNA during assembly.

Instead of the large deletions, such as the deletions of residues 2-31 and 1-50, which were investigated for FHV coat protein (Dong *et al*., 1998; Marshall and Schneemann, 2001), it was decided to restrict the deletions to DI and DII of p71. Therefore, three types of deletion within the coding sequence of p71 were constructed, deleting either DI (residues 10-18) or DII (residues 32-38) or both. These three deletions (referred to as p71 mutants) were proposed to have an effect on VLP assembly.

Since it was not known whether the P_{AG} expression of p71 mutants would be expressed and if so, whether they would be processed to form p64, it was decided to analyse the expression system over three time intervals of 16, 24 and 32 hours. To align the position of the 64-kDa mature HaSV capsid protein (p64), native HaSV obtained from its natural host was used. In addition, the detection of the p64 band confirmed the anti-HaSV polyclonal antibody detection.

For the p71 mutants, expression of p71 ΔDI had similar expression levels to that of the wild type at 16 and 24 hours (Figure 3.6, panel B, compare lanes 2, 8 with 1 and 7, respectively). Although, there was no basal expression of p71 containing double deletion of both clusters (ΔDI&DII) (Figure 3.6, panel B, lane 4), there was a similar expression level with the wild type and with p71 ΔDI, at 24 hours (compare lanes 10 with 1 and 2, respectively). Apart from the expression of p71 ΔDI and p71 ΔDI&DII, p71 containing deletion of DII (ΔDII) was not expressed (Figure 3.6, panel B, lanes 3, 9 and 14). The lack of p71 ΔDII expression was not due an error within its coding sequence, since no mutation was found other than the deletion of the coding sequence for residues 32-38. This was the objective in using the sequencing primer MT4. However, a reverse primer should have been used to sequence the 3' end of the P_{AG}, so as to detect any mutation within the promoter region preceding the *Bcl* I restriction site. Such mutations are likely to have been introduced by inverse PCR, but the sequencing of this region was not done since it relied on the low error rate (approximately 4.8×10^{-6}) of the Expand High Fidelity PCR system (Roche). Another possible reason for the lack of p71 ΔDII expression could be that this deletion of the second cluster might have had a profound effect on the expression of it. This is because p71 ΔDI&DII was expressed, suggesting that the deletion of both clusters might have rescued the expression deficit of p71 ΔDII.

Referring to the expression of p71 ΔDI and p71 ΔDI&DII at 32 hours (Figure 3.6, panel B, lanes 13 and 15, respectively), the amount of protein expressed was significantly lower than that of the wild type (Figure 3.6, lane 12). This difference in the amount of protein expressed was not due to the amount of protein loaded onto the gels, since the yeast crude extracts were estimated so as to ensure that equal protein concentrations were loaded (Figure 3.6, panel A, lanes 1-4, 5-8, 9-12). The drop in the protein amount

could be attributed to the significance of the deleted peptides from p71. In addition, there was no extra band co-migrating with the 64-kDa band denoting mature capsid protein, p64 (Figure 3.6, panel B, compare lanes 13 and 15 with 11). This implies that the mutants were not processed to form p64. Data obtained from the p71 mutants at 32 hours is discussed next.

Canady *et al*. (2000) and Taylor *et al*. (2002) demonstrated that maturation cleavage of NωV capsid pre-cursor protein α to β and γ proteins is assembly-dependant, which means there is only cleavage of the capsid protein once the particles are assembled. For HaSV, Tomasicchio (unpublished results) has showed that the detection of p64 indicated the presence of mature VLPs, therefore the absence of p64 for the p71 mutants suggested that no mature VLPs were present. Due to the proposal that excess viral mRNA achieved by the P_{AG} expression system could be a required factor for VLP assembly, data obtained for the p71 mutants suggests that the peptides deleted from p71 might be of significance. As a result, these deletions probably prevented the interaction between p71 and mRNA, thus preventing VLP assembly. This agrees with the experimental attempt where no VLPs were purified from the cells expressing p71 mutants. These p71 mutants might have assembled inefficiently or formed into unstable VLPs.

A proposal was put forward to explain the decreased amount of expressed p71 mutants observed at 32 hours (Figure 3.6, panel B, lanes 13 and 15). It was known that from the P_{GAL} expression system of *P71*, the insoluble aggregates disappeared after 12 hours post-induction. Venter (2001) proposed that these aggregates were degraded by yeast proteolytic activity, which was based on the demonstration by Finley (1992) that the presence of protein aggregates in yeast induced the ubiquitin pathway. Considering these facts, it seems that the solubility of p71 is dependent on the presence of sufficient viral RNA as proposed for the P_{AG} expression system. It is proposed that these p71 mutants might have decreased solubility that would render them more susceptible to insoluble aggregate formation, and probably proteolytic degradation. Although the solubility of the p71 mutants was not investigated, two conclusions could be derived from this experiment. Firstly, the expression of p71 mutants decreased at 32 hours and secondly, the proteins were not processed into p64.

Since only the expression of p71 mutants was investigated, more work is required to characterise the expressed protein. Two questions remain unanswered; whether there are soluble p71 mutants being expressed, and whether the expressed proteins assemble into ordered structures representing VLPs under certain conditions.

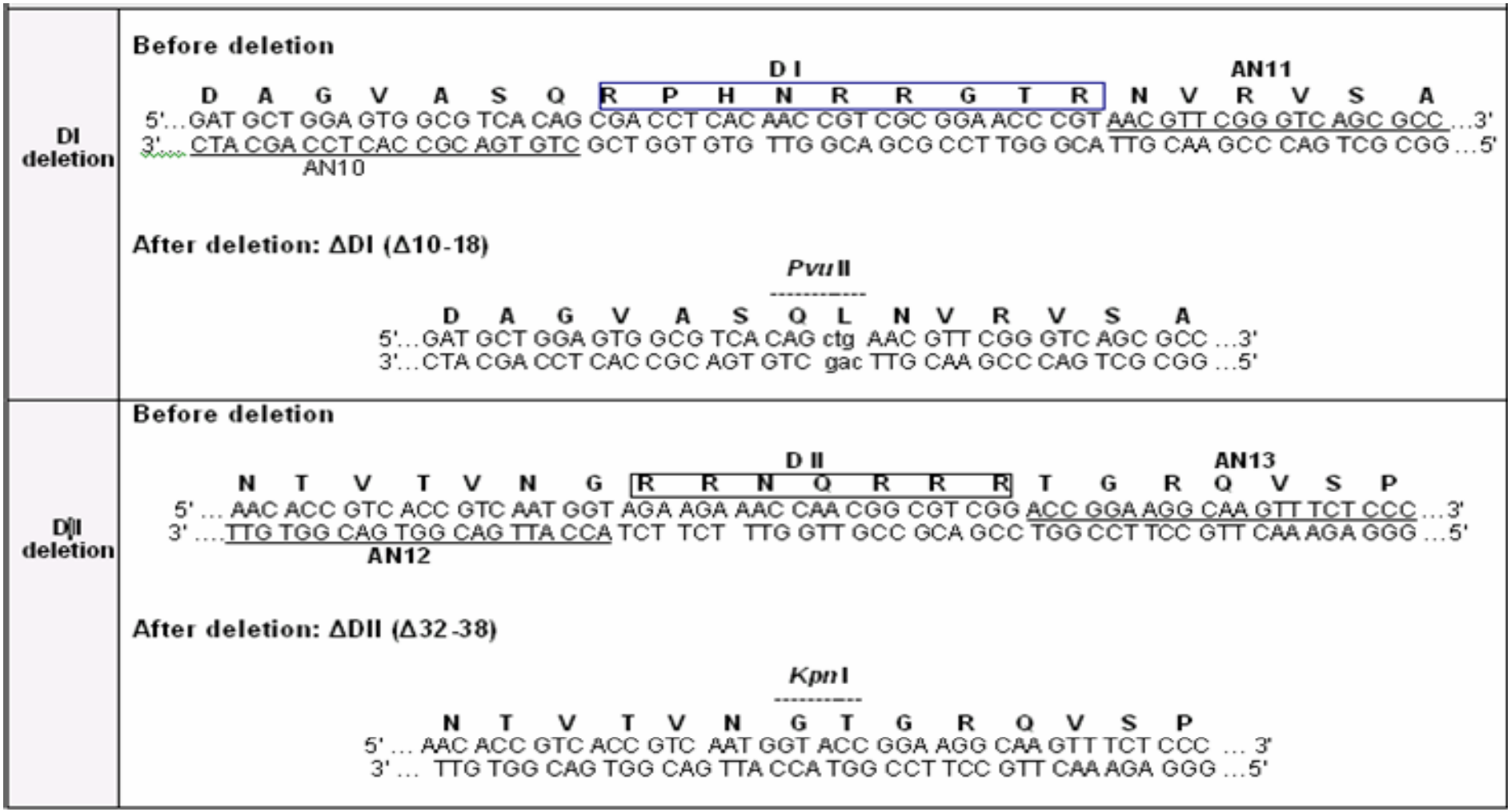


Figure 3.2 Coding sequence of *P71* with primers designed to delete DI and DII. Inverse PCR was used with the following primer pairs AN10 & AN11, and AN12 & AN13 to delete DI (residues 10-18) and DII (residues 32-38), respectively. Lower case letter within the coding sequence of p71 with DI deleted refers to the heterologous sequence introduced by the primer AN10. This introduced the recognition sequence for the restriction digest, *Pvu* II. The recognition sequence for the restriction digests *Pvu* II and *Kpn* I was employed for the identification of DI and DII deletions, respectively.

Table 3.2 Alignment of the region spanning the N-terminal domain of the capsid precursor protein of two omegatetraviruses and two nodaviruses.

Nudaurelia ω virus (NωV)	MDSNSASG**KRR**S**R**NV**R**IAANTVNVAP**K**Q**R**QA**R**G**RR**A**R**S**R**AN…	Omegatetraviruses
Helicoverpa armigera stunt virus (HaSV)	MGDAGVASQ**R**P**H**N**RR**GT**R**NV**R**VSANTVTVNG**RR**NQ**RRR**T**R**Q…	
Flock house virus (FHV)	MVNNN**R**P**RR**Q**R**AQ**R**VVVTTTQTAPVPQQNVP**R**NG**RRRR**N**R**T**RR**N**RRR**V**R**GMN…	Nodaviruses
Pariacoto virus (PaV)	MVS**R**T**K**N**RR**N**K**A**RK**VVS**R**STALVPMAPASQ**R**TGPAP**RK**P**RKR**NQALV**R**NP**R**LTA…	

The basic amino acids, Arginine (R), Histidine (H) and Lysine (K), are in blue bold.
The clusters of basic amino acids that are observed to be conserved amongst the two groups of viruses are outlined in a box.

Chapter 4: Conclusion & Future investigations

The principal objective of this study was to develop an experimental system to investigate the interaction between the HaSV capsid protein and viral RNA during virus assembly. Such an experimental system would provide an understanding of the mechanism for the recognition and packaging of the viral genome, which in the case of tetraviruses, is poorly understood.

Comparative studies between the well characterized tetravirus, NωV, and the structurally analogous nodaviruses suggest that the N- and C-terminal regions of the viral coat protomers which lie in close proximity to the encapsidated viral RNA, are similar in both virus families. The N terminal domain is the region of the coat protomer between the N terminus and the β-barrel domain. According to the literature, the N terminal domains of the two nodaviruses PaV and FHV, and the tetravirus NωV, share similarities due to their highly positive charge (Agrawal and Johnson, 1995; Marshall and Schneemann, 2001; Tang *et al*., 2001). This similarity is attributed to the presence of numerous positively charged amino acids, which are predominantly arginine residues, found clustered proximally to the amino end of this protein. Secondly, the disordered residues 1-48 of the NωV N terminus are analogous to residues 2-31 of the nodavirus, FHV, which have been proposed to project towards the interior of the capsid shell, possibly organising the central core of RNA during assembly (Helgstrand *et al*., 2004; Tihova *et al*., 2004). In addition, these positively charged residues residing in the N terminal domain were found to contact viral RNA as depicted in the crystal structure of PaV (Tang *et al*., 2001) and are required for the specific encapsidation of RNA2 by FHV particles (Marshall and Schneemann, 2001).

Evidence supporting further similarities between the N terminal domains of the two viral families was obtained from an alignment of this region. This alignment was compiled using the coat protein sequences of two omegatetraviruses, HaSV and NωV and two nodaviruses, FHV and PaV (Table 3.2). It was found that the organisation of the positively charged arginine residues is conserved between these

groups. These residues are grouped into two clusters with an intersecting region of aliphatic residues. Most importantly, due to this degree of conservation, the two basic clusters of NωV and HaSV, overlapped those of nodaviruses that had been studied. More specifically, the positively charged residues residing in the two clusters of nodaviral coat proteins were observed to contact RNA within PaV particles (Tang *et al*., 2001). In addition, the five arginine residues residing in the first cluster of FHV N terminal domain were required for the specific encapsidation of viral RNA2 (Marshall and Schneeman, 2001). A deletion that abolished both clusters from FHV capsid protomer inhibited viral assembly, thus suggesting that both clusters were essential for particle assembly (Dong *et al*, 1998). Due to the evidence of a degree of conservation existing between the N terminal domains of tetra- and nodavirus capsid protomers, it could be proposed that these two viral families share similar assembly pathways.

To fulfil the research objective, two experimental systems were developed *in vitro* and *in vivo* to investigate the interaction between the N terminal domain of HaSV p71 and viral RNAs. The decision to develop an *in vitro* experimental system was based on the results obtained from the yeast P_{GAL} expression of the gene encoding HaSV capsid precursor protein (p71) (Venter, 2001). This expression system resulted in a low yield of assembled HaSV VLPs, which was affected by the formation of insoluble aggregates (Venter, 2001), and was therefore unsuitable for *in vivo* studies of the assembly of tetraviral particles. The IMPACT^{TM}-CN system was therefore used to develop an *in vitro* RNA-protein pull down assay. Residues 1-121 of the N terminal domain of HaSV capsid protein were fused to the chitin binding affinity tag (intein peptide) to create a fusion protein. The adaptation of this protein purification system involved the affinity matrix, which was purposely used for the purification of fusion protein and to pull down the fusion protein together with any bound RNA. To add to the simplicity of this *in vitro* experimental system, the interaction between the fusion protein and RNA was confirmed by gel electrophoresis analysis of precipitated bound RNA on an agarose gel containing ethidium bromide.

Despite attempts at optimisation, major problems associated with the development of an *in vitro* experimental system were encountered. These included inconsistent results partly attributed to the state of the fusion protein expressed (in particular *in*

vivo cleavage), and also to the interaction occurring between the intein peptide and nucleic acids. In addition, the excessive fluorescence, present in the gels and the presence of bands not representing nucleic acids, made it difficult to interpret the results. Thus, the current binding assay was not effective in answering the question of whether the N terminal domain of p71 specifically binds HaSV RNAs. It is suggested that either an alternative tag should be used to construct the fusion protein, or that an alternative RNA-protein interaction assay be used, such as North-western analysis or electrophoretic mobility shift assay (EMSA). The latter two assays are common approaches used to investigate RNA-protein interactions for RNA viruses. Since much information on the assembling of virus particles and packaging of the viral genome by the nodavirus, FHV, has been studied *in vivo,* it was decided to investigate the assembly of HaSV VLPs *in vivo,* in order to draw comparisons.

Until recently, the solubility of HaSV p71 expressed in yeast was improved by using the P_{AG} expression system *in vivo* (Tomasicchio, unpublished results). Not only was the solubility of p71 improved, but also the yield of assembled HaSV VLPs was increased and these VLPs were shown to mature *in vivo*. These VLPs were found to be structurally indistinguishable from the wild type virus (Tomasicchio, unpublished results), which led to the suggestion that this expression system was suitable for *in vivo* studies. This was based on the fact that much information was garnered from *in vivo* studies on NωV VLPs produced in insect cell lines. NωV VLPs were produced using the baculoviral expression of its capsid precursor protein gene in insect cell lines (Agrawal and Johnson, 1995). The production of NωV VLPs has promoted *in vivo* studies on the structure, assembly and maturation of NωV coat protein (Munshi *et al*., 1996; Helgstrand *et al*., 2004; Canady *et al*., 2000, 2001; Taylor *et al*., 2002).

The second part of this study employed the P_{AG} expression system to investigate the effect of expressed p71, containing deletions of either the first or second basic cluster, or both, within the N terminal domain, on assembly into VLPs *in vivo*. It was found that p71 containing deletions of the first and both basic clusters were expressed. However, the construct encoding p71 containing an N terminal deletion of the second basic cluster failed to be expressed, most possibly this deletion results in expression of an unstable protein. This observation requires further investigation. The p71 mutants which were expressed were not processed into p64, which

suggests that there was no assembly of VLPs, since it has been shown that the presence of p64 indicates the presence of mature VLPs (Tomasicchio, unpublished results). In other words, cleavage is indicative of assembled particles as opposed to protein (Canady *et al.,* 2000; Taylor *et al.,* 2002)). This proved to be true, since no VLPs were purified when an experiment to do a large scale preparation for the purification of VLPs for p71 mutants was attempted. Not only were p71 mutants not processed, but there was evidence of decreased expression for the p71 mutants at the time of the expected processing of p71. It was proposed that the p71 mutants might have decreased solubility and might have been subjected to proteolytic degradation. However, the solubility of p71 mutants still remains to be investigated. The lack of processed p71 mutants, their decreased expression and the lack of purified VLPs, all emphasized the significance of the positively charged clusters within the N terminal domain of p71. It is proposed that the deleted region affected the efficiency of interaction between the p71 mutants and viral RNAs, thus affecting VLPs assembly.

A correlation could be made between the assembly defect of p71 mutants and FHV VLPs assembled from the coat protein of FHV containing deletions within the N terminal domain. The effect of two different types of N terminal domain deletions within the FHV coat protomer on assembly is discussed next. In addition, only the *in vivo* studies of FHV VLPs assembly that were carried out under non-viral replication conditions are referred to. Expression of FHV coat protein containing 2-31 residues deleted (Δ31), which removed the first basic cluster as well as the intersecting region of amino acids, produced FHV VLPs that possessed highly heterogeneous structures (Dong *et al*., 1998). These structures not only possessed wild type like *T*= 3 particles, but also small bacilliform-like structures and irregular structures. The second deletion of 1-50 residues, which removed both clusters including the intersecting region, abolished the assembly of VLPs. This led Dong *et al*. (1998) to propose that the 17 arginine residues residing within this deleted region were significant for the neutralisation of encapsidated RNA. This hypothesis could be true for the assembly defect of p71 mutants, more specifically for the deletion of both clusters of arginine residues. A comparison of assembly defects of the p71 deletion mutant of the first basic cluster with the assembly of Δ31 FHV VLPs, suggests that this p71 mutant could be assembled into unstable particles.

It would be interesting to investigate the effect on assembly of p71 mutants in the presence of co-expressed full length RNA2. This is because the only viral RNA present in the expression system was the mRNA encoding the p71 mutants and this might have had an RNA encapsidation signal removed. Although there were assembled Δ31 FHV VLPs which suggested that this might not have been the case, future investigation may reveal a difference between the assembly pathway requirements between noda- and tetraviruses.

Not only the co-expression of RNA2 with p71 mutants should be investigated, but also the co-expression of full-length HaSV RNAs. The co-expression of full length HaSV RNAs may circumvent this (the defect of p71 mutant VLPs) as these RNAs may bind elsewhere on the protein. This suggestion is based on the studies of Δ31 FHV VLPs assembly, where the effect of VLPs assembled differ in the presence or absence of a biologically active viral replicase (Dong *et al*., 1998; Krishna *et al*., 2003). Under non-replication conditions, which means low viral RNA concentrations, the structures of assembled Δ31 FHV VLPs possessed true *T*= 3 quasi-symmetry in addition to predominantly heterogeneous particles (Dong *et al*, 1998). In contrast, when FHV Δ31 VLPs were assembled in the presence of active viral replicase (which has been shown to replicate both genomic FHV RNAs *in vivo*) their capsid shells were indistinguishable from the wild type (Krishna *et al*., 2003). Most importantly, these VLPs assembled, allowed further investigation of the encapsidated nucleic acid of these virus particles. These studies proved that the C-terminus specifically encapsidates both viral RNAs, of which the N terminus is dependant for the specificity for viral RNA2 (Marshall and Schneeman, 2001; Schneeman and Marshall, 1998). Although there is a lack of a suitable cell line that supports tetraviral replication (Bawden *et al*., 1999), the next few paragraphs will provide insight into an alternative expression system that is currently being investigated.

One major drawback of tetravirus research is that there is no available cell line that supports replication, except in the case of the *Providence virus* (Pringle *et al*., 2003). In the laboratory, an investigation is being carried out to determine whether the co-expression of the gene encoding FHV viral replicase with HaSV full-length RNAs, would replicate HaSV RNAs. If there is replication of HaSV RNAs by FHV replicase, the system could subsequently be used for *in vivo* studies. Since PrV is the only tetravirus known to replicate in a cultured cell line (Pringle *et al*., 2003), this virus

would be ideal for investigations in order to provide insight into the tetraviral encapsidation process. Although PrV is a betatetravirus that encapsidates a monopartite genome, RNA-encapsidation analysis using this virus model *in vivo* may provide clues on how tetraviruses interact with their genomes.

An alternative experimental system to investigate assembly of tetraviral VLPs *in vitro*, could be based on a study by Choi *et al*. (2002). In this study, the coat proteins of the *T*= 3 plant virus, *Brome mosaic virus* (BMV) particles were purified and re-assembled *in vitro* in the presence of its viral RNAs. Further experiments revealed that a region of the viral RNA was required for initiation of BMV VLP assembly. Based on this study, HaSV p71 could be purified under denaturing conditions, and allowed to re-assemble under native conditions in the presence of increasing concentrations of *in vitro* synthesised viral RNA transcripts. This would determine whether increasing concentrations of viral RNAs would be required for HaSV VLP encapsidation, and whether further experiments could be used to determine which region of coat protein or viral RNA would be involved in RNA-protein interactions.

A similar experimental approach used for BMV (Choi *et al.,* 2002), has been employed for FHV. Purified FHV pro-virions dissociate into its capsid pre-cursor protein and viral RNA when exposed to freezing temperature at a high ionic strength and at pH 9 (Schneemann *et al*., 1994). These dissociated FHV pro-virions spontaneously re-assembled at low ionic strength and at pH 6.5 (Schneemann *et al*., 1994). In addition, these re-assembled particles exhibited normal maturation cleavage of the capsid pre-cursor protein and were infective when compared with authentic FHV. Further reconstitution experiments showed that when the dissociated viral RNA was removed, abnormal particles were obtained after re-assembly (Schneemann *et al*., 1994). It was suggested that the viral RNA might be crucial for the assembly of FHV particles.

It must be emphasised that not only the N terminal domain of HaSV p71 should be investigated, but also the γ-peptide at the C-terminal domain, since there is a basic region of predominantly arginine residues residing within this region which may also be significant. An *in vitro* RNA-protein interaction assay would be appropriate for investigation to study this interaction with RNA. A reason for the latter suggestion is that according to the crystal structure of NωV determined by Helgstrand *et al*. (2004),

the ordered γ-peptides are involved in protein-protein interactions. Thus, a deletion within this region of the protein may interfere with VLP assembly. The disordered γ-peptides have been proposed to project towards the interior of the capsid shell exposing its positively charged residues available for interaction with the viral RNAs. However, it is known from studies on FHV C-terminal domain under viral replication conditions (Schneemann and Marshall, 1998) that this is required for the specific encapsidation of viral RNAs to which the N termini confers specificity for RNA2. In other words, the N terminal domain is dependent on the specific encapsidation of viral RNAs by the C-terminal domain. Due to the degree of similarities, in particular between the N terminal domains, of tetra- and nodaviruses, it could be suggested that the specific encapsidation of viral RNAs by nodaviruses (FHV) C-terminal domain might be similar for tetraviruses as well.

Since, the P_{AG} expression of p71 in yeast has not been fully investigated more studies are required to determine whether abundant viral RNA promotes HaSV VLP assembly. Although, there is still much work to be done on these simple tetraviruses, these RNA viruses will continue to challenge us.

Appendices

Appendix 1 *E. coli* strains, preparation of media and protocols for the preparation and transformation of *E. coli* competent cells

Recombinant DNA techniques

The *E. coli* strains used for plasmid maintenance are: DDH5α (Hanahan, 1983) as a host for recombinant plasmids; and JM110 (Yanisch-Perron *et al*., 1985) a *dam*$^-$, *dcm*$^-$ methylase deficient strain for certain recombinant plasmids containing methylase protected cleavage sites. The bacterial strains were grown at 37^{o}C in Luria broth or on Luria agar plates supplemented with an antibiotic to which the recombinant plasmid conferred resistance. The preparation of *E. coli* competent cells and the transformation protocols are also described in this Appendix.

Isolation of recombinant plasmid DNA from *E. coli* DDH5α for diagnostic purposes was performed using the easy plasmid preparation protocol of Berghammer and Auer (1993). DNA used for sequencing, storage, and *in vitro* transcription was isolated using the High Pure Plasmid Isolation Kit (Roche) or the Quantum Prep® Plasmid Midiprep Kit (Biorad). DNA restriction and modification were performed according to the manufacturers' instructions. The resultant restriction fragments of DNA restriction analysis were separated by electrophoresis using Tris-acetate-EDTA buffer on a 1% agarose gel containing 0.5 μg/ml ethidium bromide and visualised using the Kodak DC120 gel imaging system.

The BigDye Terminator Cycle Sequencing Kit (Applied Biosystems) was used for DNA sequencing reactions. The cycle sequencing was performed on the GeneAmp 970, and DNA sequencing was determined using an ABI 3100 Genetic Analyser, which is based on capillary electrophoresis.

***E. coli* strains**

***E. coli* strain**	**Genotype**	**Reference**
BL21 (DE3)	F– *ompT gal [dcm] [lon] hsdSB* (rB– mB–; an *E. coli* B strain) with DE3, a λ prophage carrying the T7 RNA polymerase gene	(Studier and Moffatt, 1986)
DDH5 α	F´/*endA1 hsdR17* (rK–mK+) *glnV44 thi-1 recA1 gyrA* (Nalr) *relA1* Δ*(lacIZYA-argF)U169 deoR* (φ80*dlac*Δ*(lacZ)M15*)	(Hanahan, 1983)
JM110	F´ *traD36 lacIq*Δ*(lacZ)M15 proA+B+*/*rpsL* (Strr) *thr leu thi lacY galK galT ara fhuA dam dcm glnV44* Δ*(lac-proAB)*	(Yanisch-Perron *et al.*, 1985)

Preparation of media (Luria Broth (LB) and Luria Agar (LA))

Per Litre: 1% Tryptone
0.5% Yeast extract
0.5% NaCl
1.5% Agar (only for LA)

Made up to 1L with ddH_2O, and autoclaved for 20 minutes.
Antibiotics were added (if necessary) after autoclaving.

Preparation of *E. coli* competent cells

1) Prepare overnight confluence *E. coli* cultures
2) Inoculate 4 flasks containing 100 mL LB (1:5 flask liquid:vol) with 1.5, 1, 0.7 and 0.3 mL of the overnight 5 mL confluence cultures
3) Once the absorbance of the highest inoculum reached a density of OD_{600} = 0.6-0.8 cool flasks at 4°C for 5-10 minutes.
4) Harvest the cells at 5000 rpm at 4°C for 10 min
5) Decant supernatant, resuspend pellets in RF1 (100 mM KCl; 50 mM $MnCl_2$; 30 mM CH_3COOK; 10 mM $CaCl_2$;15 m/v glycerol (pH 5.8)) and incubate on ice for 20 min
6) Re-pellet cells
7) Pool together in 4 mL of RF2 (100 mM MOPS; 10 mM KCl; 75 mM $CaCL_2$; 15% m/v glycerol (pH 6.8))
8) Aliquot 150 μL into sterile eppendorfs and store at –80°C until use

Transformation of *E. coli* competent cells

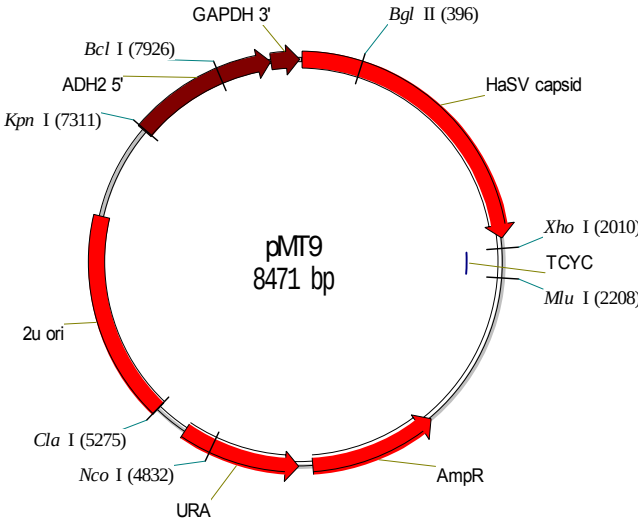
1) To 100 μL of thawed competent *E. coli* cells add 100-200 ng of plasmid DNA
2) Incubate at 4°C for 20 min
3) Heat shock at 42°C for 45 sec
4) Incubate at 4°C for 5 min
5) Add 900 μL of LB
6) Incubate at 37°C for 1 hr
7) Plate 100-200μL onto LA plates containing respective antibiotics
8) Allow plates to dry for 15 min
9) Incubate ON at 37°C

Appendix 2 Plasmids not generated in this study

Only unique restriction sites are represented

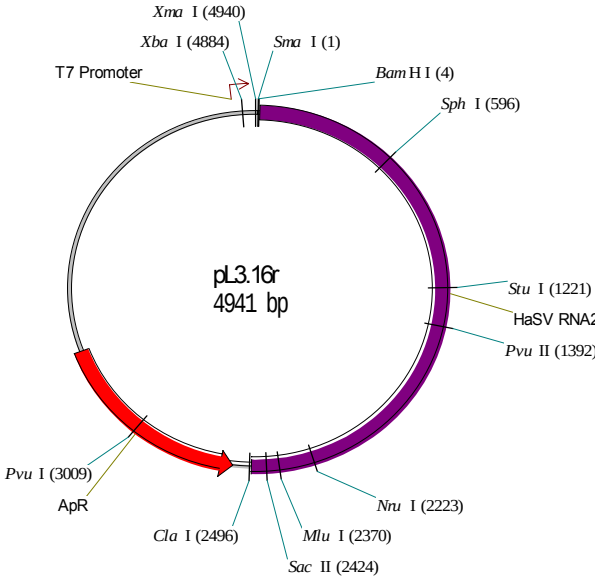
pMT9

HaSV p71 under the control of $P_{ADDH2\text{-}GAPDDH}$ constructed by Tomasicchio, M, Rhodes University)
Referred in section 3.2.1

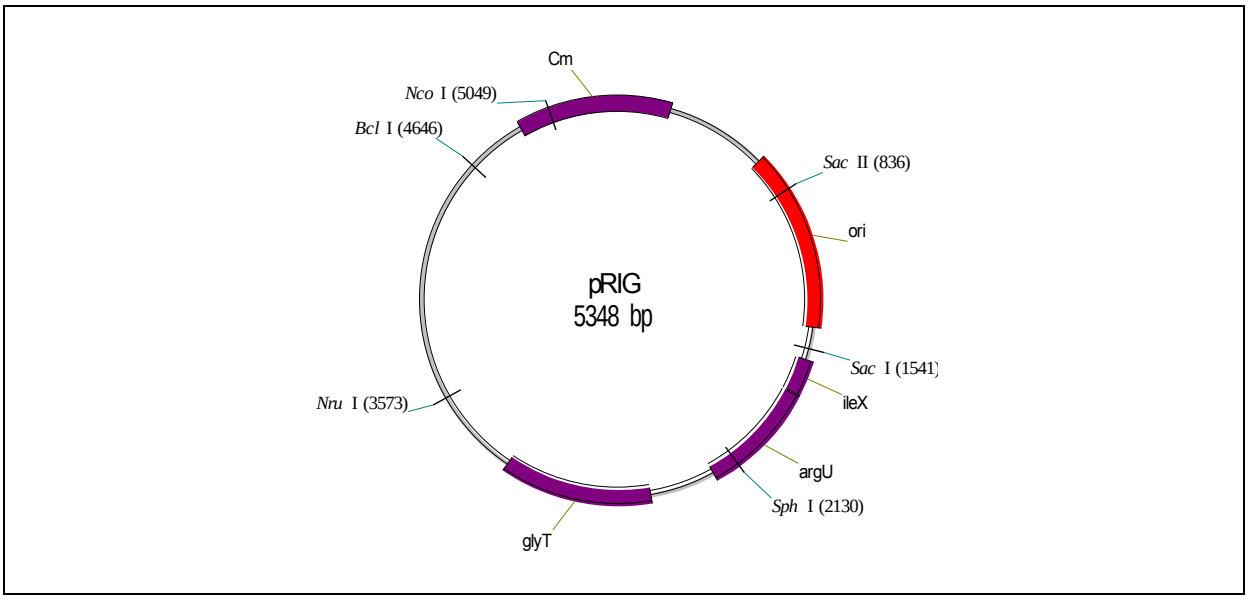


pL3.16

Full length HaSV RNA1 cDNA under control of T7 promoter (used for *in vitro* synthesis of HaSV RNA1) constructed by L. Plessis, Rhodes University. Referred in section 2.2.8.



pRIG:
Constructed by Baca and Hol (2000). Referred in section 2.3.3.



Appendix 3 Sequencing primers

Primer	Gene	Vector	D/C*	Sequence	Section ***
pUC F	*lacZ*	pGEM-Teasy	D	5' – cgc cag ggt ttt ccc agt cac gac -3'	2.2.2
InteinR	Intein	pTYB1	C	5' - acc cat gac ctt att acc aac ctc – 3'	2.2.2
MT4	$P_{ADDH2-GAPDDH}$	**	D	5' – gac ggt agg tat tga ttg taa ttc tgt aa-3'	3.2.1

*D/C – direct or complementary
** Any vector containing the $P_{ADDH2-GAPDDH}$ sequence.
*** Referred in section

Appendix 4: PCR Thermal Cycling Programmes

Programme 1		Programme 2		Programme 3	
1×	96°C for 2 min	1×	94°C for 2 min	1×	94°C for 2 min
25×	96°C for 10 sec 50°C for 5 sec 60°C for 4 min	25×	94°C for 45 sec 54°C for 45 sec 72°C for 1 min	20×	94°C for 15 sec 56°C for 30 sec 75°C for 1 min
1×	60°C for 4 min	1×	72°C for 2 min	1×	75°C for 7 min

Programme 5			Programme 6			Programme 7		
1×	95°C for 30 sec		1×	94°C for 1 min		1×	94°C for 1 min	
18×	95°C for 30 sec 52°C for 1 min 68°C for 15 min		20×	94°C for 45 sec 53°C for 90 sec 68°C for 4 min		20×	94°C for 45 sec 53°C for 90 sec 68°C for 4 min	
			1×	68°C for 8 min		1×	68°C for 8 min	

Appendix 5 SDS-PAGE & Western blotting procedures

Preparation of a 7.5% SDS-PAGE gel

Reagents	Stock solutions	Resolving gel (7.5%)	Stacking gel (5.2%)
Acrylamide stock	40%	3.75	1.3 mL
H_2O		7.5	7.25 mL
Tris pH 8.7	1M	8.33	
Tris pH 6.8	1 M		1.3 mL
SDS	10 %	0.2	100 μL
APS	10%	0.2	50 μL
TEMED		0.02	10 μL

SDS-PAGE Running buffer (×2)

57.6g glycine
12g Tris (pH 8.3)
4g SDS
Made up to 2L with ddH_2O

Bis-Acrylamide stock solution (40%)

38% Acrylamide
2% Bis-Acrylamide

Coomassie Stain
Dissolve 2g Coomassie in 200ml
Destain I, and stir overnight
Filter

Destain I (1/2 hour)
50% Methanol
10% Glacial acetic acid

Destain II (Overnight)
5% Methanol
7% Glacial acetic acid

Destain III (2 hours)

5% Methanol
7% Glacial acetic acid
3% Glycerol

Western blotting procedure

Electrotransfer/ blotting of resolved protein samples onto a Hybond-C+ membranes were performed according to manufacturers' instructions (Amersham Biosciences). Electrotransfer was carried out at 250 mA for 1 hr. Blotted membranes were stained for 1 minutes with Ponceau-S staining solution and rinsed as described by Sambrook *et al*. (1989) in order to demarcate the positions of the protein molecular weight markers. Preparation of the membranes for chemiluminescence detection was carried out according to the manufacturer's instruction of the BM Chemiluminescence Western Blotting Kit (Mouse/Rabbit) (Roche). The blocking and blotting solutions

used are described in the BM Chemiluminescence Western Blotting Kit (Mouse/Rabbit) (Roche).

The procedure of Western blotting is as follows: the membranes were incubated overnight at room temperature with blocking solution. The primary and secondary anti-bodies were incubated with the membranes in blotting solution for 3 hours and 1 hour at room temperature, respectively. Between each incubation, the membranes were rinsed with TBS-Tween buffer 4 × for 5 minutes each. Proteins were detected by exposure to X-ray film (Kodak). Standard photographic techniques and solutions (Kodak) were used to visualise the chemiluminescent exposure of the X-ray film.

Appendix 6 RNA denaturing gel

RNA denaturing gel
(1 % Agarose/formaldehyde gel)

20 mL 5 × MOPS buffer
72 mL DEPC-treated ddH_2O
1 g Agarose, molecular biology grade

Heat to boiling and cool to 55°C.
Add 17.6 mL 37% formaldehyde
and 5 μL of 10 mg/mL stock of Ethidium bromide

5 × MOPS Buffer
(Running buffer for RNA denaturing gel)

0.2 M MOPS (pH 7.0)
0.05M sodium acetate
0.005M EDTA (pH 8.0)
DEPC-treated ddH_2O

Dispense in 200 μL aliquots and autoclave

RNA sample Buffer
(RNA denaturing gel)

10 mL of De-ionized formamide
3.5 mL 37% formaldehyde
2.0 mL 5 × MOPS buffer

Aliquot and store at –20°C

RNA loading Buffer
(RNA denaturing/native gels)

50% glycerol
1 mM EDTA
0.4% bromophenol blue
DEPC-treated ddH_2O
Aliquot and store at –20°C

Appendix 7 Preparation of ssDNA (binding assays)

1) Sonicate a stock of ssDNA (10 mg/mL) until fragments are 100 bp
2) Phenol extract and ethanol preciptate as described in Sambrook *et al*. (1989)
3) Resuspend in equal volume of RNAse free ddH_2O
4) Boil for 15 min, freeze immediately at -20°C

Appendix 8: Preparation of SSM

SMM: as described by Kaiser *et al.* (1994)

Per liter: 1.7g yeast nitrogen base without amino acids and ammonium sulfate
5 g ammonium sulfate
20g glucose
(supplemented with 20g agar for agar plates)

Make up to total volume with ddH_2O, autoclave. Once cool, add the following supplements.

Stock solution concentration	Supplement	Final concentration
1 mg/ mL	L-Tryptophan (Sigma)	20 mg/L
1 mg/ mL	L-Histidine (Sigma)	20 mg/L
1 mg/ mL	L-Leucine (Sigma)	100 mg/L

References

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. *Virology* **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.

Baca, A.M. and Hol, W.G. (2000). Overcoming codon bias: A method for high-level over-expression of *Plasmodium* and other AT-rich parasite genes in *Escherichia coli*. *Internat. J. Parasitol.* **30:** 113-118.

Baker, T.S., Olson, N.H. and Fuller, S.D. (1999). Adding the third dimension to virus life cycles: Three-dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. *Micro. Mol. Biol. Rev.* **63:** 862-922

Ball, L.A. (1995). Requirements for the self-directed replication of *Flock House virus* RNA1. *J. Virol.* **69:** 720-727.

Ball, L.A. and Johnson, K.L. (1998). Nodaviruses. In *The insect viruses*. (L.K. Miller and L.A. Ball, eds.), pp. 225-267. Plenum Press, London.

Ball, L.A. and Li, Y. (1993). Cis-acting requirements for the replication of *Flock House virus* RNA2. *J. Virol.* **67:** 3544-3551.

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

Berghammer, H. and Auer, B. (1993). "Easypreps": fast and easy plasmid minipreparation for analysis of recombinant clones in *E. coli*. *Biotechniques* **14:** 524-528.

Bong, D.T., Steinem, C., Janshoff, A., Johnson, J.E. and Ghadiri, M.R. (1999). A highly membrane-active peptide in *Flock house virus*: Implications for the mechanism of nodavirus infection. *Chemistry and Biology* **6:** 473-481.

Bothner, B.X., Dong, F., Bibbs, L., Johnson, J.E. and Siuzdak, G. (1998) Evidence of viral capsid dynamics using limited proteolysis and mass spectrometry. *J. B. Chem.* **273**: 673-676

Bothner, B., Schneemann, A., Marshall, D., Reddy, V., Johnson, J.E. and Siuzdak, G. (1999) Crystallographically identical virus capsids display different properties in solution. *Nature Struct. Biol.* **6:** 114-116

Brantley, J. and Hunt, A. (1993). The N terminal protein of the polyprotein encoded by the *Potyvirus tobacco-mottling virus* is an RNA-binding protein. *J. Gen. Virol.* **74:**1157-1162.

Canady, M.A., Tihova, M., Hanzlik, T.N., Johnson, J.E. and Yeager, M. (2000). Large conformational changes in the maturation of a simple RNA virus, *Nudaurelia capensis omega virus* (NωV). *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 omega virus* (NωV) maturation. *J. Mol. Biol.* **311:** 803-814.

Caspar, D.L.D. and Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symp. Quant. Biol.* **27:** 1-24.

Cheng, R.H., Reddy, V.S., Olson, N.H., Fisher, A.J., Baker, T.S. and Johnson, J.E. (1994). Functional implications of quasi-equivalence in a *T*= 3 icosahedral animal virus established by cryo-electron microscopy and X-ray crystallography. *Structure* **2:** 271-282.

Choi, Y. G., and Rao, A. L. N. (2000). Molecular studies on bromovirus capsid protein. VII. Selective packaging of BMV RNA4 by specific N terminal arginine residues. *Virology* **275:** 207-217.

Choi, Y.G., Dreher, T.W. and Reddy, A.L. (2002). tRNA elements mediate the assembly of an icosahedral RNA virus. *Proc. Natl. Acad. Sci. U.S.A.* **99:** 655-660.

Cousens, L.S., Schuster, J.R., Gallegos, C., Ku, L.L., Stempien, M.M., Urdea, M.S., Sanchez-Pescador, R., Taylor, A. and Tekamp-Olson, P. (1987). High level expression of proinsulin in the yeast, *Saccharomyces cerevisiae*. *Gene* **61:** 265-275

Dasgupta, R., Selling, B. and Rueckert, R. (1994). *Flock house virus***:** a simple model for studying persistent infection in cultured *Drosophila* cells. *Arch. Virol. Suppl.* **9:**121-132.

Dasmahaptra, B., Dasgupta, R., Ghosh, A. and Kaesberg, P. (1985). Structure of the *Black beetle virus* genome and its functional implications. *J. Mol. Biol.* **182:**183-189.

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

Eckerle, L, Albariño, C.G., Ball, L.A. (2003). *Flock house virus* subgenomic RNA3 is replicated and its replication correlates with transactivation of RNA2. *Virology* **317:** 95-108.

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 capsid with icosahedral surface symmetry *T*= 4. *J. Gen. Virol.* **24:** 191-200.

Finley, D. (1992). The yeast ubiquitin system. In *The molecular and cellular biology of the yeast Saccharomyces.* (E.W. Jones, J.R. Pringle and J.R. Broach, eds.), pp. 539-582. Cold Spring Harbor Laboratory Press, New York.

Fisher, A.J. and Johnson, J.E. (1993). Ordered duplex RNA controls capsid architecture in an icosahedral animal virus. *Nature* **361:** 176-182.

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.

Friesen, P.D. and Rueckert, R.R. (1981). Synthesis of *Black beetle virus* proteins in cultured *Drosophila* cells**:** differential expression of RNAs 1 and 2. *J. Virol.* **37:** 876-886.

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

Gallagher, T.M., Friesen, P.D., Ruekert, R.R. (1983). Autonomous replication and expression of RNA1 from *Black beetle virus*. *J. Virol.* **46:** 481-489

Gorbalenya, A.E., Pringle, F.M., Zeddam, J.L., Luke, B.T., Cameron, C.E., Kalmakoff, J., K, 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. and Hanzlik, T.N. (1998). Tetraviruses. In *The insect viruses*. (L.K. Miller and L.A. Ball, eds.), pp. 269-299. Plenum Press, London.

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

Gordon, K.H.J., Williams, M.R., Baker, J.S., Gibson, J.M., Bawden, A.L., Millgate, A.G., Larkin, P.J. and Hanzlik, T.N. (2001). Replication-independent assembly of an insect virus (*Tetraviridae*) in plant cells. *Virology* **288:** 36-50.

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

Hanahan, D. (1983). Studies on transformation of *Escherichia coli* with plasmids. *J. Mol. Biol.* **166:**557-580.

Hanzlik, T.N., Dorrian, S.J., Gordon, K.H. 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. (1995). Sequence of RNA2 of the *Helicoverpa armigera stunt virus* (*Tetraviridae*) and bacterial expression of its genes. *J. Gen. Virol.* **76:**799-811.

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

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. *Virology* **318:**192-203.

Hendry, D.A. (1991). *Nodaviridae* of invertebrates. In *Viruses of invertebrates*. (E. Kurstak, ed.), pp. 277-285. Marcel Dekker, New York.

Hendry, D.A., Bekker, M. and van Regenmortel, M. (1968). A non-inclusion virus of the pine emperor moth *Nudaurelia cytherea capensis* stoll. *S.Afr. Med. J.* **42:**117-121.

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.

Hosur, M.V., Schmidt, T, Tucker, R, Gallagher, T.M., Selling, B.H., Ruekert, R.R. and Johnson, J.E. (1987). Structure of an insect virus at 3Å resolution. *Proteins* **2:** 167-176.

Janshoff, A., Bong, D., Steinem, C., Johnson, J. and Reza Ghadiri, M. (1999). An animal virus-derived peptide switches membrane morphology**:** possible relevance to nodaviral transfection processes. *Biochem.* **38:** 5328-5336.

Johnson, J.E. (1996). Functional implications of protein-protein interactions in icosahedral viruses. *Proc. Natl. Acad. Sci. U.S.A.* **93:** 27-33.

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

Johnson, J.E. and V. Reddy. (**1998).** Structural studies of nodaviruses and tetraviruses. In *The insect viruses*. (L.K. Miller and L.A. Ball, eds.), pp. 171-224. Plenum Press, London.

Johnson, K.L., Price, D, Eckerle, L and Ball, LA (2004). *Nodamura virus* non-structural protein B2 can enhance viral RNA accumulation in both mammalian and insect cells. *J. Virol.* **78:** 6698-6704.

Johnson, J.E., Willits, D., Young, M. and Zlotnick, A. (2004). Interactions with capsid protein alters RNA structure and the pathway for *in vitro* assembly of *Cowpea chlorotic mottle virus*. *J. of Mol. Biol.* **335:** 455-464.

Juckes, I.R.M (1970). Viruses of the pine emperor moth. *Bull. S. Afr. Soc. Plant. Pathol. Microbiol.* **4:** 18-20.

Juckes, I.R.M (1979). Comparison of some biophysical properties of the *Nudaurelia* β and ε viruses. *J. Gen. Virol.* **42:** 89-94.

Kaesberg, P., Dasgupta, R., Sgro, J.Y., Wery, J.P., Selling, B.H., Hosur, M.V. and Johnson, J.E. (1990). Structural homology among four Nodaviruses as deduced by sequencing and X-ray crystallography. *J. Mol. Biol.* **214:** 423-435.

Kaiser, C., Michaelis, S. and Mitchell, A. (1994). Methods in yeast genetics. Cold Spring Harbor Laboratory Press, New York.

Krishna, N.K., Marshall, D. and Schneemann, A. (2003). Analysis of RNA packaging in wild-type and mosaic protein capsids of *Flock house virus* using recombinant baculovirus vectors. *Virology* **305:** 10-24.

Krishna, N.K. and Schneemann, A. (1999). Formation of an RNA heterodimer upon heating of nodavirus particles. *J. Virol.* **73:** 1699-1703.

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

Li, H., Li, W. and Ding, S. (2002). Induction and suppression of RNA silencing by an animal virus. *Science* **296:** 1319-1321.

Lu, M. and Lin, C. (2003). Involvement of the terminus of grouper betanodavirus capsid protein in virus-like particle assembly. *Arch. Virol.* **148:** 345-355.

Marino, M. (1991). Expression of heterologous proteins in yeast. In *Recombinant DNA technology and applications*, A. Prokop, R. Bajpai and C. Ho, eds (New York**:** McGraw-Hill, Inc.), pp. 29-65.

Marshall, D. and Schneemann, A. (2001). Specific packaging of Nodaviral RNA2 requires the N terminus of the capsid protein. *Virology* **285:**165-175.

Matthews, R.E.F. (1982). Classification and nomenclature of viruses. Fourth report of the international committee on taxonomy of viruses. *Intervirology* **17:** 1 – 160.

Moore, N. and Tinsley, T. (1982). The small RNA viruses of insects. *Arch. Virol.* **72:**229-245.

Mori, K.I., Nakai, T, Muroga, K, Arimoto, M, Mushiake, K and Furusawa, I (1992). Properties of a new virus belonging to *Nodaviridae* found in larval striped jack (*Pseudocaranx dentex)* with nervous necrosis. *Virology* **187:** 368-371.

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.

Murphy, F.A., Fauquet, C.M., Bishop, D.H.L., Ghabrial, S.A., Jarvis, A.W., Martelli, G.P., Mayo, M.A. and Summers, M.D. (1995). Virus Taxonomy**:** Sixth Report of the International Committee on Taxonomy of Viruses. Springer-Verlag, Vienna

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

Pierce Biotechnology Inc. Pierce.net**:** Grasp the Proteome TM (Protein interactions**:** Protein nucleic acid**:** conjugates). Grasp the ProteomeTM html**://**www.piercenet.com. (**Accessed: 2004)**.

Prasad, B.V., Hardy, M.E., Dokland T., Bella J., Rossmann M.G., and Estes M.K.(1999) X-ray crystallographic structure of the *Norwalk virus* capsid. *Science*. **286**: 287-290.

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

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

Reinganum, C. (1991). *Tetraviridae*. In *Atlas of Invertebrate Viruses*, J.Adams and J.Bonami, eds (Boca Raton, FL**:** CRC Press), pp. 533-592.

Rosenberg, A,H. Goldman, E, Dunn, J.J., Studier, F.W. and Zubay, G. (1993). Effects of consecutive AGG codons on translation in *Escherichia coli*, demonstrated with a versatile codon test system. *J. Bact* **175:**716-722.

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

Schneemann, A., Dasgupta, R., Johnson, J.E. and Rueckert, R.R. (1993). Use of recombinant baculovirus in synthesis of morphologically distinct virus-like particles of *Flock house virus*, a nodavirus. *J. Virol.* **67:** 2756-2763.

Schneemann, A., Gallagher, T.M., Ruekert, R.R. (1994). Reconstitution of *Flock House virus* provirions: A model system for studying structure and assembly. *J. Virol.* **68:** 4547-4556

Schneemann, A. and Marshall, D. (1998). Specific encapsidation of Nodavirus RNAs is mediated through the C-terminus of capsid precursor protein alpha. *J. Virol.* **72:** 8738-8746.

Schneemann, A., Reddy, V. and Johnson, J. E. (1998). The structure and function of nodavirus particles**:** a paradigm for understanding chemical biology. *Adv. Virus Res.* **50:** 381-446.

Schneemann, A., Zhong, W., Gallagher, T.M. and Rueckert, R.R. (1992). Maturation cleavage required for infectivity of a nodavirus. *J. Virol.* **66:** 6728-6734.

Struthers, J. and Hendry, D.A. (1974). Studies of protein and nucleic acid components of *Nudaurelia capensis* β *virus*. *J. Gen. Virol.* **22:** 355-362.

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.

Tan, C., Huang, B., Chang, S.F., Ngoh, G.H., Munday, B., Chen, S.C. and Kwang, J. (2001). Determination of the complete nucleotide sequences of RNA1 and RNA2 from *Greasy grouper* (*Epinephelus tauvina*) *nervous necrosis virus*, Singapore strain. *J. Gen. Virol.* **82:** 647-653.

Tan, R. and Frankel, A.D. (1995). Structural variety of arginine-rich RNA-binding peptides. *Proc. Natl. Acad. Sci. U.S.A.* **92:**5282-5286.

Tang, L, Johnson, K.N., Ball, L.A., Lin, T, Yeager, M and Johnson, J.E. (2001). The structure of *Pariacoto virus* reveals a dodecahedral cage of duplex RNA. *Nat Struct. Biol.* **8:**77-83.

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

Tihova, M, Dryden, K.A., Le, T.V., Harvey, S.C., Johnson, J.E., Yeager, M and Schneemann, A (2004). Nodavirus coat protein imposes dodecahedral RNA structure independent of nucleotide sequence and length. *J. Virol.* **78:** 2897-2905.

Towbin, H., Staehelin, T. and Gordon, J. (1979). Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets**:** procedure and some applications. *Proc. Natl. Acad. Sci. U.S.A.* **76:** 4350-4354.

van Regenmortel, M.H.V., Fauquet, C.M., Bishop, D.H.L., Carstens, E.B., Estes, E.K., Lemon, S.M., Maniloff, J., Mayo, M.A., McGeoch, D.J., Pringle, C.R. and Wickner, R.B. (2000). Virus Taxonomy**:** Seventh Report of the International Committee on Taxonomy of Viruses. Academic Press, London.

Venter, P.A. (2001) Heterologous expression of the *Helicoverpa armigera stunt virus* in *Saccharomyces cerevisiae*. Ph.D. thesis. Rhodes University, Grahamstown, South Africa.

Wery, J.P., Reddy, V.S., Hosur, M.V. and Johnson, J.E. (1994). The refined three-dimensional structure of an insect virus at 2.8 Å resolution. *J. Mol. Biol.* **235:** 565-586.

Xu, X., Severson, W., Villegas, N., Schmaljohn, C. S., Jonsson, C. B. (2002). The RNA binding domain of the *Hantaan virus* N protein maps to a central, conserved region. *J. Virol.* **76:** 3301-3308

Yanisch-Perron, C., Viera, J. and Messing, J. (1985). Improved M13 phage cloning vectors and host strains**:** nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33:** 103-119.

Zeddam, J., Rodriguez, J., Ravallec, M. and Lagnaoui, A. (1999). A noda-like virus isolated from the sweet potatoe pest *Spodoptera eridani* (*Cramer*) (Lep.; *Noctuidae*). *J. Inv. Pathol.* **74:**267-274.

Zhang, S.P., Zubay, G. and Goldman, E. (1991). Low-usage codons in *Escherichia coli*, yeast, fruit fly and primates. *Gene* **105:**61-72.

Zhong, W., Dasgupta, R. and Rueckert, R.R. (1992). Evidence that the packaging signal for nodaviral RNA2 is a bulged stem-loop. *Proc. Natl. Acad. Sci. U.S.A.* **89:**11146-11150.

Zlotnick, A., Reddy, V., Dasgupta, R., Schneemann, A., Ray, W., Rueckert, R., Johnson, J. (1994). Capsid assembly in a family of animal viruses primes an autoproteolyctic maturation that depends on a single aspartic acid residue. *J. Biol. Chem.* **269:** 13680-13684.