Characterization of VP4, a minor core protein of African horse sickness virus with putative capping enzyme activity

Jan Iman van den Bout

Submitted in fulfilment of the requirements for the degree *Magister Scientiae Genetics* in the Faculty of Natural and Agricultural Sciences, University of Pretoria

Date of submission: October 2003

"I am the wisest man alive; for I know one thing, and that is that I know nothing."
Socrates
469BC-399BC

Acknowledgements

This long and laborious journey is finally over and I can look back on a long student life at the University of Pretoria with a lot of good memories and obviously a hefty wad of knowledge under my belt. This thesis is the culmination of that journey. I was not alone on this journey and for those people who have supported me through these years I want to say thanks.

First of all to my parents who have stimulated me throughout my life to become the person I am today. Without their faith in my choice to become a scientist I would not be writing this. Thank you for your love, teaching and motivation to complete something that I will look back on with great fulfilment. Thanks also to my wife, Tessa, for understanding my long hours behind the computer and the support you have given me in this sometimes seemingly endless project. I would also like to thank all my family especially my sisters and brother and my grandmother. A special thought goes to my grandfather from who I hope I have received some of his infinite wisdom to be where I am today. You will always be in my heart.

A work like this is not performed in solitude. I have had some great colleagues and study mates who have helped me solve those problems that make science so unique. I especially want to mention Francois and Pamela who have guided me and helped me through the years in the lab. Finally, to my promoter Professor Henk Huismans with whom I have worked with to finish this thesis. Thanks for believing in me and helping me in obtaining this goal. I hope that I will make you proud with my scientific career.

I hope that this thesis will form the start of a beautiful career in the most amazing career a person can ever choose. Without science there is no knowledge and without knowledge we cannot be human.

List of keywords:

African horsesickness
Bluetongue virus
Capping enzyme
Guanylyl transferase
Leucine zipper
Methylase
Phosphatase
Solubility
Virus
VP4

List of abbreviations

Ab	Antibody
AHSV	African horsesickness virus
ATP	Adenoside triphosphate
bp	basepairs
BTV	Bluetongue virus
CsCl	Cesium chloride
ddNTP	dideoxy nucleoside triphosphate
DLP	double-layered particle
dsDNA	double-stranded deoxyribose nucleic acid
dsRNA	double-stranded ribose nucleic acid
ER	endoplasmic reticilum
GMP	guanosine monophosphate
GTP	guanosine triphosphate
kDa	kilo Dalton
Lys	lysine
MCS	multiple cloning site
ml	millilitre
MW	molecular weight
mRNA	messenger ribose nucleic acid
ng	nano gram
nm	nanometers
NMP	nucleoside monophosphate
ORF	open reading frame
pI	iso-electric point
Pi	phosphate
PPi	pyrophosphate
RNA	ribose nucleic acid
RT	room temperature
rpm	revolutions per minute
SAM	S-adenosyl methionine
ssRNA	single-stranded ribose nucleic acid
TC	transcription complex
µl	micro litre
VLP	virus-like particle
WT	wild type

Summary

African horse sickness virus (AHSV) affects equine populations around the world. It is the cause of a high rate of morbidity and associated large economic losses in affected regions. The virus is a segmented double stranded RNA virus and a member of *Orbivirus* genus in the Reoviridae family. The prototype member of the orbiviruses is bluetongue virus (BTV) and other members include Chuzan virus and St. Croix River virus. These viruses are all characterized by a genome of ten dsRNA segments that encode at least ten different proteins.

Three of the minor core proteins are found within the core of BTV. These are all associated with the RNA transcription complex and the enzymatic activities with which they are associated include an RNA polymerase (VP1), an RNA capping enzyme (VP4) and an RNA helicase (VP6). Genes homologous to the BTV genes that encode these proteins are found in all members of the *Orbivirus* genus.

The aim of this thesis is to characterize VP4 of AHSV, the capping enzyme candidate, and to compare it to other orbivirus capping enzymes. Possible functional motifs and regions of importance within the orbivirus capping enzymes will be identified. The gene will also be expressed and used to perform assays to characterize the different enzymatic activities of VP4.

The VP4 cDNA of AHSV serotype 3 was cloned and sequenced. From the full-length verified nucleotide sequence an open reading frame was identified and used to predict the amino acid sequence. These were compared to other orbivirus species including BTV, Chuzan virus and St. Croix River virus. These alignments identified a number of highly conserved regions, consisting of four or more amino acids conserved between all the sequences analyzed.

A fibronectin type 3-like motif, containing 12 conserved amino acids, was identified which could be responsible for protein binding. This motif contains 12 conserved amino acids making it a good candidate for a functional motif. Conservation does not, however, always predict regions of importance. In BTV a lysine-containing motif was identified to be responsible for GMP binding. This region is not conserved between the different viruses. AHSV has a motif containing a lysine residue similar to the motif identified in rotavirus and reovirus. Two other motifs described in BTV were also not conserved in the other viruses. One of them, a leucine zipper, was shown to dimerize BTV VP4. Phylogenetically, AHSV and Chuzan virus are the most

closely related while BTV is more distant and St. Croix River virus forms a distinct out-group when the different VP4 sequences are compared.

AHSV-3 VP4 was expressed as a histidine-tagged protein in the baculovirus expression system. Not unexpectedly, the protein was found to be insoluble, similar to BTV VP4 produced by means of the same system. However, whereas BTV VP4 could be solubilized by the addition of salt the AHSV VP4 remained insoluble at high salt concentrations. Several adjustments were made. Cells were lysed in a high salt buffer, the pH of the buffers was adjusted and sucrose cushions were used but none of the methods was found to improve the yield of soluble VP4 significantly. However, the pellet containing VP4 was relatively empty of contaminating protein and, therefore, a number of enzymatic assays were performed with the pellet. Assays for inorganic phosphatase and nucleotide phosphatase were performed. Strikingly, both assays indicated the presence of active phosphatases in the WT and VP4 pellets. Also, an assay was performed for guanylyltransferase activity but no activity was observed for this assay.

The sequence data therefore points to VP4 as the probable capping enzyme although it may have a different structural complex. The failure to produce a reliable source of soluble purified AHSV VP4 made it impossible to provide evidence to confirm the associated enzymatic activities.

Table of contents

Chapter 1.

Literature review: Structure of dsRNA virus cores, its components and RNA content.

1.1 General introduction

Viruses make out a large part of the eco system we encounter every day. They have been causing disease in all living organisms from plants to higher mammals with a variety of symptoms and severities. While human viruses affect our personal health and well being the viruses infecting plants and animals also have a large impact on human life as they can cause the loss of food through lost crop production or the loss of livestock. Some viruses are responsible not necessarily for the endangerment of human well-being but can still cause large economic effects by infecting and influencing an industry. An example of this is the horse racing industry in which millions of dollars are invested yearly.

The family *Reoviridae* consists of non-enveloped viruses with segmented dsRNA genomes of between 10 and 12 segments. There are nine genera within this family that can be distinguished on macro molecular, structural and genomic level. This family includes orthoreovirus (Reovirus type1), rotavirus (Human Rotavirus), orbivirus (Bluetongue virus), phytoreovirus (Rice dwarf virus), fijivirus (Fiji disease virus) and cypovirus (Cytoplasmic polyhedrosis). Within the orbivirus genus 21 serogroups have been identified each with a number of serotypes (table 1.1).

The Orbivirus genus of which Bluetongue virus (BTV) is the prototype consists of viruses that are characterized by ring-shaped capsomers on the surface of the core, which can be seen with negative staining under electronic microscopy (figure 1.1). These ring-shaped structures also lend the genus its name, with "Orbi" being the Latin word for ring. There are several other properties by which members of the orbivirus genus can be identified. The viral particle has a diameter of 65-80nm and a Mr of 60x 10^6 (Gould A.R., and Hyatt, 1994). These viruses are sensitive to low pH (Gould A.R. *et al.*, 1994). All orbivirus species have a double-layered shell protecting the RNA genome. The genome consists of ten segments of dsRNA of which one copy of each segment is found in the core.

Most research within this genus has been done on the prototype, Bluetongue virus, and on African horse sickness virus (AHSV). The viral proteins encoded by the genome can be divided into classes according to their localization and function. There are two outer capsid proteins that show high levels of variation between serotypes and even more so between serogroups. The core consists of two major proteins and three minor core proteins. The two major core proteins form the core shell and encapsulate the viral genome and the three minor core proteins. The three minor core proteins are involved in the enzymatic activities of the core. There are also three non-structural proteins that are not found inside the core of the virus but are expressed in infected cells.

Virus	**Number of serotypes**
African horse sickness virus	9
Bluetongue virus	24
Changuinola virus	12
Chenuda virus	7
Chobar Gorge virus	2
Corriparta virus	5
Epizootic hemorrhagic disease virus	5
Equine encephalosis virus	7
Eubenangee virus	4
Great Islands virus	36
Kemerovo virus	21
Lebombo virus	1
Leri virus	3
Orungu virus	4
Palyam virus	11
Umatilla virus	4
Wad Medani virus	2
Wallal virus	3
Warrego virus	3
Wongorr virus	8
Tentative and unassigned species	13 viruses

Table 1.1 The species within the Orbivirus genus with the number of serotypes shown for each species listed (Reoviridae study group, 1998).

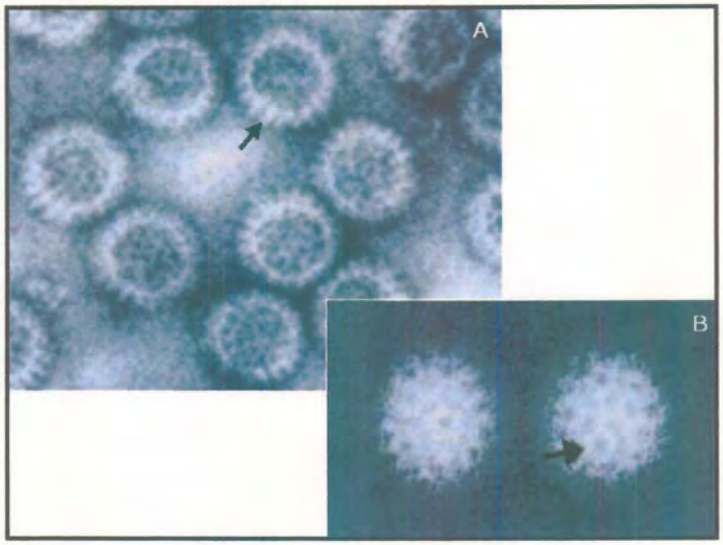


Figure 1.1 Negative staining of the core (A) and the virion (B) of BTV. Arrows indicate the VP7 trimers (A) on the core shell and the capsomers (B). Adapted from Roy P., 1996 and Gould A.R. and Hyatt, 1994.

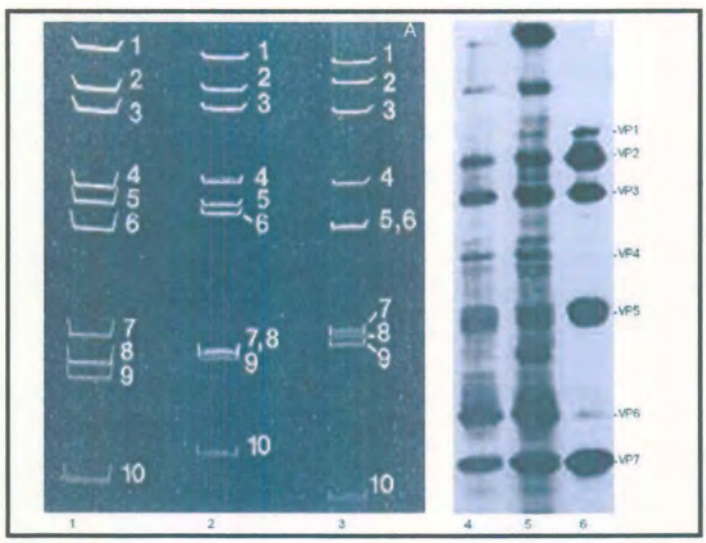


Figure 1.2 A SDS-PAGE gel of the dsRNA segments of BTV-10 (lane 1), EHDV-1 (lane 2) and AHSV-4 (lane 3). Adjacent a fluorograph of an 11% polyacrylamide gel with ^{35}S-methionine labeled proteins from AHSV-9 infected BHK cells (lane 4), cytoplasmic extracts from the infected cells (lane 5) and from the CsCl gradient isolated virus extracts (lane6). Adapted from Roy P. et al., 1994.Within the AHSV serogroup there are 9 serotypes (Roy P. et al., *1994) (table 1.1). Genes corresponding to the ten genome segments have been cloned and sequenced at least for one serotype.*

The genome of AHSV is divided in ten segments. Each segment encodes one mRNA strand that is translated by the cell's translation machinery into one protein. The genome segments are numbered according to each segments' mobility on a PAGE gel or in a CsCl gradient with the largest fragment numbered 1 and the smallest 10 (figure 1.2).

Serotypes differ mainly with respect to the serotype-specific epitopes located on outer capsid protein VP2. VP2 is encoded by large segment 2 of the viral genome. VP2 of AHSV-4 consists of 1060 amino acids and has a Mr of 124kDa (Roy P. et al., 1994)(table 1.2). The VP2 protein shows the highest level of variation in amino acid sequence between serotypes. The variability is mainly restricted to five specific regions while the carboxyl terminus is well conserved. The variability is probably due to the proteins' exposure to the environment, increasing the evolutionary pressure to be able to circumvent the immune response. Neutralizing epitopes have been identified on this protein coinciding with the regions of the protein exposed to the environment (Martinez-Torrecuadrada J.L., *et al.*, 2001; Bentley L., *et al.*, 2000; Burrage T.G. *et al.*, 1993).

VP5 is the second component of the outer shell. It has a relatively high level of variation but much less than VP2. This can be attributed to the fact that VP5 is more protected from the external environment than VP2. VP5 is encoded by medium segment 6 (table 1.2) and has a size of Mr 56,7kDa (Roy P. *et al.*, 1994). VP5 contains two hydrophobic regions that are thought to interact with the core shell protein, VP7, but have now been shown to be essential for anchorage of the protein to the membrane giving it a possible role in cell penetration (Hassan S.H., *et al.*, 2001). The sequence of VP5 also shows highly variable regions while regions like 200-270 and both termini are highly conserved (Roy P. *et al.*, 1994).

The core shell is a double layered sphere and made up of two proteins, VP3 and VP7. BTV VP3 and VP7 self-assemble into a subcore when expressed in the baculovirus expression system (French T.J., and Roy, 1990). VP3 is a hydrophobic protein with 120 copies present in the subcore layer (table 1.2). Subcore particles are similar in size to cores found *in vivo* although they are not as stable as the normal cores. Subcores may be seen as the minimal stable virus particle. This is underlined by the observation that cypovirus cores only consist of a single protein layer of 120 copies of VP3 analogs (Hill C.L. *et al.*, 1999). VP3 is coded for by large segment 3 and consists of 905 amino acids with a Mr of 103.2kDa (Roy P. *et al.*, 1994)(table

1.2). VP3 is a highly conserved protein between AHSV serotypes but also between different serogroups like BTV, EDHV and AHSV. Most core proteins are similarly conserved. The reason for this level of conservation may be the absence of selection pressure due to the fact that these proteins are not exposed to the immune competent environment outside the infected cell.

VP7 is the second major component of the core shell. The small segment 7 codes for VP7. VP7 forms capsomers that are the structures visible with negative staining under electron microscopy (Roy P. *et al.*, 1994) (figure 1.1). These capsomers are hexameric rings consisting of 4 trimers of VP7. VP7 forms the outer layer of the core shell and is attached to the VP3 lattice. There are 780 copies of VP7 found in the core. The protein comprises 354 amino acids with a molecular weight of 38.1kDa. VP7, like VP3, is hydrophobic and rich in alanine, methionine and proline (Roy P. *et al.*, 1994)(table 1.2). VP7 has a lower level of conservation than VP3 within serogroups and between serogroups within the orbivirus genus.

Segment	*Basepairs*	*Protein name*	*Size kD*	*Amino acids*	*Location*
1	3965	VP1	150	*1305*	Minor core prot.
2	3229	VP2	124	1060	Outer shell
3	2792	VP3	103	905	Major core prot.
4	1978	VP4	75	642	Minor core prot.
5	1751	NS1	63	548	Non-structural prot.
6	1566	VP5	56	505	Outer shell
7	1179	VP7	38	354	Major core prot.
8	1160	NS2	41	365	Non-structural prot.
9	1169	VP6	42	369	Minor core prot.
10	756	NS3/3A	23/22	217/206	Non-structural prot.

Table 1.2 Genome segments, size, protein names, protein size, and location in the virion of the proteins of AHSV.

Three minor core proteins are present within the core shell called VP1, VP4 and VP6. The minor core proteins have been an enigma for a long time and only recently have researchers been able to characterize their roles in virus replication. There is also as yet very little sequence data available that will allow us to compare the variation amongst different AHSV serotypes. A characteristic of all three minor core proteins

is that they are present in the virus particle in very low copy numbers as compared to the major structural proteins. Seen the small number of copies it seems unlikely that these proteins play a structural role but rather may have an enzymatic role within the viral core.

AHSV VP1 is encoded by segment 1, the largest of the genome segments, and is 3965bp in length that encodes a protein with a molecular weight of 150kDa (table 1.2). There are 12 copies of VP1 within the core. VP1 was shown to be the RNA polymerase in BTV (Urakawa T. *et al.*, 1989) and is able to catalyze the transcription and replication of RNA. In AHSV, VP1 codes for the polymerase as has been shown by Vreede F.T. and Huismans H., 1998.

The second minor core protein, VP4, is encoded by segment 4. There are 24 copies of VP4 in the core. VP4 has a size of 75.4 kDa (Roy P. *et al.*, 1994) (table 1.2). VP4 is responsible for capping nascent mRNA in BTV and this function is also though to be the responsibility of AHSV VP4. The capping of the nascent RNA includes the addition of a guanylyl moiety to the 5' end of the ssRNA segment. Capping of nascent RNA requires a guanylyltransferase and a methyltransferase catalytic domain. BTV VP4 was shown to be able to catalyze the dephosphorilation of the 5' terminal nucleotide and the transfer of a guanylyl moiety to the 5' end *in vitro* after expression in the baculovirus expression system (Ramadevi N. *et al.*, 1998a).

The third protein found in the core is VP6. Its full function is not yet fully understood. There are 72 copies of VP6 present in the core of BTV and AHSV with a size of 35.8kDa (Roy P. *et al.*, 1994) (table 1.2). The BTV VP6 protein has been shown to be a helicase and also has RNA-dependent ATPase activity (Stauber N., *et al.*, 1997).

All three components of the interior, VP1, VP4 and VP6 are found in multiples of 12. There are 12 copies of VP1, 24 copies of VP4 and 72 copies of VP6 in the BTV core (Stuart D.I., *et al.*, 1998). It can be speculated that twelve transcription complexes are formed inside the core consisting of 1 polymerase, 2 capping enzymes and 6 helicases.

The AHSV genome has three segments that code for four non-structural proteins. They are NS1, NS2, NS3 and NS3A. These are proteins not found in the virion, but only in the cytoplasm of the infected cell. Therefore, these proteins are not needed

for the virus structure but may assist in other aspects of the virus life cycle in the infected cell. Two of the four non-structural proteins (NS3 and NS3A) are coded for by the same dsRNA segment. They are produced by alternative initiation codons in the genome segment 10. Two virus-related structures are observed in virus-infected cells, tubules and virus inclusion bodies (VIB's). BTV and AHSV NS1 are responsible for the formation of tubules in the infected cell (Maree F.F., and Huismans H., 1997). The tubules consist of dimers of NS1 bound together in ribbons. NS1 is encoded by the medium segment 5 and is 548 amino acids long with a size of Mr of 65kDa (Roy P. *et al.,* 1994 and ref herein)(table 1.2). The sequence is highly conserved within serogroups while to a lesser extend between serogroups. Several cysteine residues are conserved across the serogroups and were shown to be essential for tubule formation in BTV (Monastyrskaya K. *et al.,* 1994). The small segment 8 codes for NS2. It is the only protein in BTV and AHSV that is phosphorylated (Horscroft N.J. and Roy P., 2000; Theron J., *et al.*, 1994). NS2 can bind ssRNA and shares this function with proteins such as σNS of Reovirus and NS3 of rotavirus. Its ssRNA binding ability may involve it in the targeting of the newly synthesized ssRNA from the nucleus to the newly formed cores (Horscroft N.J., and Roy P., 2000; Theron J., *et al.*, 1996). BTV NS2 was shown to form the viral inclusion body protein. NS2 in AHSV has a similar function (Horscroft N.J., and Roy P., 2000). Not only can it bind ssRNA; NS2 can also bind ATP and GTP and hydrolyze them to their respective NMP although NS2 has a higher affinity for ATP (Horscroft N.J., and Roy P., 2000). The protein consists of 365 amino acids and has a size of MW 41.1kDa (Roy P. *et al.,* 1994)(table 1.2).

The smallest segment, segment 10, codes for NS3 and NS3A. This gene is the second most variable after VP2. These proteins are formed by the use of alternative initiation codons and result in the expression of NS3 and NS3A, respectively 24 and 23 kDa in size (Van Niekerk M., *et al.*, 2003; van Niekerk M., *et al.*, 2001a; Roy P. *et al.,* 1994)(table 1.2). In BTV these two proteins are glycosylated (Wu X., *et al.*, 1992) and are involved in virus release (Beaton A.R., *et al.*, 2002). NS3 is associated with the membrane in which two hydrophobic regions in the protein are essential for anchoring. Targeting of the protein to the membrane is the function of another region. NS3 is important for virus release and therefore, for cytotoxicity (Van Niekerk M., *et al.*, 2001b; van Staden V., *et al.*, 1998).

An interesting feature of viruses within this family is that even though there may not be a high degree of sequence similarity of proteins, they show a high level of

similarity in their structure and function. This similarity indicates that these viruses may have converged through evolution to obtain the same functional and structural end product.

1.2 Infection cycle of AHSV

All viruses have a strategy to enter the cell to produce progeny and allow them to exit the infected cell. In figure 1.3 the infection cycle of viruses like BTV and AHSV is depicted. Viruses with a bi-layered shell like BTV and AHSV loose their outer shell when entering a cell. The core exists as an autonomous entity within the cytoplasm, able to transcribe its genome and produce mRNA for translation by the cellular machinery. The outer shell attaches to the cell surface via receptors located across the cell surface. In many cases these receptors are integrins. They are transmembrane proteins found as hetero dimers associated with the extra-cellular matrix on the outside of the cell while coupled to the cytoskeleton on the interior of the cell (Eaton B.T. *et al.,* 1987)(figure 1.3b). After the virus attaches to the cell via interaction of the outer capsid protein VP2 with the receptors on the cellular membrane it enters the cell through receptor-mediated endocytosis. Endocytotic vesicles are formed and these transport the virus via the cytoskeleton to a position close to the nucleus (Hyatt A.D. *et al.,* 1989; Gould A.R., and Hyatt A.D., 1994)(figure 1.3c+d). The virus may direct the movement of the vesicles to the nucleus. The vesicles fuse at this stage to form endosomes (Gould A.R., and Hyatt A.D., 1994). Additionally, VP5 has been shown to be essential for membrane puncturing at the site of virus attachment and entry while VP5 cannot by itself enter the cell (Hassan S.H., *et al.*, 2001). Endosomes have a more acidic environment that causes the disruption and removal of the outer shell. The removal of the outer shell seems to be a crucial step in the virus life cycle as it is likely to induce the activation of the core to start transcribing the genome once the core enters the cytoplasm (van Dijk A.A., and Huismans H., 1980). This may be achieved through the opening of the three-fold channels in the core shell, allowing nucleotides and other components necessary for transcription to enter the core where they are needed for transcription. The core pierces the endosome, possibly with the help of VP5, and enters the cytoplasm. Transcription commences and all ten segments are transcribed simultaneously but at different rates (figure 1.3e). The mRNA transcripts leave the core through the five-fold channels and move to the endoplasmic reticulum and the ribosomes where they are translated by the cell. Whether the cell is induced to preferentially translate viral transcripts or if viral mRNA has an inherent ability to induce higher translation levels

is not known yet. Viral inclusion bodies appear along with viral tubules (figure 1.3f). The viral inclusion body (VIB) protein, NS2, facilitates this matrix formation. These VIB's facilitate the assembly of new cores and the packaging of its genome before being formed into mature virions. The VIB contains a range of virus intermediates including core-like particles, cores and virus-like particles indicating that virus assembly takes place here (Eaton B.T. *et al.*, 1987).

Time-lapse studies of the content of VIB's in BTV infected cells have shown that over time subcores develop into core-like particles after which they develop into cores and finally into fully developed double-layered virions (Brookes S.M. *et al.*, 1993). These data support a role for NS2 and the VIB as the virus assembly center (Gould A.R., and Hyatt A.D., 1994 and ref. herein). After the virion is formed it is released from the VIB and associates with the intermediate filaments, a component of the cytoskeleton, which facilitates its transport to the cell membrane (Eaton B.T. *et al.*, 1987). Another form of transport within the cell was also observed. NS3A forms smooth vesicles that contain the virions ready for release. These two forms of transport of the virion in the cell are seen in separate serogroups (Gould A.R., and Hyatt A.D., 1994). Viruses are released from the cell either through budding or by extrusion. These two forms of virus release using the NS3 vesicle leave the cell membrane uninterrupted thus keeping the infection hidden from the immune system (van Staden V. *et al.*, 1998).

1.3 Structure of the core

Detailed knowledge of the core structure using crystallography has come to light with the improvement of resolution using cryo-electron microscopy and intensive diffraction analysis. Two viruses have been resolved in this way, BTV and Reovirus (figure 1.4). These two viruses represent two groups within the Reoviridae family as will be discussed later. The two studies resolved the core structure including partial genomic organization up to a resolution of 3.6Å.

The core exists as an autonomous particle within the cell and transcribes its own genome. The structure of the core is important to its function and several facets aid it in its role in genome transcription and replication. The core may also guide the packaging of the genome. This role of the core is currently under investigation. As previously discussed the two major proteins of the core are VP3 and VP7. Together

they form a flexible but strong shell. The BTV and Reovirus cores were crystallized and the data used to construct the molecular structure of the cores (Grimes J.M. *et al.*, 1998; Reinisch K.M. *et al.*, 2000) (figure 1.4). The structure of the core has direct bearing on its function as replication and transcription complex and this can be clearly seen from the data presented.

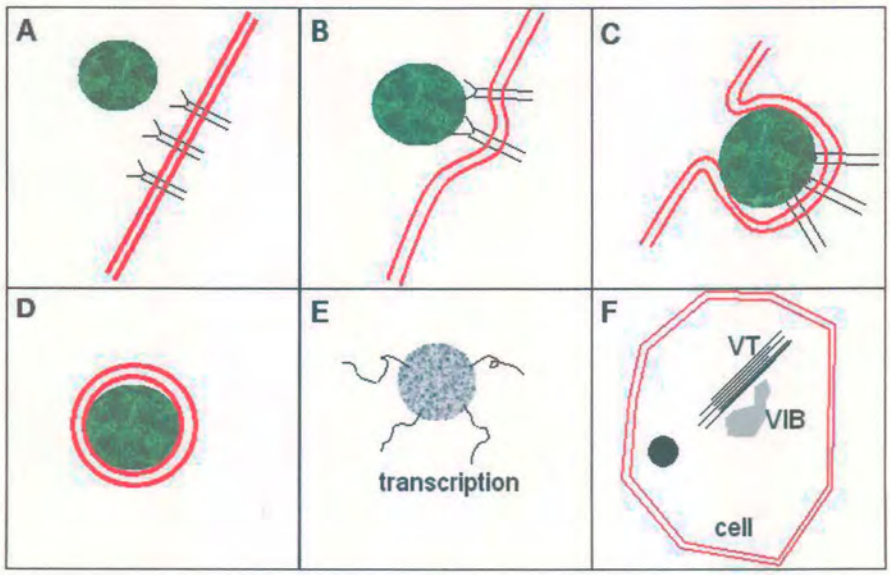


Figure 1.3 A schematic representation of the infection cycle of orbivirus members. Virions recognize ligand receptors on the target cell (A) and adhere to them. The virion attaches to the cell membrane through cell receptors like the integrins that bind a RGD motif found on many virus coats (A). The membrane bulges in (C) and the cell is endocytozed in a clathrin-coated pit. The endosome provides the correct environment for the virion to shed off its outer shell (D). The core becomes active and transcribes the mRNA for the necessary proteins (E), like the non-structural proteins that assist in making the viral tubules and the viral inclusion body (F).

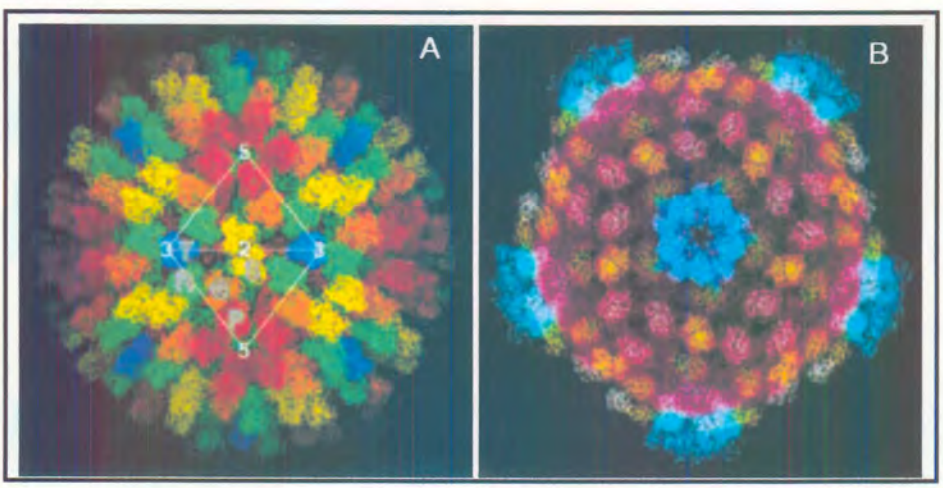


Figure 1.4. Overall core structure of the BTV and Reovirus core. The BTV core (A) has the VP7 trimers on the shell surface. The Reovirus core (B) has turrets that extent from the five-fold axis. Adapted from Grimes J.M. et al., 1998 and Reinisch K.M., 2000.

BTV and Reovirus represent two groups of viruses within the Family Reoviridae. BTV represents a group of viruses that have an icosahedral core with no turrets and a VP7 shell with T=13 symmetry with no enzymatic action (figure 1.4a). The enzymatic function lies within the subcore made up of VP3 with a T=2 symmetry (Grimes J.M. *et al.,* 1998). Reovirus, on the other hand, presents a group of viruses that have single core shells with turrets at the five-fold axis that contain enzymatic functions (figure1.4b). The core of these shells cannot self-assemble like the VP3 subcore and need a clamp subunit (Reinisch K.M. *et al.,* 2000). Except for the atomic structure of the two cores some data is also available for other serogroups like orthoreovirus, Broadhaven virus, fungal L-A virus, rotavirus and cypovirus. Earlier structural studies mostly included the expression of the different proteins in expression systems and studying their organization in solution. Early studies showed that BTV VP3 forms a subcore in solution without any assistance (French T.J. and Roy P., 1990). This gives an indication that within the cell VP3 protein in high concentration may already form the first structures needed for virus formation without the help of other proteins. This indicates the VP3 may act as the scaffold protein for the rest of the virion to be built around.

1.4 The BTV core

Since there is a high degree of similarity between BTV and AHSV, data from the atomic structure of BTV can be extrapolated to AHSV. The core was first visible with negative staining under the electron microscope and had icosahedral morphology. This was in contrast to the outer shell that had an obscure morphology. The diameter of the core was determined to be 700Å (Grimes J.M. *et al.,* 1998).

The subcore of BTV consists of VP3. VP3 forms a T=2 shell (two copies of VP3 per icosahedral subunit) of VP3 monomers found in two conformations, VP3A and VP3B (Grimes J.M. *et al.,* 1998) (figure 1.5). There are 120 copies of VP3. This is similar to many other dsRNA viruses like ϕ6 bacteriophage, reovirus, rotavirus, orthoreovirus and cypovirus. 120 units of VP3 is an unusual number for an icosahedral structure because it is impossible for 120 equal units to form a closed icosahedral structure.

Interestingly, the inner core proteins also combine into a complex formation of which twelve complexes can be formed in the core from the three minor core proteins.

Conformational switching allows a protein to move between two conformations to attach to different parts of the other protein. Conformational switching would allow for an icosahedral structure to be formed by 120 subunits. Both VP3 conformations have a triangular wedge shape. Both types have three domains called the apical, carapace and dimerization domain as seen in figure 1.5. Their positions and their secondary structures distinguish the domains from one another. The largest conformational difference between the two subtypes of VP3 is found in the dimerization domain. Small conformational changes in this domain cause a large overall change in conformation of 20° in the area at the interface of the quasi-two fold axis between the A and B subunit of two opposite icosahedral subunits (Grimes J.M. *et al.,* 1998)(figure 1.5). The difference in the conformation of VP3A and VP3B allows them to form a tight structure of ten copies around each five-fold axis shown in figure 1.6. This decamer may be seen as an early intermediate in the core assembly. Twelve such decamers will make a complete core shell. The assembly through these intermediates is likely as the associations between the subunits at the five-fold axis are the strongest while the associations furthest from the five-fold axis are the weakest. The inner wall of the VP3 core has shallow grooves that can bind dsRNA with a low affinity (Prasad B.V. *et al.,* 1996). The core is not completely closed-off. Two types of channels have been identified passing through the core shell, one at the three-fold and the other at the five-fold axes (Grimes J.M. *et al.,* 1998)(figure 1.7).

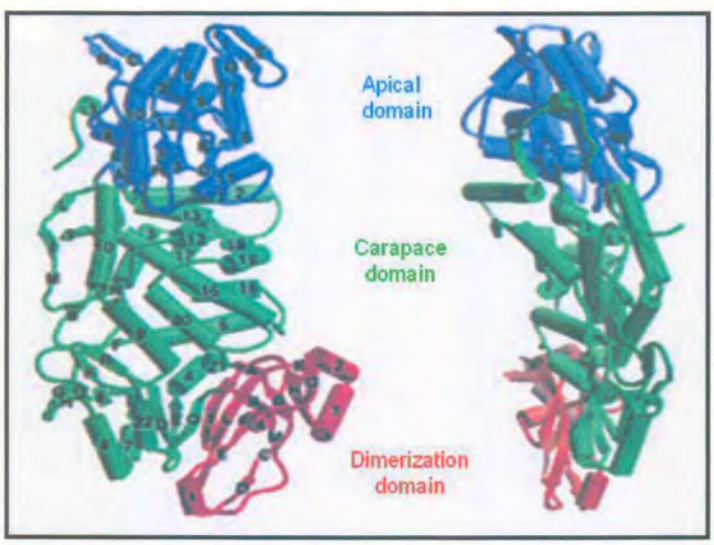


Figure 1.5. Structure of VP3 indicating the different domains it consists of. Adapted from Grimes J.M et al., 1998.

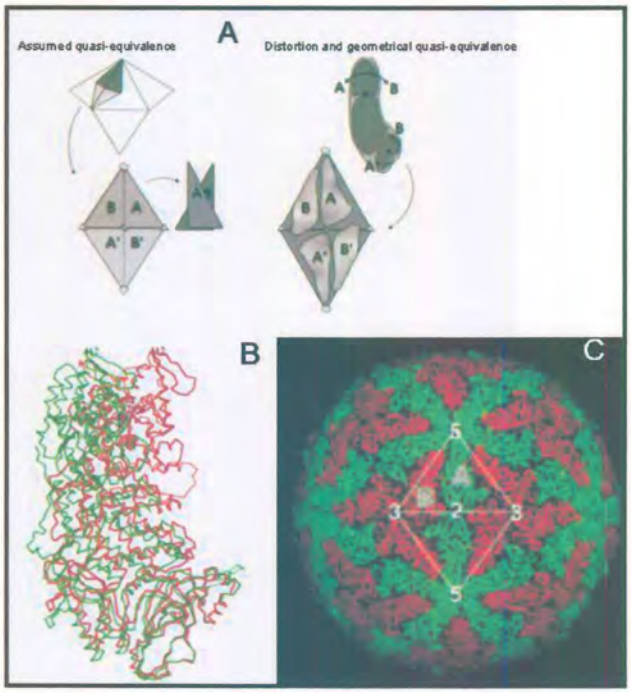


Figure 1.6. Schematic representation of the VP3 copies in an icosahedral subunit showing how the dimers are formed into a tight structure using the assumption of more extensive deformity in the VP3 structure between the A and B conformation (A). A string overlay of VP3A and VP3B shows the large amount of flexibility between the two conformations due to the flexible carapace domain (B). The icosahedral subunit is shown in context of the core shell in (C) to illustrate the effectiveness of the geometrical quasi equivalence of the subunits. Adapted from Grimes J.M. et al., 1998.

In inactive cores neither channel is big enough for ssRNA to pass through. The three-fold channel has an arginine side chain that could turn away to form a 15Å diameter channel that would be big enough. However, no data on the existence of this enlarged channel has been published and from structural studies it is evident that the VP7 T-trimer plugs this channel tightly (Grimes J.M. *et al.*, 1998).

The five-fold channel has a diameter of 9Å that is also too small to allow entry or exit of nucleotides and ssRNA. It may be assumed that the core is activated and that activation causes enlargement of the five-fold channel to allow exit of mRNA (figure 1.7). This conformational change in the protein around the channel is probable, as the structure is very flexible in the area around the 5-fold axis. This process could be

fully reversible allowing the opening and closing of the channels as needed by the transcription complex. Four arginine residues from each five-fold axis related VP3A of which three are strictly conserved in BTV (308, 317, 413), AHSV, and EHDV are found inside the five-fold channel (Grimes J.M. *et al.,* 1998). These residues give the five-fold channel an overall positive charge that may be involved in the directional movement of RNA segments out of the core. These residues may also be involved in the stripping of counter ions from the RNA entering the newly formed core.

In contrast to the bi-layered BTV and AHSV cores (VP3+ VP7), the Reovirus has a single layered core. The core comprises of three proteins, $\lambda 1$, $\lambda 2$ and $\sigma 2$. $\lambda 1$, the homologue of VP3 cannot form a subcore. For this $\sigma 2$ is needed. $\sigma 2$ subunits act as a clamp and keeps $\lambda 1$ subunits together (Reinisch K.M. *et al.,* 2000)(figure 1.8). $\lambda 2$ forms turrets over the five-fold axes, that protrude outwards from the core. These channels have enzymatic activity in the form of capping activity (Mao Z. *et al.,* 1991; Reinisch K.M. *et al.,* 2000). $\lambda 1$ is, like BTV VP3, composed of two conformational types. Additionally, $\lambda 1$ also forms decamers around each five-fold axis with equal copies of type A and B. The difference between the two subtypes (A+B) is defined by a shift of two sub domains relative to each other. Similar to BTV cores, the five-fold channels are blocked by a peripheral loop. This loop bends out of the way during active transcription opening the channel for exit of mRNA (Reinisch K.M. *et al.,* 2000). This observation is supported by the fact that the loop has a high degree of flexibility. In comparison with BTV, only the dimerization domain of Reovirus, where two subunits of adjacent decamers meet, have similar contacts (Reinisch K.M. *et al.,* 2000).

The $\sigma 2$ clamp is not present in BTV cores. $\sigma 2$ is a globular monomer that binds to three distinct sites within the icosahedral asymmetric unit. There are 150 copies of $\sigma 2$. Its clamp function is clearly shown by the binding domain situated over the middle of the one $\lambda 1$B subunit of one decamer and extending to the same region of an adjacent $\lambda 1$B subunit of a different decamer (figure 1.8). This connects the two decamers and allows for a stable sub-core (Reinisch K.M. *et al.,* 2000). As another example of quasi-equivalence, the same binding sites of $\sigma 2$ form different connections with the $\lambda 1$ subunits. This theme is found in many other viruses. This may indicate a method to save space by minimizing the need for different proteins. A characteristic of the viruses belonging to the group represented by Reovirus is the

presence of turrets at the 5-fold axis. While BTV and others have a second core layer, Reovirus and the rest of the group only have a single core layer with turrets.

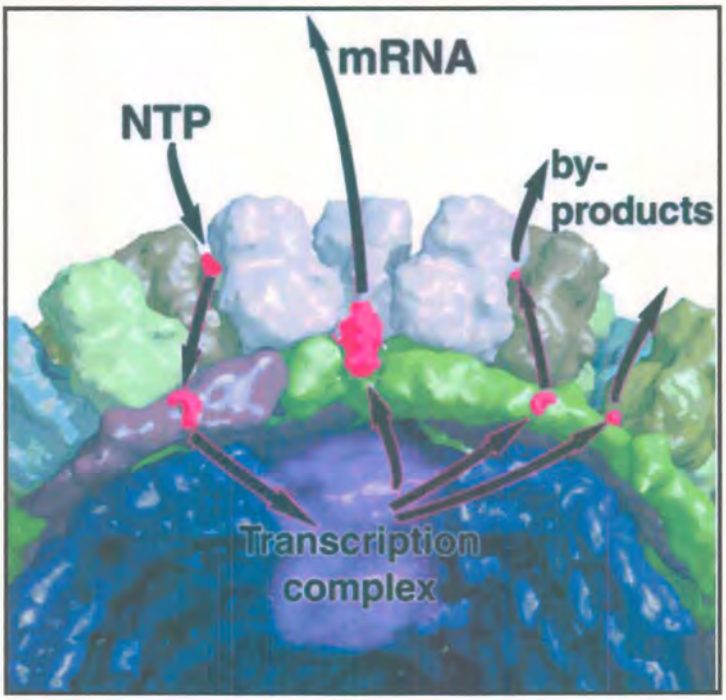


Figure 1.7. Schematic drawing of the pores that open into the core cavity. The pore at the five-fold axis is for the exit of the mRNA, while additional pores are necessary for the import of the nucleotides and ions needed for transcription. The by-products are transported out of the core through another pore. These pores are most likely opened upon activation of the core by the shedding of the outer shell. Adapted from Diprose J.M. et al., 2001.

The turrets of Reovirus consist of λ2. Five copies of the protein make up a turret. The turret has a five-part flap that acts as a lid at the extremity of the turret. The hollow cylinder can accommodate up to 300 nucleotides of RNA. The pentamer binds directly to the λ1A subunits around the axis. The turrets are enzymatically active with seven domains lining the inside of the cylinder. The first 385 residues of the turret form a cup-like structure that was identified as the guanylyl transferase activity with the essential lysine (lys) at position 190 (Reinisch K.M. *et al.*, 2000). The essential lysine residue needed for GMP transfer, Lys-190, projects into the turret. A side channel into the tunnel allows entry of nucleotides into the central cavity. The

guanylyl transferase domain is sequentially first from the core. Downstream from this domain is the methylase domain. Further into the turret a second methyltransferase is found responsible for the subsequent methylation of the 5' end of the mRNA.

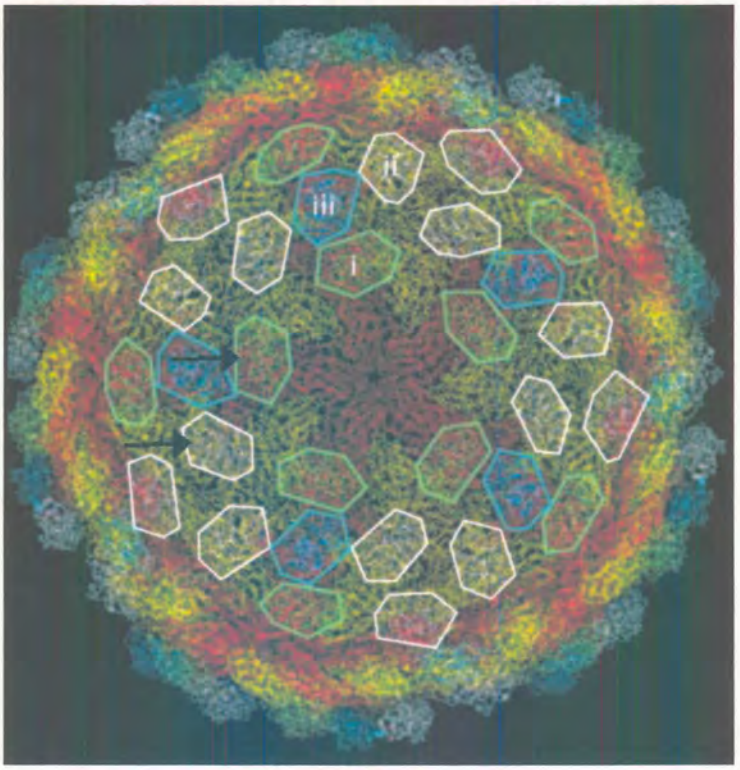


Figure 1.8. The Reovirus core, depicting the σ2 clamp on the shell surface (arrows). The five-fold axis pore is seen in the middle. Adapted from Reinisch K.M., 2000.

The second major core protein found in BTV and AHSV is VP7. VP7 exhibits more sequence variation than VP3 between serotypes. This may be due to more exposure to selective pressures from the cell environment. A feature unique to AHSV is the ability of VP7 to form flat crystalline structures in solution (Chuma T. *et al.*, 1992). Whether this ability has any structural or functional relevance is not known at this stage. It is, however, interesting to note that BTV, EDHV and AHSV VP7 are hydrophobic and that in the protein sequence of BTV and EDHV there is only one lysine residue found on residue 255 (Roy P. *et al.*, 1994). This residue is replaced by an arginine in AHSV. It is not known whether this residue is responsible for the AHSV protein's ability to form the structures. It is known that the mutation of this lysine residue to a leucine in BTV prevents the formation of core-like particles, indicating the importance of the residue in BTV (Roy P., 1996). VP7 is found in a T=13 conformation on the VP3 sub-core. One icosahedral subunit consists of 4 trimers and one monomer. The trimers have general positions on the subunit while

the monomer's molecular 3-fold axis extends along the core icosahedral 3-fold axis. Although the trimers are equivalent, they have different interactions with VP3 (Roy P., 1996). In total there are 13 sets of distinct contacts between the VP7 and VP3 copies in one icosahedral subunit. The four trimers are virtually identical while making distinct contacts with the VP3A and VP3B molecules. This is indeed another example of quasi equivalence of the core. It seems that there is a hierarchy of interactions of the different trimers with the VP3 molecules. This is substantiated by the observation that the T trimer closest to the five fold axis has the strongest interaction with the sub-core while the trimers furthest away from the axis has the least favorable contacts. The interaction between VP7 and VP3 is contained in an area between residue 336 and 349 of VP7 as shown by mutation and truncation studies (Le Blois H., and Roy P., 1993; Roy P., 1996; Grimes J.M. *et al.*, 1998). This area is highly hydrophobic and forms flat surfaces. The ability of VP7 to form trimers in solution indicates that the core may form through crystallization of the VP7 trimers on the sub-core shell.

In conclusion, it is interesting to note that ϕ6 bacteriophage shows many similarities with BTV as well as with Reovirus and L-A virus. All have a major shell protein of 120 copies surrounding a transcriptionally active interior and show flexibility around the five-fold axis. (Butcher S.J. *et al.*, 1997) This indicates that the basic structure of the icosahedral core with the enzymatic proteins inside the core may serve as a basic template for a wide variety of viruses.

1.5 Inside the core

The cores of all the Reoviridae members investigated so far remain intact after cell entrance to protect the dsRNA genome against degradation from cellular endonucleases. The core was shown to be transcriptionally active inside the cell extruding mature mRNA from its pores that won't attract cellular defense mechanisms. Butell *et al*, 1974 showed that multiple RNA's are extruded from the intact core indicating simultaneous transcription. This data indicated that transcription of the genome takes place inside the core and, therefore, that the machinery has to be present inside the core. The three proteins present in the BTV and AHSV core are VP1, VP4 and VP6. Functions were assigned to these proteins but have not been proven in all viruses. In BTV, the VP1 component is the polymerase (Rou P., *et al.*, 1988); the VP4 is the capping enzyme (Mertens P.P.C., *et al.*, 1992), while VP6 is a helicase (Stauber N., *et al*, 1997).

Only lately has the organization of the inner core proteins become known. There are twelve copies of the RNA polymerase, VP1, in the core or alternatively, one copy for every five-fold axis. Also, the turrets of the turreted Reoviridae viruses have RNA capping activity, guanylyl transferase and methylase functions. Core proteins, like λ2 and λ3 from orthoreovirus, interact in solution with each other indicating their ability to form complexes (Dryden K.A. *et al.,* 1998). These observations may indicate their association as a complex inside cores. There are twelve polymerase units that can associate with each five-fold axis and there are no more than twelve genome segments. This leads to the conclusion that each genome segment is associated with a single transcription complex. High-resolution cryo-electron microscopy confirmed this hypothesis of a single transcription complex at the five-fold axis associated with a single genome fragment (Gouet P. *et al.,* 1999) (figure 1.9).

The transcription complex (TC) may also be involved in the organization of the RNA genome. Orthoreovirus cores deficient in dsRNA provided a view into the core cavity. In a cut-away of the core a channel can be seen passing through the core shell.

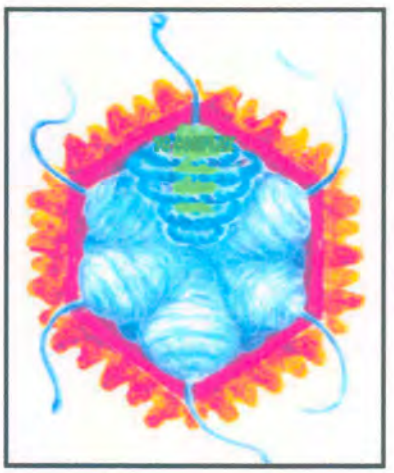


Figure 1.9. The association of TC complex with individual dsRNA segments in the core. A single dsRNA segment in blue is wound around the TC complex extending from the five-fold axis. Adapted from Gouet P. et al., 1999.

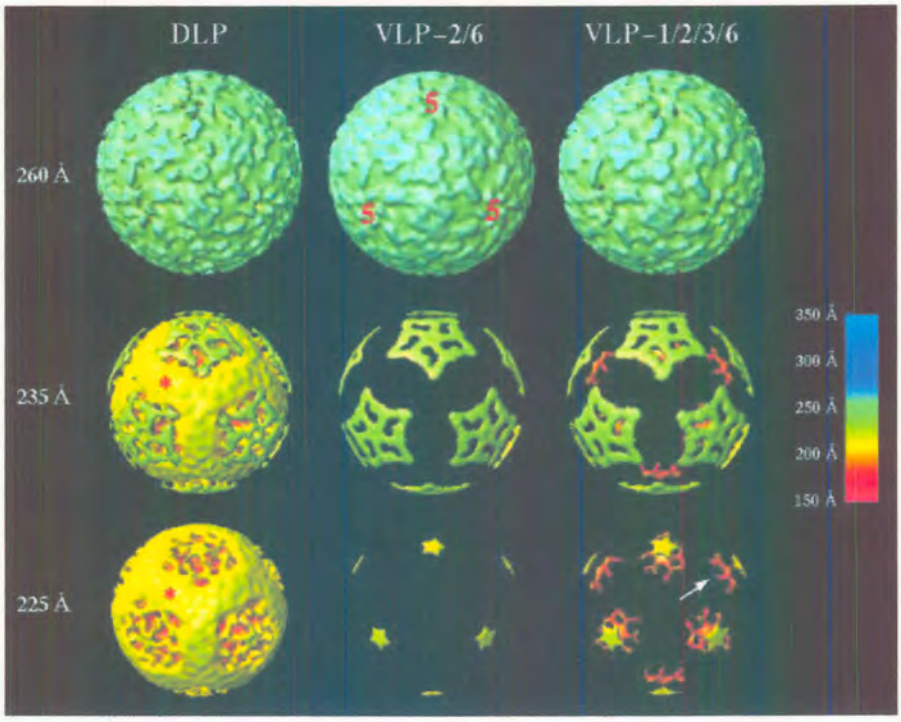


Figure 1.10. The structure of different rotavirus core-like particles at different radii showing the flower-like structures extending into the core cavity (arrow). Adapted from Prasad et al., 1996.

At the site of the channel inside the core, a structure protrudes into the cavity along the five-fold axis. The structure has a flower-like appearance with five petals attached to the tip of a stalk (Dryden K.A. *et al.,* 1998)(figure 1.10).

Surprisingly, similar structures were observed in BTV and rotavirus even though they have a capping enzyme as part of their TC. Co-expression of the rotavirus core proteins along with VP1 and VP3, the polymerase and capping enzyme, result in a core with flower-like structures at the five-fold axis inside their cores. The stalk length differs between BTV and Reovirus by about 30Å. This difference could account for the addition of the capping enzyme (Hill C.L. *et al.,* 1999). The flower-like structures are only those parts of the TC that have icosahedral symmetry and are therefore visible. Present studies can only be done with icosahedral averaging not allowing for structures with different symmetries to be visible. The structure seen only accounts for 25% of the total mass of the TC complex if normal packing densities are assumed (Gouet P. *et al.,* 1999). In this study, the structure of the TC complex was examined under high-resolution microscopy. The capping enzyme, VP4, lies closest to the inner wall of the core with the polymerase more into the core cavity (figure 1.11). This conformation would allow sequential transcription and

capping of the genome with the newly synthesized mRNA being expelled from the core through the channel. This is in accordance with the turreted viruses that carry the capping activity in the turret and the polymerase just inside the core.

Another feature found in all *Reoviridae* viral cores is that the core shell protein C-terminal extends into the cavity with ~50 amino acids around the five-fold axis. Five such tails extend into the core at the five-fold axis (Grimes J.M. *et al.*, 1998)(figure 1.11). They may act as nucleation points for the TC complex and help keep it in the correct orientation. In orthoreovirus, chymotrypsin treatment of the core results in cleavage of the tail of the core shell protein extending into the core cavity (Dryden K.A. *et al.*, 1998). The fixation of the TC complex at the five-fold axis underwrites the moving-template hypothesis for the transcription of the genome segments as will be discussed later.

Less ordered densities are also observed at the three-fold axis. What these are is not known yet. The first observations of the inside of the rotavirus DLP indicated that at least part of the genome is organized into icosahedral symmetry (Prasad B.V. *et al.*, 1996)(figure 1.12).

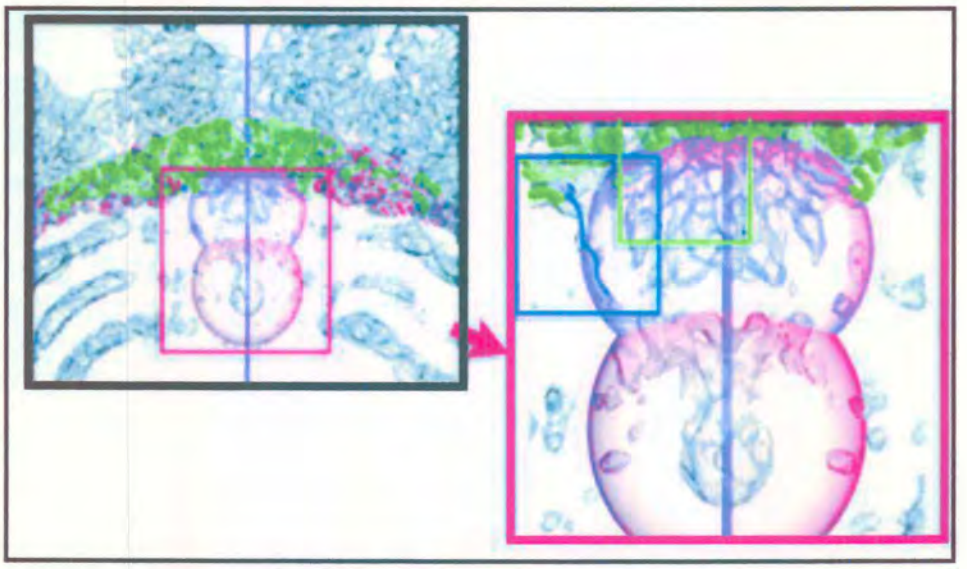


Figure 1.11. The TC complex extends along the five-fold axis into the core cavity. RNA is wound around the complex. The C-terminal of VP3 extends into the core possibly anchoring the complex to the inner wall of the core. Adapted from Gouet P. et al., 1999.

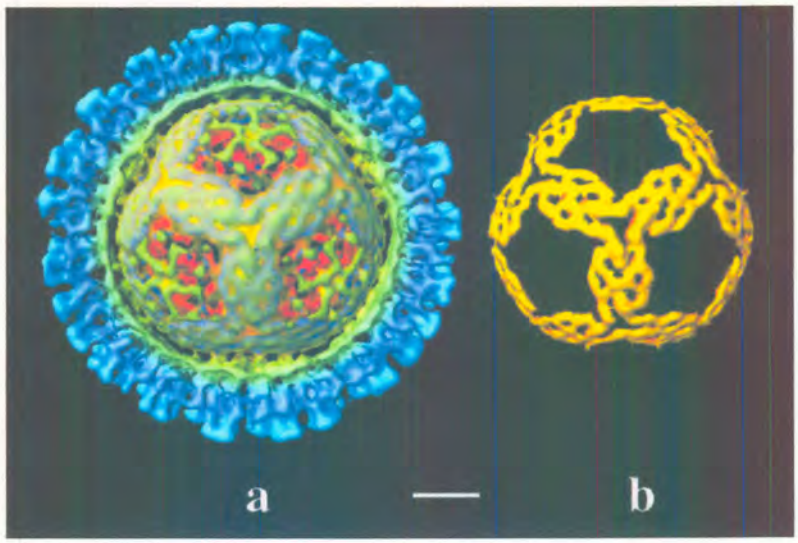


Figure 1.12. The structural organization of RNA within the Rotavirus core. The gray shaded areas in a represent the outer ordered layer of RNA that is showed alone in b for clarity. The flower-like structure is also seen in green around the five-fold axis. Adapted from Prasad et al., 1996.

High densities were observed at 230Å radii below the core shell in the RNA-containing DLP while this density was absent in the RNA-free VLPs of rotavirus. At high resolution this density consists of parallel tube-like strands 30 Å apart (Prasad B.V. *et al.*, 1996). With these images 25% of the genome was visualized (figure 1.12).

Icosahedral ordering of parts of the genome was also observed in BTV and cypovirus (Gouet P. *et al.*, 1999; Hill C.L. *et al.*, 1999). High-resolution images and refined averaging mathematics allowed 80% or 19219bp of the BTV genome to be visualized (Gouet P. *et al.*, 1999). dsRNA has no natural tendency to conform to icosahedral symmetry. Therefore, it is assumed that the dsRNA has to be forced into this symmetry by a template. The inner side of the core shell is chemically featureless with shallow grooves. The core shell has some points of low affinity dsRNA binding that would be able to implement its own icosahedral order on at least the outer layer of RNA in the core (Prasad B.V. *et al.*, 1996)(figure 1.12). There are only a few points of RNA/protein interactions possible. In BTV, a RNA binding motif is found on the $\beta J/\beta K$ loop of the VP3 unit at residues 790-817. Only one residue, Lys-807, is present in this loop while the rest of this motif has no charge (Grimes J.M. *et al.*, 1998). The presence of very few interaction points combined with a chemically inert

groove would allow dsRNA to be packed along the grooves while allowing movement of the RNA, possibly facilitated by the specific RNA/protein interactions.

The inner layers of RNA in the BTV core seem to be less ordered. If only the core shell imposes icosahedral symmetry, this would be expected, as the shell would have no effect on the inner parts of the core. The amount of RNA packed into the core would necessitate a high level of compaction. This assumption was used in the study on BTV to be able to discover the characteristics of the genome in the core using X-ray crystallography (Gouet P. *et al.*, 1999). The concentration of the RNA in the core gave several clues into the possible behavior of the genome in the core. The volume of the core is 60.6 x 10^6Å^3. The volume of the inner core proteins is 7.58 x10^6 Å^3. This is deducted from the copy number of the different minor proteins (VP1=12, VP4=24, VP6=72) and their molecular mass (VP1=150kDa, VP4=76.4kDa, VP6=35.8kDa). The residual volume of the core is 53 x 10^6 Å^3. The genome has a molecular mass of 13.1 x 10^6 kDa. This results in the concentration of dsRNA to be 410mg/ml. Characteristics of dsDNA at high concentrations of ~400g/ml include the DNA being in a liquid crystalline state that can be visualized with X-ray diffraction as powder rings. These powder rings were also visible when BTV cores were examined with X-ray diffraction (Gouet P. *et al.*, 1999)(figure 1.13). This packaging of RNA is also a feature in rotavirus (Prasad B.V. *et al.*, 1996) and cypovirus (Hill C.L. *et al.*, 1999). Therefore, it indicated that dsRNA can be densely packed like DNA and that this occurs inside the BTV core. Cores alone are transcriptionally active. To investigate the effect of transcription activation, components essential for transcription were added to the core containing crystals and the diffraction patterns were observed. The powder rings seem to be enhanced for several minutes before disappearing (Gouet P. *et al.*, 1999)(figure 1.13c). This data suggests that activation of transcription may induce a higher level of ordering before transcription actually commences.

Densely packed dsRNA with a spacing of 30Å does present some problems. Steric interference can hamper movement of RNA strands relative to each other. This would inhibit the ability of the RNA to be moved through the fixed transcription complex. To counteract steric interference, counter ions can be used. Concentrations of possible counter ions, like Zn^{2+} and Mg^{2+}, were too low to act as efficient counter ions. Therefore, an organic counter ion, like spermidine, may be involved (Gouet P. *et al.*, 1999).

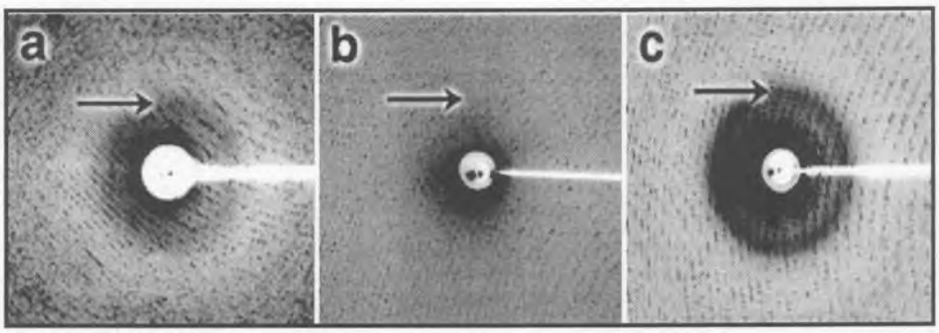


Figure 1.13. RNA powder rings in low-resolution diffraction patterns of BTV-1. Quiescent cores show a ring (a) while in 25% CsCl (b) the ring is lost due to the contrast matching of the RNA and the mother liquor. The ring is enlarged in active cores (c) via the addition of transcription reaction mix. Adapted from Gouet P. et al., 1999.

Gouet *et al*, 1999 proposed a working model of the packing of the genome inside the core. As already established, a TC complex is associated with a single genome fragment. Spiraling of the genome is seen in the core. It is proposed that RNA strands enter the core through the channel at the five-fold axis. As RNA enters the core cavity it spirals along the inner wall of the core around and away from the TC complex situated just below the channel. When the RNA strand clashes with another genome segment it switches to a second more inward layer and spirals around and back towards the TC complex. Redirection of the RNA may also be attributed to the Lys-807 residue on the long, flexible loop of VP3. While steric clashes may limit the lateral expansion of genome segments, difference in segment size is compensated for in an unknown manner.

It is apparent that the cores of AHSV, BTV and other related viruses are optimized for the organization of their genome while retaining their ability to autonomously transcribe the genome inside the infected cell. The role of the core shell in the movement of the RNA is ingenious and fits well with the current model of moving template transcription.

1.6 Enzymatic events in the core

Two events take place affecting the RNA genome. Firstly, each genome segment is transcribed to produce a positive strand mRNA that is extruded into the cytoplasm and secondly replication takes place in newly formed cores to produce dsRNA from the positive ssRNA template.

Transcription from a dsRNA origin can take place either through fully- or semi-conservative transcription. The more complex members of the *Reoviridae* family, including AHSV, transcribe the genome in a fully conservative manner while simple viruses, like ϕ6, transcribe the genome in a semi-conservative manner (Makeyev E.V. *et al.*, 2000)(figure 1.14). Fully conservative transcription follows the following steps. The dsRNA genome segment is split, possibly by a helicase, and the negative strand is used as template for positive strand synthesis. The newly formed strand is dissociated from its template and extruded away from the polymerase. Here again the helicase may play a role. The separated genomic strands are once again re-annealed (figure 1.14b). For simple viruses, semi-conservative transcription is employed. The two strands of the genomic segment are separated and a new positive strand is synthesized on the negative strand template. The original positive strand is removed, capped and extruded from the core (figure 1.14a). The first cycle of strand separation may be helicase-independent. A plough-like protrusion with a highly positive environment may be able to force the two strands apart as shown for the simple virus ϕ6 (Makeyev E.V. *et al.*, 2000).

One may also expect the other *Reoviridae* viruses to have this system for strand separation. This leads to the hypothesis that the helicase is primarily involved in the second part of strand re-assortment. This means that the more complex viruses need two rounds of strand separation through helicase activity while the simple viruses only need one strand separation event (figure 1.15). Transcription takes place after the virus has entered the cell and shed its outer capsid. In some species, like BTV and possibly AHSV, the 5' ends of the positive strands of each genome segment are attached to its respective transcription complex (Butcher S.J. *et al.*, 2001). The attachment of the terminus keeps the initiation site for transcription on the negative strand 3' end close to the polymerase and thus enhances transcription efficiency. The replication of the genome only occurs once the viral mRNA moves into the newly formed core.

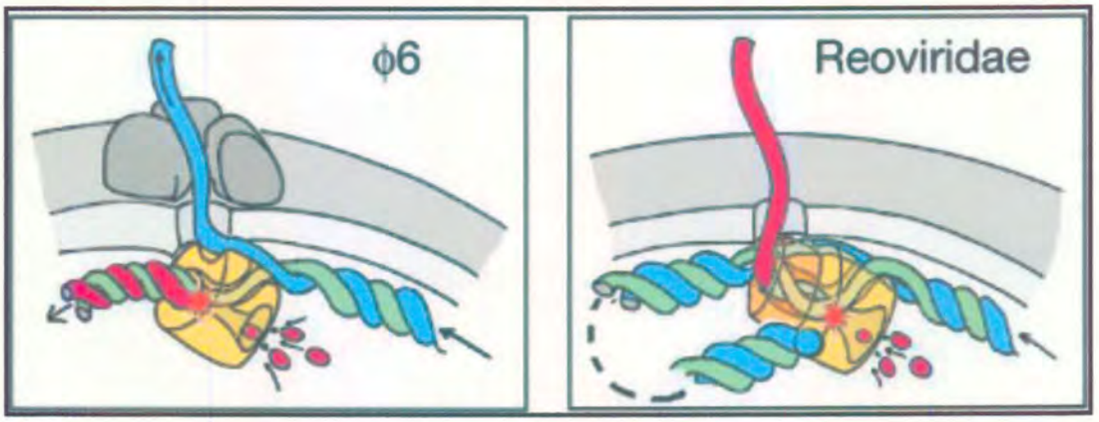


Figure 1.14. The differences between semi- and fully conservative transcription are described according to the transcription mechanism of ϕ6 and Reoviridae. While one mother strand is adapted into mRNA during semi-conservative transcription (left), a new strand is synthesized as mRNA during fully conservative transcription (right). Adapted from. Makeyev E.V and Bamford H., 2000.

The core assembles from the decameric units probably with the TC complex already attached to the five-fold axis of each decameric subunit. After core formation the mRNA is targeted to the core through a, yet unknown, mechanism. Subsequent negative strand synthesis results in a complete genome inside the core.

The regulation of replication and transcription may rely on the RNA sequence itself. As shown in Rotavirus the 3' terminals play an important role in the efficiency of negative strand synthesis (Chen D. *et al.*, 1998). The 3' terminal and 5' untranslated region form modified panhandle structures with the cis-acting replication signals present in the single-stranded tail from the panhandle. Mutation that caused complementarities between the 5' and 3' tail significantly inhibited replication. This data suggests that the 3' single stranded terminal is essential for replication and that the specific binding of the polymerase alone is not sufficient for replication (Chen D. and Patton., 1998).

For transcription initiation the 3' ends of the negative strand may be essential. In AHSV, but also in most other *Reoviridae* species the termini are highly conserved. At the 3' end of AHSV a CAATT motif is present (Roy P. *et al.*, 1994).

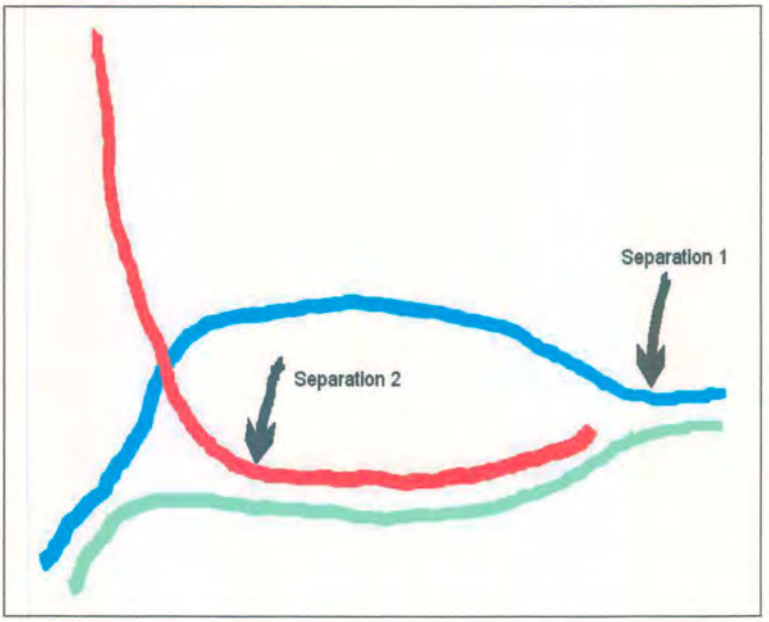


Figure 1.15. Two sites of strand separation result in a "transcription bubble". Two separation events occur. The first occurs when the two mother strands are separated before transcription is started and the second when mRNA is released.

The transcription of the genome results in the synthesis of nascent positive strand RNA's. Because the cores do not contain their own translation machinery the cellular translation machinery in the form of the endoplasmic reticulum has to be coaxed into translating the viral RNA. For translation to occur the RNA must be recognized. A large part of the recognition lies in the presence of a cap structure at the 5' end of the RNA. Beyond this essential function the cap structure has been implicated in many processes and many functions have been ascribed to it.

1.6.1 Viral polymerase

The viral polymerase fulfills two functions. It transcribes the viral genome producing a positive single strand RNA using the negative genome strand as template and it replicates the genome using the positive strand as template. For these functions a RNA dependent RNA polymerase is needed (Urakawa T., *et al.*, 1989). The polymerase must also be able to initiate and polymerize the new strand without the help of a primer. BTV VP1 is a basic protein with 7 highly conserved hydrophobic domains. It is the largest protein in the core at 146kDa. In ϕ6 cores the polymerase has a four amino acid loop (630-633) that is though to stabilize the first two incoming

bases during transcription by stacking (Laurila M.R. *et al.*, 2002). Stacking interactions are lost if the Y630 residue is mutated. As mentioned earlier this protein can also separate the dsRNA segment with a plough-like protrusion. This polymerase and probably other dsRNA virus polymerases are able to separate the incoming dsRNA segment, initiate transcription and do so without the help of a primer.

1.6.2 Viral helicase

In BTV and AHSV VP6 is thought to be the helicase. As mentioned before the first strand separation is probably performed by the polymerase itself but because these viruses undergo fully-conservative transcription a second helicase activity is needed to drive the genomic strand re-assortment. BTV VP6 can bind ATP and has ATPase activity. The ATPase activity is RNA dependent (Stauber N. *et al.*, 1997). . This suggests that the protein utilizes the ATP as energy source when unwinding the dsRNA strand. The ATP binding region has been partially identified in BTV. A motif seen in other ATPase active helicases was identified in this region. The motif (G/AxxGxGKS/T) is found back twice in the region identified with a single amino acid change in both possible motifs (AxxGxGKV and AxxGxGAT) (Stauber N. *et al.*, 1997). Additionally, several RGG RNA binding motifs have been identified. They are clustered around positively charged arginine-rich motifs.

1.6.3 Viral mRNA cap synthesis

a. Cap structure

The 5' guanosine cap structure consists of an inverted guanosine residue. The 7-methyl-guanosine is linked to the first nucleotide of the ssRNA segment via a 5'-5' triphosphate bridge (m^7GpppN)(Banjeree A.K., 1980; Shuman S. *et al.*, 1994). Viral and cellular capping enzymes can produce the cap structure.

As mentioned earlier the cap structure has several possible functions. Of importance is the recognition of viral mRNA by the cellular translation machinery. The cap structure also camouflages the viral mRNA from intracellular defense mechanisms. The effects of siRNA or small interfering RNA that can inhibit the translation of a

given gene highlight that these cellular mechanisms do exist. Cellular endonucleases are thought to be involved in these processes. Other functions may include processing and transport from the nucleus to the endoplasmic reticulum (Reddy R. *et al.*, 1992). The cap may even be involved in the termination of transcription. Certain proteins bind specifically to the cap and are activated by this binding. These include proteins like eIF4E that, in complex with another protein, mediate the initiation of translation (Lewis J.D., and Izaurelda., 1997). In the nucleus another complex mediates the stimulatory effects of the cap on pre-mRNA splicing, RNA 3' end formation and RNA nuclear export (Lewis J.D., *and Izaurelda.*, 1997)

The complete cap structure not only includes the guanine moiety at the 5' end but also the methylation of specific groups on the guanine residue and adjacent residues.

b. Capping process

Furuchi and Shatkin., 1976 summarized the formation of the complete cap structure in the following steps:

pppGpXp...RNA 5' triphosphatase ppGpXp... + Pi

ppGpXp... + GTP guanylyltransferase GpppGpXp... + PPi

GpppGpXp... + SAM methyltransferase 1 ^{m7}GpppGpXp...

^{m7}GpppGpXp... + methyltransferase 2 ^{m7}GpppG^{m}pXp...

The guanylyl transferase step is additionally broken down into the following reaction:
GMP + transferase → GMP-transferase complex → GpppGpXp... + transferase

This basic reaction still stands as working theory although some exceptions to this process do exist. A multi-step process like this requires different catalytic entities. As seen in the summary above an RNA triphosphatase is required to remove the 5' terminal γ-phosphate. Subsequently, the guanylyl transferase binds a GMP moiety and transfers it to the biphosphate end of the RNA. Lastly, the cap and adjacent nucleotides are methylated. The cap itself is methylated on position 7 by a methylase, yielding a cap 0 structure. Frequently, a second methylation step methylates the 2' hydroxyl group of the ribose on the adjacent nucleotide resulting in a cap 1 structure (Martin S.A., and Moss., 1975; Mizumoto K., and Kaziro., 1987).

The different catalytic entities may be situated on different proteins as found in most mammals or, alternatively, some or all of the catalytic sites are situated on one protein as seen in many viruses. In viruses several capping pathways have been studied. These include studies on complex viruses like vaccinia virus and on simple viruses like BTV. The study of BTV RNA capping has revealed that all the catalytic steps are performed by one single protein, VP4 (Ramadevi N. *et al.,* 1998a).

1.6.4 Specific enzymatic reactions

a. RNA-triphosphatase

A 5' RNA-triphosphatase catalyses the first step in the capping reaction. Nascent transcribed RNA fragments have a three-phosphate chain at the 5' terminal nucleotide. The triphosphatase excises the γ-phosphate by hydrolyzing the γ-β bond leaving behind a double phosphate chain. Excising the γ-phosphate ensures the formation of a triple phosphate bridge. Exceptions to this structure can be found in some mammalian cells and eukaryotic embryos where four-phosphate bridges have been observed (Bisaillon M., and Lemay., 1997b) and ref therein. The function of a triphosphate bridge as opposed to a tetra phosphate bridge is not entirely clear. Two possible functions for a triphosphate bridge may be stability of the cap or the recognition of the cap by the translation machinery. Emphasizing the importance of the triphosphate-bridge is the fact that RNA-triphosphatase activity has a 100-fold higher velocity than GMP-transfer in vaccinia virus. The higher catalytic speed of the RNA-triphosphatase ensures that there are only biphosphate ends available for GMP transfer (Venkatesan S. *et al.,* 1980).

Interestingly, an additional nucleoside triphosphatase activity was also described for several viral capping enzymes (Martinez-Costas J. *et al.,* 1995)(Ramadevi N. *et al.,* 1998c). Competitive inhibition studies showed that both the RNA- and nucleoside tri-phosphatase activities reside at the same catalytic site. The function of the nucleoside tri-phosphatase in the virus environment or life cycle is still unknown. Additionally, it was shown for λ1 capping protein of Reovirus that the affinity for NTP's is much lower than for RNA, indicating a preference for the RNA triphosphatase function. It has not been established whether the nucleoside triphosphatase is active *in vivo*. (Bisaillon M., and Lemay., 1997b). A possible role in providing GMP for the capping reaction does not seem likely. When incubating the Reovirus λ1 protein with

GTP very little GMP is found, indicating that the nucleoside triphosphatase is not involved in the generation of GMP (Bisaillon M., and Lemay., 1997b).

Two possible motifs for tri-phosphatase activity were identified in several viruses. The first is a four amino acid motif (LxP/IR). This motif was identified in West Nile virus NS3 (LRIR), reovirus λ1(LRPR) and Vaccinia virus D1 (LKPR) (Bisaillon M. *et al.*, 1997b), (Bisaillon M. *et al.*, 1997a). The other motif is more degenerate. In West Nile virus the motif is RTNTILE, in vaccinia virus RPNTSLE and Reovirus RDETGLM. This motif may be essential as a substitution of the glutamic acid renders the triphosphatase activity inactive in Reovirus (Yu L., and Shuman., 1996; Bisaillon M. *et al.*, 1997c).

b. Guanylyltransferase

We now turn our attention to the central and essential part of the cap structure formation. The guanylyl transferase catalyzes the transfer of GMP from GTP to the 5' diphosphoryl end of the nascent RNA. This process revolves around a GMP-enzyme intermediate. This intermediate occurs through the presence of a covalent phospho-amide bond between the GMP and a lysine residue in the active site of the protein. (Fausnaugh J., and Shatkin., 1990; Luongo C.L. *et al.*, 2000). After the formation of this intermediate transfer of the GMP occurs from the enzyme to the 5' diphosphoryl RNA. The exact mechanism for this process is still unknown but structural studies of cypovirus PBCV-1 capping enzyme has revealed some interesting data on the mechanics underlying the enzymatic actions (Hakansson K. *et al.*, 1997)(figure 1.16).

This capping enzyme consists of two domains separated by a deep cleft. The GTP-binding site and the active site lysine are located in domain 1. Some conserved motifs are also located around the GTP binding site and across the cleft. The cleft is thought to bind the ssRNA. Interestingly, upon GTP binding the enzyme undergoes a substantial conformation change through which the cleft is closed off (Hakansson K. *et al.*, 1997). This change in conformation allows the active site lysine residue to move close to the GTP molecule and force the hydrolization of the two phosphates and the binding of the monophosphate to the active lysine.

Additionally, this specific enzyme needs a cation like magnesium or manganese (Le Blois H. *et al.*, 1992; Bisaillon M., and Lemay., 1997b). This is deduced from the

observation that only the closed conformation binds the cation. It also means that GTP binding is needed for cation binding and presumably subsequent transfer. Upon hydrolysis of the α-β phospho bond the enzyme re-opens and RNA is now able to bind into the cleft and transfer of the GMP (figure 1.17).

This basic mechanism will undoubtedly be tested with other viral and eukaryotic transferases as several differences in motifs are observed. While no universal transferase motif exists a lysine residue is essential for binding of the GTP. The KxDG motif is found in a variety of DNA virus transferases like vaccinia virus, African swine fever, Shope fibroma virus and the yeasts *S. cerevisiae*, *S. pombe* (Bisaillon M. *et al.*, 1997c; Shuman S. *et al.*, 1994 and ref. herein). Mutation of the lysine but also the glycine resulted in loss of function underwriting the importance of this motif in binding the GMP moiety (Shuman S. *et al.*, 1994).

Other conserved sequence elements (II-V) were identified in DNA viruses and yeasts. These elements were proposed to indicate structural conservation among the viruses (Shuman S. *et al.*, 1994), but the structural determination of PBCV-1 indicates that these elements are located around the GTP binding site in the open conformation (Hakansson K. *et al.*, 1997). Therefore, a possible functional role for these elements cannot be disregarded as individual amino acids within these elements were shown to be essential for capping *in vivo* in yeast (Shuman S. *et al.*, 1994)
Together, these data suggest that parts of these elements are indeed necessary for GTP binding.

The KxDG motif is not found in a substantial group of viral guanylyltransferase. These include reovirus λ2, rotavirus VP3, BTV VP4 and presumably all members of the Reoviridae family (Bisaillon M., and Lemay., 1997c) and ref herein. The active site sequence of the reovirus λ2 (KPTNG) and BTV VP4 (KLTGD) diverge from the KxDG motif although retaining some similarity. Another motif is conserved between reovirus λ2 (YVRKN), murine rotavirus SA-11VP3 (YVRKN) and BTV VP4 (YKRKN) (Bisaillon M., and Lemay., 1997c).

c. Methyltransferase

Methylation of the cap and the adjacent nucleotides is the final step in the capping process of RNA. An exception is the capping of RNA of the alpha viruses like

Sindbis and Semliki Forest virus where the GMP moiety is methylated prior to transfer to the RNA strand (Aloha T., and Kaarianinen *et al.,* 1995)

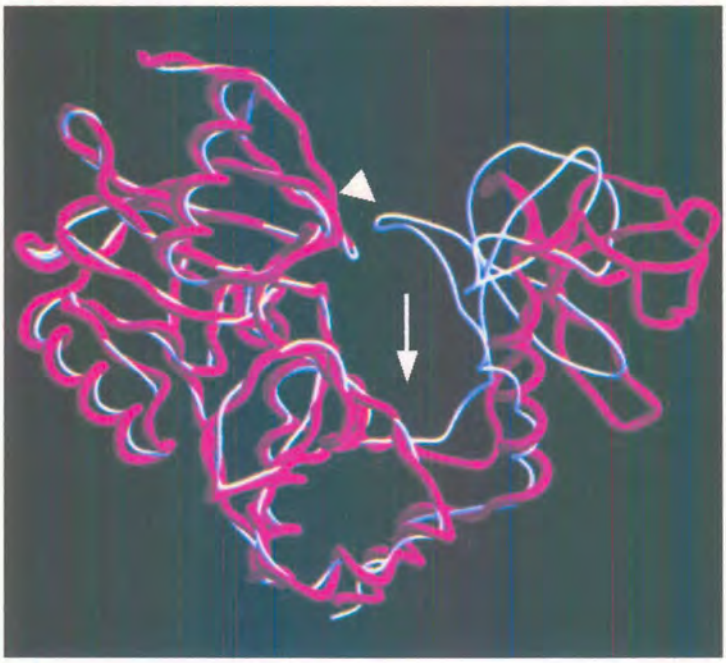


Figure 1.16. A string diagram of the PBCV-1 enzyme. The arrow indicates the cleft. The white overlay depicts the change in conformation upon activation with the arrowhead indicating the area with the largest change. Adapted from Hakansson K. et al., 1997.

Different levels of methylation are observed and are referred to by cap 0, cap 1 and cap 2. These cap numbers correspond to the number of methylated ribose residues present adjacent to the cap (Furuichi Y., and Shatkin., 1989). The cap 0 structure therefore has an $m^7GpppGp$... while the cap 1 has an m^7GpppG^mp...and cap 2 an $m^7GpppG^mpG^mp$...structure. Many eukaryotic and viral mRNAs contain a cap 1 structure (Ramadevi N. *et al.,* 1998a and ref herein). This is in contrast with Bisaillon and Le May, 1997 who claim that higher eukaryotes have cap 2 structures while lower eukaryotes contain mRNAs with cap 0 structures. It is thought that methylation, although not essential in many organisms, is important for efficient initiation of translation by ribosomes (Bisaillon M., and Lemay., 1997c). Other, as yet unknown, functions are not excluded.

For methylation the universal methyl donor, S-adenosyl-L-methionine, is used. In reovirus the active site for SAM binding was localized to the $\lambda 2$ protein, the reovirus capping enzyme (Luongo C.L. *et al.,* 1998).

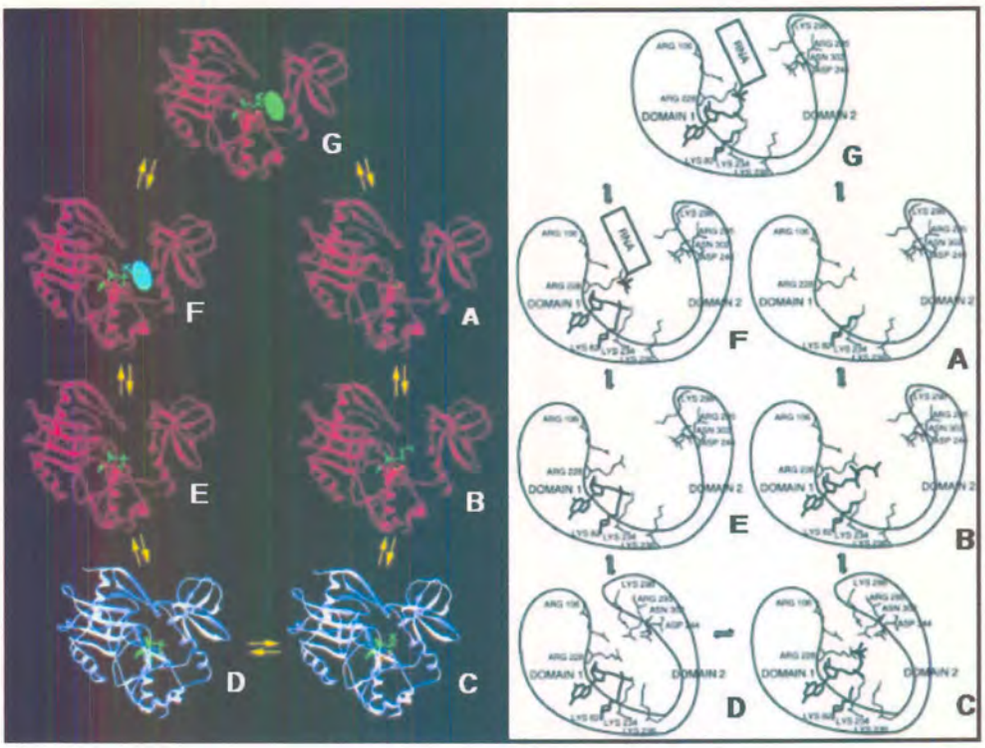


Figure 1.17. A proposed mechanism for the nucleotidyl transfer. The red wire diagrams represent the enzyme in open conformation while the white diagrams are closed. The nucleotide moiety (green) is attached to the lysine side chain (yellow). The apo-enzyme (a) binds to the nucleotide (b) that results in a closed conformation (c). The GTP is catalyzed to GMP (d) and this causes the conformation to return to open (e). RNA is attached to the GMP moiety (f). Adapted from Hakansson K. et al., 1997.

Furthermore, cleavage of the protein indicated that the site is located in the region between residue 792 and 1100. This area includes a smaller region that contains a SAM binding pocket similar to other methyl transferases. Alanine substitutions at two positions, 827 and 829, greatly reduced the ability of baculo-expressed protein to undergo UV cross-linking to SAM (Luongo C.L. *et al.,* 1998), (Koonin E.V., 1993). The central region of the λ2 protein overlaps with the sequences identified by Koonin *et al*, 1993.

The alanine substitutions were made in the BBB-D (to A)-B-G (to A)-XG motif (where B is a hydrophobic residue and X is any residue) as identified by Koonin *et al*, 1993. Ramadevi *et al*, 1998 showed that the guanylyl transferase and triphosphatase, VP4, catalyze both methylation steps in formation of cap 1 structures. Vaccinia virus also contains all catalytic activities on the two subunits of the capping enzyme (Ramadevi N. *et al.,* 1998a). This means that all three enzymatic reactions needed for cap

formation are compacted on one protein. In contrast, eukaryotic proteins are only responsible for one of the three reactions. The second methylation step by VP4 is limited when purified VP4 is used indicating that the conformation of VP4 may only be optimal in the presence of the other core proteins and the scaffold, VP3.

Ramadevi *et al*, 1998a suggested a possible mechanism in which capping of the first nucleotide and methylation initiates the further transcription of the RNA is put forward by. This is supported by the fact that VP4 is able to transform GTP to GpppG dinucleotide and that transcription and RNA modification occurs within the activated core with the core always extruding mature mRNAs. This hypothesis must still be proven. In contrast, rotavirus requires a 3' ssRNA strand for efficient transcription (Chen D., and Patton., 1998)

1.7 Model for transcription in core of segmented RNA viruses

While the mechanism for the transcription in a segmented RNA virus is far from resolved several pieces of data allow the proposition of a hypothetical model for the transcription of an RNA segment and the associated capping and reorganization of the dsRNA. A general mechanism is possible for the whole family of segmented dsRNA viruses even though it is known that the more simple viruses like ϕ6 do not have a helicase activity and perform semi-conservative transcription. On the other hand, the more complex viruses like BTV and AHSV have helicase activity and perform fully conservative transcription, two characteristics that are related to each other.

The dsRNA segments are individually wound around each TC complex located at the five-fold axis. The 5' end of the positive strand of each segment is attached firmly to its associated TC complex. This may be needed for efficient transcription as the negative strand 3' end is positioned close to the active site of the polymerase. After the virus enters the cytoplasm, and in the process has released it capsid, it becomes activated. Activation is probably due to the release of the constraints imposed on the core by the capsid and the influx of ions like Mg^{2+}. Diprose *et al.* 2001, suggested, on basis of crystallography data, that Mg^{2+} plays a structural role in the core. This might indicate that Mg^{2+} assists in the activation of the core by binding components of the core into their active conformation. The necessary substrates like NTP's and Mg^{2+}

will be able to enter the core through the "N"pore (Diprose *et al.*, 2001). Activation of the core will lead to the initiation of transcription.

For the separation of the two strands of RNA a helicase was always thought to be needed. Data from the ϕ6 virus suggests that no helicase is needed for the initial separation. A plough-like protrusion coupled with a highly positively charged environment around it, physically shears the RNA and directs the negative strand towards the catalytic site (Butcher *et al.*, 2001). The positive strand 5' end would remain attached to the polymerase protein. Ramadevi *et al.*, 1998a postulated that capping of a guanine residue takes place before transcription is actually initiated and that this GpppG moiety is attached to the first nucleotides to be transcribed. Although possible, this hypothesis does not explain how this additional guanine residue is removed once this mRNA is packaged again in a newly synthesized virion. An, as yet unknown, mechanism would have to exist to remove this additional guanine. In the model proposed here only Mg^{2+} and conformational changes in the core structure initiate transcription. In this model the helicase would only be responsible for the separation of the newly formed strand from the negative template. The ssRNA is directed towards the active site of the capping enzyme lying adjacent to the polymerase. Capping takes place while transcription is in progress and the transcript is probably not longer than 10 or 20 bases. Therefore, a helicase has to be situated on the interphase between the polymerase and the capping enzyme to be able to separate the two strands as they leave the polymerase catalytic site. The helicase may also help in directing the negative mother strand back towards the positive strand, still in place because of the attachment of the 5' end. A reversibly placed helicase may actually drive the re-annealing of the two strands although the density of the RNA could also assist in this event. The newly formed mRNA is extruded through the five-fold pore directed by the charged amino acids inside the pore. Once the mRNA appears outside of the core it may be recognized by certain cellular transport proteins that will actively transport the mRNA out of the core towards the nucleus along the microtubules. This action of these transport proteins might actually provide the energy needed for the mechanism proposed. In this model the TC complex is a revolving unit. Viruses like Herpes Simplex virus and adenovirus use the cellular transport pathways to get their mRNA into the ER (Lewis J.D., and Izaurelda., 1997; Koffa M.D. *et al.*, 2001). The turning of the TC complex will create the momentum needed for the RNA segment to move through the polymerase and re-anneal after transcription.

Firstly, it is important to note that the negative mother strand will be separated from the mother positive strand at two places and annealed to the newly formed daughter stand once as can be seen in the figure 1.15.

Once the newly formed strand is separated from the mother negative strand the negative strand is directed towards the positive strand and re-annealed. This already happens while transcription is in progress creating a kind of adapted transcription bubble. It makes sense that re-annealing is most important for the organization of the RNA inside the core as we already discussed that the spatial restrictions on the genome in the core are severe. In this way as little as possible single stranded RNA will be present in the core minimizing the presence of space-intensive ssRNA. The winding of the RNA around the TC complex will be redirected so that the post-transcription part of the genome segment is wound around the TC complex again. For all these actions the simplest way to achieve them is for the polymerase to turn around its axle. To underlie this model it is interesting to remember that VP3 allows its C-terminal to extend into the core at the five-fold axis forming a possible axle for the polymerase to turn on. It was shown that this polypeptide is indeed attached to the polymerase. The attached 5' end of the positive strand would create the necessary tension in the dsRNA strand to allow for a tight winding of the segment around the polymerase.

Turning of the polymerase would therefore pull the negative strand through the polymerase catalytic sites, driving transcription. This would also include separation of the strands at both the beginning and the end of transcription. To revolve the polymerase would need energy. Two options exist. The first would include the function of the motor proteins to actively transport the mRNA towards the nucleus. This would create tension in the ssRNA strand and like a fan belt of a car pull the wheel that is the polymerase. The other option is the use of the NTPase activity of the capping enzyme as shown in BTV and other viruses to produce the energy needed for such movement. Data, however, suggests that the ATPase activity is not very efficient and therefore this seems less likely.
This model does not explain some aspects of transcription. For instance, how is initial transcription taking place while the mRNA has not exited the core? A role for NS2 might be possible here, as we know that NS2 does have ssRNA binding ability (Horscroft N.J., and Roy, P., 2000). Unfortunately, this would be impossible as NS2 first needs to be expressed and for that mRNA is necessary.

1.8 Aims

As shown in this review several questions about the structure and complex enzymatic activity of the transcription machinery within the core of the dsRNA virus remain unanswered. The mechanistic features of the transcription complex and the co-operation between the different subunits need to be further investigated. In the laboratory of genetics at the University of Pretoria African horse sickness virus is studied. To investigate the transcription complex, the putative capping enzyme VP4 needs to be characterized. Only one serotype of AHSV VP4 has been sequenced and a second is needed to derive general sequence motifs and possible physical features from the amino acid sequence. This is especially important, as several key motifs identified in BTV VP4 are not seen in the AHSV VP4. To compare the protein further with BTV it is important to characterize the enzymatic activities the AHSV VP4 protein comprises. To this end soluble pure protein samples are needed. In the following chapters we will show data from sequencing AHSV-3 VP4 and the efforts of expressing VP4 and purifying it for functional studies.

Chapter 2.
The analysis of AHSV serotype 3 VP4.

2.1 Introduction

In the case of BTV, the three minor core proteins have been characterized with respect to some of the enzymatic activities needed to produce mRNA. With respect to VP4 of BTV, this protein was first identified as the candidate protein to catalyze the capping of RNA in the core (Le Blois and Roy.P., 1992). Later VP4 was positively identified as the capping enzyme of BTV (Martinez-Costas J. *et al.,* 1998).

In case of guanylyl transferases of DNA viruses and yeast the proteins were characterized by a conserved lysine-containing motif, KxDG. Mutational studies showed that the lysine residue is essential for capping (Bisaillon M. and Le May G., 1997c). The lysine was subsequently shown to be the attachment point for the GMP moiety used in the addition of the guanine cap. In a paper by Huang *et al.,* 1993 five BTV serotype's M4 segments were sequenced and analyzed. It was found that the sequences were very conserved between the different serotypes. A number of conserved motifs with potential functional relevance were identified. From the alignment study it was shown that the lysine-containing motif as identified in DNA viruses does not exist but a divergent motif was found. BTV has a KLTGD motif that is more similar to the motifs present in reovirus (KPTNG) and rotavirus (KPTGN) (Bisaillon M. and Le May G., 1997c). Bisaillon M. and Le May G., 1997c also showed that another motif is conserved between viral capping enzymes from BTV, reovirus and rotavirus and may prove to be important in the capping process. In BTV this motif is YKRKM, while in Reovirus it is YVRKN and in rotavirus YYRYN. The functions of these motifs are not known yet.

Moreover, a leucine zipper motif is present in the BTV VP4 amino acid sequences. A leucine zipper motif consists of several repeats of a leucine residue at every seventh position and a short stretch of charged residues adjacent to the motif. Leucine zippers were first described in DNA binding proteins c-Myc and c-Jun (Dang C.V. *et al.,* 1989). In BTV the motif was found in serotypes 2, 10, 11, 13 and 17 at amino acid residue 523-551 and consists of five leucine residues (Huang I-J *et al.,* 1993). The zipper identified in BTV is modified and has a proline residue in the

middle that might influence its normal coiled-coil conformation. Upstream of this motif a stretch of prolines (5/9aa's) was found.

Leucine zippers are structural features that can assist in the dimerization of proteins (Dang *et al.*, 1989). Indeed, in BTV, mutational studies showed that the leucine zipper is responsible for dimerization of the VP4 protein and that dimerization is essential for its encapsidation and capping function (Ramadevi *et al.*, 1998b).

While AHSV is a family member of the orbivirus genus, its infection range is very different from BTV. For AHSV the segment 4 of serotype 4 has been sequenced (Yu Y. *et al.*, 1987). No analysis has been done to identify possible motifs that BTV and AHSV may have in common. It is likely that there will be shared motifs between AHSV and BTV. The reason for this is the relatively high level of conservation between the two viruses' gene segments. It is not known if AHSV VP4 is indeed the capping enzyme.

The aim of this chapter is to sequence another AHSV serotype's VP4 gene segment. The two serotypes will be compared and used in alignment studies with other members of the orbivirus genus. We will also analyze the sequence to identify possible motifs and compare the sequence with BTV VP4 to determine whether the motifs identified in BTV VP4 are also present in AHSV VP4.

2.2 Materials and methods

1. Cloning of VP4 in the pGEM vector

Previously, AHSV-3 VP4 cDNA was cloned into the *PstI* site of the bacterial cloning vector pBR322 by S. Cormack in the department of genetics, University of Pretoria. In pBR322 *Bgl II* sites are present on both sides of the gene. For sequencing purposes VP4 is recloned into the pGEM vector.

1.1 Restriction enzyme digestions

To liberate VP4 from the pBR322, a restriction enzyme digestion was performed. 1µg of VP4-pBR322 was digested with 0.5 units *Bgl II* in buffer M for 1hour. pGEM-3Z is a bacterial vector commonly used in sequencing experiments. A multiple cloning site (MCS) is located within the lacZ gene allowing colour selection of

recombinants. The plasmid also contains an ampicillin resistance gene and the universal M13 –21 and reverse primer sequences are present on either side of the MCS. 1µg pGEM vector was digested with 0.5 units *BamHI* in buffer B for 1 hour.

1.2 Dephosphorilation of plasmid

To inhibit the ability of digested pGEM vector to self-anneal during ligation, the vector was dephosphorilated to remove all phosphate groups from the 5' ends of the linear vector. After restriction enzyme digestion of the plasmid, it was dephosphorilated by adding the digested vector to a reaction mixture consisting of 1 unit calf intestinal alkaline phosphatase, 4µl buffer (0.5M Tris-HCl, 1mM EDTA, pH 8.5), making a final volume of 40µl and incubated for 15 minutes at 37°C. To stop the reaction, the DNA was separated by electrophoresis on an agarose gel.

1.3 Electrophoresis of DNA

DNA fragments were separated on a 1% agarose gel by electrophoresis using a E-C Minicell electrophoretic gel system under 75 Volts for an hour in the presence of EtBr in 1xTAE buffer (0.4M Tris-acetate, 2mM EDTA, pH 8.5). The DNA was visualized using a Vilber Lourmat UV trans illuminator and a Sony video graphic printer.

1.4 DNA purification from gel

After electrophoresis of pGEM and VP4, both DNA fragments were recovered from the agarose gel. This was done using the GeneClean kit from Bio101. The gel slice containing the DNA of interest was melted in 4 volumes of 6M NaI at 55°C for 5 minutes. 5µl DNA binding matrix, Glassmilk, was added to the solution and incubated with shaking at room temperature (RT) for 15 minutes. After pelleting the glassmilk for 5 seconds, the pellet was washed twice with 0.5ml NewWash solution (NaCl, Tris, EDTA, ethanol, and water) with intermediate centrifugation of five seconds before the DNA was eluted by addition of 10µl double-distilledH_2O to the pellet and incubated for 2 minutes at 55°C. The glassmilk was separated from the dissolved DNA by centrifugation.

1.5 Ligation

The full-length VP4 fragment was annealed to the pGEM vector by combining the insert and vector in a ratio of 3:1 and adding 1 unit of T4 ligase (Boehringer Mannheim) in the presence of 1µl of 10x ligase buffer (66mM Tris-HCl, 5mM $MgCl_2$, 1mM DTT, 1mM ATP, pH 7.5) in a total volume of 10µl. The reaction was incubated for 16 hours at 16°C.

1.6 Making competent cells

For bacterial cells to be able to take up foreign DNA they have to be made competent. *E.coli* XL1Blue bacterial cells were made competent using the $CaCl_2$ method (Sambrook *et al.*, 1979). 100ml Luria-Bertani broth (1% w/v tryptone, 0.5% w/v yeast, 1% w/v NaCl) was inoculated with 1ml of an overnight bacterial cell culture. The culture was grown to an optical density of 0.45 at 550nm wavelength (OD_{550}). Cells were collected through centrifugation at 5000 rpm for 5 minutes at 4°C and then resuspended in 10ml ice cold, fresh 50mM $CaCl_2$. The cells were collected by centrifugation and resuspended in 1ml 50mM $CaCl_2$. Cells were incubated on ice for at least 1 hour before use.

1.7 Transformation

The ligation reaction was added to 100µl of competent cells. The cells and DNA were mixed gently and incubated on ice for 30 minutes after which the cells were heat-shocked for 90 seconds at 42°C. After the cells cooled on ice for 2 minutes, 800µl LB-broth was added and the mixture incubated at 37°C for 1 hour. Cells were plated out in 150µl quantities on LB agar plates containing 50mg/ml ampicillin, 12.5mg/ml tetracycline, 24µg/ml IPTG and 50µl X-gal (2% solution in N, N' dimethyl formamide). Plates were incubated at 37°C for 24 hours. After 24 hours incubation, plates were inspected for blue and white colonies. All white colonies were picked up and restreaked on fresh plates. Test tubes with 3ml LB broth containing 50µg/ml ampicillin and 12.5µg/ml tetracycline were also inoculated with samples from the white colonies. These inoculations were grown overnight at 37°C and used for isolation of plasmid DNA.

2. Selection of recombinant colonies containing pGEM with full-length VP4

2.1Miniprep DNA isolation

Plasmid purifications were performed according to the alkaline lysis method described by Birnboim and Doly, 1979. Overnight cultures were centrifuged in a tabletop centrifuge for 1 minute to collect the cells. Briefly, cell pellets were resuspended in 100µl Solution I (50mM glucose, 1mM EDTA, 25mM Tris; pH8). Freshly made Solution II (0.2N NaOH, 1% SDS) was added to the cell suspension to lyse the cells after which the suspension was incubated on ice for 5 minutes. Afterwards, 150µl 3M sodium acetate, pH 4.8 was added, to precipitate all proteins, and incubated for 10 minutes on ice. The solution was centrifuged at 14000rpm for 10 minutes in an IEC MicroMac benchtop microcentrifuge. The supernatant was retrieved and incubated for 10 minutes at RT with 320µl 100% isopropanol to precipitate all DNA before being centrifuged at 14000rpm for 10 minutes. The resulting pellet was washed with 80% ethanol and vacuum-dried in a Savant Speedvac concentrator. The pellet was resuspended in 30µl ddH_2O and stored at 4°C.

2.2 Restriction enzyme digestion

To determine recombinancy of the isolated plasmid DNA samples, they were digested with *EcoRI* and *XbaI*. The recognition site for *EcoRI* lies downstream of the *BamHI* site in the MCS. The *XbaI* site lies in the VP4 gene 620bp downstream from the 5' end of VP4. 1µg of DNA was incubated with 0.5units *EcoRI* and 0.5 units *XbaI* for 1 hour at 37^0C. Afterwards the digestions were analysed on a 1% agarose gel.

3. Sequencing

3.1DNA extraction and purification

For sequencing purposes high quality and very pure DNA samples are needed. DNA was extracted using the miniprep method as described in B.1 and further purified using the High Pure filter columns of Boehringer Mannheim's High Pure plasmid Purification™ kit. The kit uses filter tubes, packed with glass fibers, to bind nucleic acids specifically, thereby eliminating all contaminants. After DNA from miniprep isolation was resuspended in UHQ it was applied to the filter tube. The filter tube was inserted in an Eppendorf and centrifuged for a few seconds. Wash buffer (Tris-HCl, pH6.6, ethanol) was added to the filter tube and centrifuged. Finally, ddH_2O was added and the DNA eluted through centrifugation. Extracted DNA was analysed on a 1% agarose gel and the concentration measured using a Beckman DU® 64

spectrophotometer. The samples were measured at wavelengths 260nm, 280nm and 320nm. The concentration was calculated using the extinction coefficient for dsDNA as $1A_{260}$=50µg/µl, while the 260nm to 280nm ratio gives an indication of sample purity. The 320nm reading gives an indication of the contaminating salt concentration.

3.2 Big dye termination reaction

To sequence DNA a set of fragments corresponding to each possible length of the fragment is needed. This fragment must be labelled with a fluorescent dye on the nucleotide at its 3' end using a labelled dideoxynucleotide (ddNTP). The different fragments can be separated on an acrylamide gel and the different fluorescent dyes read by exitation of the separate dyes and the detection of the different emission frequency. From this data a contig of sequence fragments can be assembled and the final DNA sequence can be determined. To produce the labelled fragments the BigDye terminator mix is used. 250ng DNA was incubated with 3.2 pmole of M13 forward primer or M13 reverse primer (primers bind to the 5' and 3' side of the multiple cloning site (MCS) respectively), or a VP4 internal primer (5'-ACGAGAACGAGAGCGG-3')(VP4int) and 4µl Terminator Ready Mix all in a final volume of 10µl. Thermal cycle reactions were performed in the P.E. 9600 thermal cycler with heated lid. The reaction was heated to 96°C with incubation for 10 seconds to denature the DNA before ramping to 50°C and incubation for 5 seconds to allow annealing of the primer.

A final ramp to 60°C and incubation for 4 minutes followed to allow extension of the new poly-nucleotide. This cycle was repeated 25 times followed by cooling to 4°C. Unincorporated nucleotides and primers were removed with an ethanol precipitation procedure as described in the Dye terminator-sequencing manual. The reaction volume was added to 16µl ddH_2O and 64µl 98% ethanol. The DNA was precipitated for 10 minutes at RT. The precipitate was collected through centrifugation at 10000 rpm for 10 minutes. The pellet was washed using 80% ethanol and air-dried.

3.3 Separation of sequencing fragments

The resulting sequencing reactions were analysed using the ABI PRISM 377 automated sequencer. Samples were resuspended in 3µl loading buffer (5:1 ratio of deionised formamide to 25mM EDTA pH8.0 containing 50mg/ml dextran blue). The

samples were denatured at 95°C for 2 minutes prior to loading. 1.5µl of each sample was loaded on a 4% denaturing polyacrylamide gel and run for 7 hours at 1.6kV.

4. Analysis

To interpret the raw data (nucleotide sequence) and to identify possible relationships with other genes as well as identify potential regions of high conservation between related genes several analyses were performed. Raw sequence data was processed using ABI PRISM Sequencing Analysis™ and further analysed using the ABI PRISM Sequence Navigator™ software. Possible open reading frames were identified using the ORF finder program from the National Library for Biotechnology Information (NCBI). The ORF was also translated into an amino acid sequence with the same program. The nucleotide sequence and the amino acid sequence are used to identify other genes and polypeptides with high homology using the online program BLAST (http://www.ncbi.nlm.nih.gov/blast). The amino acid sequences of closely related proteins were aligned using the web-based Clustal W multiple alignment program (http://npsa-pbil.ibcp.fr/NPSA/npsa_clustalw.html) and trees were drawn using the alignment with the program Phylodendron. Motifs are regions of high conservation between related proteins that may indicate areas with functional significance. To identify possible motifs within several of the amino acid sequences the web based analysis tool Block Maker was used. Block maker generates possible motifs from the amino acid sequences using two distinct methods, MOTIF and GIBBS. While the Motif method detects possible conserved motifs by locating spaced triplets, the GIBBS method uses a more statistical method and does not rely heavily on the amino acid identities. Spaced triplets are two sets of three amino acids that are found with the same interval in several amino acid sequences. When results from these two methods were compared a prediction with high significance was made. To identify possible motifs already identified in other proteins and indicative of possible functions a meta-search engine (eMOTIF) (http://motif.genome.ad.jp) and also Predprot (http://cartan.gmd.de/TargId-bin/) was used that includes several different motif search analysis programs. These include prosite pattern, prosite profile, BLOCKS, ProDom, PRINTS, Pfam and Pfam_fs. Proteins contain two secondary structure forms, β–sheets and coiled coils. While sheets are more stable, the helices are mostly associated with active catalytic residues. The secondary structure of the protein can be estimated using several web based analysis tools. For this prediction an analysis program from the NPSA (Network Protein Sequence Analysis) server (http://npsa-pbil.ibcp.fr) was used that

creates a consensus sequence from 8 different prediction programs. Additionally, the amino acid sequence was fitted onto the 3D structure of possible aligned proteins using the 3D-PSSM server.

2.3 Results

To facilitate the sequencing AHSV-3 VP4 had to be cloned into the sequencing vector pGEM. VP4 was re-cloned into the pGEM vector that has recognition sites for the universal primers on both sides of the MCS. With the use of full-length and subcloned inserts of VP4 and an internal primer we were able to sequence the full-length VP4 of AHSV-3

1. Cloning

For sequencing purposes the full-length cDNA of VP4 was cloned into the pGEM vector. Additionally, subclones were made to allow complete coverage of the gene. The full-length cDNA copy of AHSV-3 M4 was cloned from pBR322 into the pGEM vector using the *Bgl II* sites present on either side of the VP4 gene. The pGEM sequencing plasmid was digested with *BamHI* to linearize the vector. The virus segment was ligated into the plasmid and recombinant plasmids were selected through restriction enzyme digestion with enzymes *EcoRI* and *Hind III*. Subclones of the VP4 gene were made by digestion with *XbaI* (figure 2.1) and cloning the two fragments into the *BamHI* and *XbaI* sites of pGEM as described in the cloning diagram in figure 2.2. Two subclones (0.7kb and 2kb) and the full-length clone were obtained all in the pGEM vector. These three constructs were used in further sequencing reactions.

2. Sequencing of AHSV-3 VP-4

The M13 forward and reverse primers were available for sequencing in the pGEM vector. Additionally, an internal primer (ACGAGAACGAGAGCGG) was used to sequence the middle part of the large subclone as depicted in figure 2.1. Sequencing was performed as described above in 2.2.3.1. On average 500bp were sequenced in one run and all runs were performed at least twice. The fragments were put together in a contig as shown in figure 2.3. and used to produce a final nucleotide sequence (figure 2.4) using the ABI Prism Navigator analysis program.

The nucleotide sequence was subsequently used for a number of analyses described further in this section.

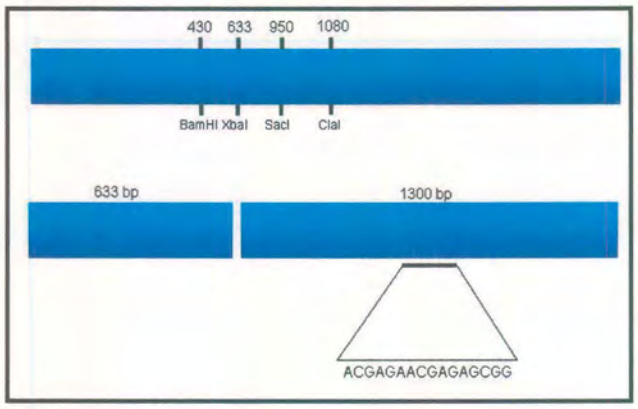


Figure 2.1. Restriction map of AHSV VP4 indicating the two subclones and the site of the internal primer used for sequencing.

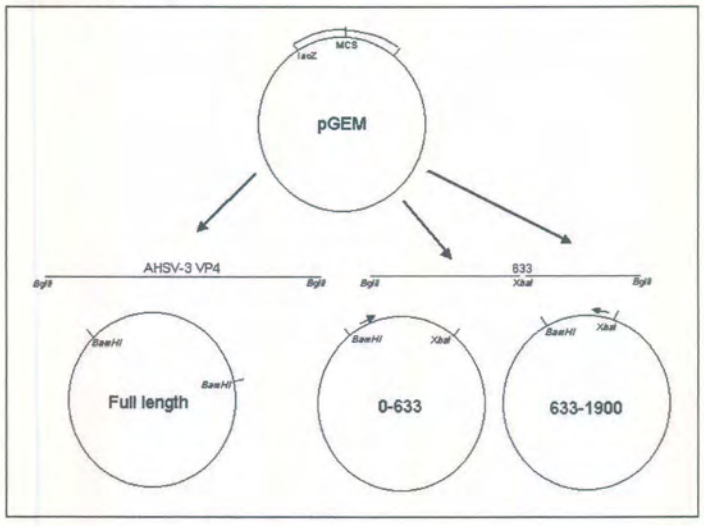


Figure 2.2 Cloning diagram for the cloning of the full-length VP4 cDNA and the subclones into the pGEM vector.

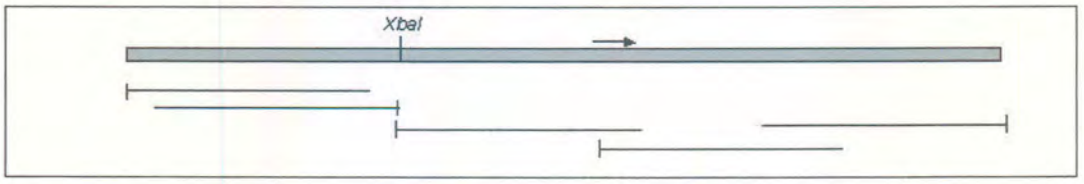


Figure 2.3 Sequence contig of the AHSV VP4 indicating that the complete gene was covered with the sequencing runs. The XbaI site used for subcloning is indicated and the site of the internal primer (arrow).

3. Analysis

3.1 Sequence analysis

In figure 2.4 the complete nucleotide sequence is shown as determined by sequencing experiments. The full-length sequence of AHSV-3 VP4 is 1972 nucleotides in length. There are very few differences between the different serotypes within AHSV for the nucleotide sequence of VP4. As shown in the nucleotide sequence comparison in figure 2.5, the VP4 sequence of BTV differs from AHSV in length by 6 nucleotides. This difference is attributed to five deletions and three insertions from the AHSV sequence relative to the BTV sequence. Compared to AHSV-3 VP4, the Chuzan VP4 gene is 6 nucleotides shorter but St. Croix virus 3 nucleotides longer. Interestingly, all sequences shown in figure 2.5 have the same 5' and 3' dinucleotide (5'GT-AC3'). Not including Chuzan virus, the first three nucleotides are the same between the sequences. The same is true for the last three nucleotides. From the sequence presented in figure 2.4 possible open reading frames were identified using ORF Finder and the most probable ORF was used to produce an amino acid sequence. The longest open reading frame starts at nucleotide 39 and extends up to nucleotide 1938 resulting in an ORF of 1926 bases long with an 11 bp long 5' untranslated region and a 38bp 3' untranslated region (underlined regions in figure 2.4). The 5' untranslated region of AHSV is three nucleotides longer the BTV genes, 2 nucleotides longer than St. Croix River virus VP4 and 1 nucleotide longer than Chuzan virus (figure 2.5). Within the BTV and AHSV serogroups the 5' untranslated region is completely identical while between the two species only 3 of the 8 nucleotides (BTV untranslated region) are the same. Between AHSV and Chuzan there are 4 similar while the UTR in AHSV is two nucleotides longer. The stop codon is the same for AHSV and BTV (TAA). Both have the same length 3' untranslated region. In this region 21 of the 38 bp are identical between the two species. Thus, the 3' region is more conserved (55%) than the 5' region (38%). The resulting polypeptide shown in figure 2.4 is 642 peptides in length and has a theoretical size of 75.781kDa with a pI of 8.33.

gtttatttagg

atggaaccttacgcaatattgtatgttacgcaggagatcgaatacctactcaaagatagttttctt

M E P Y A I L Y V T Q E I E Y L L K D S F L

ccaaagtgggaacttgatggaattaaggatcttaatacattatggttagagaggggcagaatggca

P K W E L D G I K D L N T L W L E R G R M A

tgtgacacatacgcaactggaaaaattgaacaatggtcggtgcggcagctacgcgcacatagattt

C D T Y A T G K I E Q W S V R Q L R A H R F

ttgtttataagcacgaaaaggaaaatcaggctgaaggattgcacgatttcacccgacatattcatt

L F I S T K R K I R L K D C T I S P D I F I

ttgaaaaaagaattaaaggaatataatatgaagagatttgaaacgttaattggcagaagaagggta

L K K E L K E Y N M K R F E T L I G R R R V

acgttaagaaagagtttcgggaatatgttaagggcctatgctttccaacacgtgactgttttacac

T L R K S F G N M L R A Y A F Q H V T V L H

ggaagtgaggcggagacgttaagttatgcggatccgaagagacacgtcgtgaaaggccagcccaaa

G S E A E T L S Y A D P K R H V V K G Q P K

gctgctccaatgtatgatcatccagacagatggtggcgggacgttgatgatggaccaaccgataag

A A P M Y D H P D R W W R D V D D G P T D K

aaattagttagtatgcttgattacattatatatagtgcggatgaagtgtattatgtcggatgtggt

K L V S M L D Y I I Y S A D E V Y Y V G C G

gatttaaaaacacttgaacaattcgcgtctagagacaggaagcggtttgacaggattaaatggata

D L K T L E Q F A S R D R K R F D R I K W I

tgcatagatccaatcgctccggaaacatcgtacgctaatgtaaaagttgtaaaagaaaaagttgtg

C I D P I A P E T S Y A N V K V V K E K V V

tctgcgcgtgatttgaaacattatttgatgcgcgatgaggttgagcgattattaatttgggatgtg

S A R D L K H Y L M R D E V E R L L I W D V

agtgcagatgggctgaaggggacaattgagtgggagaagcggaggtttaaggaggatcgtaatggt

S A D G L K G T I E W E K R R F K E D R N G

gaaaacatcgcggaagcgttgtgtgcagattttgctttagccttgattaaacatcgaataccagaa

E N I A E A L C A D F A L A L I K H R I P E

gagagtgatgaatatatttgcaggagctcttggctgctaccacagcctggggctccaataacgatg

E S D E Y I C R S S W L L P Q P G A P I T M

tatgagctgcgaaatctcatgcgtttggacggatattcacacgtcgaaaggaaacacataccaaga

Y E L R N L M R L D G Y S H V E R K H I P R

gcgtacgttcggaaaatcgatgcggaggttgcgagaagattagttgaagagtatcatggagaggat

A Y V R K I D A E V A R R L V E E Y H G E D

gtagggcgattgttgaaacgatctttatatgaagatatacatattgagcgcgctgatgggttaacg

V G R L L K R S L Y E D I H I E R A D G L T

gatggtgacgagagaacgagagcggatctcttttacttgacaaatatccggaatgctgcgtttatg

D G D E R T R A D L F Y L T N I R N A A F M

catgacgtatatggagtagttgaaaaaagttttatttcaaccctgtgggtttcgaacaggcagaat

H D V Y G V V E K S F I S T L W V S N R Q N

ttcacatatgatgatgtcccggttaataggaattttattacattacgcttttcgaagaagaatcga

F T Y D D V P V N R N F I T L R F S K K N R

cgagtgctcgactggaatggagcgatcttgtttctgatgtggcaacattcgaaggattttccaaaa

R V L D W N G A I L F L M W Q H S K D F P K

actatgaactatgaccccagttgggcgaagaactatgctgttattttctatcatgcgttaacgagt

T M N Y D P S W A K N Y A V I F Y H A L T S

ccggttccggacctttcattatgcagatttattggattaaggttgatgtcttcgactttgaggata

P V P D L S L C R F I G L R L M S S T L R I

aattcagatcgtgcacatcaagtgactgatattttgaagaaattgggcttgaacgtatcgggtcat

N S D R A H Q V T D I L K K L G L N V S G H

ctatttatttgtttaatgtcaaattcatatgttgctgatcttgattggtggtttcgtatgatcctg

L F I C L M S N S Y V A D L D W W F R M I L

gagtggtccgtcaaggatagagaagggaagctggcggcgttaagtgaagcgaaggcggaactcatt

E W S V K D R E G K L A A L S E A K A E L I

gagtggaaggatgagaaagcggacgaaccttggcacataaagaatgatttgctggcggcgttgttt

E W K D E K A D E P W H I K N D L L A A L F

gagtttatgtactttgcgaaacattttgagataaacgaggggtatgtcgagtcctggatacaatat

E F M Y F A K H F E I N E G Y V E S W I Q Y

ttgcgtaacgcttaacgcttaaaggtgacacttagcgaggggggaatcctaataaccttac

L R N A *

Figure 2.4. The nucleotide sequence of AHSV-3 VP4. The 5' and 3' untranslated regions are indicated in blue underlined. The possible start codons are shown in light blue with the probable open reading frame and resulting polypeptide shown in single letter amino acid nomenclature.

3.2 Homology search

The genes and amino acid sequences used in the comparison in figure 2.5 and figure 2.6 were identified through a homology search using the BLAST program from NCBI. For the nucleotide sequence the n-BLAST version was used while for the amino acid sequence the psi-BLAST version was used. The results of the n-BLAST search are summarized in table 2.1. The nucleotide and amino acid sequences were identified as that of AHSV VP4. The E-value used in the table is a parameter describing the chance hits one can expect when searching the database with this query. The lower the E-value the more significant the match is. The score is calculated by summing the scores of each letter-to-letter and letter-to-null position. These scores are derived from a substitution matrix that in this case is BLOSUM. The higher the score the better the alignment of the query with the identified database sequence. Therefore, the high score and low E-value give the highest homology and statistical value to each found sequence. AHSV-3 showed a very high level of homology with AHSV serotype 4 VP4 (score 1212). Other genes that showed significant homology were Chuzan virus VP4 (score 630), BTV serotype 2 VP4 (score 577) and Saint Croix River virus VP4 (score 349). From figure 2.6 it is calculated that the two AHSV serotypes have a 97% identity. In comparison, AHSV-3 VP4 has only a 49% amino acid identity with BTV VP4. Between all the orbivirus species there is a 24% amino acid conservation

Accession number	**Name**	**Score**	**E-value**
D14402	AHSV-4 VP4	1212	0.0
AB018087	Chuzan virus VP4	630	e-179
L13726	BTV-10 VP4	573	e-162
L08638	BTV-11 VP4	570	e-161
L08639	BTV-17 VP4	569	e-161
L08640	BTV-13 VP4	563	e-159
L08637	BTV-2 VP4	562	e-159
AF145401	St.Croix river VP4	349	9e-95

Table2.1. The results of BLAST searches with the nucleotide sequence of AHSV-3 VP4 showing the homology between related genes. The level of homology can be judged using the scores in the table. High scores indicate high levels of homology between two sequences.

A

	AHSV3	AHSV4	Chuzan	BTV10	BTV11	BTV17	BTV13	BTV2	St.Croix
AHSV3	100								
AHSV4	96.8	100							
Chuzan	58.1	58.3	100						
BTV10	57.4	57.2	56.9	100					
BTV11	57.4	57.6	56.6	98.7	100				
BTV17	57.1	57.2	56.4	98.5	99.4	100			
BTV13	57.2	57.4	56.8	98.4	98.4	98.1	100		
BTV2	56.8	56.5	57.1	89.4	89.6	89.4	89.2	100	
St.Croix	50.3	50	48.2	49.9	50.3	50.2	50.1	49.5	100

B

	AHSV3	AHSV4	Chuzan	BTV10	BTV11	BTV17	BTV13	BTV2	St.Croix
AHSV3	100/100								
AHSV4	97/99	100/100							
Chuzan	50/70	51/70	100/100						
BTV10	50/66	50/66	50/67	100/100					
BTV11	49/66	49/66	50/67	99/100	100/100				
BTV17	49/66	49/66	50/67	99/100	100/100	100/100			
BTV13	48/65	49/65	50/67	98/99	98/99	99/99	100/100		
BTV2	49/64	49/65	49/66	95/97	95/98	95/98	94/97	100/100	
St.Croix	36/52	37/53	34/55	36/52	35/52	35/52	35/52	35/52	100/100

Table2.2. Level of homology between the VP4 genes of all the orbivirus species for which the gene has been sequenced. Percentage homology is given for the nucleotide sequence of the different VP4 sequences in A and for the amino acid identity/similarity in B.

```
L13726.1 GGATCCGTATGCGCAAGGGATTTGGTGATGCATTGAGAAATTACGCCTTTAAGATG-GCT
L08640.1 GGATCCGTATGCGCAAGGGATTTGGTGATGCATTGAGAAATTACGCCTTTAAGATG-GCT
L08638.1 GGATCCGTATGCGCAAGGGATTTGGTGATGCATTGAGAAATTACGCCTTTAAGATG-GCT
L08639.1 GGATCCGTATGCGCAAGGGATTTGGTGATGCATTGAGAAATTACGCCTTTAAGATG-GCT
L08637.1 GAATCCGTATGCGTAAGGGATTTGGTGATGCGCTGAGGAATAAGCTCTTTAAGATG-GCT
AF246225.GGGTAACGTTAAGAAAGAGTTTCGGGAATATGTTAAGGGCCTATGCTTTCCAACAC-GTG
D14402.1 GGGTAACGTTAAGGAAGAGTTTCGGGAATATGTTAAGGGCTTATGCTTTCCAACAC-GTG
AB018087 GAATTGGATTACGGAAAAACTTCGGGGATCAGTTACGGAAGTACGCACTTTTGA---GTT
AF145401 GCGTCCCGCTTCGCCGAACTTTTGGTGACGAGATCCGGAAACACGCTGCCACCTTTTGTG
          *  *     *  *        ** **  *     *  *     * *

L13726.1 ATCGAATTTCA--CGGGTCGGAGGCGGAAACGTTGA--ACGATGCAAATCCTCGGTTA-C
L08640.1 ATCGAATTTCA--CGGATCGGAGGCGGAAACGTTGA--ACGATGCAAATCCTCGGCTA-C
L08638.1 ATCGAATTTCA--CGGGTCGGAGGCGGAAACGTTGA--ACGATGCAAATCCTCGGCTA-C
L08639.1 ATCGAATTTCA--CGGGTCGGAGGCGGAAACGTTGA--ACGATGCAAATCCTCGGCTA-C
L08637.1 ATCGAGTTTCA--CGGGTCGGAAGCGGAGACGTTAA--ACGATGCAAATCCTCGATTG-C
AF246225.ACTGTTTTACA--CGGAAGTGAGGCGGAGACGTTAA--GTTATGCGGATCCGAAGAGA-C
D14402.1 ACCGTTCTACA--CGGAAGTGAGGCGGAAACGTTGA--GTTATGCGGATCCGAAGAGA-C
AB018087 GTCGGTTCTTAAACGGTTCTGAAGCTGAGACATTAA--ATGTAGCTGAGCCAAATGTC-C
AF145401 ATAGCTTCTCA---GGATCGGAGGCCGAAACGATCATGATGCTGCGAACTGATAGGTATC
            *      *   **    ** ** ** **  * *       **  *           *

L13726.1 ATAAAATTTATGGAATGCCGGAAATACCGCCATTATACATGGAATATGCGGAAATAGGGA
L08640.1 ATAAAGTTTATGGAATGCCGGAAATACCGCCATTATACATGGAATATGCGGAAATAGGGG
L08638.1 ATAAAATTTATGGAATGCCGGAAATGCCGCCATTATACATGGAATATGCGGAAATAGGGA
L08639.1 ATAAAATTTATGGAATGCCGGAAATACCGCCATTATACATGGAATATGCGGAAATAGGGA
L08637.1 ATAAAATTTATGGAATGCCGGAAATGCCGCCACTATACATTGAATATGCGGAAATAGGGA
AF246225.ACGTCGTGAAAGGCCAGCCCAAAGCTGCTCCAATGTATG-ATCATCCAGACAGATGGTGG
D14402.1 ACGTCGTGAAAGGCCAGCCTAAAGCTGCTCCAATGTATG-ATCATCCAGACAGATGGTGG
AB018087 ATAAGGTGCGTGGGCTCACAATGGGTCCACCAACACACTTAAATCTTCTGGGAA--GTTA
AF145401 ACATCTCAGGTCGCACTCCGAACCCTCCGAATTCCAATTTTCCTTATCCTTCACTTCCCT
          *          *     *        *        *

L13726.1 CTAGAT-TTGACGATGAGCCGACTGATGAAAAGTTAGTATCAATGCTTGATTATATCGTT
L08640.1 CTAGAT-TTGACGATGAGCCGACTGATGAAAAGTTAGTATCAATGCTTGATTATATCGTT
L08638.1 CTAGAT-TTGACGATGAGCCTACTGATGAAAAGTTAGTATCAATGCTTGATTATATCGTT
L08639.1 CTAGAT-TTGACGATGAGCCTACTGATGAAAAGTTAGTATCAATGCTTGATTATATCGTT
L08637.1 CTAGAT-TTGACGATGAGCCTACTGATGAAAAGTTAGTATCGATGCTTGATTATATCGTT
AF246225.CGGGACGTTGATGATGGACCAACCGATAAGAAATTAGTTAGTATGCTTGATTACATTATA
D14402.1 CGGGATGTTGATGATGGACCAACCGATAAGAAATTAGTTAGTATGCTTGATTACATTATA
AB018087 TGAGA-----ATGATGGACCGACTGATGAAAAACTTGTTTCAATGATTGATTATATTACA
AF145401 TCAAG----GGGGATAGGGAGTGTGGTGACAAACTTGTTGCCATGCTCGACTATCTTATC
                  ***         * * * *** *  *    *** * ** **  *

L13726.1 TACAGTGCCGAAGA---GGTGCACTATGTTGGATGTGGTGACCTACGTACCCTAATGCAG
L08640.1 TACAGTGCCGAAGA---GGTGCACTATGTCGGATGTGGTGACCTACGTACCCTAATGCAG
L08638.1 TACAGTGCCGAAGA---GGTGCACTATGTTGGATGTGGTGACCTACGTACCCTAATGCAG
L08639.1 TACAGCGCCGAAGA---GGTGCACTATGTTGGATGTGGTGACCTACGTACCCTAATGCAG
L08637.1 TACAGCGCAGAGGA---GGTACACTATGTAGGATGCGGTGACCTACGTACCATTATGCAG
AF246225.TATAGTGCGGATGA---AGTGTATTATGTCGGATGTGGTGATTTAAAAACACTTGAACAA
D14402.1 TATAGTGCGGATGA---AGTGTATTATGTCGGATGTGGTGATTTAAAAACACTTGAACAA
AB018087 TATAGTGCTGATCA---AATATATTATATTGGGTGTGGGGATTTAAGAACTTTGAAATTG
AF145401 CCTCATGCTGAGTATCGAGCCATTTACGTAGGTTCTGGTCGCGGAGACACAGTTCAGTCT
            ** **  *          **  * ** *  **      *   **  *

L13726.1 TTCAAGAAACGATCACCAGGACGGTTTAGAAGGGTGTTATGG-CACGTATATGACCCAAT
L08640.1 TTCAAGAAACGATCACCAGGACGGTTTAGAAGGGTGTTATGG-CACGTATATGACCCAAT
L08638.1 TTCAAGAAACGATCACCAGGACGGTTTAGAAGAGTGTTATGG-CACGTATATGACCCAAT
L08639.1 TTCAAGAAACGATCACCAGGACGGTTTAGAAGAGTGTTATGG-CACGTATATGACCCAAT
L08637.1 TTCAAAAAACGTTCACCAGGACGGTTTAGAAGGGTATTATGG-CACGTATATCATCCAAT
AF246225.TTCGCGTCTAGAGACAGGAAGCGGTTTGACAGGATTAAATGGATATGCATA-GATCCAAT
D14402.1 TTCGCGTCTAGAGATAGGAAGCGGTTTGACAGAATTAAATGGATATGCATA-GATCCAAT
AB018087 TTTGCAAAAAGAGATAAGAGAAGATTCAATCGCGTGCGTTGGTTTTGC-TTTGATCCTAT
AF145401 TTCGCACGCCGTTCCCCGCGAAGGTTCTCACTCGGCCAGTGGGTTCTCATTGACGCTAAT
          **        *          * **             ***       *     *  **
```

```
L13726.1 AG--CACCTGAGTGTTCAGATCCAAACGTTATAGTTCATAATATTATGGTGGATTCAAAG
L08640.1 AG--CACCTGAGTGTTCAGATCCAAACGTTATAGTTCATAATATTATGGTGGATTCAAAG
L08638.1 AG--CGCCTGAGTGTTCAGATCCGAACGTTATAGTTCATAATATTATGGTGGATTCAAAG
L08639.1 AG--CACCTGAGTGTTCAGATCCGAACGTTATAGTTCATAATATTATGGTGGATTCAAAG
L08637.1 AG--CACCGGAATCTTCAGATCCAAACGTTATAGTTCATAATGTTATGGTGGATTCAAAG
AF246225.CG--CTCCGGAAACATCGTACGCTAATGTAAAAGTTGTAAAAGAAAAAGTTGTGTCTGCG
D14402.1 CG--CTCCGGAAACATCGTACGCTAATGTAAAGGTTGTAAGAGAAAAAGTTGTGTCTGCG
AB018087 CG--CTCCGGAAAGTTTTGCTGAGAACGTATTAGTGAGGAAAACAAAGATAGAAAATTAT
AF145401 GTGTCATGGGAGGCCCCCCCTTCTAACGTACTTCTGTATCAGGGGTATGTCCATTCTGTC
             *    **             ** **     *            *

L13726.1 AAAGACATTTTGAAACAT-ATGAATTTTTTGAAACGT-GTTGAGAGACTTTTCATATGGG
L08640.1 AAAGACATTTTGAAATAT-ATGAATTTTTTGAAACGT-GTTGAGAGACTTTTCATATGGG
L08638.1 AAAAACATTTTGAAACAT-ATGAATTTTTTGAAACGT-GTTGAGAGACTTTTCATATGGG
L08639.1 AAAGACATTTTGAAACAT-ATGAATTTTTTGAAACGT-GTTGAGAGACTTTTCATATGGG
L08637.1 AAAGACATTTTGAAACAT-ATGAACTTTTTAAAACGT-GTTGAAAGGCTCTTCATATGGG
AF246225.CGTGA---TTTGAAACATTATTTGATGCGCGATGAG--GTTGAGCGATTATTAATTTGGG
D14402.1 CGTGA---TTTGAAACATTATTTAATGCGCGATGAG--GTCGAGCGATTATTGATCTGGG
AB018087 AAAGA---TCTGAAAAAATATATGGATAGCGGCTCG--GTAGAAAGAGTATTGATATGGG
AF145401 GACGA-----CGTGGCCCAGTTCCTTGTTCCAGGCGCCCTTAACCAAATTCTCATTTGGG
             *      *        *                *   * *     *  * ** ****

L13726.1 ATGTATCGTCAGATCGATCGCAGATGAATGATCATGAGTGGGAGACGACCAGGTTTG-CG
L08640.1 ATGTATCGTCAGATCGATCGCAGATGAATGATCATGAGTGGGAAACAACCAGGTTTG-CG
L08638.1 ATGTATCGTCAGATCGATCGCAGATGAATGATCATGAGTGGGAGACAACCAGGTTTG-CG
L08639.1 ATGTATCGTCAGATCGATCGCAGATGAATGATCATGAGTGGGAGACAACCAGGTTTG-CG
L08637.1 ATGTGTCATCAGATCGATCGCAGATGGATGATGATGAATGGGAATCGACCAGATTCG-CA
AF246225.ATGTGAGTGCAGATGGGCTGAAGG-GGACAAT--TGAGTGGGAGAAGCGGAGGTTTA-AG
D14402.1 ATGTGAGTGCAGATGGGCTGAAGG-GGACAAT--TGAGTGGGAGAAGCAGAGGTTTA-AG
AB018087 ATGTGAGTGGAGATGGAAAGAAAG-GGACGAT--TGAATGGGA-ACAACAACGTGCATCA
AF145401 ACGTCAGGACTGACAATGTCGGTTTGAGTAAGTTCCAATGGGAGG-AGCGCGCCATGGAG
         * **       **             *    *     * *****

L13726.1 GAGGATAGATTGGGTGAGGAAATAGCTTATGAAATGGGTGGTGCATTTTCCAGTGCATTG
L08640.1 GAGGATAGATTGGGTGAGGAAATAGCTTATGAAATGGGTGGTGCATTTTCCAGCGCATTG
L08638.1 GAGGATAGATTGGGTGAGGAAATAGCTTATGAAATGGGTGGTGCATTTTCCAGTGCATTG
L08639.1 GAGGATAGATTGGGTGAGGAAATAGCTTATGAAATGGGTGGTGCATTTTCCAGTGCATTG
L08637.1 GAGGATAGGTTGGGTGAGGAGATAGCTTATGAAATGGGCGGTGCGTTCTCCAGTGCATTG
AF246225.GAGGATCGTAATGGTGAAAACATCGCGGAAGCGTTGTGTGCAGATTTTGCTTTAGCCTTG
D14402.1 GAGGATCGTAATGGTGAAAACATCGCGGAAGCGTTGTGTGCAGATTTTGCTTTAGCTTTG
AB018087 GAAGACAGGATGGGGGAGCAAATTGCAAAAGGATTAACTAAATTTTTTAGTTTTGCTGTA
AF145401 CAGGACATGTTAGGCGAGAACACCGCTGAGGCCTTGCAGTCGGAACTCGCTGCCGCACTT
         * **       ** **  * *  **  * *   *                    **  *

L13726.1 ATCAAGCACAGGATACCGAATTCAAAAGACGAATATCACTGCATTTCGACCTAC---TTA
L08640.1 ATTAAACACAGGATACCGAATTCAAAAGACGAATATCACTGCATTTCGACCTAC---TTA
L08638.1 ATTAAGCACAGGATACCGAATTCAAAAGACGAGTATCACTGCATTTCGACTTAC---TTA
L08639.1 ATTAAGCACAGGATACCGAATTCAAAAGACGAGTATCACTGCATTTCGACTTAC---TTA
L08637.1 ATAAAACATAGGATACCAAATTCAAGAGACGAGTATCACTGCATTTCGACTTAC---CTT
AF246225.ATTAAACATCGAATACCAGAAGAGAGTGATGAATATATTTGCAGGAGCTCTTGG---CTG
D14402.1 ATTAAACATCGAATACCAGAAGAGAGTGATGAATATATTTGCAGGAGCTCTTGG---CTA
AB018087 ATTAAACATCGAATTCCAAAGGATAACGAAGTCTATAGTTGCTATTCATCTCTA---ATT
AF145401 CTTAAACATAGAATCCCGCAGTTCTCAGATAATCGGGATGAAGTGTTCACTTCCACTCTT
         * ** **  * ** **  *       **                    *       *

L13726.1 TTCCCCCAACCGGGGGCCGATGCGGATATGTATGAATTAAGGAATTTCATGAGATTGAGA
L08640.1 TTCCCCCAACCGGGGGCCGATGCGGATATGTATGAATTAAGGAATTTCATGAGATTGAGA
L08638.1 TTCCCCCAACCGGGGGCCGATGCGGATATGTATGAATTAAGGAATTTCATGAGATTGAGA
L08639.1 TTCCCCCAACCGGGGGCCGATGCGGATATGTATGAATTAAGGAATTTCATGAGATTGAGA
L08637.1 CTCCCCCAGCCGGGGGCTGATGCGGATATGTACGAATTAAGGAATTTCATGAGATTGAAA
AF246225.CTACCACAGCCTGGGGCTCCAATAACGATGTATGAGCTGCGAAATCTCATGCGTTTGGAC
D14402.1 ATACCACAGCCCGGGGCTCCAATAACGATGTATGAGCTGCGAAATCTCATGCGTTTGGAC
AB018087 ATCCCACAACCGGGTGCGGCGCAAGATATGTATGAATGTAGGAATATAATAAAATTAGAA
AF145401 CTGCCTCAGCCCGGAGCGCCCGCTGGCATGTATGAAATGCGGAACTTTTGTCGCCTCGAC
         * ** ** ** ** **          ***** **     * **  *        *
```

```
L13726.1 GGTTACTCACACGT-GGATCGCCACATGCAT-CCAGACGCTTCCGTGACGAAAGTTGTTT
L08640.1 GGTTACTCACACGT-GGATCGCCACATGCAT-CCAGACGCTTCCGTGACGAAAGTCGTTT
L08638.1 GGTTACTCACACGT-GGATCGCCACATGCAT-CCAGACGCTTCCGTGACGAAAGTTGTTT
L08639.1 GGTTACTCACACGT-GGATCGCCACATGCAT-CCAGACGCTTCCGTGACGAAAGTTGTTT
L08637.1 GGTTACTCACATGT-AGATCGCCATATGCAC-CCAGATGCATCCGTAATGAAAGTTGTTT
AF246225.GGATATTCACACGTCGAAAGGAAACACATAC-CAAGA-GCGTACGTTCGGAAAATCGATG
D14402.1 GGATATTCACACGTCGAAAGGAAACACATAC-CAAGA-GCGTACGCTCGGAAAATCGATG
AB018087 GGATTTTCAAAAGTCGATAGGACACATTTGG--GAGAGGCGAGAATTCATTATGTTTCAC
AF145401 GGTC-CCCGTACTTTTAACCGCAGCATCCCCACCGCGCAAAGGCATCCCATCAAATACGA
         **     *  *  *      *    *

L13726.1 CACGTGATGTGCGCAAAATGGTCGAATTGTATCACGGTCGCGATCGTGGTAGATTTCTAA
L08640.1 CTCGTGATGTGCGCAAAATGGTCGAATTGTATCACGGTCGCGATCGTGGTACATTTCTAA
L08638.1 CGCGTGATGTGCGCAAAATGGTCGAATTGTATCACGGTCGCGATCGTGGTAGATTTCTAA
L08639.1 CGCGTGATGTGCGCAAAATGGTCGAATTGTATCACGGTCGCGATCGTGGTAGATTTCTAA
L08637.1 CGCGTGATGTACGCAAAATGGTCGAATTGTATCATGGTCGCGATCGTGGCAGATTTGTGA
AF246225.CGGAGGTTGCGAGAAGATTAGTTGAAGAGTATCATGGAGAGGATGTAGGGCGATTGTTGA
D14402.1 CGGAAGTTGCGAGAAGATTAGTTGAAGAGTATCATGGAGAGGATGTAGGGCGGTTGTTGA
AB018087 CCAAAGATTTGAGACTGTTGATACACCGATTTCATGGCTTTGGGAGAGGGAGGAAATTGA
AF145401 TGAGTGTCGCGC-TATGGTGGAAGAGCTGCATC--GATCGCG-TAGGGGCGCCAATCTTA
              *             *     *      **  *     *     **         * *

L13726.1 AAAAGAGACTATTTGAACACCTTCATATTGTACGTAAGAATGGATTGTTACACGAAAGTG
L08640.1 AAAAGAGACTATTTGAACACCTTCATATTGTACGTAAGAATGGATTGTTACACGAAAGTG
L08638.1 AAAAGAGACTATTTGAACACCTTCATATTGTACGTAAGAATGGATTGTTACACGAAAGTG
L08639.1 AAAAGAGACTATTTGAACACCTTCATATTGTGCGTAAAAATGGATTGTTACACGAAAGTG
L08637.1 AGAACAGGCTATTTGAACACCTTCATATCGTACGTAAGAATGGATTGCTGCATGAAAGTG
AF246225.AACGATCTTTATATGAAGATATACATATTGAGCGCGCTGATGGGTTAACGGATGGTGACG
D14402.1 AACGATCTTTATATGAAGATATACATATTGAGCGCGCTGATGGGTTAACGGATGGCGACG
AB018087 AGAAAAGTATTTTTGAATATTTACATATTGAGAGAGTAAATGGACT---AGATTCTCTAG
AF145401 GACGGCGCATCTTTGAGTTTCTGCACATACAGGACGAGGACGGGCTGTTGCATTTCGGAG
                  * * ***     * *****           * **  *             *

L13726.1 ATGAGCCACGAGCGGATCTGTTTTATTTGACCAATCGGTGCAATATGGGATTGGAGCCTA
L08640.1 ATGAGCCACGAGCGGATCTGTTTTATTTGACCAATCGGTGCAACATGGGATTGGAGCCTA
L08638.1 ATGAGCCACGAGCGGATCTGTTTTATTTGACCAATCGGTGCAACATGGGATTGGAGCCTA
L08639.1 ATGAGCCACGAGCGGATCTGTTTTATTTGACCAATCGGTGCAACATGGGATTGGAGCCTA
L08637.1 ATGAACCACGAGCAGATCTGTTTTATTTGACCAATCGGTGCAACATGGGACTGGAACCAA
AF246225 AGAGAACGAGAGCGGATCTCTTTTACTTGACAAATATCCGGAATGCTGCGTTTATGCATG
D14402.1 AGAGAACGAGAGCGGATCTCTTTTACTTGACAAATATGCGGAATGTTGCGTTTATGCATG
AB018087 ACGAGCCAAGAGCTGATTTGTTTTACCTCACAAATCATAGAAATGCAGCGCGAGTTGAGG
AF145401 AAAAGACC---GCGCATCTTTTCTATTTGACCAATTCATGTAATGAGGAACACATCGGTG
         *     *    **  ** * ** **  * ** ***    * **    *

L13726.1 GTATTTATGAGGTAATGAAGAAGTCGGTGATAGCTACTGCTTGGGTGGGCCGTGCGCCTT
L08640.1 GTATTTACGAGGTAATGAAGAAGTCGGTGATAGCTACTGCTTGGGTGGGCCGTGCGCCTT
L08638.1 GTATTTATGAGGTAATGAAGAAGTCGGTGATAGCTACTGCTTGGGTGGGCCGTGCGCCTT
L08639.1 GTATTTATGAGGTAATGAAGAAGTCGGTGATAGCTACTGCTTGGGTGGGCCGTGCGCCTT
L08637.1 GCATTTATGAAGTAATGAAGAAGTCGGTGATCGCTACAGCCTGGGTAGGCCGTGCTCCTC
AF246225 ACGTATATGGAGTAGTTGAAAAAAGTTTTATTTCAACCCTGTGGGTTTCGAACAGGCAGA
D14402.1 ATGTATATCGAGTAGTTGAAAAAAGTTTTATCTCTACCCTGTGGGTTTCGAACAGGCAGA
AB018087 ACATTAGAAGAGTAGTTGAACAAAGTACGATCTCAACATTATGGGTAGGAAAGAGACCAC
AF145401 ACCTACGCGCCATCGTTCGTAGTGCTGGAATCGCAACTCTGTGGGTTGGAGGCGAGATTT
            *        *  *              **  * **    *****

L13726.1 TATATGATTATGATGATTTCGCGTTACCCAGATCTACCGTTATGCTCAACGGATCCTACC
L08640.1 TATATGATTATGATGATTTCGCGTTACCCAGATCTACCGTTATGCTCAACGGATCCTACC
L08638.1 TATATGATTATGATGATTTCGCGTTACCCAGATCTACCGTTATGCTCAACGGATCCTACC
L08639.1 TATATGATTATGATGATTTCGCGTTACCCAGATCTACCGTTATGCTCAACGGATCCTACC
L08637.1 TATATGATTATGATGATTTCGCGTTACCCAGATCTACTGTCATGCTCAATGGATCTTACC
AF246225.ATTTCACATATGATGATGTCCCGGTTAATAGGAATTTTATTACATTACGCTTTTCGAAGA
D14402.1 ATTTCACATATGATGATGTCCCGGTTAATAGGAATTTTATTACATTACGCTTTTCGAAGA
AB018087 TATATGATTATCCTGATTTTAGATATCCGAGATGTGATGCCATGTTAAGATTTAGCAATA
AF145401 T---CGGCTATCCGGATTTCTCGTACGATAGACGGCTCGCTATTGCTGATTTCTGCACGA
                ***   *** *          **          *
```

```
L13726.1 GCGATATCAGAATTCTAGATGGCAATGGCGCAATCCTATTCCTAATGTGGAGGTACCCGG
L08640.1 GCGATATCAGTATTCTAGATGGCAATGGCGCAATCCTATACCTAATGTGGAGGTACCCAG
L08638.1 GCGACATCAGAATTCTAGATGGCAATGGCGCAATCCTATTCCTAATGTGGAGGTACCCAG
L08639.1 GCGACATCAGAATTCTAGATGGCAATGGCGCAATCCTATTCCTAATGTGGAGGTACCCAG
L08637.1 ATGATATCAGAATTCTAGATGGTAATGGCGCAATCCTATTTCTAATGTGGAAGTACCCAG
AF246225.AGAATCGACGAGTGCTCGACTGGAATGGAGCGATCTTGTTTCTGATGTGGCAACATTCGA
D14402.1 AGAATCGACGAGTGCTCGACGGGAATGGAGCAATCCTGTTTCTGATGTGGCAGCATCCGA
AB018087 TTACCAACAGAGTCTTCGATGGCAATGGAGCTTTATTATATTTGATGTGGAAGTATCCTG
AF145401 AGGACAGAATGGTGTATTCCGGTCTAGGATACATCCTCTTACTTATGTGGGAAGGTTCAG
                  *       *    **     *  * *   * ******       *

L13726.1 ATATCGTAAAGAAAGATTTGACGTATGACCCTGCGTGGGCGATGAATTTTGCTGTT-TCG
L08640.1 ATATCGTAAAGAAAGATTTGACGTATGACCCTGCATGGGCGATGAATTTTGCTGTT-TCG
L08638.1 ATATCGTAAAGAAAGATTTGACGTATGACCCTGCGTGGGCGATGAATTTTGCTGTT-TCG
L08639.1 ATATCGTAAAGAAAGATTTGACGTATGACCCTGCGTGGGCGATGAATTTTGCTGTT-TCG
L08637.1 ACATTGTGAAGAAAGATTTAACATATGACCATGCGTGGGCAATGAATTTCGCTGTC-TCG
AF246225.AGGATTTTCCAAAAACTATGAACTATGACCCCAGTTGGGCGAAGAACTATGCTGTT-ATT
D14402.1 AGGATTTTCCAAAAACTATGAACTATGACCCCAGTTGGGCGGAGAACTATGCTGTC-ATT
AB018087 AAAAATTCAGTAAGAAGATTAATTATGACCCAGCGTGGGCAGAAAATTATTGTGTC-ATT
AF145401 TTCCCACTTCG------CTGCCTTTCGACCCCTGGTGGGCCGACTCTTTTGCCGTTATCG
                         *    *  ****     *****       *     **

L13726.1 CTAAAGGAGCCGATACCTGATCCTCCTGTGCCTGATATTTCTTTGTGTAGGTTCATCGGA
L08640.1 CTAAAGGAGCCGATACCTGATCCTCCTGTGCCTGATATCTCTTTGTGTAGGTTCATCGGA
L08638.1 CTAAAGGAGCCGATACCTGATCCTCCTGTGCCTGATATTTCTTTGTGTAGGTTCATCGGA
L08639.1 CTAAAGGAACCGATACCTGATCCTCCTGTGCCTGATATTTCTTTGTGTAGGTTCATCGGA
L08637.1 TTAAAGGAGCCGATACCCGATCCCCCTGTGCCCGATATCTCCCTATGCAGATTTATCGGA
AF246225.TTCTATCATGCGTTAACGAGTCCG---GTTCCGGACCTTTCATTATGCAGATTTATTGGA
D14402.1 TTCTATCATGCGTTAACGAGTCCA---GTTCCGGATCTTTCATTATGCAGATTTATTGGA
AB018087 TTAAAAGAGGATATTCCAAGAATACCTGTTCCAGAGCTTTCACTATGTAGATTTATTGGT
AF145401 TCAAGAGACGCGAGGTTCCCTTCC----TCCCAGACGTCTCTTTGTGCAGGTTTATCGGT
           *                        * ** **  * **  * ** ** ** ** **

L13726.1 CTGCGCGTCGAATCATCCGTGCTGAGGGTCCGAAACCCAA--CAT-TACATGAGACGGCT
L08640.1 CTGCGCGTTGAATCATCCGTGCTGAGGGTCCGAAACCCAA--CAT-TACATGAAACGGCT
L08638.1 CTGCGCGTTGAATCATCCGTGCTGAGGGTCCGAAATCCAA--CAT-TACATGAAACGGCT
L08639.1 CTGCGCGTGGAATCATCCGTGCTGAGGGTCCGAAATCCAA--CAT-TACACGAAACGGCT
L08637.1 CTGCGCGTTGAGTCATCTGTGTTGAGGGTCCGAAACCCAA--CAT-TACATGAAACGGCT
AF246225.TTAAGGTTGATGTCTTCGACTTTGAGGAT---AAATTCAGATCGTGCACATCAAGTGACT
D14402.1 TTAAGGTTGATGTCTTCGACTTTGAGGAT---AAATTCAGATCGTGCACATCAAGTGACT
AB018087 TTGAGGACAGTAAGCACTATGATGAGAGTTC--AAACAAATTCAG-TTCATCAGATTTCC
AF145401 ATTAGAGAACGGTCCTCTCAACTACGTCTCCGCTTTCGAG---AGGTGCACGATGTTATG
          *  *          *     *  *  *          *          **  *

L13726.1 GACGAACTGAAACGGATGGGATTGGATTTGTCTGGTCATTTATATGTCACATTAATGTCC
L08640.1 GACGAACTGAAACGGATGGGATTGGATTTGTCTGGTCATTTATATGTCACATTAATGTCC
L08638.1 GACGAACTGAAACGGATGGGATTGGATTTGTCTGGTCATTTATATGTCACATTAATGTCC
L08639.1 GACGAACTGAAACGGATGGGATTGGATTTGTCTGGTCATTTATATGTCACATTAATGTCC
L08637.1 GATGAACTAAAACGGATGGGATTGGATTTATCTGGTCATCTATACGTCACATTGATGTCC
AF246225.GATATTTTGAAGAAATTGGGCTTGAACGTATCGGGTCATCTATTTATTTGTTTAATGTCA
D14402.1 GATATCTTGAAGAAATTGGGCTTGGACGTATCGGGTCATCTATTTATTTGTTTAATGTCA
AB018087 GATAAGGTGAAGAAGATGGGGTTAGATCTGTCAGGACATTTGTTTATTGCTTTGATGTCT
AF145401 GATGTAGTCAAGTCCCTGGGCGTCGACGCTTCAGGCCATCTGTTCATCGGGATTCTCTCC
         **     * **     ****  *  *    ** ** *** * *   *     *  * **

L13726.1 GGCGCTTATGTCACAGATCTGTTTTGGTGGTTTAAGATGATTCTAGATTGGTCTGCGCAA
L08640.1 GGCGCTTATGTCACAGATCTGTTTTGGTGGTTTAACATCATTCTAGATTGGTCTGCGCAA
L08638.1 GGCGCTTATGTCACAGATCTGTTTTGGTGGTTTAAGATGATTCTAGATTGGTCTGCGCAA
L08639.1 GGCGCTTATGTCACAGATCTGTTTTGGTGGTTTAAGATGATTCTAGATTGGTCTGCGCAA
L08637.1 GGCGCTTATGTCACAGATCTGTTCTGGTGGTTCAAGATGATTCTCGATTGGTCTGCGCAA
AF246225.AATTCATATGTTGCTGATCTTGATTGGTGGTTTCGTATGATCCTGGAGTGGTCCGTCAAG
D14402.1 AATTCATATGTTGCTGATCTTGATTGGTGGTTTCGTATGATCCTGGAGTGGTCCGTCAAG
AB018087 GATAGTTACATTTCGGATATTATGTGGTGGTTTGACATGATATTAAATTGGTCGGTCTTA
AF145401.TCCAACTATGTCTTTGATCCGTACATCTGGGTCAACATGATCATGCTGTGGTCACAGCAA
           **  *    ***         *** *     ****  *    *****
```

```
L13726.1 AACAGGGAGCAAAAACTACGCGATCTAAAGAGGTCTGCGGCGGAAGTAATTGAGTGGAAG
L08640.1 AACAAGGAACAAAAACTACGCGATCTAAAGAGGTCTGCGGCGGAAGTAATTGAGTGGAAG
L08638.1 AACAAGGAACAAAAACTACGCGATCTAAAGAGGTCTGCGGCGGAGGTAATTGAGTGGAAG
L08639.1 AACAAGGAACAAAAACTACGCGATCTAAAGAGGTCTGCGGCGGAGGTAATTGAGTGGAAG
L08637.1 AGCAAGGAGCAAAAATTACGTGATTTAAAGAGGTCTGCAGCGGAGGTAATCGAATGGAAG
AF246225.GATAGAGAAGGGAAGCTGGCGGCGTTAAGTGAAGCGAAGGCGGAACTCATTGAGTGGAAG
D14402.1 GATAGAGAAGGGAAGCTGGCGGCGTTAAGTGAAGCGAAGGCGGAACTTATCGAGTGGAGG
AB018087 AATCGTGAAGAGAAAAAGAGGAAGTTATTTGAAATGAATGCGGAAGTTATTGAATGGAAA
AF145401.TCTAGAACGGACAAGATCAGAGACATAGAGTCACACGGTGCGGCGGTCATCGAGTGGGAT
                  **          **           ****   * ** ** ***
L13726.1 GAGCAGATGGCCGAGCGTCCATGGCATGTGAGAAATGATCTGACTCGT---------GCG
L08640.1 GAGCAGATGGCCGAGCGTCCATGGCATGTGAGAAATGATCTGATTCGT---------GCG
L08638.1 GAGCAGATGGCCGAGCGTCCATGGCATGTGAGAAATGATCTGATTCGT---------GCG
L08639.1 GAGCAGATGGCCGAGCGTCCATGGCATGTGAGAAATGATCTGATTCGT---------GCG
L08637.1 GAGCAGATGGCCGAGCGTCCATGGCACGTGAGAAATAGTCTAATTGCA---------GCG
AF246225.GATGAGAAAGCGGACGAACCTTGGCACATAAAGAATGATTTGCTGGCG---------GCG
D14402.1 GATGAGAAAGCGGATGAGCCTTGGCACATAAAGAATGATTTGCTGGCG---------GCG
AB018087 GAAGATCGAGCGAATGAGCCTTGGCACATTAAACCAGATTTAATTGCA---------GCG
AF145401 CGGTCCCACATGGATAAACCGTGGCACTCGTTCGATGATCTTACTGCTTCCCTCGTGGCT
                    *    ** *****          * * *                  *
L13726.1 CTAAGGGAA--------TACAAA-------CGGAAAATGGGGATGAGAGAGGGAGCCTCG
L08640.1 CTAAGGGAA--------TACAAA-------CGGAAAATGGGGATGAGAGAGGGAGCCTCG
L08638.1 CTAAGGGAA--------TACAAA-------CGGAAAATGGGGATGAGAGAGGGAGCCTCG
L08639.1 CTAAGGGAA--------TACAAA-------CGGAAAATGGGGATGAGAGAGGGAGCCTCG
L08637.1 TTGAGGGAG--------TACAAA-------CGGAAAATGGGGATAAGAGAGGGAGCCTCG
AF246225.TTGTTTGAG--------TTTATGTACTTTGCGAAACATTTTGA-GATAAACGAGGGGTAT
D14402.1 TTGTTTGAG--------TTTATATACTTTGCGAAGCATTTTGAT-ATAAACGAGAGGTAT
AB018087 TTGATGGAA--------TTTTCAAGATTT-ATTAATGTCATGATGACGGATGAGGCCTGT
AF145401 TTAGGGAAAATCCTGCCTCCTCAACGTTATCGATCCTTTCTGAGTCGAGTACGAGGATGG
          *    *         *                  *   **                *
L13726.1 ATTGATTCGTGGCTAGAATTACTGCGTCACTTATAATGCGTGACTGCTAGG--TGAGGGG
L08640.1 ATTGATTCGTGGCTAGAATTACTGCGTCACTTATAATGCGTGACTGCTAGG--TGAGGGG
L08638.1 ATTGATTCGTGGCTAGAATTACTGCGTCACTTATAATGCGTGACTGCTAGG--TGAGGGG
L08639.1 ATTGATTCGTGGCTAGAATTACTGCGTCACTTATAATGCGTGACTGCTAGG--TGAGGGG
L08637.1 ATCGATTCGTGGCTAGAATTACTGCGTCACTTATAATGCGTGACTGCTAGG--TGAGGGG
AF246225.GTCGAGTCCTGGATACAATATTTGCGTAACGCTTAAAG-GTGACACTTAGC--GAGGGGG
D14402.1 GTCGAGTCCTGGATACAATATTTGCGTAACGCTTAAAG-GTGACACTTAGC--GAGGGGG
AB018087 GTCACAAAGTGGATACTGTACCTCCGTTCAAAAGAGTAGGTGA-GACTCGA--AGAGGCT
AF145401 CTGGACTCCTATGCGTAGACTAACCCTCGCGCATAGTGGGCTCTCCTTAGAACCCGAGTG
          *       *                 * *       *    *       * *       *
L13726.1 GGCATGTACAAC-T----------------TAC
L08640.1 GGCATGTACAAC-T----------------TAC
L08638.1 GGCATGTACAAC-T----------------TAC
L08639.1 GGCATGTACAAC-T----------------TAC
L08637.1 GGCATGTACAAC-T----------------TAC
AF246225.AATCCTAATAACCT----------------TAC
D14402.1 AATCCTAATAACCT----------------TAC
AB018087 TTCATGAAACAACT---------------C-AC
AF145401 GGTTTGTACGGTGTGAAGGCCTCTTTTACCTAC
               *     *                **
```

Figure 2.5 The nucleotide sequences of all orbivirus species VP4 are aligned using Clustal software. Conserved nucleotides are marked by a star underneath the alignment. In the nucleotide alignment the start codon is shown in blue and the stop codon in pink

```
AHSV-4   M-EPYAILYVTQEIEYLLKDSFLPKWELDGIKDLNTLWLERGRMACDTYAIGKIEQWSVR
AHSV-3   M-EPYAILYVTQEIEYLLKDSFLPKWELDGIKDLNTLWLERGRMACDTYATGKIEQWSVR
BTV-11   MPEPHAVLYVTNELSHIVKNGFLPIWKLTGDESLNDLWLENGKYATDVYAYGDVSKWTIR
BTV-17   MPEPHAVLYVTNELSHIVKNGFLPIWKLTGDESLNDLWLENGKYATDVYAYGDVSKWTIR
BTV-10   MPEPHAVLYVTNELSHIVKDGFLPIWKLTGDESLNDLWLENGKYATDVYAYGDVSKWTIR
BTV-13   MPEPHAVLYVTNELSHIVKNGFLPIWKLTGDESLNDLWLENGKYATDVYAYGDVSKWTIR
BTV-2    MPEPHAVLYVTNELSHIVKSGFLPIWRLTGVESLNVLWLENGKYATDVYAYGDVSKWTIR
Chuzan   M-EPCAVLHVSGDIVPLIEKGFLPIFDVNRISTLNDLWIERGKFGTECYCYGTLTKWTIR
StCroix  MDPSHVVLYLSRSLCRYLGAG-ADLLTLRGAEDPQWLWQASSTRFKDFYCTGPIHNFSIR
         *  . .:*::: .:   :  .     :   .  : **   .    : *. * : ::::*

AHSV-4   QLRAHRFLFISTKRK-IRLKDCTISPDIFILKKELKEY-DMKRFETLIGRRRVTLRKSFG
AHSV-3   QLRAHRFLFISTKRK-IRLKDCTISPDIFILKKELKEY-NMKRFETLIGRRRVTLRKSFG
BTV-11   QLRGHGFIFISTHKN-VQLADIIKTVDVRIPREVARSH-DMKAFENEIGRRRIRMRKGFG
BTV-17   QLRGHGFIFISTHKN-VQLADIIKTVDVRIPREVARSH-DMKAFENEISRRRIRMRKGFG
BTV-10   QLRGHGFIFISTHKN-VQLADIIKTVDVRIPREVARSH-DMKAFENEIGRRRIRMRKGFG
BTV-13   QLRGHGFIFISTHKN-VQLADIIKTVDVRIPREVARSH-DMKAFENEIGRRRIRMRKGFG
BTV-2    QLRGHGFIFISTHKN-IQLADIIKTVDVRIPREVAKSQ-DMKAFENEIGRRRIRMRKGFG
Chuzan   QLRGHRFIFVSTKKR-IHLKDTIITPDVVISQQVLSG--GLRELETIIGFKRIGLRKNFG
StCroix  QLRGYSFLFVVYEGDRVRCLDGDVPRDVVITPHQLSSLKGQKDLESLIGKARVPLRRTFG
         ***.: *:*:  .   ::  *   . *: *  .      . : :*. *.  *: :*: **

AHSV-4   NMLRAYAFQHVTVLHGSEAETLSYADPKRHVVKGQPKAAPMYDHPDRWWRDVDDGPTDKK
AHSV-3   NMLRAYAFQHVTVLHGSEAETLSYADPKRHVVKGQPKAAPMYDHPDRWWRDVDDGPTDKK
BTV-11   DALRNYAFKMAIEFHGSEAETLNDANPRLHKIYGMPEMPPLYMEYAEIGTRFDDEPTDEK
BTV-17   DALRNYAFKMAIEFHGSEAETLNDANPRLHKIYGMPEIPPLYMEYAEIGTRFDDEPTDEK
BTV-10   DALRNYAFKMAIEFHGSEAETLNDANPRLHKIYGMPEIPPLYMEYAEIGTRFDDEPTDEK
BTV-13   DALRNYAFKMAIEFHGSEAETLNDANPRLHKIYGMPEIPPLYMEYAEIGTRFDDEPTDEK
BTV-2    DALRNKLFKMAIEFHGSEAETLNDANPRLHKVYGMPEIPPLYMEYAEIGARFDDEPTDEK
Chuzan   DQLRKYALLSCRFLNGSEAETLNVAEPNVHKVRGLTMGPPT--HLNLLGSYENDGPTDEK
StCroix  DEIRKHAATFCDSFSGSEAETIMMLRTDRYHISGRTPNPPNSNFPYPSLPFKGDRECGDK
         : :*        : ******:    .  : : * .  .*           .*   ..*

AHSV-4   LVSMLDYIIYSAD-EVYYVGCGDLKTLEQFASRDRKRFDRIKWICIDP-IAPETSYANVK
AHSV-3   LVSMLDYIIYSAD-EVYYVGCGDLKTLEQFASRDRKRFDRIKWICIDP-IAPETSYANVK
BTV-11   LVSMLDYIVYSAE-EVHYVGCGDLRTLMQFKKRSPGRFRRVLWHVYDP-IAPECSDPNVI
BTV-17   LVSMLDYIVYSAE-EVHYVGCGDLRTLMQFKKRSPGRFRRVLWHVYDP-IAPECSDPNVI
BTV-10   LVSMLDYIVYSAE-EVHYVGCGDLRTLMQFKKRSPGRFRRVLWHVYDP-IAPECSDPNVI
BTV-13   LVSMLDYIVYSAE-EVHYVGCGDLRTLMQFKKRSPGRFRRVLWHVYDP-IAPECSDPNVI
BTV-2    LVSMLDYIVYSAE-EVHYVGCGDLRTIMQFKKRSPGRFRRVLWHVYHP-IAPESSDPNVI
Chuzan   LVSMIDYITYSAD-QIYYIGCGDLRTLKLFAKRDKRRFNRVRWFCFDP-IAPESFAENVL
StCroix  LVAMLDYLIPHAEYRAIYVGSGRGDTVQSFARRSPRRFSLGQWVLIDANVSWEAPPSNVL
         **:*:**:   *: .  *:*.*   *:  *  *.  **    *   .. :: *    **

AHSV-4   VVREKVVSARDLKHYLMRDE-VERLLIWDVSAD-GLKGTIEWEKQRFKEDRNGENIAEAL
AHSV-3   VVKEKVVSARDLKHYLMRDE-VERLLIWDVSAD-GLKGTIEWEKRRFKEDRNGENIAEAL
BTV-11   VHNIMVDSKKNILKHMNFLKRVERLFIWDVSSDRSQMNDHEWETTRFAEDRLGEEIAYEM
BTV-17   VHNIMVDSKKDILKHMNFLKRVERLFIWDVSSDRSQMNDHEWETTRFAEDRLGEEIAYEM
BTV-10   VHNIMVDSKKDILKHMNFLKRVERLFIWDVSSDRSQMNDHEWETTRFAEDRLGEEIAYEM
BTV-13   VHNIMVDSKKDILKYMNFLKRVERLFIWDVSSDRSQMNDHEWETTRFAEDRLGEEIAYEM
BTV-2    VHNVMVDSKKDILKHMNFLKRVERLFIWDVSSDRSQMDDDEWESTRFAEDRLGEEIAYEM
Chuzan   VRKTKIENYKDLKKYMDSGS-VERVLIWDVSGD-GKKGTIEWEQQRASEDRMGEQIAKGL
StCroix  LYQGYVHSVDDVAQFLVPGA-LNQILIWDVRTDNVGLSKFQWEERAMEQDMLGENTAEAL
         : .  : .  :: :.:     :::::**** *    .  :**     :*  **: *  :
```

```
AHSV-4   CADFALALIKHRIPEESDE-YICRSSWLIPQPGAPITMYELRNLMRLDGYSHVERKHIPR
AHSV-3   CADFALALIKHRIPEESDE-YICRSSWLLPQPGAPITMYELRNLMRLDGYSHVERKHIPR
BTV-11   GGAFSSALIKHRIPNSKDE-YHCISTYLFPQPGADADMYELRNFMRLRGYSHVDRHMHPD
BTV-17   GGAFSSALIKHRIPNSKDE-YHCISTYLFPQPGADADMYELRNFMRLRGYSHVDRHMHPD
BTV-10   GGAFSSALIKHRIPNSKDE-YHCISTYLFPQPGADADMYELRNFMRLRGYSHVDRHMHPD
BTV-13   GGAFSSALIKHRIPNSKDE-YHCISTYLFPQPGADADMYELRNFMRLRGYSHVDRHMHPD
BTV-2    GGAFSSALIKHRIPNSRDE-YHCISTYLLPQPGADADMYELRNFMRLKGYSHVDRHMHPD
Chuzan   TKFFSFAVIKHRIPKDNEV-YSCYSSLIIPQPGAAQDMYECRNIIKLEGFSKVDRTHLGE
StCroix QSELAAALLKHRIPQFSDNRDEVFTSTLLPQPGAPAGMYEMRNFCRLDGPRTFN-RSIPT
           :: *::******:  :      :: ::******   *** **: :* *   .:

AHSV-4   AYARKIDAEVARRLVEEYHGEDVGRLLKRSLYEDIHIERADGLTDGDERTRADLFYLTNM
AHSV-3   AYVRKIDAEVARRLVEEYHGEDVGRLLKRSLYEDIHIERADGLTDGDERTRADLFYLTNI
BTV-11   ASVTKVVSRDVRKMVELYHGRDRGRFLKKRLFEHLHIVRKNGLLHESDEPRADLFYLTNR
BTV-17   ASVTKVVSRDVRKMVELYHGRDRGRFLKKRLFEHLHIVRKNGLLHESDEPRADLFYLTNR
BTV-10   ASVTKVVSRDVRKMVELYHGRDRGRFLKKRLFEHLHIVRKNGLLHESDEPRADLFYLTNR
BTV-13   ASVTKVVSRDVRKMVELYHGRDRGTFLKKRLFEHLHIVRKNGLLHESDEPRADLFYLTNR
BTV-2    ASVMKVVSRDVRKMVELYHGRDRGRFVKNRLFEHLHIVRKNGLLHESDEPRADLFYLTNR
Chuzan   ARIHYVSPKDLRLLIHRFHGFGRGRKLKKSIFEYLHIERVNGLDSLDE-PRADLFYLTNH
StCroix AQRHPIKYDECRAMVEELHRSRRGANLRRRIFEFLHIQDEDGLLHFGE-KTAHLFYLTNS
         *    :     * ::.  *    *  ::. ::* :**   :**   .:   *.******

AHSV-4   RNVAFMHDVYRVVEKSFISTLWVSNRQNFTYDDVPVNRNFITLRFSKKNRRVLDGNGAIL
AHSV-3   RNAAFMHDVYGVVEKSFISTLWVSNRQNFTYDDVPVNRNFITLRFSKKNRRVLDWNGAIL
BTV-11   CNMGLEPSIYEVMKKSVIATAWVGRAPLYDYDDFALPRSTVMLNGSYRDIRILDGNGAIL
BTV-17   CNMGLEPSIYEVMKKSVIATAWVGRAPLYDYDDFALPRSTVMLNGSYRDIRILDGNGAIL
BTV-10   CNMGLEPSIYEVMKKSVIATAWVGRAPLYDYDDFALPRSTVMLNGSYRDIRILDGNGAIL
BTV-13   CNMGLEPSIYEVMKKSVIATAWVGRAPLYDYDDFALPRSTVMLNGSYRDISILDGNGAIL
BTV-2    CNMGLEPSIYEVMKKSVIATAWVGRAPLYDYDDFALPRSTVMLNGSYHDIRILDGNGAIL
Chuzan   RNAARVEDIRRVVEQSTISTLWVGKRPLYDYPDFRYPRCDAMLRFSNITNRVFDGNGALL
StCroix CNEEHIGDLRAIVRSAGIATLWVG-GEIFGYPDFSYDRRLAIADFCTKDRMVYSGLGYIL
          *     .:  ::..: *:* **.    : * *.   *       .     : .  * :*

AHSV-4   FLMWQHPKDFPKTMNYDPSWAENYAVIFYHALT-SPVPDLSLCRFIGLRLMSSTLRINSD
AHSV-3   FLMWQHSKDFPKTMNYDPSWAKNYAVIFYHALT-SPVPDLSLCRFIGLRLMSSTLRINSD
BTV-11   FLMWRYPDIVKKDLTYDPAWAMNFAVSLKEPIPDPPVPDISLCRFIGLRVESSVLRVRNP
BTV-17   FLMWRYPDIVKKDLTYDPAWAMNFAVSLKEPIPDPPVPDISLCRFIGLRVESSVLRVRNP
BTV-10   FLMWRYPDIVKKDLTYDPAWAMNFAVSLKEPIPDPPVPDISLCRFIGLRVESSVLRVRNP
BTV-13   YLMWRYPDIVKKDLTYDPAWAMNFAVSLKEPIPDPPVPDISLCRFIGLRVESSVLRVRNP
BTV-2    FLMWKYPDIVKKDLTYDHAWAMNFAVSLKEPIPDPPVPDISLCRFIGLRVESSVLRVRNP
Chuzan   YLMWKYPEKFSKKINYDPAWAENYCVILKEDIPRIPVPELSLCRFIGLRTVSTMMRVQTN
StCroix LLMWEG--SVPTSLPFDPWWADSFAVIVK-RREVPFLPDVSLCRFIGIRERSSQLRLRFR
          ***.    . . : :*  ** .:.* .        :*::********:*  *: :*:.

AHSV-4   RAHQVTDILKKLGLDVSGHLFICLMSNSYVADLDWWFRMILEWSVKDREGKLAALSEAKA
AHSV-3   RAHQVTDILKKLGLNVSGHLFICLMSNSYVADLDWWFRMILEWSVKDREGKLAALSEAKA
BTV-11   TLHETADELKRMGLDLSGHLYVTLMSGAYVTDLFWWFKMILDWSAQNKEQKLRDLKRSAA
BTV-17   TLHETADELKRMGLDLSGHLYVTLMSGAYVTDLFWWFKMILDWSAQNKEQKLRDLKRSAA
BTV-10   TLHETADELKRMGLDLSGHLYVTLMSGAYVTDLFWWFKMILDWSAQNREQKLRDLKRSAA
BTV-13   TLHETADELKRMGLDLSGHLYVTLMSGAYVTDLFWWFNIILDWSAQNKEQKLRDLKRSAA
BTV-2    TLHETADELKRMGLDLSGHLYVTLMSGAYVTDLFWWFKMILDWSAQSKEQKLRDLKRSAA
Chuzan   SVHQISDKVKKMGLDLSGHLFIALMSDSYISDIMWWFDMILNWSVLNREEKKRKLFEMNA
StCroix EVHDVMDVVKSLGVDASGHLFIGILSSNYVFDPYIWVNMIMLWSQQSRTDKIRDIESHGA
          *:  * :* :*:: ****:: ::*. *: *   *. :*: **  .:  *   :   *
```

```
AHSV-4   ELIEWRDEKADEPWHIKNDLLAALFEFIYFAKHFDINERYVE-SWIQYLRNA
AHSV-3   ELIEWKDEKADEPWHIKNDLLAALFEFMYFAKHFEINEGYVE-SWIQYLRNA
BTV-11   EVIEWKEQMAERPWHVRNDLIRALREYKR--K-MGMREGASIDSWLELLRHL
BTV-17   EVIEWKEQMAERPWHVRNDLIRALREYKR--K-MGMREGASIDSWLELLRHL
BTV-10   EVIEWKEQMAERPWHVRNDLTRALREYKR--K-MGMREGASIDSWLELLRHL
BTV-13   EVIEWKEQMAERPWHVRNDLIRALREYKR--K-MGMREGASIDSWLELLRHL
BTV-2    EVIEWKEQMAERPWHVRNSLIAALREYKR--K-MGIREGASIDSWLELLRHL
Chuzan   EVIEWKEDRANEPWHIKPDLIAALMEFSRFINVMMTDEACVTKWILYLRSKE
StCroix AVIEWDRSHMDKPWHSFDDLTASLVALGKILPP-QRYRSFLSRVRGWLDSYA
          :***  .  :.***   .*  :*         :        :   :
```

Figure 2.6. The amino acid sequences of all orbivirus species VP4 are aligned using Clustal software. Conserved amino acids are marked by a star underneath the alignment, when amino acids differ but are both polar or non-polar a colon is used (:) while if one amino acid belongs to another group a period (.) is used.

3.3 Terminal hexanucleotides

In BTV all gene fragments have a conserved 5' and 3' terminal hexanucleotide. When all sequenced segments of AHSV were compared a conserved terminal region was identified as shown in table 2.3. The overall consensus sequence for all serotypes is GUUA/UAA/U--A/CCA/UUAC. From this data it is clear that at least the three terminal nucleotides at the 5' and 3' end are conserved between all the serotypes in AHSV unlike the terminal six nucleotides conserved in BTV.

Serotype	Number of sequences	Consensus sequence
1	1	GUUUAA --ACUUAC
2	1	GUUUAA --ACUUAC
3	5	GUU$_{A/U}$AA$_{/U}$--A$_{/C}$CUUAC
4	10	GUU$_{A/U}$AA$_{/U}$--AC$_{A/U}$UAC
5	1	GUUUAA--ACUUAC
6	5	GUUAA$_{A/U}$--ACU$_{A/U}$AC
7	2	GUUUAA--ACUUAC
8	2	GUUUAA--ACUUAC
9	8	GUU$_{A/U}$AA$_{/U}$--ACA$_{A/U}$UAC

Table 2.3. The consensus sequence of the 5' and 3' terminal hexanucleotides of each serotype within the AHSV serogroup are shown. The first three nucleotides on the 5' end are always conserved as well as the last three nucleotides on the 3' end among the different serotypes.

3.4 Phylogenetic relationships

To determine the evolutionary relationship between the different VP4 genes and proteins identified by BLAST a multiple alignment was performed using the Clustal W program. The results of this alignment are seen in figure 2.5 and figure 2.6. From both the alignments phylogenetic trees were drawn. The trees are represented in figure 2.7. It is clear from figure 2.7 that Chuzan and AHSV are the most closely related serogroups within the orbivirus genus and that BTV is more distant to both of them while ST. Croix River virus forms a distinct out group. The trees for the nucleotide alignment and the amino acid alignment are very similar in terms of grouping of the different serogroups. Branch lengths in the amino acid comparison

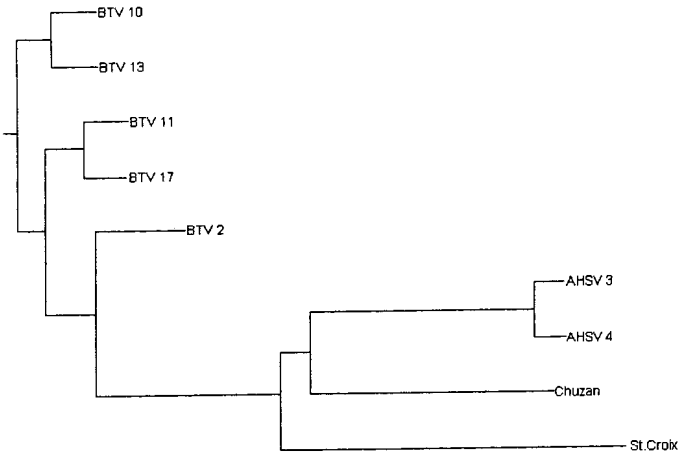
are shorter due to the lesser influence of silent mutations in the DNA sequence that do not translate into amino acid mutations.

3.5 Identifying regions of high homology within the orbivirus genus VP4

Using the Clustal protein alignment (figure 2.5) it is possible to identify regions of four or more completely conserved amino acids. These can be classified as highly conserved because they are also conserved in the more divergent St. Croix virus. Eight such regions were found as shown in table 2.3. These regions have a 62% homology on nucleotide level while the average homology for the complete gene sequence is 24%. The first region is found at 60-RQLR-63, the second 135-(H)GSEAET(L)-142, the third at 267-IWDV-271, the fourth at 309-(I)KHRIP-314, the fifth at 330-PQPGA-334, the sixth at 412-A(D)LFYLTN-419, the seventh at 521-SLCRFIG(L)R-529 and the eighth at 565-SGHL-568. When excluding St. Croix river virus from this alignment it is possible to make these regions even larger with more conserved amino acids. The seventh region is found right in front of the leucine zipper motif identified in BTV of which many amino acids are not conserved between all the orbivirus species. Apparently the sequence represented in region 7 has some importance. The relevance of these conserved motifs will be discussed in the next part when known motifs are identified.

Conserved amino acids are found throughout the protein sequence. However, there is only a very low level of conservation in the C-terminal 30 amino acids between all the species and also between AHSV and BTV. Of the 30 amino acids only 2 (6% identity) amino acids are conserved between all the species and 8 amino acids (26% identity) between AHSV and BTV. The insertion/deletion is also present in this region.

A.



B.

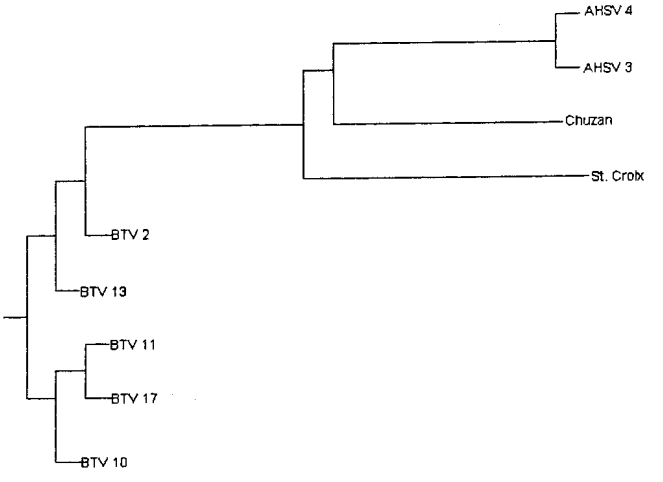


Figure 2.7. The phylogenetic trees of the nucleotides (A) and amino acids (B) alignment are represented comparing VP4 of AHSV, BTV, Chuzan and St. Croix River virus. The trees indicate that AHSV and Chuzan virus are closely related while BTV is more distant with St. Croix as an outgroup.

	Position	aa sequence	No. conserved nucleotides	Percentage conservation
1	60-63	RQLR	7/12	58
2	135-142	(H)GSEAET(L)	10/18	55
3	267-271	IWDV	9/12	75
4	309-314	(I)KHRIP	9/15	60
5	330-334	PQPGA	10/15	66
6	412-419	A(D)LFYLTN	11/18	61
7	521-529	SLCRFIG(L)R	13/21	62
8	565-568	SGHL	8/12	66

Table 2.4 Regions of high homology were identified when comparing all nine VP4 sequences available for orbivruses. The conservation on nucleotide level is always higher than the 24% average conservation over the complete gene sequence between the nine viruses. The percentage conservation is that of the nucleotides making up the region in the amino acid sequence.

3.6 Motif identification

Motifs are short regions within a protein that have been linked to a function of the protein. These motifs can be searched for using algorithmic search engines. The meta search site eMotif finder was used to identify possible functional motifs within the AHSV VP4 sequence. The meta search engine uses several analysis programs to identify functional motifs. Three of these programs did not identify any motifs in the AHSV amino acid sequence. These were prosite pattern, prosite profile and Pfam. BLOCKS identified many possible motifs. However, only one motif was found in two programs. The motif is a fibronectin type III repeat signature found between amino acid 325 and 342 (figure 2.8) and was identified by BLOCKS and PRINTS. When the region was compared between all nine sequences available for orbivirus, 7 of the 11 residues (63% conservation) in the motif were completely conserved. Moreover, the region surrounding this motif has a high level of conservation between all species with 12 out of 24 residues (50% conservation) conserved. Between BTV and AHSV serotypes the level of conservation is even higher with 17/24 amino acids conserved resulting in a 71% conservation.

When the conserved regions identified in section 3.4 and shown in table 2.4 are compared to the motif identification results of BLOCKS only one region contains a motif. That is region 2 135-(H)GSEAET(L)-142. The SEAE motif fits the description for the casein kinase II phosphorylation site S/T.xx.D/E. However, this motif is small and in the complete protein sequence of AHSV-3 VP4, 12 such motifs were identified.

The BLOCKS program alone identified a number of motifs previously identified in BTV were also found in AHSV. These are a myristoylation site at residue 544 GLNVSG and an amidation site IGFRR at residue 106 as shown in figure 2.9d. The myristoylation site is partially conserved between all the species used in the alignment. Three of the six amino acids are conserved between all the species. The amidation site is completely conserved between AHSV and BTV and 3 out of 5 amino acids are conserved between BTV and Chuzan virus and St.Croix river virus.

Two motifs have been identified in BTV VP4 that have known functions. The first is the consensus motif for GMP binding by a viral capping enzyme and is KxDG. This motif is found in rotavirus and reovirus while a divergent motif (KxxGD) is found in BTV. When the BTV motif is compared to the AHSV and other orbivirus sequences it is evident that the motif present in AHSV is more like the KxDG present on the capping enzymes in viruses like reovirus (λ2) and rotavirus (VP3). The lysine residue is separated by 3 amino acids from a 'DG' amino acid pair versus the BTV lysine separated by two amino acids from a GD amino acid pair. Additionally, from a region of 10 amino acids surrounding the motif 6 are conserved between AHSV and BTV (figure 2.9a). While some amino acids around the motif are also conserved in Chuzan virus, both St.Croix virus and Chuzan lack a lysine residue that is essential for GMP binding in this region. Therefore, we can identify a possible KxxxDG motif in AHSV-3 VP4 that is more like the KxDG motif described in other viruses than the KxxGD motif found in BTV.

The leucine zipper motif (523-LRVESSVLRVRNPTLHETADELKRMGLDL-551) identified in BTV is only partially present in AHSV. Of the five leucine residues involved in the zipper, two leucine residues (3rd and 5th leucine) are not present but replaced with an alanine and valine respectively (519-LRLMSSTLRINSDRAHQVTDILKKLGLDV-547). Also, the Chuzan virus sequence has no leucine zipper motif with only the first and last leucine present in the sequence. Interestingly, the region surrounding the motif contains many conserved

amino acids. Immediately in front of the motif is a region completely conserved (CRFIG). Following the zipper region is also a region of 4 amino acids completely conserved (SGHL). Both are shown in red in figure 2.9c. Out of 34 amino acids including the leucine zipper, 18 are conserved between all species (52% identity). When only BTV and AHSV are compared the conservation is even higher with 23 out of 34 amino acids conserved (67% identity). In both cases this is higher than the conservation observed over the complete sequence length. Upstream of the leucine zipper in BTV there is a region with a high number of proline residues (5/9aa). From these five proline residues only two are found in AHSV while three are found in Chuzan virus and one in St. Croix River virus. However, there is a second proline in St. Croix virus in this region on a different position compared to the BTV sequences.

The last motif conserved between BTV, rotavirus and reovirus is found at residue 621 in AHSV, in the least conserved variable region between AHSV and BTV. This motif has no known function. The alignment in figure 2.9b shows that the motif YKRKM is only found in the BTV sequence and not in the other three orbivirus species. This motif falls within the region described above with only 6% identity between all species. Additionally, this is also the region in which the insertion/deletion between BTV and AHSV is present. The region in AHSV similar to BTV contains an YFAKH sequence. Only the tyrosine and the lysine are conserved between BTV and AHSV. The first lysine residue in BTV is replaced by phenylalanine (positive, polar to non-polar, hydrophobic), the arginine is replaced by an alanine residue (positive, polar to non-polar) and lastly the methionine residue in BTV is replaced in AHSV by a histidine (non-polar to positive, polar). The region found in AHSV also does not resemble the regions found in rotavirus (YYRYN) or Reovirus (YVRKN).

```
AHSV-3          LIPQPGAPITMYELRNLMRLDGYSHV
AHSV-4          LLPQPGAPITMYELRNLMRLDGYSHV
BTV-11          LFPQPGADADMYELRNFMRLRGYSHV
BTV-17          LFPQPGADADMYELRNFMRLRGYSHV
BTV-10          LFPQPGADADMYELRNFMRLRGYSHV
BTV-13          LFPQPGADADMYELRNFMRLRGYSHV
BTV-2           LLPQPGADADMYELRNFMRLKGYSHV
Chuzan          LLPQPGAPAGMYEMRNFCRLDGPRTF
StCroix         IIPQPGAAQDMYECRNIIKLEGFSKV
                ::*****   *** **: :* *   .
```

Figure 2.8. The possible motif for fibronectin type III repeat signature was identified by the eMOTIF software. 7 of the 11 residues are conserved between the orbivirus members. Conserved residues are shown in purple.

A. Lysine-containing GMP binding motif

```
AHSV-3    21-FLPKWELDGI
AHSV-4    21-FLPKWELDGI
BTV-2     22-FLPIWKLTGD
BTV-10    22-FLPIWKLTGD
BTV-11    22-FLPIWKLTGD
BTV-13    22-FLPIWKLTGD
BTV-17    22-FLPIWKLTGD
Chuzan    21-FLPIFDVNRI
St.Croix  23-DL--LTLRGA
              *
```

B. Conserved motif in BTV, Reovirus and Rotavirus

```
AHSV-4   LLAALFEFIYFAKHFDINERYVESWIQYLRNA-
AHSV-3   LLAALFEFMYFAKHFEINEGYVESWIQYLRNA-
BTV-11   LIRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-17   LIRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-10   LTRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-13   LIRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-2    LIAALRE--YKRKMGIREGASIDSWLELLRHL-
Chuzan   LIAALMEFSRFINVMMTDEACVTKWILYLRSKE
StCroix  LTASLVALGKILPPQRYRSFLSRVRGWLDSYA-
         *  :*         :        :   :
```

C. Leucine zipper motif

```
AHSV-3    ALT-SPVPDLSLCRFIGLRLMSSTLRINSDRAHQVTDILKKLGLNVSGHLFICLMS
AHSV-4    ALT-SPVPDLSLCRFIGLRLMSSTLRINSDRAHQVTDILKKLGLDVSGHLFICLMS
BTV-10    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
BTV-2     PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHEAADELKRMGLDLSGHLYVTLMS
BTV-11    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
BTV-13    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
BTV-17    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
Chuzan    DIPRIPVPELSLCRFIGLRTVSTMMRVQTNSVHQISDKVKKMGLDLSGHLFIALMS
St.Croix  RREVPFLPDVSLCRFIGIRERSSQLRLRFREVHDVMDVVKSLGVDASGHLFIGILS
                  *  ******* *  *   *      *   *  *  * * ****     *
```

```
AHSV-3          LIPQPGAPITMYELRNLMRLDGYSHV
AHSV-4          LLPQPGAPITMYELRNLMRLDGYSHV
BTV-11          LFPQPGADADMYELRNFMRLRGYSHV
BTV-17          LFPQPGADADMYELRNFMRLRGYSHV
BTV-10          LFPQPGADADMYELRNFMRLRGYSHV
BTV-13          LFPQPGADADMYELRNFMRLRGYSHV
BTV-2           LLPQPGADADMYELRNFMRLKGYSHV
Chuzan          LLPQPGAPAGMYEMRNFCRLDGPRTF
StCroix         IIPQPGAAQDMYECRNIIKLEGFSKV
                ::*****   *** **: :* *   .
```

Figure 2.8. The possible motif for fibronectin type III repeat signature was identified by the eMOTIF software. 7 of the 11 residues are conserved between the orbivirus members. Conserved residues are shown in purple.

A. Lysine-containing GMP binding motif

```
AHSV-3    21-FLPKWELDGI
AHSV-4    21-FLPKWELDGI
BTV-2     22-FLPIWKLTGD
BTV-10    22-FLPIWKLTGD
BTV-11    22-FLPIWKLTGD
BTV-13    22-FLPIWKLTGD
BTV-17    22-FLPIWKLTGD
Chuzan    21-FLPIFDVNRI
St.Croix  23-DL--LTLRGA
             *
```

B. Conserved motif in BTV, Reovirus and Rotavirus

```
AHSV-4   LLAALFEFIYFAKHFDINERYVESWIQYLRNA-
AHSV-3   LLAALFEFMYFAKHFEINEGYVESWIQYLRNA-
BTV-11   LIRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-17   LIRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-10   LTRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-13   LIRALRE--YKRKMGMREGASIDSWLELLRHL-
BTV-2    LIAALRE--YKRKMGIREGASIDSWLELLRHL-
Chuzan   LIAALMEFSRFINVMMTDEACVTKWILYLRSKE
StCroix  LTASLVALGKILPPQRYRSFLSRVRGWLDSYA-
         *  :*         :        :   :
```

C. Leucine zipper motif

```
AHSV-3    ALT-SPVPDLSLCRFIGLRLMSSTLRINSDRAHQVTDILKKLGLNVSGHLFICLMS
AHSV-4    ALT-SPVPDLSLCRFIGLRLMSSTLRINSDRAHQVTDILKKLGLDVSGHLFICLMS
BTV-10    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
BTV-2     PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHEAADELKRMGLDLSGHLYVTLMS
BTV-11    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
BTV-13    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
BTV-17    PIPDPPVPDISLCRFIGLRVESSVLRVRNPTLHETADELKRMGLDLSGHLYVTLMS
Chuzan    DIPRIPVPELSLCRFIGLRTVSTMMRVQTNSVHQISDKVKKMGLDLSGHLFIALMS
St.Croix  RREVPFLPDVSLCRFIGIRERSSQLRLRFREVHDVMDVVKSLGVDASGHLFIGILS
                * ******* *  *   *     *   *  *  * * ****     *
```

D. N-myristoylation site

```
AHSV-4   GLDVSG
AHSV-3   GLNVSG
BTV-11   GLDLSG
BTV-17   GLDLSG
BTV-10   GLDLSG
BTV-13   GLDLSG
BTV-2    GLDLSG
Chuzan   GLDLSG
StCroix  GVDASG
         *   **
```

E. Amidation site

```
AHSV-4   IGRRR
AHSV-3   IGRRR
BTV-11   IGRRR
BTV-17   ISRRR
BTV-10   IGRRR
BTV-13   IGRRR
BTV-2    IGRRR
Chuzan   IGFKR
StCroix  IGKAR
         *   *
```

Figure 2.9. All the possible motifs identified in AHSV or BTV and their counterparts in other viruses are shown. In A the lysine-containing motif important for BTV VP4 function is shown along with the possible motif present in AHSV VP4. In B the motif identified in BTV, rotavirus and reovirus is compared with the other orbivirus members. In C the leucine zipper and the regions upstream of it is shown indicating the presence of leucine residues in the zipper and proline residues in the upstream region. Completely conserved regions up and down stream of the leucine zipper are shown in red. In D the N-myristoylation site identified first in BTV and also identified with similar software in AHSV are shown. In E the amidation site is shown. Residues part of the motif are shown in purple while the possible functional lysine residue in A are colored in blue

2.4 Discussion

An essential part of the survival of a number of RNA viruses is their ability to produce mRNA that can be used by the host cells translation machinery to produce viral proteins. Viruses that use this survival mechanism include rotavirus, reovirus and orbivirus. The cellular translation machinery of mammals only recognizes RNA strands with a guanidine cap attached to the 5' end of the strand. Adding the guanine cap is the task of the capping enzyme. Members of the Reoviridae have a capping enzyme within their viral core structure.

The protein identified as the capping enzyme in BTV is VP4 (Le Blois H. *et al.*, 1992; Martinez-Costas J. *et al.*, 1998; Ramadevi N. *et al.*, 1998a). In other viruses capping enzymes have also been identified and characterized. In reovirus, lamda2 is responsible for capping (Mao Z. and Joklik W.K., 1991; Luongo C.L. *et al.*, 1998) while West Nile virus NS3 and vaccinia virus D1 all have motifs correlating with capping activity (Bisaillon M. *et al.*, 1997b).

The VP4 sequence of AHSV-4 was already known (Mizukoshi N. and Sakamoto K., 1993). To determine whether this sequence was correct and representative of the serogroup AHSV and determine the level of variation for VP4 within the AHSV serogroup we sequenced the VP4 gene of AHSV-3. When a BLAST search was done with the obtained sequence, AHSV-4 VP4 aligned to the AHSV-3 sequence with the highest homology (table 2.1). Additionally, the VP4 protein of different BTV strains aligned as well as the VP4 genes of Chuzan and St. Croix River virus. These were less homologous to AHSV-3 than AHSV-4. The homologous sequences identified all belong to serogroups within the orbivirus genus and therefore we could positively establish the authenticity of the AHSV-3 VP4 sequence.

When orbivirus members are compared on nucleotide level we see a strong relationship between the different species. For VP3, the major core protein, there is a 79% homology between AHSV and BTV nucleotide sequences resulting in a 58% identity of amino acid while for the major outer capsid protein VP2 there is only about 20% amino acid identity between AHSV and BTV (Roy P., 1992). Here it is shown that VP4 has a 49% homology on amino acid level. Thus, VP4 is more conserved than VP2 but less than VP3. The VP4 nucleotide sequence of the two AHSV serotypes only differed in 55 nucleotides (2.7%) of which only seven resulted in an amino acid change within the same class of amino acid (polar to polar) and seven

changing the class of amino acid (non-charged to charged). These changes do not cause any differences in the conserved regions identified in BTV or those identified between all orbiviruses in table 2.4. This level of conservation is comparable to that of VP4 in BTV where between five serotypes there is not more than 2.5% difference in the nucleotide sequence (Huang I-Jen *et al.*, 1993).

When the nucleotide sequences of all the orbivirus capping enzymes are compared a small size difference is apparent. This size difference between AHSV and BTV and also between Chuzan and BTV and St. Croix River virus and BTV is only a few nucleotides which is much smaller than other genes within the genus. Moreover, the insertions/deletions take place in regions where no known motifs have been identified and take mostly place in the 3'end that already has a lower level of conservation than the rest of the protein.

When AHSV and BTV amino acid sequences are compared the C-terminal region is less conserved than the rest of the protein. In the last 30 amino acids only 21% of the amino acids are conserved while over the complete protein 48% are conserved. This can be extrapolated to the other species in the comparison. While 21% of amino acids are identical between all the species only 6% are conserved in the last 30 amino acids. One possible reason for this may be the location of the C-terminus. It could be in such a position that it makes contact with the environment and as such is under more evolutionary pressure. Alternatively, no amino acids are found in this region that have a role in the function of the protein and therefore mutations can aggregate in this region.

When the alignments of the nucleotides and the amino acid sequences are used to construct a Phylogenetic tree we see that the BTV is more distinct from AHSV than Chuzan virus. Chuzan virus and AHSV have in terms of VP4 the least number of differences while St. Croix River virus forms a distinct out group. This confirms previous data where analysis of St. Croix River virus VP1 shows it to be distantly related to the other orbivirus species (Attoui H. *et al.*, 2001). Moreover, analysis of Chuzan virus proteins showed it is most closely related to AHSV (Yamakwa M. *et al.*, 1999).

From the amino acid comparison of all orbivirus species we were able to identify some regions of high conservation between all the species. The result of this comparison was shown in table 2.4. Eight regions of conservation of four amino

acids or longer were identified. When also looking at similar amino acids, these regions can be extended further. These regions do not contain motifs that were previously identified as important for the capping functions of VP4. Some of the conserved regions do, however, contain motifs for amidation and glycosylation also previously identified in BTV (Huang I-Jen. *et al.*, 1993). These are shown in figure 2.9d+e. The fact that these regions are conserved in all serogroups that have only 21% identity between them may indicate that they have some importance in the structure or function of the protein. Whether the protein is glycosylated or amidation takes place is not known yet. When the AHSV-3 VP4 sequence was analyzed using the different motif identifier programs only one motif was found in two different programs. The fibronectin type III repeat motif identified in VP4 is responsible for protein-protein interactions. In this motif seven of the eleven amino acids are conserved between all the orbivirus VP4 sequences indicating that this regions is conserved. The hypothesis is that the three minor core proteins form a complex at the five-fold axis in the core (Diprose J.M. *et al.*, 2001). For this, protein interactions are essential and this motif may be involved in such an interaction.

When VP4 was analyzed within the BTV serogroup and compared to other Reoviridae members a number of conserved motifs were identified. The lysine containing motif, KxDG, is found in the sequences of capping enzymes from eukaryotes and DNA viruses and was shown to be essential for the binding to the GMP moiety (Bisaillon M and Le May G., 1997c). A divergent motif (KLTGD) was found in BTV (Yu Y. *et al.*, 1987). In AHSV the motif, (KWELDG), is closer related to the original motif as the 'DG' residues are in the correct order as opposed to the BTV 'GD'. However, there is an extra amino acid between the lysine residue and the 'DG' pair including the large, hydrophobic tryptophan residue. Whether this motif is the possible GMP binding motif is unknown. Strikingly, both Chuzan virus and St. Croix River virus do not have any motif or even diverging motif in this region (Chuzan FLPIFDVNRI; St. Croix DL—-LTLRGA). The Chuzan virus sequence shown here is an exact match for the first four amino acids of BTV but the rest is completely different. Another motif, conserved between BTV, rotavirus and reovirus, found in AHSV or Chuzan virus is not present in AHSV. Therefore, a possible GMP binding motif can be identified in AHSV but a similar motif is not seen in Chuzan or St. Croix River virus.

Strikingly, the leucine zipper identified in BTV (Huang I-J *et al.*, 1993) that is essential for dimerization and encapsidation is interrupted in AHSV and Chuzan virus (figure

2.9c). The motif itself is interrupted in the third leucine residue in both AHSV serotypes and in Chuzan and St. Croix River virus. Huang I-J. *et al.*,1993 showed that the leucine zipper is essential by using mutational studies. The third leucine residue of the motif was mutated to either an alanine or a proline. These mutants were not able to dimerize VP4 and led to its exclusion from viral bodies. Therefore, this indicates that it is likely that AHSV, Chuzan virus and St. Croix virus may not have a functional leucine zipper and that the dimerization step may be independent of the leucine zipper. It is interesting that immediately in front of the motif there is a long stretch of amino acids conserved between all species (SLCRFIG(L)R). While a motif search did not reveal any functional motifs in this region it may indicate that the leucine zipper region is indeed important for the function of the capping enzyme in the different viruses. Functional studies of the leucine zipper region will have to point out whether this region is needed for dimerization in other orbiviruses and if other orbiviruses have VP4 dimers at all. If dimerization is zipper independent then the fibronectin repeat motif found in AHSV and conserved among the species may prove important as a protein-binding region.

In conclusion, we have identified a possible lysine-containing motif in AHSV VP4 that may be responsible for the GMP transfer central to the capping process. We show that there are eight regions of conservation within the orbivirus VP4 gene of which one contains a possible phosphorylation motif. It is known, however, that BTV NS2 is the only phosphorylated protein and thus it is unlikely that VP4 is phosphorylated (Thomas C.P. *et al.*, 1990). Lastly, the leucine zipper identified in BTV as essential for dimerization of VP4 is interrupted in AHSV and may not be functional. A possible fibronectin type III repeat motif for protein-protein interaction may be important for the formation of protein complexes seen at the fivefold axis of virus cores. The function of these motifs is not known yet and will be determined by performing assays with the biologically active protein.

Chapter 3.

Expression and purification of AHSV serortype 3 VP4.

3.1 Introduction

Preciously few minor core proteins within the orbivirus group have been characterized up to date. The only protein studies done on the viral capping enzyme were performed with purified BTV VP4. From the sequence analysis in the previous chapter it becomes clear that there may be several important differences between the BTV capping enzyme and that of viruses like AHSV and Chuzan virus.

The VP4 proteins are assumed to have similar functions in the different orbiviruses. The VP4 gene of AHSV-3 has about 30-40% homology with cognate VP4 genes of other members of the orbivirus group. While there is homology between VP4 from BTV and AHSV there are some essential differences in their nucleotide sequences as discussed in the previous chapter. BTV VP4 has already been studied extensively with several papers published on the function and activities of the protein. Le Blois *et al.*, 1992 presented data indicating that when VP4 was expressed in core-like particles consisting of VP3 and VP7, GTP-binding was observed. Similarly, virus-like particles (VLP's) with VP6 but without VP4 did not have this activity. As GTP binding is one of the prerequisites for capping activity, VP4 became the candidate-capping enzyme. Two papers appeared in 1998 describing the activities of purified BTV VP4. (Ramadevi N., *et al.* 1998a) and (Martinez-Costas J., *et al.* 1998). VP4 was identified as the viral capping enzyme. In these papers VP4 was expressed using the baculo-virus expression system and subsequently purified. Catalytic assays were performed for GTP binding, guanylyl transferase activity, inorganic pyro-phosphatase as well as methyl transferase activity. The assays were done with single nucleotides as well as ssRNA templates indicating that the enzymatic activity is physiologically relevant. Additionally, VP4 was also able to remove the 5' terminal λ-phosphate from an ssRNA substrate as well as transfer a guanylyl moiety to the terminal nucleotide (Martinez-Costas J., *et al.* 1998). Later in 1998, Ramadevi *et al* showed that VP4 catalyses additional steps including the methylation of the cap and adjacent nucleotides to produce a type 2 cap structure. Together these data indicated that the VP4 protein performs the complete capping process. In addition to the previously mentioned activities, VP4 was also shown to

be an NTPase. This ability allows VP4 to produce GMP that may be needed for the extraction of RNA molecules from the core (Ramadevi N. and Roy P. 1998). Moreover, it was shown that BTV VP4 is found in the core as a dimer and that the leucine zipper domain is responsible for dimerization (Ramadevi N. *et al.*, 1998b). Dimerization of BTV VP4 allows encapsidation of the protein by the VP3/VP7 decamers that form the core during virus production in the cell. No capping enzyme of any other member of the orbivirus genus has been characterized yet. The effects of the absence of the motifs, identified in BTV, in AHSV and Chuzan are not known.

This chapter describes the expression of AHSV VP4, attempts to purity the recombinant protein and the use of functional assays to elucidate the role of the protein in the virus core. For this aim VP4 was expressed as a histidine tagged fusion protein in the baculovirus expression system. The Bac-to-Bac baculovirus expression system is used for the expression of viral proteins for a number of reasons. The system allows high-level expression of proteins in the *Spodoptera frugiperda* cell line. Additionally, these proteins are mostly soluble and folded in the correct way. Expression is achieved by the use of recombinant baculovirus that infects these cells and initiates viral protein expression that also includes the recombinant gene of interest. Recombinant virus is produced by cloning the gene of interest into a vector based on the pUC vector, containing site-specific transposition motifs around the expression cassette. These Tn7 sites are inserted in the N-terminus of the lacZ gene without causing a reading frame shift. The vector also contains a gentamycin resistance marker. A recombinant bacmid is produced when the expression vector is transformed into the E.coli DH10Bac cell line that contains the bacmid inside. Transposition occurs when the mini-Tn7 element from the pUC based vector is transposed into the mini-*att*Tn7 attachment site on the bacmid. The donor plasmid used in this chapter is called pFastBac Hta. Except for the features already mentioned these vectors also harbour a 6xhistidine repeat at the 5' end of the MCS. Expression of the protein and tag are under control of the polyhedron promoter. Three different plasmids are available having staggered endonuclease sites to allow fro in-frame cloning of the gene of interest with the histidine tag. Histidine tag purification is a much used method to obtain pure samples of exogenously expressed protein. The histidine tag through its highly charged side chains can bind with great affinity to positive ionic molecules like Nickel. Ni^{2+} is immobilized on a column through which protein samples can be added. Proteins with a histidine repeat will adhere to the column allowing it to be washed and finally can

be eluted by adding chelating compounds like EDTA. Through this method we wanted to produce pure protein samples for our in vitro assays.

3.2 Materials and Methods

1. Viruses and cells

Spodoptera frugiperda (Sf9) cells were used for transfections and infections. Cells were grown in Grace's medium containing 10% v/v fetal calf serum and penicillin/streptomycin. For the expression of the tagged his-VP4 insert, the recombinant baculovirus included in the Bac-To-Bac baculovirus system was used. For expression of the wild type full length AHSV 3 VP4, a recombinant *Autographa californica* nuclear polyhedrosis virus was used.

2. Cloning of the fusion construct:

To obtain the full length VP4 DNA, a pBR vector containing the full-length gene, was digested with *Bgl II* while the HtC expression vector was digested with *BamHI* (figure 3.1). The vector was dephosphorilated by adding 10 x alkaline phosphatase buffer and 0.5 units alkaline phosphatase to the digestion mixture and incubating it for 15 minutes at 37°C. The mixture was purified using the Geneclean kit immediately afterwards. The full-length VP4 fragment was ligated into the HtC vector using T7 ligase. The ligated plasmids were transformed into *E.coli* XL1 Blue cells and plated out on LB plates with ampicillin. Recombinancy and orientation of the insert were checked with a *BamHI/EcoRI* digestion.

3. Transposition and screening of positive clones:

Both recombinant vectors were first propagated in *E.coli* before they were transposed into the bacmid containing dH10 *E.coli* cells. Transposition as discussed in the introduction involved the site-directed transposition of the expression cassette in the pFastBac vector to transposed into the attTn7 site of the bacmid to allow incorporation into the baculovirus genome. The antibiotic resistance gene $Gent^r$ allows for the selection of bacmid containing bacterial colonies that have inserted the expression cassette. Once integrated, the recombinant bacmid DNA was isolated using the alkaline lyses method and used to transfect Sf9 cells. The bacmid containing cells were made competent using DMSO as described by Miller, 1988. A

culture of cells was grown to an OD of 0.3-0.6 at a wavelength of 600. Cells were collected and resuspended in cold TSB (10ml TSB= 1g PEG-8000, 0.5ml DMSO, 0.1ml $MgCl_2$, 0.1ml $MgSO_4$).

The plasmid was transposed by addition of 1ng of plasmid to 100µl competent cells and the mixture was heat-shocked for 90 seconds. After incubation in TSBG (TSB medium with 25% glucose) medium for 4 hours, the cells were plated out on LB plates containing tetracycline, kanamycin, gentamycin, X-gal and IPTG. Positive cells were selected on colour (white colonies).

Recombinancy of bacmids was confirmed using PCR. PCR was performed with the polH forward primer and the M13 reverse primer. The standard protocol for PCR with Taq polymerase was used. Amplified fragments were analysed on 1% agarose gel.

4. Transduction and amplification of recombinant bacmid

Recombinant bacmids were isolated and used to transfect SF9 insect cells. Insect cells were seeded in 35mm well plates in Grace's insect medium with streptomycin and penicillin and 10% fetal calf serum. Cellfectin was used as cationic lipid to transfect the bacmid. The cells were incubated with the transfection mixture for 4 hours after which FCS containing medium was added to the cells. After four days the supernatant containing the recombinant virus was collected. The supernatant was filtered to remove all cells. The virus titer of the primary virus stock was amplified by adding virus at a MOI of 1 to a T75 culture flask of Sf9 cells and leaving it for four days. The supernatant containing the virus was harvested and used to infect SF9 cells for protein expression.

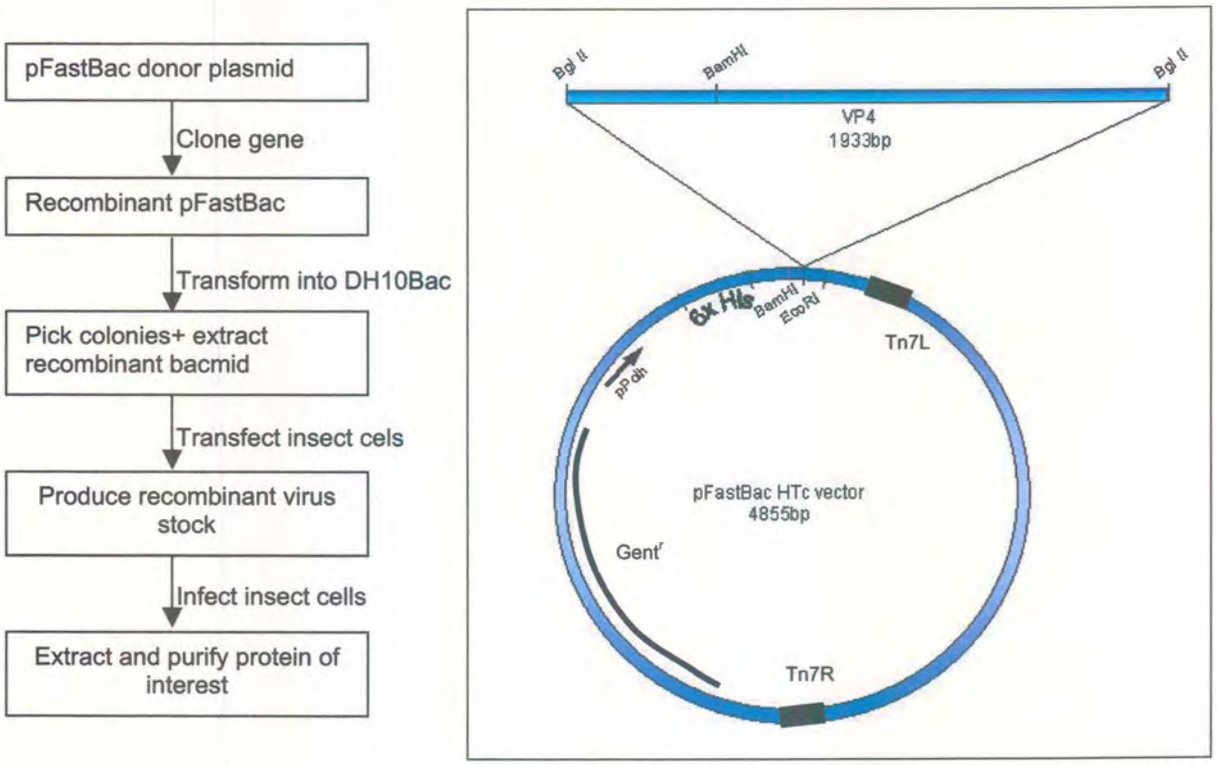


Figure 3.1. Cloning diagram for producing protein in the Bac-To-Bac baculovirus expression system and for the cloning of VP4 into the pFastBac HTc vector for expression of VP4 with a histidine tag.

5. Protein expression and identification

For production of VP4 protein, recombinant baculoviruses his-VP4 and Ac.AHSV3.VP4 were used to infect Sf9 cells at a MOI of 5-10. Ac.AHSV3.VP4 is a VP4 containing expression vector (S.Cormack, department of Genetics, University of Pretoria). It contains the full-length VP4 gene and can be expressed in the baculovirus expression system.

After 72 hours cells were washed once in PBS, dislodged from the flask and concentrated by centrifugation. Collected cells were lysed for further use. Cells infected with recombinant virus were harvested and lysed in 2x protein solvent buffer (0.125M Tris-HCl, pH 6.8, 4% SDS, 20% glycerol, 10% mercapto ethanol) along with a mock-infected sample and wild type infected cells 72 after infections. Cell lysates were boiled and centrifuged before supernatant samples were loaded on a stacked 10% SDS-PAGE gel according to the method of Laemmli, 1970. Samples were loaded on a polyacrylamide gel with a 5% staking part

(30%acrylamide/0.8%bisacrylamide, 4x stacking buffer pH 6.8, 10% ammonium phosphate and 1% TEMED) on top of a 12% separating gel (30% acrylamide/8% bisacrylamide, 4x separating buffer pH8.8, 10% ammonium phosphate, 10% TEMED) Electrophoresis was performed at 130V for 2 hours. Gels were stained with Coomasie Blue for 20 minutes and destained in destaining buffer (10% methanol, 10% acetic acid) for 30 minutes.

6. Western immuno-blot

To identify the VP4 protein, a Western blot was performed. After samples are separated on SDS-PAGE gel, they were blotted onto a nylon membrane using transfer buffer (0.025M Tris-HCl, 0.15M glycine and 20% methanol). Transfer was performed with a submerged blotting system with a 130mA current for 1.5 hours. After transfer, membranes were blocked in 1% ELK (1% dried milk powder in TBS-T) in PBS for 30 minutes at RT. The membrane was incubated with guinea pig anti-AHSV 3 anti-serum and subsequently washed three times in TBS-T washing buffer (0.05% Tween in 1x PBS). Secondary antibody was added (protein A conjugated to horseradish peroxidase) after which the membrane was again washed. Finally, the membrane was incubated with 60mg 4-chloro-1-naphtol in 20ml methanol with hydrogen peroxide. Membranes were incubated in this solution until bands became visible.

7. Purification of his-tagged protein

Cells were collected in ice-cold 1x binding buffer (5mM imidazole, 0.5M NaCl, 20mM Tris-HCl, pH 7.9) and dounced. The cell lysate was sonicated and centrifuged at 10000rpm for 20 minutes. For binding of histidine-tagged proteins Ni-NTA agarose beads were used charged with positive Ni ions in a spin column of Qiagen (www.qiagen.com). Lysates were loaded onto the column and the column was washed first with 1x binding buffer and subsequently with 1x wash buffer (10mM imidazole, 0.5M NaCl, 20mM Tris-HCl, pH7.9). Protein was eluted with 1x elution buffer (1M imidazole, 0.5M NaCl, 20mM Tris-HCl, pH7.9). Fractions were collected of all steps of the procedure and analysed on SDS-PAGE gel. The binding buffer was adapted to contain less imidazole (2.5mM) and also the wash buffer contained less imidazole (5mM)

8. Solubilization of VP4

8.1 Solubilization from the pellet

Cells were collected by centrifugation at 2400*g* and resuspended in a buffer containing 10mM Tris-HCl pH 7.6, 0.5% Nonidet. The cell solution was dounced repeatedly to lyse the cells. The cell extracts were centrifuged at 2000*g* for five minutes and the pellet was reconstituted in buffer containing 10mM Tris-HCl pH7.6 and 500mM $MgCl_2$.

8.2 Solubilization using high salt lysis buffers

Lysis buffer (50mM HEPES, pH7.4, 0.5% Nonidet with different concentrations NaCl [0M, 0.1M, 0.25M, 0.5M and 1M]) was added to the cell pellet and after 30 minutes on ice dounced and centrifuged at 2000rpm for 5 minutes, before being loaded onto a SDS-PAGE gel.

8.3 Role of pH in solubility

To investigate pH levels, lysis buffer as described in 3.8.2 with either 0M NaCl or 0.5M NaCl was adjusted for pH (pH 7.8, 8.1 or 8.6) using HCl or NaOH. Cells were lysed as described above and centrifuged at 2000rpm for 5 minutes. Pellets and supernatants were analysed on SDS-PAGE gel. All samples were kept on ice to keep the pH stable.

9. Sucrose cushion and sucrose gradient centrifugation

Cells were lysed by douncing in 0.5M NaCl with 0.5% Nonidet. The cell lysate was loaded on top of a 50% sucrose cushion in 250mM NaCl and 50mM HEPES, pH7.4. The cushion was centrifuged at 30000rpm for 3 hours. Separate fractions were extracted and analysed on 10% SDS-PAGE gel. A 20-50% sucrose gradient was made in 250mM NaCl, 50mM HEPES, pH 7.4. After cell lysate was loaded on top of gradient, it was centrifuged at 35000rpm for 17 hours. Fractions were collected and analysed on SDS-PAGE gel.

10. Nucleotide phosphatase assay

The protein was added to reaction buffer (15mM $MgCl_2$, 90mM HEPES, pH 7.4) along with 2mCi α ^{32}P ATP. The reaction was incubated for 30 minutes at 37°C. Alkaline phosphatase was included in the assay as positive control. After stopping the reaction by adding 0.5M EDTA, a drop of each reaction was spotted on a PEI cellulose chromatography plate. The plate was developed in 0.375M KH_2PO_4. Plates were autoradiographed overnight at -70°C.

11. Inorganic pyrophosphatase assay

The protein was added to a reaction mixture containing 50mM Tris-HCl, pH8, 60mM NaCl, 6mM $MgCl_2$ and 0.6µCi ^{32}P PPi. The reaction was incubated for 30 minutes at 37°C. The reaction mixture was put on ice to stop the reaction after which drops were spotted on cellulose plates and developed as described above.

12. GTP-PPi exchange assay

Protein was added to 60mM Tris-HCl, pH 8, 10mM DTT, 5mM $MgCl_2$, 0.05µmol GTP and 0.26µCi ^{32}P PPi. The mixture was incubated for 30 minutes at 25°C. To stop the reaction 100ul 100mM PPi, 2µl 2% BSA and 500µl cold tri-chloro-acetic acid (TCA) was added. After 30 minutes on ice the reaction was centrifuged and the supernatant added to 30% (w/v) activated charcoal and incubated on ice for 15 minutes. After centrifugation the pellet was washed three times with 5mM HCl. Finally, the pellet was resuspended in 200µl 50% ethanol, 2% NH_4OH and incubated on ice for 15 minutes. After centrifugation the supernatant was freeze-dried overnight. The resulting pellet was resuspended in de-ionized water and spotted onto cellulose plates and developed with NH_4HCO_3.

3.3 Results

1. Cloning of VP4 into the histidine tag vector

To express the full-length VP4 gene with the histidine tag, the VP4 gene was cloned into the HtC histidine expression vector as described in section 3.2.1 of materials and methods. Full-length AHSV 3 VP4 was recovered from the initial cloning vector, pBR 322, through *Bgl II* digestion and cloned into the *Bam HI* site of the HtC baculo

expression vector. The pFastBac HtC vector was used to allow the VP4 ORF to be in-frame with the histidine tag upstream of the MCS.

After digestion of the different plasmids, both the HtC vector and VP4 were isolated from the agarose gel and ligated using T4 ligase. Figure 3.1 shows the plasmid map of HtC vector containing VP4 in the correct orientation indicating the expected fragment size after digestion with *EcoRI* and *BamHI* as a 1500bp fragment. After miniprep DNA extraction and digestion of the ligated and transformed samples, they were analysed on agarose gel. In figure 3.2 lane three and four two samples are analysed. In both lanes a band is visible at the correct size indicating that these two clones (HtC-VP4a and HtC-VP4b) have the VP4 genes incorporated into the HtC expression vector in the correct orientation. To produce virus containing the recombinant protein, the HtC-VP4a recombinant construct was first transposed into the bacmid DNA containing the genes for all the virus proteins necessary to produce virus. After transposition cells were plated on agar plates and colonies were inoculated. DNA isolated from these colonies was used to perform a PCR to determine the recombinancy of the bacmid DNA. For the PCR, primers for the polyhedron promoter upstream of the inserted gene and the M13 reverse primer were used. A PCR with these primers will yield a 2430bp fragment if only the HtC vector was transposed into the bacmid while a HTC-VP4 containing bacmid will yield a fragment of 4330bp (2430bp + 1900bp). In figure 3.3, lane 2, the negative control of an empty bacmid vector results in a band on gel of 2400bp in size while the presence of a 4300bp band is shown in lane 3 and 4 for two recombinant bacmids hereafter called HIS-VP4bac-a and HIS-VP4bac-b.

The resulting HIS-VP4bac-a was transfected into cells to produce recombinant HIS-VP4 virus. Resulting HIS-VP4 virus supernatant was harvested and used to infect to S. frugiperda allow VP4 expression.

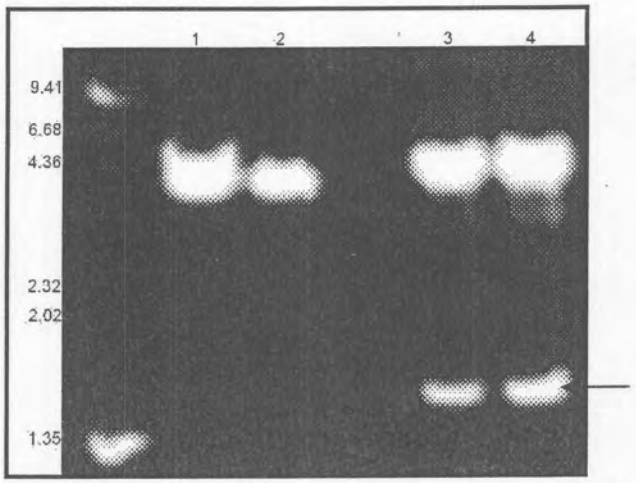


Figure 3.2. Restriction enzyme digestion indicating the presence and correct orientation of the VP4 fragment in the HtC expression vector. Lane 1 and 2 indicate the result of digestion of wild-type HtC vector with BamHI and EcoRI while lane 3 (HtC-VP4a) and 4 (HtC-VP4b) represent two colonies selected containing a 1500bp fragment which represents the VP4 cDNA in the correct orientation as indicated in figure 3.1.

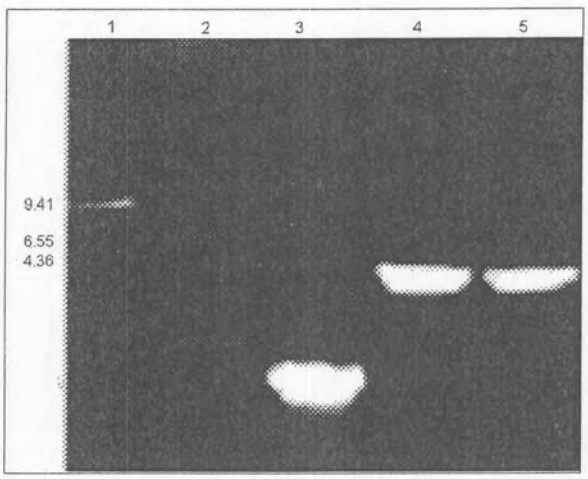


Figure 3.3. Result of PCR of his-VP4 bacmid DNA with the polH and M13 primers. Lane 1, molecular markers. Lane 2, negative control. Lane 3 contains an empty HtC vector transposed bacmid. Lane 4 (HIS-VP4bac-a) and 5 (HIS-VP4bac-b) contain a 4.3kb band representing the Ht-vector with the VP4 gene.

2. Expression and identification of his-VP4

To produce VP4 protein, S. frugiperda cells were infected with the HIS-VP4 virus or with a his-empty virus stock. After cells were collected by centrifugation, proteins were extracted from cells by adding 1x Protein Solvent Buffer. Supernatants of wild type HtC vector- and recombinant VP4 virus-infected cells were loaded on a 10% SDS-PAGE gel to be analysed by Coomassie staining. In figure 3.4 the Coomassie stained gel is shown with the empty vector expressing cell extract in lane two and the VP4 expressing cell extract in lane three. When the two lanes are compared a band is visible in lane three at about 75kD that is not present in lane two. This is a good indication that VP4 is expressed. To identify VP4 specifically a more accurate method was employed.

3. Western blot analysis of his-VP4 expression

With a Western blot a protein can be specifically identified with the use of antibodies (Ab) that are directed against that protein. A secondary antibody recognizing the Fc part of the primary antibody is linked to a peroxidase that converts its substrate to form a coloured residue where the antibody has bound to the membrane or to the protein. To positively identify VP4, a Western blot was performed with full-length and histidine tagged VP4. VP4 was visualized using the guinea pig AHSV-3 antiserum as primary antibody. In figure 3.5 the results of the Western blot are represented. In lane 2, loaded with full-length VP4 and lane 3, loaded with his-tag VP4, a band is visible on the membrane after antibody incubation. This band corresponds to the 75kD band predicted for VP4. There is a slight size difference between the two bands as the full-length VP4 in lane 2 is smaller than the histidine tag fusion protein in lane 3. With this Western blot VP4 was positively identified in cells expressing it as full-length and as histidine tag fusion protein in the baculovirus expression system.

4. Purification

For purification the his-tag attached to VP4 was used. The tag allows the protein to bind to Ni^{2+} loaded resin. The standard protocol was used as described in Material and Methods. To purify VP4, cells were lysed in ice-cold binding buffer and dounced. After sonication the resulting protein solutions were centrifuged to remove the cell debris. The supernatant was allowed to bind to the Nickel beads before the column was washed twice. Purified protein was eluted with elution buffer.

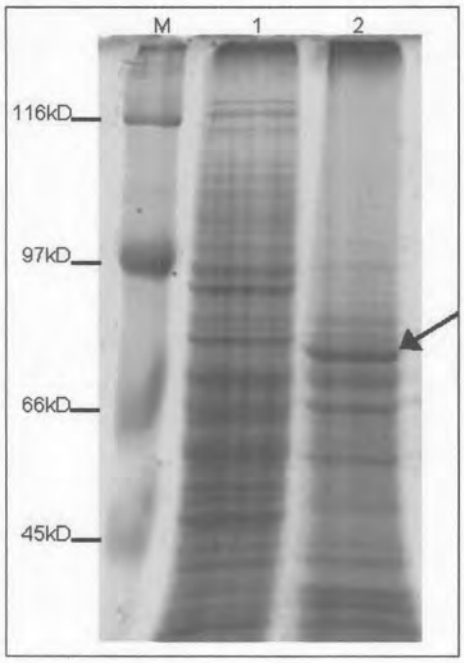


Figure 3.4. A Coomassie stained gel indicating the expression pattern of HIS-VP4 virus infected insect cells (lane 2). The pattern is compared with that of wild-type insect cells (lane 1). The arrow indicates the band representing his-VP4.

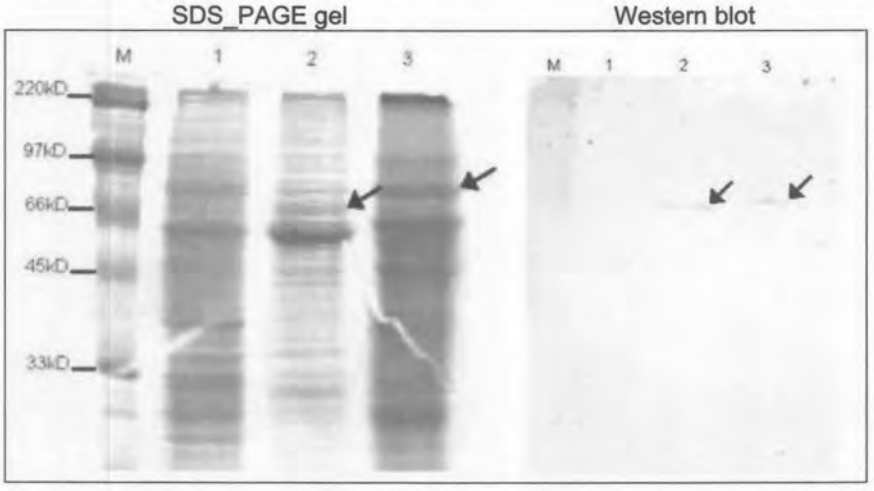


Figure 3.5. Western blot of cell lysates from wild type infected Sf9 cells (lane1), Sf9 cells infected with the AcMNPV-VP4 virus (lane2), and cells infected with the HIS-VP virus (lane3). Arrows indicate the presence of VP4 protein. The size difference between the full-length and the histidine tagged VP4 can be explained by the presence of the his-tag added to the full-length gene.

A sample of unpurified WT and VP4 cell extract were loaded onto a 10% SDS-PAGE gel along with samples of all the steps in the purification. These fractions include the flow-through, pellet, washes, and final elution steps as shown in figure 3.6. When analyzed on SDS-PAGE gel, VP4 was present in the fractions representing the flow-through in lane 3 but most of the protein was present in the pellet as seen in lane 6 of figure 3.6. This means that most of the VP4 was already present in the pellet before the extract was applied to the column. The VP4 left in solution did not bind to the beads in the column as no protein was found in lane 4 after washing the column or in lane 5 after elution from the column. Recommended adjustments including changes in the imidazole concentrations did not lead to a better purification. This data indicates that VP4 is found in the pellet after cell lysis and may therefore be considered insoluble.

5. Solubilization of VP4 from the pellet

In previous articles describing the expression and isolation of BTV VP4 high salt buffers were used to make the insoluble BTV VP4 soluble. In figure 3.6 it is apparent that AHSV VP4 is also insoluble. Different strategies were employed to obtain a soluble fraction. Cell extracts of WT and VP4 expressing cells were pelleted and the pellets were subsequently reconstituted in a high salt buffer. The high salt buffer used contained 500mM $MgCl_2$. Fractions of the different steps in the process were taken for analysis. After the initial centrifugation most VP4 is present in the pellet as can be seen in lane 3 of figure 3.7. A very small amount of VP4 is present in the supernatant after this initial centrifugation step (lane 4). When the pellet is reconstituted in high salt buffer a band corresponding to VP4 is seen in lane 5 representing the supernatant with very few contaminating proteins. However, most VP4 is still found in the pellet after reconstitution in high salt buffer and subsequent centrifugation (lane 6). This experiment was also repeated with 0.5M and 1M NaCl in the suspension buffer but similar results were obtained. It is clear from figure 3.7 that VP4, in contrast to BTV VP4, does not become soluble after the addition of high concentrations of salt to the pellet.

6. Lysis of cells in high salt concentration buffer

In the previous section it was apparent that VP4 stays insoluble after it has been pelleted even when reconstituted in high salt buffers. The centrifugation step may influence the solubility of VP4 by creating an environment in the pellet where it will be

even more hydrophobic and thus make VP4 more insoluble. To avoid this, the cell lysis buffer was adjusted to contain a high salt concentration that could keep VP4 in solution before centrifugation was performed on the cell extract. The lysis buffer containing either 0.5M NaCl or 1M NaCl was used to lyse VP4 expressing cells after which the samples were centrifuged at low speed. Cells were dounced to mechanically break them open as the high salt buffer might cause the cell to shrivel up due to osmosis. The results of this experiment are shown in figure 3.8. VP4 is present in the VP4 expressing cell extract (lane 2) as indicated by the arrow. The pellet from cells lysed in either 0.5M or 1M NaCl is shown in lane 3 and 5 respectively. The supernatants of both the treatments are shown in lane 4 and 6. It is clear that most VP4 is present in lane 3 and 5 (pellets) while a small fraction is visible in lane 4 and 6 (supernatants).

This indicated that VP4 was also not soluble under these conditions. Additionally, it must be mentioned that these results were obtained after low speed centrifugation. Extract centrifuged at higher speed had even less protein in the soluble fraction. This data leads us to believe that VP4 is insoluble even after initial cell lysis with high salt concentrations. The data indicates that VP4 was already insoluble before centrifugation. Therefore, adding salt to the lysis buffer had no effect on VP4 solubility.

7. Effect of pH on solubility

The effect of different pH values in the lysis buffer on the solubility of VP4 after cell lysis was investigated. The iso-electric point (pI) is the pH at which the overall charge of the protein is closest to zero and therefore will precipitate most readily. Therefore, proteins in a pH solution around their pI value are most likely to be insoluble (Whittles K. 1993). The pI of VP4 is 8.33. Lysis buffers were adjusted for pH to be below the pI (pH 7.8 and 8.1) or above the pI (pH 8.6). The lysis buffers contained 0.5M NaCl. The resulting cell extract was centrifuged and the different samples analysed on SDS-PAGE gel as shown in figure 3.9. Expression of VP4 is evident in the band indicated by an arrow in lane 3 when compared to the WT expressing cells loaded in lane 2 and the mock infected cells loaded in lane 1. The pellet recovered after lysis with the different buffers all contains the majority of the VP4.

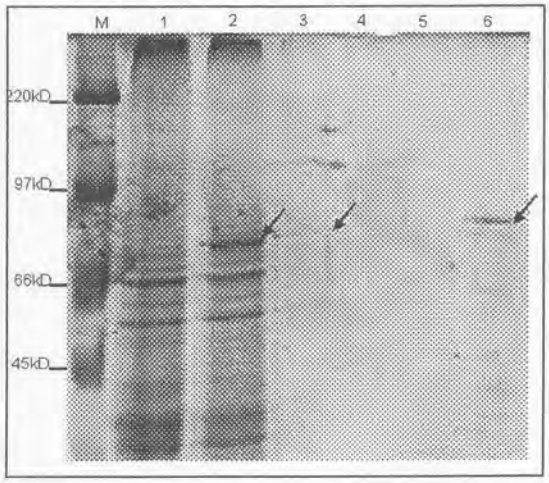


Figure 3.6. Histidine purification of VP4. Wild-type cell extract was loaded in lane 1 with cell extract from the VP4 expressing cells loaded in lane 2. A sample of the flow-through from the column was loaded in lane 3, showing a faint band at the correct height. A sample from the wash step was loaded in lane 4 and a sample of the final elution of VP4 was loaded in lane 5. The pellet from the initial cell lyses and centrifugation was loaded in lane 6. VP4 is present in the flow-through but mostly in the pellet.

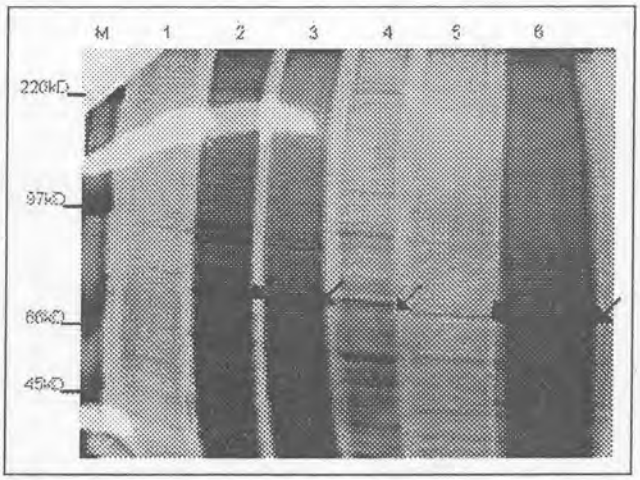


Figure 3.7. SDS-PAGE gel indicating the presence of VP4 in the pellet after reconstitution in high salt buffers. Lane 1 contains the cell extract from infected cells as control. In lane 2 VP4 is visible and indicated by the arrow. This is the initial cell extract of his-VP4 infected cells. Lane 3 contains the proteins present in the initial pellet after cell lyses indicating the presence of VP4 (arrow). Lane 4 is the supernatant retrieved after the initial centrifugation where a minor VP4 protein band is visible. Lane 5 is the supernatant after the pellet was reconstituted in high salt buffer and lane 6 is the pellet formed after the initial pellet was reconstituted in high salt buffer.

Lane 4 and 5 represent the supernatant and pellet from lysis in pH 7.8 respectively, lane 6 and 7 represent the supernatant and pellet from lysis in pH 8.1 respectively and lane 8 and 9 represent the supernatant and pellet from lysis in pH8.6 respectively. A well-defined band is visible in every lane containing a pellet fraction while there is no or very little VP4 present in the supernatant. Adjusting the pH to this degree did not induce solubility of VP4. Whether a more extreme pH for the lysis buffer would alter the result is not known.

8. Effect of sucrose cushion on VP4 solubility

From the previous experiments it seems evident that very little VP4 is present in the soluble cell supernatant fraction after cell lysis. It is possible that the inclusion of a sucrose cushion might improve the recovery of the VP4. To test this, cell lysates (in 0.5M NaCl) of cells infected earlier with the VP4 baculovirus recombinant and a WT control were loaded on top of a 50% sucrose cushion and centrifuged. Sucrose gradient fraction were analysed on SDS-PAGE gel as shown in figure 3.10. Lane 1 and 2 represent the initial cell extract of WT and VP4 expressing cell extracts. A band corresponding to the size of VP4 is present in lane 2 while absent in lane1. Lane 3 and 4 represent WT and VP4 expressing cell extract samples taken from the solution above the sucrose cushion in the tube after centrifugation and represent the soluble proteins in the cell extract. In lane 4 a band is visible at the same position as the VP4 band seen in lane 2. This indicates that at least a fraction of the VP4 in the cell lysate is present in the soluble fraction. The sucrose cushion itself contains very little protein as seen in lane 5 (WT) and 6 (VP4). The pellets were loaded in lane 7 (WT) and 8 (VP4). The VP4 expressing lysate pellet (lane 8) shows a band indicated by the arrow corresponding to the size of VP4. From this experiment it is clear that most VP4 is aggregated and moves through the sucrose cushion. A possible band representing VP4 is present in lane 4 and may represent a small fraction of VP4 that is soluble. However, this band seems similar to the faint WT band present just above the VP4 band in the initial cell extract in lane 2. This data indicates that VP4 is still mostly insoluble even when centrifuged in the presence of a sucrose cushion. The pellet is retains very few contaminating proteins when compared to the initial cell lysate (lane 2).

From the previous experiments it is clear that baculovirus expressed VP4 is largely insoluble. Adjusting the salt concentration or pH n did not improve VP4 solubility significantly. However, the pellet containing VP4 after centrifugation over a sucrose

gradient is relatively free of contaminating proteins and the pellet fraction was therefore reconstituted and used in a number of assays to determine if VP4 has any biological activity similar to its BTV counterpart.

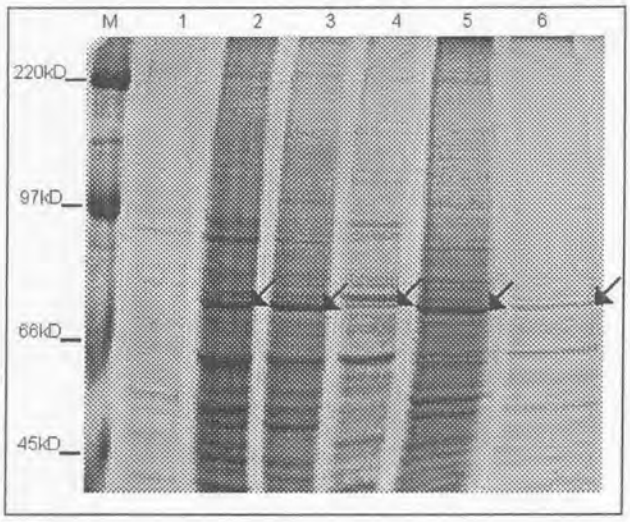


Figure 3.8. Analysis of the different fractions after lysis of VP4 expressing cells with high salt buffers. His-VP4 expressing cells were lysed in buffer containing 0.5M (lane 3 and 4) or 1M (lane 5 and 6) NaCl. Both conditions result in a large amount of VP4 in the pellets (lane 3 and 5) while a small amount of VP4 is found in the supernatant (lane 4 and 6). Lane 1 contains wild type infected cells while lane 2 contains a complete cell extract of VP4 expressing cells.

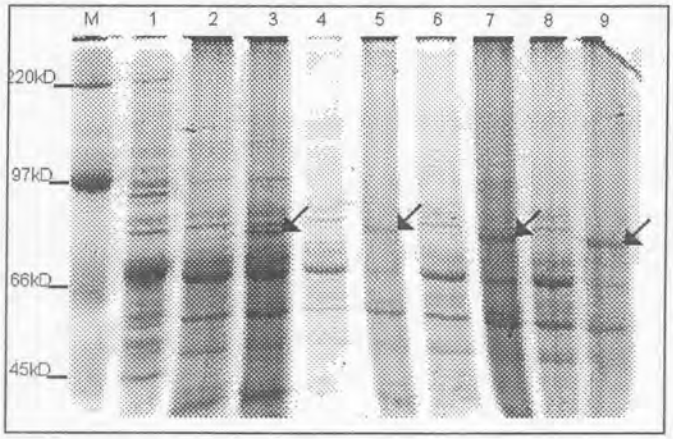


Figure 3.9. The effect of pH on the solubility of his-VP4. Mock (lane1) and wild type (lane2) infected cells are included as control. VP4 is found in the pellet (lane5) after lyses in pH 7.8 with no VP4 present in the supernatant (lane 4). Moreover, VP4 is almost exclusively present in the pellets of cells treated with pH 8.1 (lane7) and pH 8.6 (lane 9). A small amount of VP4 is evident in the supernatant of cells treated with pH8.1 (lane 6) but no VP4 visible in the supernatant of cells lysed at pH 8.6 (lane 8).

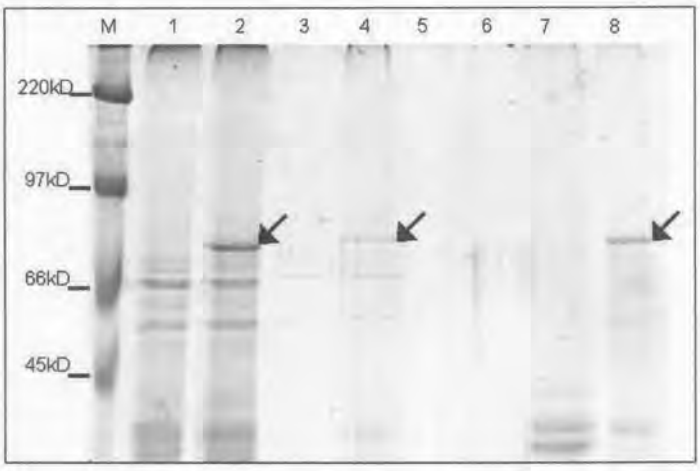


Figure 3.10. Centrifugation of the soluble fraction of VP4 through a sucrose cushion. Lane 1 contains the control wild type while lane 2 contains the control total protein from VP4 expressing cells. The top fractions of wild type and VP4 are shown in lane 3 and 4 respectively. A band that seems a bit larger than VP4 is seen in the top fraction of VP4 (lane 4). The sucrose cushion fraction of wild type (lane 5) and VP4 (lane 6) contain very few proteins. The pellet of wild type (lane 7) seems relatively free of protein while the VP4 pellet contains VP4 (lane 8)

9. Inorganic phosphatase assay

Most capping enzymes have inorganic pyrophosphatase activity. Determining whether VP4 has pyrophosphatase activity will give a good indication that VP4 is a capping enzyme. Pellets from both wild type and VP4 were reconstituted. Additionally, the supernatant after cell lysis of wild type and VP4 expressing cells were used in the reaction mixture. The protein samples were added to a reaction mixture containing ^{32}P-labelled PPi (pyrophosphate). The mixtures were blotted on TLC plates and developed. If phosphatase activity is present a spot representing inorganic phosphate will be present. In figure 3.12 the result of the incubation is shown. In lane 1 and 2 the supernatants of WT and VP4 expressing cells were loaded and have a band present at the height of inorganic phosphate indicating that there is phosphatase activity in both WT and VP4 samples. No distinction can be made between these two samples making it impossible to indicate that VP4 has any activity.

The pellets reconstituted from WT and VP4 expressing cells are represented in lane 3 and 4 respectively. In both lanes a band is present representing phosphate. These data indicate that both the WT and VP4 samples have pyrophosphatase

activity. It is impossible to identify VP4 as a possible pyrophosphatase from the pellet or the supernatant fractions due to the background activity seen in the WT fractions.

10. Nucleotide phosphatase assay

While the inorganic pyrophosphatase assay gives an indication of possible capping activity, nucleotide phosphatase represents the first actual step in the capping process. To determine whether VP4 has nucleotide phosphatase activity, it is incubated with γ-phosphate ^{32}P labelled ATP. If VP4 has nucleotide phosphatase activity a ^{32}P P_i fraction will be detected after thin layer chromatography. Reconstituted pellets from WT and VP4 expressing cells were added to the reaction buffer together with labelled ATP and incubated before being spotted on the nitro-cellulose plates and developed. In addition, a negative control consisting of VP6 protein was added as well as a positive control, alkaline phosphatase. In figure 3.13 the result of this experiment is shown. In lane 1 and 2 the WT and VP4 reaction mixture was loaded. In both lanes no labeled ATP is left while a small amount of pyrophosphate is present with the majority of ^{32}P label present as phosphate.

Since the protein solution was also not pure from contaminating proteins a band is present at the height of phosphate indicating the presence of phosphatases. This was not a good negative control. The data points out that there is active protein present in the pellet of both WT and VP4 containing fractions. This data is not able to identify VP4 as a potential phosphatase although it cannot exclude it either. However, it is interesting that PPi was produced in addition to Pi indicating that both pellets contain a protein catalysing the breakdown of ATP to AMP and PPi in contrast to alkaline phosphatase that only removes the γ-phosphate group.

11. The GTP-PP_i exchange assay

IF VP4 is a capping enzyme it has to be able to transfer the GMP moiety to the terminal pyrophosphate group of the RNA strand. This assay indicates whether a protein is able to transfer a labelled PPi to an unlabelled GTP moiety. The previous nucleotide phosphatase assay indicated that ATP is broken down to ADP and AMP indicating that there is a phosphatase that can initiate the first step in the capping reaction.

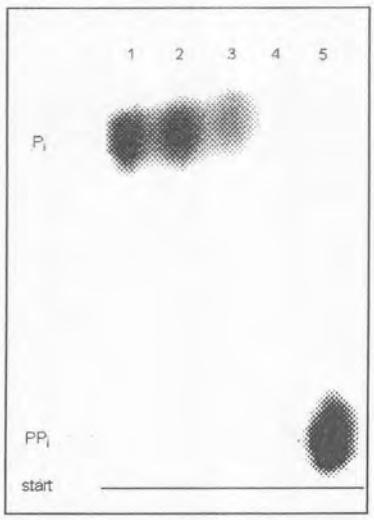


Figure 3.11. Inorganic pyrophosphatase assay. The supernatants of wild type and VP4 are shown in lane 1 and 2 respectively. The pellet of wild type is shown in lane 3 while the VP4 pellet is in lane 4. As negative control the reaction was incubated without added protein and resulted in only PP_i. All four fractions have pyrophosphatase activity.

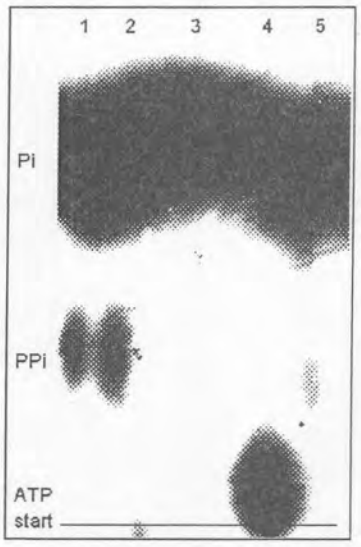


Figure 3.12. Nucleotide phosphatase assay of VP4. VP4 (lane2) along with wild type (lane 1) were incubated with labelled ATP. Alkaline phosphatase was included as positive control (lane 3) and labelled ATP was incubated with the reaction buffer alone as negative control (lane 4). Additionally, a second negative control was added consisting of VP6 (lane 5). The positive control (lane 3), alkaline phosphatase, was more efficient in producing phosphate since no labeled pyrophosphate is present but only labeled phosphate. In lane five VP6 was used as a negative control.

The second step, transfer of PPi to GMP can be visualized with this experiment. Labelled pyrophosphate and GMP is added to a reaction mixture containing the different protein samples. If enzymatic activity is present labelled GTP should be present after analysis on TLC plates. VP4 and wild type fractions reconstituted from the pellet were used for this assay. Initially, cleaning steps with activated charcoal were included in the assay to reduce the amount of labelled PPi in the sample. After the fractions were purified samples were loaded on cellulose plates and developed. No spots could be detected after the cleaning steps that are indicated in the protocol described in 3.2.12. Therefore, the assay was also performed without cleaning the reaction mix after incubation and under those conditions spots became visible as can be seen in figure 3.14. In lane 1 and 2 supernatants of WT and VP4 expressing cell lysates were incubated in the reaction mixture. The pellets of WT and VP4 were loaded in lane 3 and 4 respectively. In all four lanes only labelled phosphate is present while no spot was seen on the level of ATP loaded in lane 7. In lane 5 the negative control only contains labelled PPi in the reaction mixture similar to lane 6 where only PPi was loaded without reaction mix.

From this experiment it is clear that there is no labelled GTP present in the lanes containing either WT or VP4 containing protein samples. The absence of labelled GTP can be due to the absence of an active GTP-PPi exchange enzyme or the presence of a phosphatase that converts all labelled GTP into GMP and labelled phosphate. From previous experiments we know that a phosphatase is present and active.

3.3 Discussion

The aim of the work described in this chapter was to express AHSV-3 VP4, purify it and use to characterize the enzymatic activities it may possess. AHSV-3 VP4 was expressed as a fusion protein with a histidine tag. This tag would allow for easy one-step purification of the protein. VP4 was expressed in the baculovirus expression system. The baculovirus expression system makes use of a recombinant baculovirus that infects insect cells and in the cells allows the expression of the recombinant protein. The advantage of this system over bacterial expression systems is the ability of the cellular machinery of the insect cells to fold the protein in the correct conformation and in this way produce active protein.

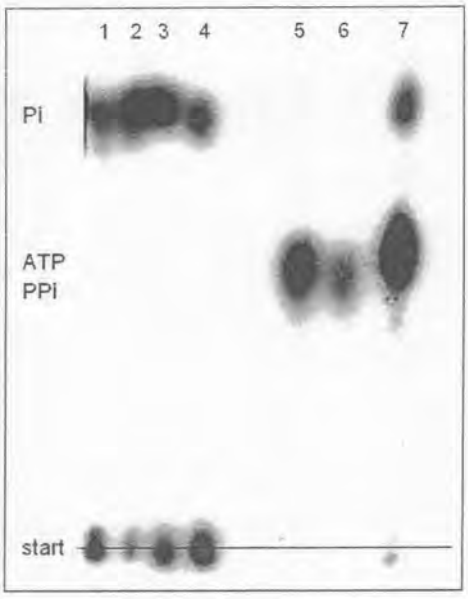


Figure 3.13. GTP-PPi exchange assay. After the incubation the solutions were developed. Wild type (lane3) and VP4 (lane 4) pellets are developed along with the supernatant of wild type (lane 1) and VP4 (lane 2). No labelled GTP is found, as GTP should be present on the same distance from the start as ATP (lane 7). Only labelled phosphate is present in these lanes. Lane 5 contains a negative control while lane 6 contains only labelled PPi.

This system was also used in previous studies identifying the different functions of BTV VP4 (Ramadevi N and Roy P., 1998; Martinez-Costas J., *et al.*, 1998; Le Blois T., *et al.*, 1992). In contrast to the BTV VP4 expression studies a histidine tag was added to AHSV VP4 to aid purification of the protein after cell lysis. The histidine tag has a high affinity for positively charged ions like Nickel. These can be used in a column to bind the protein while letting contaminating proteins pass through. However, it is not unlikely that the tag has some effect on the solubility and conformation of the protein. It has been published that using a histidine tag either directly coupled or via a linker sequence that includes a protease site, a protein can obtain different properties (Rumlova M. *et al.*, 2001). When a protein was expressed as wildtype or with a histidine tag or with a histidine tag and protease site different effects were observed. The histidine tag allowed the protein to form organized structures in the bacterial expression system while the wildtype protein formed amorphous aggregates. The WT and histidine protease site tagged protein formed inclusion bodies while the histidine tagged protein was soluble. This paper indicates that adding the histidine tag can have an effect on the properties of the protein. BTV VP4 was already found to be insoluble after cell lysis (Ramadevi N and Roy P., 1998; Martinez-Costas J., *et al.*, 1998; Le Blois T., *et al.*, 1992). In all three papers a high

salt buffer was used to solublilize VP4 from the initial pellet. However, it is interesting to note that the first paper (Le Blois T., *et al.*, 1992) used a much lower salt concentration of 100mM $MgCl_2$ or 200mM NaCl to solublilize VP4 while the two later papers reported using 1M NaCl to achieve the same effect. The reason for this difference is not clear. When AHSV VP4 was expressed as a fusion protein with a histidine tag it was also mostly found in the insoluble fraction of the cell lysate. From the results in this chapter it is clear that solubilizing VP4 with high salt buffers was not successful. This was the case for adding high salt buffer to the pellet or alternatively, lysing the cells in high salt buffer.

The pH of the lysis buffer can also assist in making a protein soluble. As mentioned earlier the pI of a protein is the pH value at which the protein is least soluble. Using a more acidic or basic lysis buffer may make the protein soluble. In this chapter the lysis buffer was adjusted to a pH below and above the pI value of VP4 (8.3). This experiment did not produce any soluble VP4. Changing the pH too much may have negative effects. An alkaline pH may induce protein degradation due to alkaline proteases and can also allow non-specific binding of proteins to the resin used for his-tag purification (BD Biosciences). In literature a new method for producing purified protein has been described using the pI of the protein as a first purification step. Here it is described that using liquid phase iso-electric focusing, protein can be concentrated within a narrow acidic pH range (Yvon S., *et al.*, 1998). While this is used for soluble proteins it may be possible to adapt this type of protocol to solublilize protein by changing the pH to a much more acidic environment. Adjusting the pH is not a new method in producing soluble protein and has been widely used.

When using a bacterial expression system one problem encountered is the production of too high levels of proteins. These highly concentrated proteins form aggregates and inclusion bodies that are hard to bring into solution and are often inactive. The expression of VP4 as judged from the different experiments performed in this chapter is not high. It does occur that low protein expression may also induce the formation of aggregates (personal communication, P. Knipscheer). Inducing higher expression by the use of a different promoter or reducing the effect on transcription of the 5' UTR by deleting it from the cDNA used may help in obtaining higher expression levels.

Other expression systems may also be utilized to express VP4. One is the mammalian expression system. The advantage of using this system is that the

proper folding enzymes are available and in this way may reduce the change of aggregation. Alternatively, a yeast based expression system may be used. For the identification of the Bamboo Mosaic virus guanylyltransferase a yeast based expression system was used. The protein, VP6, was found mostly in the pellet initially, but after centrifugation on a sucrose gradient, soluble fraction were purified from the gradient and used in functional assays (Li Y-I, *et al.*, 2001).

Another expression system has been described that is able to express human cytomegalovirus (HCMV) DNA replication proteins that are lowly expressed in other systems and insoluble (McCue L.A. and Anders D.G., 1998). The Semliki Forest virus was used to express these proteins in mammalian cells. All of the proteins expressed in this way were soluble. This system would also be an alternative for the baculo virus expression system used in this chapter.

Another method to produce soluble protein is used extensively in the structural biology. Instead of using the complete insoluble protein, fragments of the protein are produced that are soluble and still have the functional characteristics of their full-length counterparts. For instance it was possible to identify the avian reovirus guanylyltransferase after expressing the full-length L3 genome segment. The resulting protein was partially digested in a 100kDa and a 42kDa fragment with the smaller fragment retaining the activity of the full-length protein. This was also true for reovirus lambda2 (Hsiao J., and Martinez-Costas J., 2002; Luongo C.L. *et al.*, 2000). In this way deductions can still be made about the protein activity while a soluble and therefore usable protein is produced. Making VP4 subclones and expressing them may be a useful tool in characterizing its functions. A motif for the lysine containing GMP transfer activity has been identified in the amino acid sequence (figure 2.9A) and therefore a more targeted approach can be taken when making shorter protein fragments.

The his-tag has undergone a lot of refinements since its early use. Several commercial systems have adapted the histidine tag to be more physiological. For instance, instead of using a stretch of six histidines an amino acid sequence from chicken lactate dehydrogenase is used that contains six histidines within a sequence 18 amino acids. This tag has less effect on the solubility of proteins tagged (BD biosciences).

In literature no references were found for the use of the insoluble protein for activity assays. It is assumed that insoluble proteins are inactive due to their aggregated form. Surprisingly, both the nucleotide phosphatase and the inorganic phosphatase assays indicated that there were active phosphatases in the insoluble samples. It was impossible to discriminate between the WT and the VP4 insoluble fractions in terms of phosphatase activity. It can be concluded that there are active phosphatases present in the pellet. Whether VP4 is also responsible for this activity is not known. It is possible that some soluble protein is caught up in the aggregates that migrate to the pellet during centrifugation. This protein may be reconstituted and remain active and thus give a positive result in the functional assays.

The guanylyltransferase assay indicated no formation of labelled GTP that would point to the presence of an active transferase. It is possible to deduce from this that VP4 is not active in the pellet and as such no GTP was formed. However, it is also possible that even though GTP was formed the active phosphatases present in the pellet removed the labelled phosphate from the newly synthesized GTP and therefore no labelled GTP was found.

Using the baculovirus expression system and the histidine fusion tag it was impossible to provide evidence that AHSV VP4 is the capping enzyme in AHSV. To produce soluble, active protein a number of alternative approaches are suggested. These alternative methods will have to be investigated in order to successfully produce soluble protein and subsequently use it to characterize the enzymatic activities of AHSV VP4.

Chapter 4
Conclusions

AHSV is a member of the orbivirus genus of which BTV is the prototype. Viruses in this genus have segmented RNA genomes from which mRNA segments are produced inside the core that are extruded into the cytoplasm of the infected cell to be used as templates for the expression of viral proteins. The minor core proteins are responsible for the production of mRNA. A large body of research is available on the activity of BTV core proteins. The polymerase, capping enzyme and helicase are all identified and for the capping enzyme the specific activities have been established. It is clear from the second chapter and other published data that there are some differences between the capping enzymes of BTV and other orbivirus members. The sequencing and subsequent analysis of the nucleotide and amino acid sequence of AHSV-3 VP4 has pointed out these differences but also the similarities. A possible lysine-containing motif has been identified in AHSV that may be responsible for the GMP binding essential to the transfer step in capping of nascent RNA. Additional differences have been identified and will warrant further research.

Although the different enzymatic activities necessary for successful capping of RNA have been identified in BTV, the specific motifs responsible for these activities are still largely unknown. Using a soluble AHSV VP4 protein and fragments of the protein will help in determining the regions responsible for these function. Several conserved regions have been found and these should receive special attention when trying to identify the motifs involved. However, it is clear from the alignments performed in chapter 2 that the lysine motif is not conserved between the different species. This means that there may be unique motifs in the different proteins responsible for the enzymatic activities. These can be identified by mutational analysis as previously performed on other viral capping enzymes (Ramadevi N., *et al.*, 1998).

It remains important to produce a soluble, pure protein that can be used in these functional assays. The use of a his-tag may have proven unsuccessful but other methods may be successful. Alternatively, adjustments as discussed above may be sufficient to produce such a soluble protein. The purification and production of such fusion proteins has been evolving over past few years and today there are many options available to alter the solubility of the protein. A important future push should

be in the production of protein fragments that may have one or two enzymatic activities as opposed to the complete protein. These protein fragments will be much more informative.

Another area of interest is the effect of the 5' and 3' UTR regions on expression. Proteins like VP2, VP5 and VP7 have much higher expression levels in the same baculo virus expression system than VP4. This may seem logical given the fact that in the viral core very few copies of VP4, VP1 or VP6 are found in comparison to the outer core proteins or the major core proteins. In other viruses like rotaviruses these differences have been ascribed to the untranslated regions. A similar mechanism may also be responsible for the translational regulation in orbivirus.

Understanding the mechanisms through which viruses like AHSV replicate and proliferate may help in the design of specific drugs inhibiting the expression of viral protein needed for virus survival.

Reference List

1. Aloha T. and Kaarianinen L. (1995). Reaction in alpha virus mRNA capping: Formation of a covalent complex of non-structural protein nsP1 with 7-methyl-GMP. *Proc.Natl.Acad.Sci.USA*. **92,** 507-511.

2. Attoui H., Stirling J.M., Munderloh U.G., Billoir F., Brookes S.M., Burroughs J.N., de Micco P., de Lamballerie X. (2001). Complete sequence characterization of the genome of the St Croix River virus, a new orbivirus isolated from cells of Ixodes scapularis. *J. Gen. Virol.* **82,** 795-804.

3. Banjeree A.K. (1980). 5'-Terminal cap structure in eukaryotic messenger ribonucleic acids. *Microbiol. Rev.* **44,** 175-205.

4. Beaton A.R., Rodriguez J., Reddy Y. K., Roy P. (2002). The menbrane trafficking protein calpactin forms a complex with bluetongue virus protein NS3 and mediates virus release. *PNAS* **99**, 13154-13159.

5. Bentley L., Fehrsen J., Jordaan F., Huismans H., du Plessis D.H. (2000). Identification of antigenic regions on VP2 of African horsesickness virusserotype 3 by using phage-displayed epitope libraries. *J. Gen. Virol.* **81,** 993-1000.

6. Birnboi H.C., and Doly J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acid Res.* **7**, *1513-1523.*

7. Bisaillon M., Bergeron J., and Lemay G. (1997a). Characterization of the NTPase and helicase activities of reovirus lambda 1 protein. *J. Biol. Chem.* **272,** 18298-18303.

8. Bisaillon M. and Lemay G. (1997b). Characterization of the Reovirus Lambda 1 protein RNA 5'-triphosphatase activity. *J. Biol. Chem.* **272,** 29954-29957.

9. Bisaillon M. and Lemay G. (1997c). Viral and cellular enzymes involved in synthesis of mRNA cap structure. *Virology*. **236,** 1-7.

10. Brookes S.M., Hyatt A.D., and Eaton B.T. (1993). Characterization of virus inclusion bodies in bluetongue virus-infected cells. *J. Gen. Virol.* **74,** 525-530.

11. Burrage T.G., Trevejo R., Stone-Marschat M., and Laegreid W.W. (1993). Neutralizing epitopes of African horsesickness virus serotype 4 are located on VP2. *Virology*. **196,** 799-803.

12. Butcher S.J., Dokland T., Ojala P.M., Bamford D.H., and Fuller S.D. (1997). Intermediates in the assembly pathway of the double-stranded RNA virus phi6. *EMBO.* **16,** 4477- 4487.

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

14. Chen D. and Patton J.T. (1998). Rotavirus RNA replication requires a single-stranded 3' end for efficient minus-strand synthesis. *J. Virol* **72,** 7387-7396.

15. Chuma T., Le Blois H., Sanchez-Vizcaino J.M., Diaz-Laviada M., and Roy P. (1992). Expression of the major core antigen VP7 of African horsesickness virus by a recombinant baculovirus and its use as a group-specific diagnostic reagent. *J. Gen. Virol.* **73,** 925-931.

16. Dang C.V., McGuire M., Buckmire M., and Lee W.M.F. (1989). Involvement of the "leucine zipper" region in the oligomerization and transforming activity of c-Myc protein. *Nature.* **337,** 664-660.

17. Diprose J.M., Burroughs J.N., Sutton G.C., Goldsmith A., Gouet P., Malby R., Overton I., zientara S., Mertens P.P.C., Stuart D.I., and Grimes J.M. (2001). Translocation portals for the substrates and products of a viral transcription complex: the bluetongue virus core. *EMBO.* **20,** 7229-7239.

18. Dryden K.A., Farsetta D.L., Wang G., Keegan J.M., Fields B.N., Baker T.S., and Nibert M.L. (1998). Internal structures containing transcriptase-related proteins in top component particles of mammalian orthoreovirus. *Virology.* **245,** 33-46.

19. Eaton B.T., Hyatt A.D., and White J.R. (1987). Association of bluetongue virus with the cytoskeleton. *Virology.* **157,** 107-116.

20. Fausnaugh J. and Shatkin A.J. (1990). Active site localization in a viral capping enzyme. *J. Biol. Chem.* **265,** 7669-7672.

21. Vreede F.T., Huismans H. (1998). Sequence analysis of the RNA polymerase gene of African horse sickness virus. *Arch. Virol.* **143,** 413-419.

22. French T.J. and Roy P. (1990). Synthesis of bluetongue virus (BTV) core-like particles by a recombinant baculovirus expressing the two major structural core proteins. *J. Virol.* **64,** 1530-1536.

23. Furuichi Y. and Shatkin A.J. (1989). Characterization of cap structures. *Meth. Enz.* **180,** 164-177.

24. Gouet P., Diprose J.M., Grimes J.M., Malby R., Burroughs J.N., Zientara S., Stuart D.I., and Mertens P.C. (1999). The highly ordered double-stranded RNA genome of Bluetongue virus revealed by crystallography. *Cell.* **97,** 481-490.

25. Gould A.R. and Hyatt A.D. (1994). The orbivirus genus. Diversity, structure, replication and phylogenetic relationships. *Comp. Immun. Microbiol. Infect. Dis.* **17,** 163-188.

26. Grimes J.M., Burroughs J.N., Gouet P., Diprose J.M., Malby R., Zientara S., Mertens P.C., and Stuart D.I. (1998). The atomic structure of the blutongue virus core. *Nature.* **395,** 470-478.

27. Hakansson K., Doherty A.J., Shuman S., and Wigley D.B. (1997). X-ray crystallography reveals a large conformational change during guanyl transfer by mRNA capping enzymes. *Cell.* **89,** 545-553.

28. Hassan S.H., Wirblich C., Forzan M., Roy P. (2001). Expression and functional characterization of bluetongue virus VP5 protein: role in cellular permeabilization. *J. Virol.* **75,** 8356-6367.

29. Hill C.L., Booth T.F., Prasad B.V., Grimes J.M., Mertens P.C., Sutton G.C., and Stuart D.I. (1999). The structure of a cypovirus and the functional organization of dsRNA viruses. *Nat. Struct. Biol.* **6,** 565-568.

30. Hsiao J. and Martinez-Costas J. (2002). Cloning, expression, and characterization of avian reovirus guanylyltransferase. *Virology.* **296,** 288-299

31. Horscroft N.J. and Roy P. (2000). NTP binding and phosphohydrolase activity associated with purified bluetongue virus non-structural protein NS2. *J. Gen. Virol.* **81,** 1961-1965.

32. Huang I-J, Hayama E., Jeong Y-J, and Li J.K-K. (1993). Conservation of the segment 4 gene sequence and of a leucine zipper motif in VP4 among five US bluetongue viruses. *Virology.* **195,** 772-779.

33. Hyatt A.D., Eaton B.T., and Brookes S.M. (1989). The release of bluetongue virus from infected cells and their super infection by progeny virus. *Virology.* **173,** 21-34.

34. Koffa M.D., Clements J.B., Izaurralde E., Wadd S., Wilson S.A., Mattaj I.W., and Kuersten S. (2001). Herpes simplex virus ICP27proteins provides viral mRNAs with the access to the cellular mRNA export pathway. *EMBO.* **20,** 5769-5778.

35. Koonin E.V. Computer-assisted identification of a putative methyltransferase domain in NS5 protein of flaviviruses and 2 proteins of reovirus. *J. Gen. Virol.* **74,** 733-740.

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

37. Laurila M.R., Makeyev E.V., and Bamford D.H. (2002). Bacteriophage phi6 RNA-dependent RNA polymerase: Molecular details of initiating nucleic acid synthesis without primer. *J.Biol.Chem.* **277,** 17117-17124.

38. Lawton J.A., Estes M.K., and Prasad B.V.V. (2000). Mechanism of genome transcription in segmented dsRNA viruses. *Adv. Vir. Res,* **55,** 185-229.

39. Le Blois H., French T., Mertens P.C., Burroughs J.N., and Roy P. (1992). The expressed VP4 protein of bluetongue virus binds GTP and is the candidate guanylyl transferase of the virus. *Virology.* **189,** 757-761.

40. Le Blois H. and Roy P. (1993). A single point mutation in the VP7 major core protein of BTV prevents the formation of core-like particles. *J. Virol.* **67,** 353-359.

41. Lewis J.D. and Izaurralde E. (1997). The role of the cap structure in RNA processing in nuclear export. *Eur. J. Biochem.* **247,** 461-469.

42. Li Y-I., Chen Y-J., Hsu Y-H. and Meng M. (2001). Characterization of the AdoMet-dependent guanylyltransferase activity that is associated with the N terminus of bamboo mosaic virus replicase. *J. Virol.* **75,** 782-788.

43. Luongo C.L., Contreras C.M., Farsetta D.L., and Nibert M.L.(1998). Binding site for S-adenosyl-L-methionine in a central region of mammalian reovirus in lambda2 protein. *J. Biol. Chem.* **273,** 23773-23780.

44. Luongo C.L., Reinisch K.M., Harrison S.C., and Nibert M.L. (2000). Identification of the guanylyltransferase region and active site in reovirus mRNA capping protein lambda2. *J. Biol. Chem.* **275,** 2804-2810.

45. Makeyev E.V. and Bamford D.H. (2000). The polymerase subunit of a dsRNA virus plays a central role in the regulation of viral RNA metabolism. *EMBO.* **19,** 6275-6284.

46. Mao Z. and Joklik W.K. (1991). Isolation and enzymatic characterization of protein lambda2, the reovirus guanylyltansferase. *Virology.* **185,** 377-386.

47. Maree F.F., Huismans H. (1997). Characterization of tubular structures composed of nonstructural protein NS1 of African horsesickness virus expressed in insect cells. *J. Gen. Virol.* **78**, 1077-1082.

48. Martin S.A. and Moss B. (1975). Modification of RNA by mRNA guanylyltransferase and mRNA (guanine-7-)methyltransferase from vaccinia virions. *J. Biol. Chem.* **250,** 9330-9335.

49. Martinez-Costas J., Sutton G.C, Ramade, N., and Roy P. (1998). Guanylyltransferase and RNA 5'-triphosphatase activities of the purified expressed VP4 protein of bluetongue virus. *J. Mol. Biol.* **280,** 859-866.

50. Martinez-Costas J., Varela R., and Benavente J. (1995). Endogenous enzymatic activities of the avian reovirus s1133: Identification of the viral capping enzyme. *Virology.* **206,** 1017-1026.

51. Martinez-Torrecuadrada J.L., Langeveld J.P., Meloen R.H., Casal J.I. (2001). Definition of neutralizing sites on African horse sickness virus serotype 4 VP2 at the level of peptides. *J. Gen. Virol.* **82,** 2415-2424.

52. McCue L.A., Anders D.G. (1998) Soluble expression and complex formation of proteins required for HCMV DNA replication using the SFV expression system. *Prot. Expr. Purif.* **13,** 301-312.

53. Mertens P.P.C., Burroughs J.N., Wade-Evans A.M., Le Blois H., Oldfield S., Basak A., Loudon P., and Roy P. (1992). Analysis of guanylyltransferase and transmethylase activities associated with bluetongue virus cores and recombinant baculovirus-expressed core-like particles. In Walton T.E., and Osburn B.I. (eds). *Bluetongue, African horsesickness and related orbiviruses: Proceedings of the second international symposium. CRC Press, Boca Raton, FL, p404-415.*

54. Miller L.K. (1988). Baculoviruses for foreign gene expression in insect cells. *Biotechnology.* **42**, 177-199.

55. Mizukoshi N., Sakamoto K. (1993). The complete nucleotide sequence of African horsesickness virus serotype 4 (vaccine strain) segment 4, which encodes the minor core protein VP4. *Virus. Res.* **28,** 299-306

56. Mizumoto K. and Kaziro Y. (1987). Messenger RNA capping enzymes from eukaryotic cells. *Prog Nucleic Acid Res Mol Biol.* **34,** 1-28.

57. Monastyrskaya K., Booth T.F., and Roy P. (1994). Mutation of either two cysteine residues, or deletion of the amino or carboxyl termini of a non-structural protein NS1 of bluetongue virus abrogates virus-specified tubule formation in insect cells. *Virology.* **68,** 2169-2178.

58. O'Shea E.K., Klemm J.K., and Alber T. (1991). X-ray structure of the GCN4 leucine zipper, a two stranded, parallel coiled coil. *Science.* **254,** 539-544.

59. Patton J.T., Jones M.T., Kalbach A.N., He Y-W. and Xiabo J. (1997). Rotavirus RNA polymerase requires the core shell protein to synthesize the double-stranded RNA genome. *J. Virol.* **71,** 9618-9626.

60. Prasad B.V., Nothnagel R., Zeng C.Q., Jakana J., Lawton J.A., Chiu W., and Estes M.K. (1996). Visualization of ordered genomic RNA and localization of transcriptional complexes in rotavirus. *Nature.* **382,** 471-473.

61. Ramadevi N., Burroughs N.J., Mertens P.C., Jones I.M., and Roy P. (1998a). Capping and methylation of mRNA by purified recombinant VP4 protein of bluetongue virus. *Proc.Natl.Acad.Sci.USA.* **95,** 13537-13542.

62. Ramadevi N., Rodriguez J., and Roy P. (1998b). A leucine zipper-like domain is essential for dimerization and encapsidation of bluetongue virus nucleocapsid protein VP4. *J. Virol.* **72,** 2983-2990.

63. Ramadevi N. and Roy P. (1998c). BT virus core protein VP4 has nucleoside triphosphate phosphohydrolase activity. *J. Gen. Virol.* **79,** 2475-2480.

64. Reddy R., Singh R., and Shimba R. (1992). Methylated cap structures in eukaryotic RNAs: structure, synthesis and functions. *Pharmacol. Ther.* **54,** 249-267.

65. Reinisch K.M., Nibert M.L., and Harrison S.C. (2000). Structure of the reovirus core at 3.6A resolution. *Nature.* **404,** 960-967.

66. Reovirdae study group. (1998). Reoviridae: Draft Taxonomy report from the Reoviridae Study Group for ICTV 10/11/98.

67. Roy P., Fukusho A., Ritter G.D., and Lyon D. (1988). Evidence for genetic relationship between RNA and DNA viruses from the sequence homology of a putative polymerase gene of bluetongue virus with that of vaccinia virus: conservation of RNA polymerase genes from diverse species. *Nucleic Acids Res.* **16**, 11759-11767.

68. Roy P. (1992). Bluetongue virus proteins. *J. Gen. Virol.* **73,** 3051-3064.

69. Roy P. (1996). Orbivirus structure and assembly. *Virology.* **216,** 1-11.

70. Roy P., Mertens P.C., and Casal I. (1994). African horse sickness virus structure. *Comp. Immun. Microbiol. Infect. Dis.* **17,** 243-273.

71. Rumlova M., Benedikova J., Cubinkova R., Pichova I., Ruml T. (2001) Comparison of classical and affinity purification techniques of Mason-Pfizer monkey virus capsid protein: the alteration of the product by an affinity tag. *Prot. Expr. Purif.* **23,** 75-83.

72. Sambrook J., Fritsch E.F., and Maniatis T. (1989). Molecular cloning, a laboratory manual. Second Edition Cold Spring Harbour Laboratory Press.

73. Shuman S., lin Y, and Schwer B (1994). Covalent catalysis in nucleotidyl transfer reactions: essential motifs in *S. cerevisiae* RNA capping enzyme are conserved in *S. pombe* and viral capping enzymes and among polynucleotidyl ligases. *Proc. Natl. Acad. Sci. USA.* **91,** 12046-12050.

74. Shuman S. and Ru X-m (1995). Mutational analysis of vaccinia DNA ligase defines residues essential for covalent catalysis. *Virology.* **211,** 73-83.

75. Stauber N., Martinez-Costas J., Sutton G.C, Monastyrskaya K., and Roy P. (1997). Bluetongue virus VP6 protein binds ATP and exhibits an RNA-dependent ATPase function and a helicase activity that catalyzes the unwinding of double-stranded RNA substrates. *J. Virol.* **71,** 7220-7226.

76. Staurt D.I., Gouet P., Grimes J.M., Malby R., Diprose J.M., Zientara S., Burroughs J.N., and Mertens P.P.C. (1998). Structural studies of orbivirus particles. *Arch. Virol.* **14**, 235-250.

77. Theron J., Huismans H., Nel L.H. (1996). Site-specific mutations in the NS2 protein of epizootic haemorrhagic disease virus markedly affect the formation of cytoplasmic inclusion bodies. *Arch. Virol.* **141,** 1143-1151.

78. Theron J., Uitenweerde J.M., Huismans H., Nel L.H. (1994). Comparison of the expression and phosphorylation of the non-structural proteinNS2 of three different orbiviruses: evidence for the involvement of anubiquitous cellular kinase. *J. Gen. Virol.* **75,** 3401-3411.

79. Thomas C.P., Booth T.F., and Roy P. (1990). Synthesis of blue-tongue virus-encoded phosphoprotein and formation of inclusion bodies by recombinant baculovirus in insect cells. *J. Gen. Virol.* **71,** 2073-2083.

80. Urakawa T., Ritter G.D., and Roy P. (1989). Expression of largest RNA segment and synthesis of VP1 protein of bluetongue virus in insect cells by recombinant baculovirus: association of VP1 protein with RNA polymerase activity. *Nuc. Acid. Res.* **17,** 7395-7401.

81. van Dijk A.A. and Huismans H. (1980). The *in vitro* activation and further characterization of the bluetongue virus-associated transcriptase. *Virology.* **104,** 347-356.

82. van Niekerk M., Freeman M, Paweska J.T., Howell P.G., Guthrie A.J., Potgieter A.C., van Staden V., Huismans H. (2003). Variation in the NS3 gene and protein in South African isolates of bluetongueand equine encephalosis viruses. *J. Gen. Virol.* **84,** 581-590.

83. van Niekerk M., van Staden V., van Dijk A.A., Huismans H. (2001a). Variation of African horsesickness virus nonstructural protein NS3 in southern Africa. *J. Gen. Virol.* **82,** 149-158.

84. van Niekerk M., Smit C.C., Fick W.C., van Staden V., Huismans H. (2001). Membrane association of African horsesickness virus nonstructural protein NS3 determines its cytotoxicity. *Virology.* **279,** 499-508.

85. van Staden V., Smit C.C., Stoltz M.A., Maree F.F., and Huismans H.(1998). Characterization of two African horse sickness virus nonstructural proteins, NS1 and NS3. *Arch. Virol. Suppl.* **14,** 251-258.

86. Venkatesan S., Gershowitz A., and Moss B. (1980). Modification of the 5' end of mRNA. *J. Biol. Chem.* **255,** 903-908.

87. Whittles K. and Goldie A. Concise Dictionary of Biology. Published 1993. Tiger Books International. ISBN 1-85501-366-5.

88. Wu X., chen S.Y., Iwata H., Compans R.W., Roy P. (1992). Multiple glycoproteins syhtesized by the smallest RNA segment (s10) of bluetongue virus. *J.Virol.* **66**, 7104-7112.

89. Yamakawa M., Kubo M., and Furuuchi S. (19990. Molecular analysis of the genome of Chuzan virus, a member of the Palyam serogroup viruses, and its Phylogenetic relationships to other orbviruses. *J. Gen. Virol.* **80,** 937-941.

90. Yu L. and Shuman S. (1996). Mutational analysis of the RNA triphosphatase component of vaccinia virus mRNA capping enzyme. *J. Virol.* **70,** 6162- 6168.

91. Yu Y., Kukusho A. and Roy P. (1987). Nucleotide sequence of the VP4 core protein gene (M4 RNA) of US bluetongue virus serotype 10. *Nuc. Acid. Res.* **15,** 7206.

92. Yvon S., Rolland D., Charrier J.P., Jolivet M. (1998). An alternative for purification of low soluble recombinant hepatitis C virus core protein: preparative two-dimensional electrophoresis. *Electrophoresis.* **19,** 1300-1305