CHAPTER 2:CLONING AND CHARACTERIZATION OF THE GENOME SEGMENT ENCODING VP6 OF AHSV

2.1 INTRODUCTION

Very little is known about the three minor structural proteins of AHSV. Prior to this study, only partial clones of the AHSV VP6 gene were available which provided some 5' and 3' sequence data. Unexpectedly, it was found to be extremely difficult to obtain full-length VP6 gene clones (Cormack, 1996).

Variation at nucleotide level provides the most direct quantitative measurement of the differences between related viruses. It is, therefore, useful to compare the sequence data of closely related viruses such as BTV (the prototype virus of the *Orbivirus* genus) and AHSV as well as viruses within the same serogroup. A comparison of the gene sequence is valuable for the identification of regions of importance that may have been conserved between different orbiviruses or indicate specific regions of interest that are unique to AHSV VP6. In BTV VP6 notable primary structural features include an unusually high glycine content and the highest number of charged amino acids of all the BTV proteins (Fukusho *et al*., 1989). BTV VP6 has nucleic acid binding characteristics and some motifs have been proposed (Roy *et al*, 1990, Hayama and Li, 1994; Kar and Roy, 2003). In 1997, Stäuber *et al*. demonstrated the ATPase and helicase activities of BTV VP6. They reported little similarity to other helicases but some similarity to common helicase motifs. Kar and Roy (2003), identified three motifs consistent with SF2 helicases namely, the Walker A and B motifs and an RNA binding motif similar to a conserved helicase motif. The genome nucleotide sequences of two other orbiviruses have been published; Chuzan virus which is a member of the Palyam serogroup (Yamakawa *et al*., 1999) and the tick borne St Croix River virus (Attoui *et al*., 2001). The availability of this data will facilitate analysis of the sequence of AHSV VP6.

Numerous conserved motifs relating to nucleic acid binding, NTP binding and hydrolysis and helicase activity have been identified. The identification of such motifs are used to make predictions about the functions of proteins.

Eight protein families recognized by common RNA binding motifs have been identified namely: the **RNP** (or otherwise known as the RNA recognition motif - RRM; the arginine-rich motif **(ARM)**; the **RGG** box; the **KH** motif; the **double-stranded RNA binding motif**; the **Zinc finger-knuckle**; the cold shock domain (**CSD)** and finally, the **GRPs** (Glycine-rich proteins) (Mattaj, 1993; Burd and Dreyfuss; 1994; Graumann and Marahiel, 1996; Sachetto-Martins *et al*., 2000).

Another motif important to helicase type proteins is the NTP binding motif. This motif is common to all the helicase superfamilies defined on the basis of a series of conserved motifs (Kadaré and Haenni, 1997).

Prior to the initiation of this study, no information about AHSV VP6 sequence was available because the gene had never been cloned. The aim of the work described in this chapter is to characterize the primary structure of the genome segment encoding VP6 of AHSV. In order to achieve this, the genome segment encoding VP6 of both serotypes 3 and 6 were cloned and sequenced. The resultant nucleotide and amino acid sequences were analyzed.

2.2 MATERIALS AND METHODS

2.2.1 PREPARATION OF dsRNA FOR cDNA SYNTHESIS

cDNA was produced using the cloning method of Cashdollar (1982) with modifications described by Bremer *et al*. (1990). DsRNA of AHSV-9 was purified as described by Van Staden *et al*. (1995) and further purified on sucrose gradients. The gradients consisted of 5 steps from 5 – 40% sucrose made up in 1 x TE (10mM Tris pH 7.4, 1mM EDTA pH 8.0) in 0.1% DEPC (BDH) water. It was centrifuged for 16hrs at 38 000rpm. Sixteen fractions of 10 drops each were collected from the bottom of the gradient using a gradient tube fractionator (Hoefer Scientific Instruments). 2μl of fractions 3 - 14 were run on a 1% TAE agarose gel in the presence of EtBr (Roche). Fractions containing segments 4 – 8 were pooled and precipitated by the addition of NaAc pH 7 to the final concentration of 0.3M and 2 volumes of 96% ethanol. Precipitated dsRNA was resuspended in DEPC UHQ to the final concentration of 1μg/μl as determined by optical density measurements at 260nm in a Du-64 Beckman spectrophotometer.

For polyadenylation of the dsRNA, a mix consisting of 10μg of purified dsRNA diluted in DEPC UHQ to a final volume of 32μl was prepared. In another tube the polyadenylation mix was assembled with the following final concentrations: 60mM Tris pH 8.0; 10mM $MgCl_2$; 3mM $MnCl_2$; 300mM NaCl; 0.3mM ATP (Gibco BRL). 43μl of the polyadenylation mix was added to a tube containing 10μCi of lyophilized ^{3}H-ATP (Amersham). After the addition of the dsRNA mix, the polyadenylation reaction was incubated at 37°C for 50min. The polyadenylated dsRNA was separated from unincorporated free nucleotides by G-75 Sephadex column chromatography.

2.2.2 SEPHADEX COLUMN CHROMATOGRAPHY

Sephadex G-75 (Sigma) swollen overnight in 0.1% DEPC UHQ was packed into a glass Pasteur pipette with a glass bead blocking the mouth of the pipette and allowed to settle. The column was rinsed with 3ml of 1mM Tris pH 8.0. The polyadenylation reaction was layered onto the column and washed in with 600μl of Tris buffer. 12 fractions of 100μl each were collected in separate Eppendorf tubes. 2μl of each fraction was counted in 2.5ml Ready Value™ liquid scintillation cocktail in a Beckman LS 3801 scintillation counter. The peak fractions were pooled in a corex tube and lyophilized. The lyophilized polyadenylated dsRNA was resuspended in 20μl of DEPC UHQ with a concentration of 0.75μg/μl.

2.2.3 cDNA SYNTHESIS

Three μg of polyadenylated dsRNA was denatured by adding 100mM MMOH (methyl mercuric hydroxide) to a final concentration of 20mM and incubating at RT for 30min. Standard cDNA mix A was prepared to the following final concentrations in 1% DEPC UHQ: 64mM Tris pH 8.3; 13mM $MgCl_2$; 150mM KCl; 0.6mM dNTP mix (Roche) and 4mM β-mercaptoethanol incubated at

RT for 30min. Standard mix B was prepared as follows: 47μl of mix A; 20μCi α^{32}P-dCTP; 120U HPRI (human placental RNase inhibitor) (Amersham); 16U AMV reverse transcriptase (Promega), 1μg Oligo dT (Roche) and the denatured polyadenylated dsRNA. The reaction was incubated at 42°C for 60min. The reaction volume was increased to 100μl and passed through a Sephadex G-75 column to separate the cDNA from unincorporated free nucleotides (as described 2.2.2). The fractions were counted in a Beckman LS3801 scintillation counter and peak fractions were pooled. The pooled fractions were lyophilized and resuspended in 20μl of 10 x alkaline buffer (300mM NaOH, 20mM EDTA) and 5μl tracking dye (40% sucrose in alkaline buffer with Blue/Orange loading dye (Promega).

2.2.4 ALKALINE AGAROSE GEL ELECTROPHORESIS

For the purposes of this investigation, specific cDNA segments namely segments 7 – 9 which run together on an agarose gel were required. cDNA was visualized by electrophoresing the above sample in a vertical agarose gel under denaturing conditions. A 1.5% agarose gel was prepared in 1 x Alkaline buffer (30mM NaOH; 2mM EDTA). The sample was loaded on the gel and electrophoresed for 4hrs at 40V, 105mA in a SE 400 Hoefer Scientific Instruments gel electrophoresis unit. Following “wet” gel autoradiography, the band containing segments 7 – 9 was excised from the gel.

2.2.5 GLASSMILK PURIFICATION

The cDNA segments were purified from the agarose by the Glassmilk (BIO 101) purification procedure. The manufacturers protocol was followed with some modifications. Briefly, the excised agarose band was melted in 3 volumes of 6M NaI at 50°C until completely dissolved. 5μl of glassmilk was added and the reaction was incubated at RT with shaking for 15min. The glassmilk-cDNA pellet was collected by centrifugation in a benchtop microcentrifuge at 15 000rpm for 1min. The pellet was washed 3 times in NEW (NaCl, ethanol and water) wash and eluted twice for 10min at 50°C in UHQ.

2.2.6 PREPARATION AND TRANSFORMATION OF COMPETENT *E. COLI* CELLS

Bacterial cells were made competent by means of the $CaCl_2$ method (Mandel and Higa, 1970; Cohen *et al*., 1972). An overnight culture of *E. coli* cells was prepared by inoculating 3ml of LB-broth (1% w/v bacto-tryptone, 0.5% w/v bacto-yeast extract, 1% NaCl pH 7.4) with a single colony and incubating for 16hrs at 37°C with shaking. 100ml of LB-broth was inoculated with 1ml overnight culture and incubated until log phase (OD_{550} approximately 0.5). 20ml of cells were collected by centrifugation in a Beckman JS-7.5 swing bucket rotor at 4000rpm for 3min at 4°C. Cells were resuspended in 10ml ice cold 50mM $CaCl_2$ solution and incubated on ice for 30min. Cells were collected as before, resuspended in 1ml ice cold 50mM $CaCl_2$ solution and incubated on ice for 1hr before use.

Fifteen μl of annealing mix was added to 200μl of cells in a glass test tube. A test transformation with supercoiled pBR322 was also performed. Transformations were incubated on ice for 30min followed by heat-shock for 90sec at 42°C. After 2min cooling on ice, LB-broth was added to a final volume of 1ml and incubated for 1hr at 37°C with shaking. Transformation mixes were plated out on LB-agar plates containing 12.5μg/ml tetracycline (Roche).

2.2.7 PLASMID ISOLATION

Plasmid DNA was isolated using the alkaline lysis method developed by Birnboim and Doly (1979). Large scale isolations were modified according to the manufacturers protocol for nucleobond plasmid isolation (Macherey-Nagel). Cells grown in overnight culture were collected by centrifugation in a benchtop microcentrifuge (miniprep) or a Beckman J2-21 floor centrifuge (large scale) at 5000 rpm, depending on the volume of cells. Volumes for large scale plasmid isolation are indicated in brackets. The pellet was drained and resuspended in 100μl (4ml) solution I (50mM glucose; 10mM EDTA and 25mM Tris pH 8). Cells were lysed by the addition of 200μl (4ml) alkaline-SDS buffer (0.2N NaOH, 1% SDS) followed by incubation for 5 min on ice. 150μl cold 3M NaAc pH 4.8 (4ml 2.5M KAc pH 5.2) was added and incubated for 5min on ice. Chromosomal DNA and SDS precipitated proteins were removed by centrifugation for 10 min (45 min) at 15 000 – 16 000 rpm. The plasmid DNA in the clear supernatant was precipitated by the addition of 2 volumes of 96% ethanol or 0.7 volumes of isopropanol. In the miniprep, the precipitated DNA pellet collected by centrifugation was washed with 80% ethanol, air-dried and resuspended in 20μl of UHQ. The large scale was completed in one of two ways. For nucleobond purification, the supernatant was passed through a nucleobond AX100 column (as per manufacturers recommendations). The DNA trapped in the column was washed and eluted. The eluted DNA was precipitated by the addition of 0.7 volumes of isopropanol followed by a washing step with 80% ethanol. Alternatively, subsequent to precipitation, the pellet was resuspended in 6ml 1 x TE and the RNA was precipitated by the addition of 5ml 7.5M ammonium acetate with incubation for 1hr on ice. The RNA was removed by centrifugation and the plasmid DNA in the supernatant precipitated by the addition of 2 volumes of 96% ethanol. Contaminating protein was removed by phenol / chloroform extraction. The plasmid solution was made up to a volume of 400μl by the addition of 1 x TE. 200 μl of Tris buffered phenol (pH 7.8) and 200 μl of chloroform. The solution was mixed and incubated at RT for 2min followed by centrifugation in a microcentrifuge for 2min. The aqueous phase was transferred to another tube and treated with 200μl of chloroform to remove residual phenol. The centrifuge procedure was repeated and the DNA in the aqueous phase was precipitated by the addition of 2 volumes of 96% ethanol in the presence of 0.3M NaAc pH 7. Following an 80% ethanol washing step, the purified plasmid DNA was dried and resuspended in 300μl UHQ.

2.2.8 PREPARATION OF RECOMBINANT PLASMID DNA BY CESIUM CHLORIDE GRADIENT PURIFICATION

Closed circular plasmid DNA may be separated from contaminants by buoyant density centrifugation in CsCl-ethidium bromide gradients developed by Clewell and Helinski (1969). The bouyant density of most dsDNA in CsCl is 1.70g/cm^3. RNA has a bouyant density of 1.80g/cm^3 and forms a pellet at the bottom of the CsCl gradient while protein (bouyant density of 1.30g/cm^3 in CsCl) usually floats on top (Sambrook and Russell, 2001). 1.617g/ml solutions of CsCl were made up in 1 x TE buffer in 5ml Beckman ultra-clear tubes. The plasmid DNA in a volume of 1ml and 3mg from a 10mg/ml EtBr solution was added to the tubes. The gradients were centrifuged for 40hr at 38 000rpm in a SW50 rotor in a Beckman ultracentrifuge. Under UV light the bottom band (of two, top band was linear plasmid DNA) representing supercoiled plasmid DNA was removed using a bent needle and syringe. EtBr was removed by several washes with UHQ saturated n-buthanol. The washed supernatant was diluted with 1 x TE (to prevent CsCl crystal precipitation) and plasmid DNA was precipitated with 2 volumes of 96% ethanol followed by an 80% ethanol wash. The dried pellet was resuspended in 300µl of UHQ. Concentration and purity were determined using optical density measurements at 260 and 280nm respectively.

2.2.9 AMPLIFICATION BY PCR

5' and 3' sequence data were available for the genome segment encoding VP6 of AHSV-3 and these were used for the design of primers (Cormack, 1996). *Bam*HI sites were incorporated on the ends of the primers to facilitate cloning into expression vectors. Primers were designed with similar melting temperatures calculated using the following formula: $T_m = [4(G+C) + 2(A+T)] – 5$. Three extra bases were added to the 5' ends of the primers to ensure enzyme recognition of the *Bam*HI sites. The sequence of the forward primer was: **VP6.1** 5' *CAG*GGATCCGTTAAATAAGTTGTCTCATGTC 3' and the reverse primer sequence was: **VP6.2** 5' *CAG*GGATCCGTAAGTTTTAAGTTGCCCTC 3'. The PCR reaction was assembled as follows: 100pmol each of primers VP6.1 and VP6.2, 400ng of cDNA, 2.5U TaKaRa Ex Taq and reaction buffer with 2mM $MgCl_2$ and 2.5mM of each dNTP in sodium salts (TaKaRa) according to manufacturers recommendations. The reaction was performed in a Hybaid thermocycler using the following programme: 1 cycle of 95°C for 5min; 51°C for 45 sec; 72°C for 3min followed by 28 cycles of 95°C for 1min; 51°C for 45 sec; 72°C for 3min and completed with 1 cycle of 95°C for 1min; 51°C for 45 sec; 72°C for 5min. Reactions were terminated by incubating on ice. PCR products were separated on a 1% 1 x TAE agarose gel. Using a control of a PCR reaction with constructed VP6 gene as template and molecular weight markers ∅X [PhiX174 / HaeIII] (Promega) and DNA molecular weight marker II (Roche), the amplified VP6 gene was excised from the gel and purified by the glassmilk procedure (2.2.5).

2.2.10 CLONING OF THE PCR PRODUCT

The genome segment encoding VP6 of AHSV-3 and AHSV-6 were cloned into the pBS bluescribe RNA transcription vector (Stratagene). pBS is a 3.2kb colony producing phagemid vector derived from phage f1 and pBS of Stratagene. It has a high copy number with T3 and T7 bacteriophage promoters. Selection of recombinants is facilitated by blue/white selection as a result of α-complementation (Ullmann *et al*., 1967) and ampicillin resistance (Sambrook and Russell, 2001). The PCR products and vector were digested with restriction endonuclease

*Bam*HI. The linearized vector was dephosphorylated using 1U of alkaline phosphatase (Roche) according to manufacturers protocols. Digestion products / dephosphorylation products were purified from contaminating enzymes by the glassmilk procedure. The genome segment encoding VP6 genes were ligated into the prepared pBS bluescribe vector by ligation with T4 DNA ligase (Roche). Bacteriophage T4 DNA ligase catalyzes the formation of phosphodiester bonds between 3'-hydroxyl and 5'-phosphate termini in DNA molecules located adjacent to each other (Weiss *et al*., 1968). Vector and insert were ligated at a ratio of 1 vector : 3 insert in 1 x ligase buffer (66mM Tris-HCl; 1mM DTT; 5mM $MgCl_2$ and 1mM ATP pH 7.5) with 1U of T4 DNA ligase. After 16hrs incubation at 16°C, half of the ligation mixtures were used to transform competent XL1blue cells (Stratagene) made competent using the $CaCl_2$ method described above (2.2.7). Transformation mixes were plated onto plates containing 12.5μg/ml tetracycline (XL1blues have tetracycline resistance) and 100μg/ml ampicillin containing LB-agar plates. Blue/white selection was performed by the addition of the chromogenic substrate X-gal (50μl of a 2% solution per plate) and the lactose analogue IPTG (10μl of a 2% solution per plate) which inactivates the *LacZ* repressor (Sambrook and Russell, 2001). White colonies were selected and incubated as overnight colonies in LB-broth with antibiotics. Putative recombinants were screened after miniprep plasmid DNA extraction on a 1% 1 x TAE agarose gel with pBS as a control. Plasmids with a larger size than the control were digested with *Bam*HI to screen for an insert. Recombinants which yielded the correct insert were digested with *Eco*RI to determine the orientation of the insert. All restriction endonuclease digestion products were analyzed by 1% 1 x TAE agarose gel electrophoresis. Recombinant plasmids containing the genome segment encoding VP6 of AHSV-3 and 6 were selected.

2.2.11 SUBCLONING OF THE GENOME SEGMENT ENCODING VP6

Based on sequence data from AHSV-3 VP6 gene available from Cormack (personal communication), numerous restriction endonuclease digestions were performed. Three sites within the genome segment encoding VP6 namely *Bgl*II, *Pst*I and *Eco*RI were selected for subcloning purposes. The AHSV-3 VP6 gene was subcloned into M13 and single strand sequenced. In order to generate sequence data for characterization of the genome segment encoding VP6 of AHSV including assessment of intra- and interserogroup conservation, the AHSV-6 VP6 gene was subcloned into pBS bluescribe and double strand sequenced. Several internal primers as described by Cormack (1996) were used to overcome difficulties arising from secondary structures.

2.2.11.1 Subcloning into M13

The M13mp range of vectors allows for dsDNA manipulation in the replicative form (RF) and the subsequent generation of ssDNA ideal for manual sequencing. The methods for cloning and manipulation of the M13 vectors were obtained from Sambrook *et al*. (1989). Infection of bacterial cells by M13 bacteriophage requires an intact F pilus. JM109 *E. coli* cells carry an F' episome and will, therefore, support the growth of M13. Genes for enzymes involved in proline biosynthesis have been deleted from the JM109 genome but are supplied by the F' plasmid. In order to maintain the F' plasmid, propagation occurs on medium lacking proline, in this instance M9 minimal medium (1 x M9 salts; 1mM $MgSO_4$; 1mM thiamine HCl; 0.1mM $CaCl_2$ and 0.2% glucose) so that only bacterial cells with the F' plasmid will be proline phototrophs (Sambrook and Russell, 2001).

Large amounts of RF M13mp19 were prepared as follows. An overnight culture of JM109s was

diluted $^{1}/_{40}$ and inoculated with a M13mp19 bacteriphage plaque. After 5hrs incubation at 37°C with shaking, 1ml of the culture was centrifuged for 5min in a benchtop microcentrifuge. 500µl of the supernatant was added to 2.5ml JM109 overnight culture, mixed and incubated at room temperature for 5min. Following dilution to a final volume of 250ml in LB-broth, the cells were incubated for 5hrs with shaking at 37°C. Cells were collected by centrifuging at 4000g for 15min at 4°C. The supernatant could be stored for single-stranded DNA extraction. The pellet was washed in 100ml ice-cold STE buffer (0.1m NaCl; 10mM Tris-HCl pH 8 and 1mM EDTA). Closed circle bacteriophage RF DNA was isolated by the alkaline lysis method. Subcloning proceeded using standard restriction endonuclease and ligation methods.

Competent JM109 cells were prepared as previously described (2.2.7) except that cells were grown to OD_{550} = 0.3 before treatment with $CaCl_2$ commenced. 200µl of competent JM109s were incubated with half the ligation mixes of each subclone of the genome segment encoding VP6 of AHSV-3 and incubated on ice for 40min. The transformation mixes were heat shocked for 75sec at 42°C and cooled on ice. 40µl of a 2% solution of both X-Gal and IPTG were added to 200µl log phase JM109 cells (OD_{550} = 0.3) in a test tube. The competent cells and ligation mixture were added to the test tube together with 3ml of soft agar. The soft agar mix was poured on top of agar plates and incubated at 37°C overnight.

White recombinant plaques were picked using a sterile tip. The plaque was rinsed off into 5ml LB-broth containing 50µl of JM109 overnight culture. This was followed by incubation for 5hrs at 37°C with shaking. 1ml of culture was centrifuged for 5min at RT. SDS was added to a sample of the supernantant to a final concentration of 0.1% and incubated at 65°C for 5min. Recombinant plaques were analyzed by 0.7% 0.5 x TBE agarose gel electrophoresis.

Single-stranded DNA for sequencing was prepared as follows: 1ml of recombinant culture was centrifuged for 5min in a microcentrifuge. DNA was precipitated by treating the supernatant with 200µl of a 20% PEG 6000 and 2.5M NaCl solution. After collecting the pellet by centrifugation, it was resuspended in 1 x TE. Protein contamination was removed by phenol / chloroform extraction as described above followed by ethanol precipitation. After a 70% ethanol wash, the pellet was resuspended in 30µl of UHQ.

2.2.11.2 Subcloning into pBS

The AHSV-6 VP6 gene was subcloned into pBS for double-stranded DNA sequencing. The same restriction enzyme sites were used namely *Bam*HI on the termini, *Bgl*II, *Pst*I and *Eco*RI to generate fragments for subcloning. Recombinant subclones were isolated using the alkaline lysis methods described above (2.2.9) and screened with restriction endonucleases to evaluate the nature of the inserts. Both the coding and non-coding strands were sequenced using VP6.1, VP6.2, M13 forward and reverse primers. The same internal primers as used above were used to sequence through regions with secondary structures. Primers were synthesized by Invitrogen.

2.2.12 MANUAL SEQUENCING

The method of manual sequencing employed was based on the dideoxy chain termination method of DNA sequencing (Sanger *et al*., 1977). DsDNA was denatured to provide a single-stranded template. Sequencing was performed using the Sequenase version 2 kit according to manufacturers protocols (USB). ^{35}S dATP labelled samples were electrophoresed through a 6% denaturing 1 x TBE acrylamide gel. The gel was fixed in a 10% acetic acid, 10% methanol solution and dried on a slab vacuum drier (Hoefer Scientific Instruments) followed by

autoradiography.

2.2.13 SEQUENCE ANALYSIS

Nucleotide sequences were translated using the Microgenie program (Queen and Korn, 1984). Nucleotide alignments were performed using CLUSTAL X 1.81 (Higgins and Sharp, 1988; Higgins and Sharp, 1989; Thompson *et al*., 1997). Amino acid sequences were aligned using CLUSTAL W (Thompson *et al*., 1994) available at the European Bioinformatics Institute. All available different full length orbivirus sequences for the genome segment encoding VP6 / VP6 protein (with the exception of Reovirus λ1 (Dearing strain), due the disparity in length of this putative helicase it was considered inappropriate for inclusion for the purpose of this study) were used in this study and are listed in table 2.1. As two serotypes of the AHSV VP6 gene were used, only two serotypes of BTV were used.

Table 2.1 Full-length sequences for orbivirus VP6 genes / VP6 proteins used in this study.

Virus, Serotype	Protein	GenBank accession number	Reference
AHSV-3	VP6	U19881	Turnbull *et al*., 1996
AHSV-6	VP6	U33000	Turnbull *et al*., 1996
BTV-17	VP6	U55798	de Mattos *et al*., 1996
BTV-10	VP6	D00509	Fukusho *et al*., 1989
Palyam serogroup, Chuzan virus	VP6	AB018088	Yamakawa *et al*., 1999.
St Croix River virus	VP6	AF145406	Attoui *et al*., 2001

2.2.14 PHYLOGENETIC ANALYSIS

Gene trees for the genome segment encoding VP6 / VP6 protein were constructed using PAUP version 4.0b10 (Swofford, 1999). Phylogenetic trees were drawn using an exhaustive maximum parsimony search criterion with consensus topology determined by the 50% majority rule. Statistical support for the trees was calculated using 1000 bootstrap replicates (Felsenstein, 1985).

The VP6 proteins of the viruses used in this study vary in length from 232 to 369 amino acid residues and as a result there are regions of poor alignment. In order to assess the effect of differential indel (insertion – deletion) treatment, all possible options for treatment of indels were considered. The indels were either treated as missing data, in other words ignored, where the tree is based only on fully aligned regions. Alternatively, each site at which an indel occurred was treated as a character with equal weight to any other residue (twenty first amino acid or fifth nucleotide). Finally, an indel (irrespective of size) was treated as a single event where the presence of an gap was treated as a 1 and the absence treated as a 0.

2.2.15 PRIMARY STRUCTURE ANALYSIS

Motif analysis was performed to identify regions that are conserved between different *Orbivirus* VP6 proteins. Analyses were performed using MEME [multiple EM for motif elicitation, motif discovery tool] (Bailey and Elkan, 1994) and MAST [motif alignment and search tool] (Bailey and Gribskov, 1998). Motifs were generated using the eMOTIF maker software (Nevill-Manning *et al*., 1998; Huang and Brutlag, 2001). eMOTIF search software and BLAST (Altschul *et al*., 1990; Altschul *et al*., 1997) were used to determine the significance of the identified motifs. eMOTIF software is available at Stanford University Biochemistry Department, MEME and MAST at the San Diego Supercomputer Center (SDSC) and BLAST at the National Centre for Biotechnology Information (NCBI).

Post-translational modification signals were identified using PROSITE (Bucher and Bairoch, 1994; Hoffmann *et al*., 1999) and protein molecular weight [MW] and isoelectric points [pI] were estimated using Compute pI/Mw (Bjellqvist *et al*., 1993; Bjellqvist *et al*., 1994; Wilkins *et al*., 1998) both available through EXPASY (expert protein analysis system) hosted by the Swiss Institute of Bioinformatics.

2.2.16 HYDROPHILICITY AND SECONDARY STRUCTURE

Hydrophilicity plots were determined according to the method of Kyte and Doolittle (1982) on ANTHEPROT (Deléage *et al*., 1988). The secondary structure of VP6 of AHSV-6 was determined using software available on the 3D-pssm server at the Biomolecular Modelling Laboratory hosted by the Imperial Cancer Research Fund, UK (Fischer *et al*., 1999; Kelley *et al*., 1999; Kelley *et al*., 2000). The 3D-pssm server allows the prediction of secondary structure elements and solvent accessibility of amino acid residues of a query protein. Predictions are made after comparison with master proteins in a fold library. In the library, a three state secondary structure assignment (coil; α-helix and β-sheet / strand) is made per residue using STRIDE (Frischman and Argos, 1995). The secondary structure for the query protein is predicted by PSI-Pred (Jones, 1999). Solvation potential i.e. solvent accessibility is determined by DSSP (Kabsch and Sander, 1983). Solvation potential is defined as the ratio between solvent accessible surface area and overall surface area. Values range from 0% (buried) to 100% (exposed).

2.3 RESULTS

The cloning of full-length VP6 genes was a first priority. Although the traditional procedure for cloning dsRNA genes as described by Cashdollar (1982) with modifications described by Bremer *et al*. (1990) has given satisfactory results for most of the dsRNA genes and in particular the smaller ones, the full-length cloning of the genome segment encoding VP6 remained elusive. In a typical cloning experiment, a large number of full-length copies of the VP7 and NS2 genes would be obtained, but no VP6 genes. This anomaly was investigated once more in an attempt to clone the genome segment encoding VP6 of AHSV by traditional methods.

2.3.1 CLONING OF AHSV SEROTYPE 9

In order to remove all contaminating small RNAs and to enrich the dsRNA in the segment that encodes VP6, the dsRNA isolated from CER (chicken embryo reticulocyte) cells infected with AHSV-9 was purified by two successive fractionations on sucrose gradients. The fractions were analyzed on a 1% TAE agarose gel. Two μl samples of each fraction were electrophoresed and fractions 2 - 5 containing dsRNA were pooled (figure 2.1a and b).

The pooled dsRNA was polyadenylated in the presence of ^{3}H [ATP] (adenine 5' triphosphate) and purified by Sephadex column chromatography. Two microlitre aliquots were counted in scintillation fluid and the peak fractions were pooled. Lyophilized dsRNA was used to synthesize cDNA in the presence of $\alpha^{32}P$ dCTP. The cDNA was purified from unincorporated free nucleotides by Sephadex column chromatography. Samples of each fraction were counted in a liquid scintillation counter. Peak fractions were pooled (figure 2.1c) and lyophilized.

The cDNA was analyzed by vertical agarose gel electrophoresis under denaturing conditions followed by autoradiography (figure 2.2). For the purposes of this study, the cDNA band representing segments 7 – 9 was excised from the gel and purified by the GLASSMILK procedure.

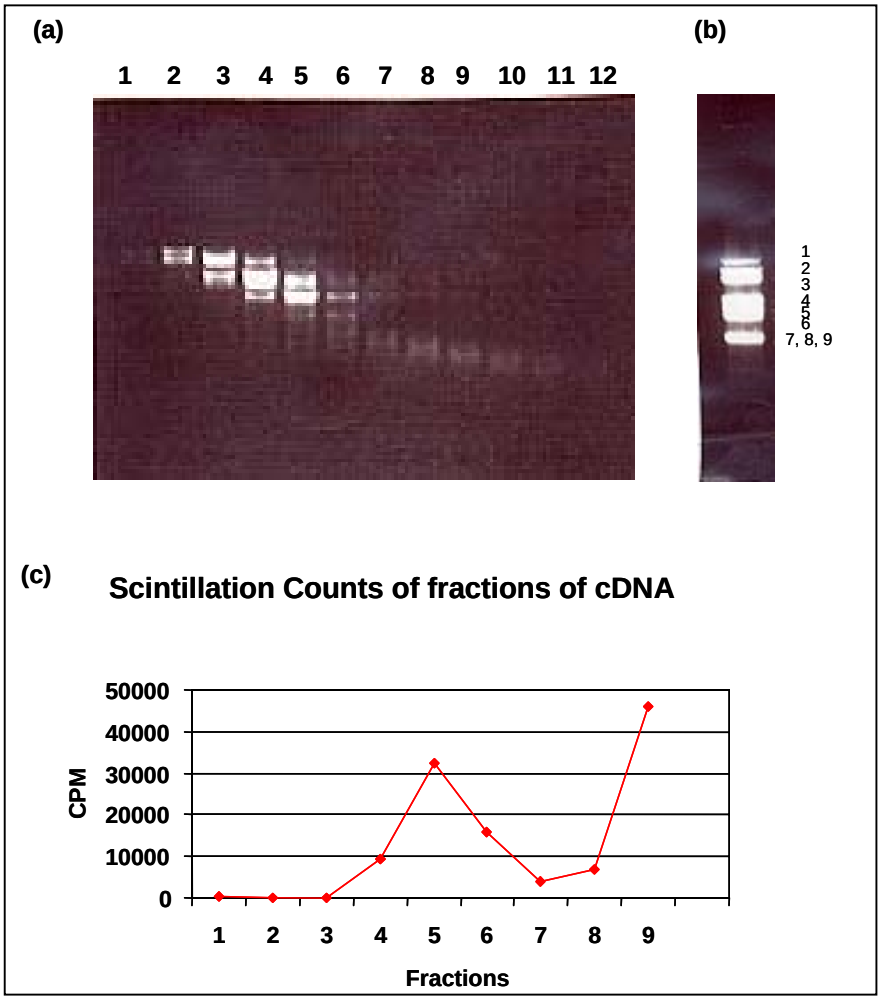

Figure 2.1 AHSV-9 dsRNA analyzed on a 1% TAE agarose gel.

AHSV-9 dsRNA was separated by sucrose gradient fractionation on a 5 – 40% sucrose gradient in a SW50 rotor for 16hrs at 38 000rpm.

(a) Lanes 1 – 12 represent samples of sucrose gradient fractionated dsRNA fractions from the bottom of the gradient (lane 1) to the top of the gradient (lane 12).

(b) Analysis of a pool of fractions (fractions 2 – 5 from 2.1a) which were used for polyadenylation.

(c) Graph depicting the scintillation counts of fractions of cDNA of AHSV-9 labelled with α^{32}P purified by G-75 Sephadex chromatography.

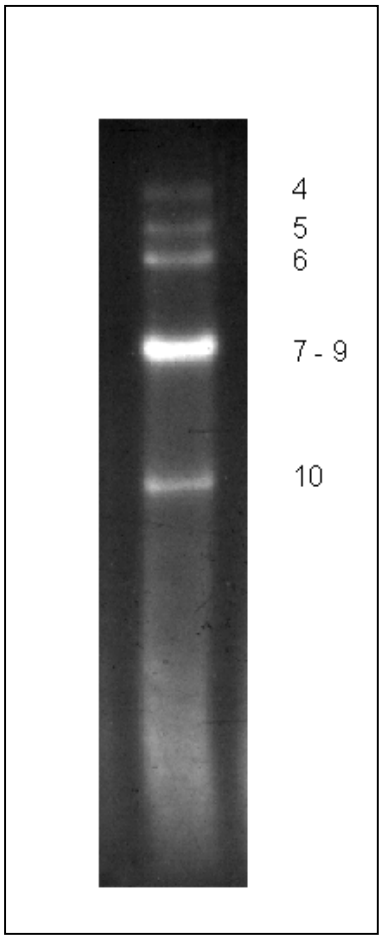

Figure 2.2 Autoradiograph of cDNA synthesized from AHSV-9 polyadenylated dsRNA.

The cDNA was analyzed by 2% alkaline agarose gel electrophoresis. Positions of the genome segments are indicated by the numbers on the right hand side.

2.3.2 AMPLIFICATION OF THE GENOME SEGMENT ENCODING VP6 OF AHSV BY POLYMERASE CHAIN REACTION FROM POOLS OF cDNA

cDNA, prepared from polyadenylated dsRNA of AHSV serotypes 3 and 6 using the method of Cashdollar (1982) with modifications described by Bremer *et al*. (1990), was used as template for a polymerase chain reaction. Primers based on sequence data provided by Cormack (1996) incorporating terminal *Bam*HI sites were used in a PCR reaction (refer to section 2.2.13). The genome segments encoding VP6 of serotypes 3 and 6 were successfully amplified as evidenced by the single amplicons corresponding in size to the 1.3kb DNA marker fragment (figure 2.3). Primer dimmers corresponded in size to the 0.31kb DNA marker fragment. The PCR products were purified from the agarose gel, digested with *Bam*HI and cloned into pBS.

Plasmid DNA from white, putative recombinants was screened against wildtype pBS. Plasmids with a larger size were digested with *Bam*HI (screening for insert) and *Eco*RI (screening for orientation) (figure 2.4). The genome segment encoding VP6 of AHSV-3 in T3 orientation (*Eco*RI digestion yielded a fragment corresponding to the expected size of 291bp) and AHSV-6 in T7 orientation (*Eco*RI digestion yielded a fragment corresponding to the expected size of 878bp) were selected for further use (figure 2.4). Large amounts of plasmid DNA were isolated using a standard large scale method of the alkaline lysis plasmid isolation procedure. Plasmid DNA was further purified by CsCl gradient centrifugation.

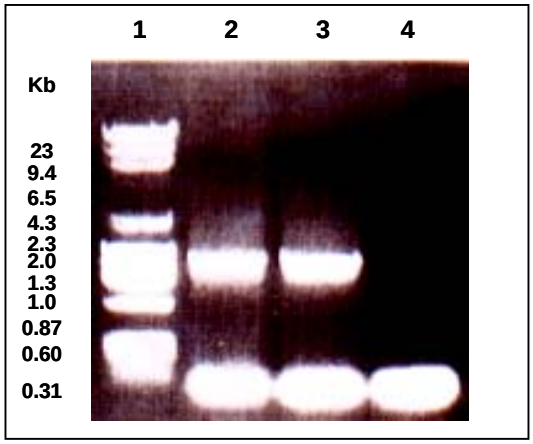

Figure 2.3 PCR amplified VP6 genes from AHSV-3 and AHSV-6 cDNA.

PCR products were separated by 1% TAE agarose gel electrophoresis.

Lanes 1: molecular weight marker.

Lane 2 and 3: PCR products using AHSV-3 and AHSV-6 cDNA as template respectively.

Lane 4: a negative control (no template).

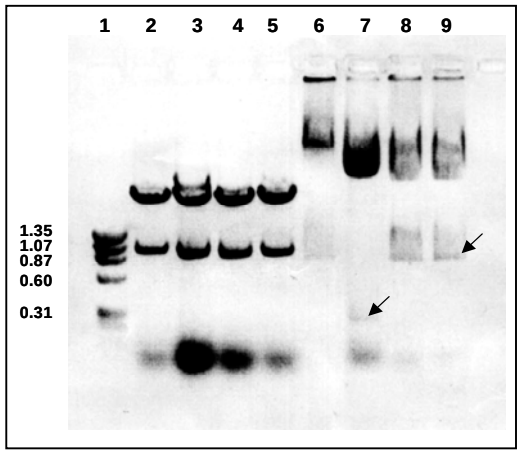

Figure 2.4 Restriction endonuclease selection for recombinant pBS clones.

Restriction endonuclease generated fragments were separated by 1% TAE agarose gel electrophoresis.

Lane 1: molecular weight marker ϕX,

lanes 2 - 3: *Bam*HI digestions of the genome segment encoding VP6 of AHSV-3 in pBS clones.

Lanes 4 - 5: *Bam*HI digestions of the genome segment encoding VP6 of AHSV-6 in pBS.

Lanes 6 - 9: orientation determination of the same clones by *Eco*RI digestions. Arrows show fragments indicating the genome segment encoding VP6 in the T3 orientation (291bp) and T7 orientation (878bp).

2.3.3 SUBCLONING AND SEQUENCING

Sequencing the genome segment encoding VP6 was an important goal as at that time no full-length sequence of the genome segment encoding VP6 of AHSV had been determined. Preliminary RE analysis using *Pst* I, *Bam*HI, *Bgl*II and *Eco*RI indicated that the genome segments encoding VP6 of serotypes 3 and 6 did not contain internal *Bam*HI sites and that this restriction endonuclease could be used to excise the entire gene (result not shown). Following double digestions, a basic restriction endonuclease map was compiled to assist in the development of a subcloning strategy for manual sequencing (figure 2.7). On the basis of these results, the genome segment encoding VP6 of AHSV-3 was subcloned into M13 for single-stranded sequencing and the genome segment encoding VP6 of AHSV-6 was subcloned into pBS bluescribe for double-stranded sequencing. All the subcloned fragments were sequenced using M13 forward and reverse primers. The sequencing results were published and submitted to GenBank (accession numbers: AHSV-3 U19881 and AHSV-6 U33000) (figure 2.5).

Predicted amino acid sequences were generated by Microgenie (Queen and Korn, 1984). The genome segment encoding VP6 of both serotypes was found to be 1169 nucleotides in length with an 1107 nucleotide open reading frame encoding a 369 amino acid polypeptide. The start codon is at nucleotide 18 and the 3' noncoding region is 42 nucleotides in length. The genome segments encoding VP6 of AHSV-3 and AHSV-6 are 96% identical. The conserved sequences GTTAAA at the 5' end and ACTTAC at the 3' end were present (Mertens and Sangar, 1985; Roy, 1989) (figure 2.7).

In order to characterize the AHSV VP6 protein, amino acid compositions of four different orbiviruses were compared in order to determine whether there were significant similarities or differences. Results are shown in table 2.2 with the residue differences between AHSV-3 and AHSV-6 indicated in parenthesis.

AHSV, BTV and Chuzan virus exhibited some similarities in amino acid composition. The most abundant amino acid residue in the VP6 proteins of AHSV, BTV and Chuzan virus is glycine, with AHSV VP6 the most glycine rich. In St Croix River virus VP6 protein, the most abundant amino acid residue is serine which is comparable to the number of serine molecules per thousand found in AHSV VP6. BTV and Chuzan virus VP6 proteins have the largest number of charged amino acid residues, both positive and negative, per thousand residues (363 and 342 respectively). This is followed by AHSV VP6 with in the order of 320 charged amino acids and St Croix River virus with 267. In all four proteins, the number of basic residues exceeds the number of acidic residues. The VP6 proteins have a low cysteine content with AHSV containing four cysteine residues, Chuzan virus has one cysteine residue and BTV and St Croix River virus have no cysteine residues. The most notable characteristics of the amino acid composition of the VP6 proteins are the large number of basic amino acids and the high glycine content. The tick borne St Croix River virus VP6 differs the most of the four in this regard.

```
AHSV-6          GTTAAATAAGTTGTCTCATGTCTTCGGCATTACTCCTTGCACCTGGCGATCTAATCGAAA
AHSV-3          GTTAAATAAGTTGTCTCATGTCTTCGGCATTACTCCTTGCACCTGGCGATCTAATCGAAA
                *******************************************************************

AHSV-6          AAGCAAAGCGCGAGCTCGAGCAACGCTCGATAACTCCGCTTTTGCGGGAGAAAGGTTCGA
AHSV-3          AAGCAAAGCGCGAGCTCGAGCAACGCTCGATAGCTCCGCTCTTGCGGGAGAAAAACTCGA
                ******************************** ******* ************   ****

AHSV-6          AAGAAGCCAAATCTAAATTAAAGGAAGACGGGGAGAAGAAGAACAAGAGTGAAAAGGAAG
AHSV-3          CAGAAGCCAAATCTAAACTAAAAGAGGATGGGGAGAAGAAGAACAAGAGTGAAAAGGAAG
                 **************** **** ** ** ********************************

AHSV-6          AGAACAAAATACATGATGATCGAAGAGTGGAGAGCCAGAAATCTGAGGGAGGCGGACCAG
AHSV-3          AGAACAAAATACATGATGATCGAAGAGTGGAGAGCCAGGAATCTGAGGGAAGCGGATCAG
                ************************************ *********** ***** ***

AHSV-6          CCGATTGTCAACGTGGCGCAGGAAGCGCAGGAGCAAATTGTGCAACATCAACAGGAGGAG
AHSV-3          CCGATTGTCAACGCGGCGCAGGAAGCCGAGGAGCAGATTGCGCAACATCAACAGGAGGAG
                ************* ************  ******* **** ********************

AHSV-6          GAGATGGAAGTGCAGGAGCAAGGACCGGGATTGGAGGGGGAGGAGTGGGAGGAGTGGATT
AHSV-3          GAGATGGAGGTGCAGGAGCAAGGACCGGGATTGGAGGGGGAGGAGTGGGAAGAGTGGATT
                ******** ****************************************** *********

AHSV-6          CGAGATCTGGAGGACATGGAGGACAGGGTGCAGCCTCGGATGGAAAGGGAGTGGGTAAAT
AHSV-3          CGAGATCTGGAGGACGTGGAGGACAGGGTGCAGCCTCGGATGGAAAGGGAGTGGGTAAAT
                *************** ************************************************

AHSV-6          CTAAGACCGGAGCAGATCGTGTCGCTAATGATGATGCAACACGCAATGTTGGTTCCAGTG
AHSV-3          CTAAGACCGGAGCAGATCGTGTCGCTAATGATGGTGCAACACGCGATGTTGGTACCAGTG
                ********************************** ********** ******** ******

AHSV-6          AGGTATCATCTGGTGGAATCACTTCAGGAGGTCTTCAAGGCCGAGGAGGACTCGTTGCAA
AHSV-3          AGGTATCATCTAGTGGAATCACTTCAGGAGGTCTTCAAGGCCGAGGAGGACTCGTTGCAA
                *********** ************************************************

AHSV-6          AGAGTGGTGAATGTGGCGGGGAATCATTGGATAGGATAGGCGGCTGCAGCGGAAATTCAA
AHSV-3          AGAGTAGTGAATGTGGCGGGGAACCATTGGATAGGACAGGCGGCTGCAGCGGAAATTCAA
                ***** ***************** ************ ************************

AHSV-6          AAACTGAGGGAGAGGAGGCGAAGGCTGGAGGGGGCGATAGACGGATTGGAGGATTAGCAA
AHSV-3          AAACTGAGGGAGAGGAGGCAAAGGTTGGAGGGGGCGATAGACGGATTGGAGGATTAGCAA
                ******************* **** ***********************************

AHSV-6          CGCAGGAGATTGCAGACTTTGTGAAGAAGAAGGTCGGAGTTGAAGTTCAGGTGTTTTCTA
AHSV-3          CGCAGGGGATTGCCGACTTTGTGAAGAAGAAGATCGGAGTTGAAGTTCAGGTGTTTTCTA
                ****** ****** ****************** ****************************

AHSV-6          AAGGAATGAGCAACTTATTTACTGTAGATAAGTCATTGCTTGAGCGGGGTGGGTTAGGGA
AHSV-3          AAGGAATGAGTAACTTTTTTACTGTAGATAAATCATTGCTTAAGCGGGGTGGGTTAGGGA
                ********** ***** ************** ********* ******************

AHSV-6          GGGAGGACATTCTACATCAATCAGATATTGTAAAAGAGATTAGAGTAAGTGATAAAAAAG
AHSV-3          GGGAAGACATTCTACATCAATCAGATATTGTAAAAGAGATTAGAGCAAGCGATAAAAAAG
                **** ****************************************** *** **********

AHSV-6          TCAAGATTATTCCTCTTTCCACAGTAAAGAGAATGATAGCGGAATTCGGAGGAACAGAGG
AHSV-3          TCAAGATTATTCCTCTTTCTACAGTGAAGAGAATGATAGCGGAATTCGGAGGAACAGAGG
                ******************* ***** **********************************

AHSV-6          AGGATGAAATCAAAGCTGTTCAAACTCAAAGTTCTTCTATCAGGTATATTAGTAATAGAA
AHSV-3          AGGATGAAATCAAAGCTGTTCAAACTCAAAGCTCTTCTATCAGGTATATTAGTAATAGAA
                ******************************** ****************************
```

```
AHSV-6          TGGAAGATGTTTCAAGAGCGAAGGCGATGTTCACAGCGCCGACGGGTGATGAGGGGTGGA
AHSV-3          TGGAAGATGTTTCAAGAGCGAAGGCGATGTTCACAGCGCCGACAGGTGATGAGGGGTGGA
                **************************************** ***************

AHSV-6          AAGAGGTCGCTAAAGCAGCAACTCAGCGTCCTAACATCATGGCGTATGTGCACGAAGGGG
AHSV-3          AAGAGGTTGCTAAAGCAGCAACTCAGCGTCCTAACATCATGGCGTATGTGCACGAAGGGG
                ******* **************************************************

AHSV-6          AAGGTGATGGATTGAAAGAGCTTTTACATCTGATTGATCATATCTAGGTCCAGGGGTAAA
AHSV-3          AAGGCGATGGATTGAAAGAGCTTTTACATCTGATTGATCATATCTAGGTCCAGGGGTAAA
                **** ****************************************************

AHSV-6          CGGCAGCTTGAGGGCAACTTAAAACTTAC
AHSV-3          CGGCAGCTTGAGGGCAACTTAAAACTTAC
                ****************************
```

Figure 2.5 Nucleotide sequences of the segment encoding VP6 of AHSV-3 (U19881) and AHSV-6 (U33000) aligned using CLUSTAL X.

The start and stop codons are indicated in bold and the conserved 5' and 3' noncoding sequences are underlined.

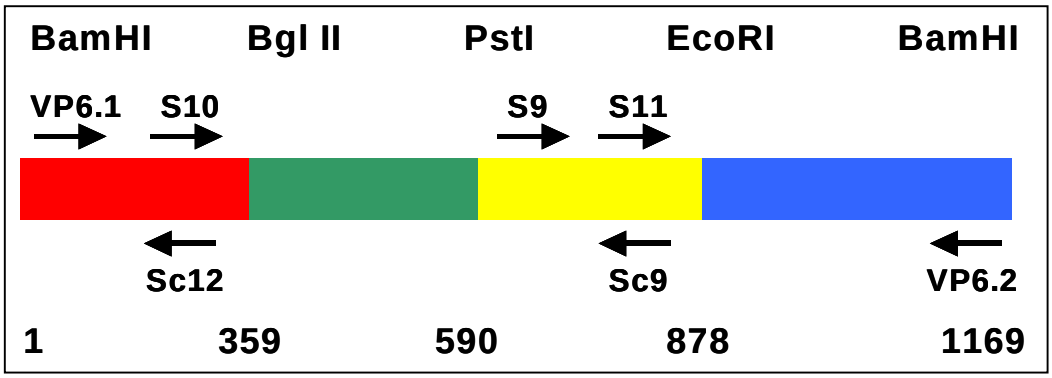

Figure 2.6 Restriction enzyme mapping of the genome segment encoding VP6 of AHSV-3 and -6.

Basic restriction enzyme map of the genome segment encoding VP6 of AHSV-3 and AHSV-6 (including non-coding regions) used for subcloning. Arrows represent the location of internal primers.

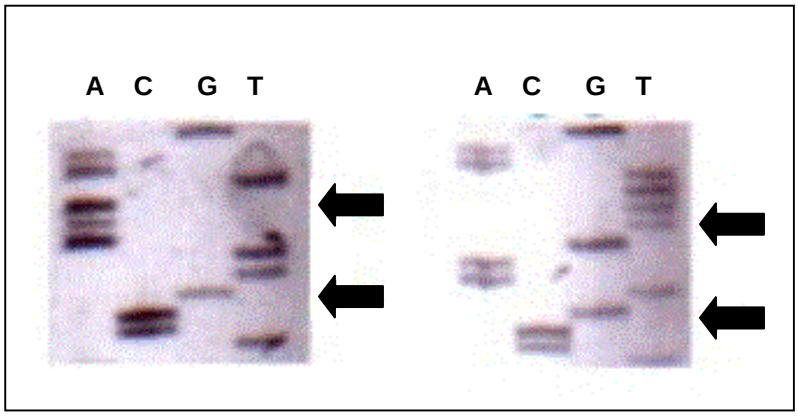

Figure 2.7 Autoradiogram of manual sequencing illustrating conserved 5' and 3' sequences of the genome segment encoding VP6 of AHSV-6.

Table 2.2 Amino acid residue frequencies of AHSV, BTV and Chuzan VP6 expressed per 1000 residues.

Amino acid residue	AHSV-6 VP6 (369)(AHSV-3)		BTV-10 VP6 (328)		CHUZAN VP6 (272)		St Croix VP6 (232)	
	N	/1000	N	/1000	N	/1000	N	/1000
Alanine (A)	30	81	29	88	16	59	19	82
Arginine (R)	22(25)	60 (68)	24	73	14	51	17	73
Asparagine (N)	9 (8)	24 (22)	8	24	8	29	8	34
Aspartic acid (D)	23(24)	62 (65)	19	58	24	88	12	52
Cysteine (C)	4	11	0	0	1	4	0	0
Glutamine (Q)	11	30	11	34	12	44	11	47
Glutamic acid (E)	33(32)	89 (87)	39	119	22	81	17	73
Glycine (G)	62(60)	168 (162)	41	125	30	110	13	56
Histidine (H)	6 (5)	16 (14)	6	18	8	29	5	22
Isoleucine (I)	20	54	19	58	17	63	11	47
Leucine (L)	22(21)	60	16	49	15	55	20	86
Lysine (K)	34(33)	92 (89)	31	95	25	92	11	47
Methionine (M)	6	16	6	18	4	15	6	26
Phenylalanine (F)	5 (6)	14 (16)	2	6	8	29	6	26
Proline (P)	6	16	7	21	3	11	16	69
Serine (S)	35(36)	95 (98)	22	67	21	77	24	103
Threonine (T)	16(18)	43 (49)	19	58	20	74	13	56
Tryptophan (W)	1	3	3	9	1	4	1	4
Tyrosine (Y)	2	5	4	12	5	18	8	35
Valine (V)	22(21)	60	22	67	18	66	14	60
R + K + H (basic)	62(63)	168 (171)	61	186	47	173	33	142
D + E (acidic)	56	151 (152)	58	177	46	169	29	125
F + Y (aromatic)	7 (9)	19 (24)	6	18	13	48	14	60

2.3.4 SEQUENCE ANALYSIS

The most common method of summarizing the relationship between two sequences is by determining their fractional or percentage similarity (Swofford and Olsen, 1990). Both identical and similar (conserved substitutions) were determined from pairwise CLUSTALW alignments (Thompson *et al*., 1994) of the amino acid sequences (table 2.3).

Table 2.3 Amino acid and nucleotide similarity between VP6 of four orbiviruses.

Values are expressed as a percentage of the total number of residues of the longer protein (identical, [similar]). Identity is defined as fully conserved residues. Similarity is fully conserved residues and strong groups which are fully conserved (i.e. conserved substitutions) as determined by CLUSTALW alignment (Thompson *et al*., 1994). Amino acid comparisons are top right of table and nucleotide comparisons are bottom left.

% identical and [similar] amino acid residues (top right of table) and % identical nucleotides (bottom left of table in italics)

	AHSV-6 VP6	AHSV-3 VP6	BTV-10 VP6	BTV-17 VP6	Chuzan VP6	St Croix VP6
AHSV-6 VP6	-	93.5 [96]	27 [46]	26 [45]	24 [42]	12 [29]
AHSV-3 VP6	*96*	-	27 [46]	26.5 [47]	25.5 [40]	12 [28]
BTV-10 VP6	*44*	*47*	-	97 [98]	26 [48]	17 [36]
BTV-17 VP6	*44*	*46*	*98*	-	26 [49]	18 [35]
Chuzan VP6	*37*	*38*	*41*	*41.5*	-	19 [37]
St Croix VP6	*29*	*28.5*	*32*	*32.5*	*37*	-

AHSV shows high interserotype conservation (93.5% identity and 96% similarity on an amino acid level) 96% identity on a nucleotide level. This corresponds to that of BTV amino acid residues showing 97% identity and 98% similarity and 98% nucleotide identity between BTV-10 and BTV-17. AHSV, BTV and Chuzan virus have amino acid identities of between 24 and 27% and similarities ranging from 42 – 49%. There is more variation on nucleotide level between AHSV and Chuzan virus (37 – 38% identity) compared with BTV and Chuzan VP6 (41 – 41.5%). St Croix is the least similar with as little as 12% amino acid identity (29% similarity) to AHSV; 17 - 18% identity (35 - 36% similarity) to BTV VP6 and 19% identity (37% similarity) to Chuzan VP6. Nucleotide identity between St Croix VP6 and the others varies between 29 – 38%. The largest size difference is between St Croix and AHSV VP6 which is reflected in the lowest percentage nucleotide and amino acid identity. Due to the large difference in size of the genes and their encoded proteins, the larger the discrepancy in size, the lower the

percentage identity calculated as number of shared residues divided by the total number of residues of the longer gene or protein.

2.3.5 PHYLOGENETIC ANALYSIS

In 1999, Yamakawa *et al*. determined the nucleotide sequence of the genome of Chuzan virus, a Palyam serogroup virus. They inferred phylogenetic relationships for the different segments of AHSV, Chuzan and BTV using the UPGMA (unweighted pairgroup method using arithmetic averages) method. On the basis of amino acid sequence data analysis, they found that all the Chuzan virus segments were more closely related to AHSV than BTV, with the exception of VP6. In their results, Chuzan virus VP6 grouped with BTV-10 VP6. This result which served as motivation for further homology studies proved problematic for a number of reasons and raised a number of concerns. Firstly, in the UPGMA method, gaps cannot contribute to the phylogenetic resolution of relationships and are excluded from the analysis. Secondly, the UPGMA method assumes a constant rate of evolution (Swofford and Olsen, 1990; Page and Holmes, 1998). This assumption is unlikely to hold true for viruses recovered from different hosts. Thirdly, poor resolution of relationships is likely to occur when gaps are excluded from the analysis where the data set is reduced from 369 amino acids in AHSV VP6 to 232 for St Croix VP6. Conversion of nucleotide data to amino acid sequences results in a three-fold reduction in the size of the data set and is commonly used to determine functional constraints as opposed to evolutionary relatedness. Lastly, there is no obvious reason for the genome segment encoding VP6 of the genome to have a different evolutionary history to the other segments. In the light of these three aspects and in an attempt to confirm or refute the results obtained by Yamakawa *et al*. (1999), both phylogenetic and functional constraint analyses were performed on nucleotide and amino acid sequence data of VP6 respectively.

The evolutionary relationship was determined for four orbiviruses by maximum parsimony methods (PAUP ver 4.0b10) (figure 2.8) using the nucleotide sequences aligned using CLUSTALX (Higgins and Sharp, 1988; Higgins and Sharp, 1989; Thompson *et al*., 1997). Based on the similarities between the *Orbivirus* VP6 genes (table 2.3), St Croix River virus was used as an outgroup. Values for g_1 are used to differentiate between phylogenetic signal and random noise in molecular data sets. If the g_1 values produced for a data set are more negative than the critical values, the data set is considered to be significantly more structured than random data (noise) (Hillis and Huelsenbeck, 1992). A single tree was obtained with a g_1 value of –0.67, $P>0.05$ indicating structured data. BTV and AHSV were found to cluster together. Due to the differences in length of the different VP6 genes, there are regions of poor alignment. A bootstrap value of 100 was obtained when each gap was treated as an individual character; when gaps were treated as individual events and when the gaps were ignored.

The results of Yamakawa *et al*., (1999) suggested that BTV VP6 and Chuzan VP6 were more closely related to one another than to AHSV VP6. In the results obtained in this study one tree was obtained which indicated that AHSV and BTV VP6 genes were more closely related to each other than either were to Chuzan virus VP6.

In order to determine the level of functional relatedness of the VP6 proteins, exhaustive maximum parsimony methods were applied to the amino acid sequences of six orbiviruses (figure 2.9). A single tree was obtained for each analysis, however, the results differed according to how the gaps in the alignments were treated. When the gaps were treated as missing data (i.e. ignored) AHSV VP6 grouped with Chuzan virus VP6 with a bootstrap confidence of 61 (figure 2.9a). Yamakawa *et al*. (1999) obtained this result for all the other virus segments except VP6. When the gaps were treated as single events, the same result was obtained, however, there is no support for the AHSV-Chuzan VP6 branch and as such the tree is considered unresolved (figure 2.9b). Each gap treated as a character generated a tree where AHSV VP6 and BTV VP6 grouped together with a bootstrap value of 74 (figure 2.9c). The four character dataset (a) has a g_1 value of –0.75, P = 0.01 and the binary data set (b) has a g_1 value of –0.79, P = 0.01. These are both relatively negative values indicating significantly more structure to the data that found in random data sets (Hillis and Huelsenbeck, 1992). None of these results correspond to the results obtained by Yamakawa *et al*. (1999).

There are large differences in the number of amino acid residues in the VP6 proteins of different orbiviruses. The number of amino acid residues varies from 369 amino acids in AHSV VP6 to 232 in St Croix River virus VP6 (see figure 2.11 for CLUSTALW amino acid alignment). This suggests that VP6 is not a good choice of gene for phylogenetic analysis and has extremely limited value in functional constraint analyses.

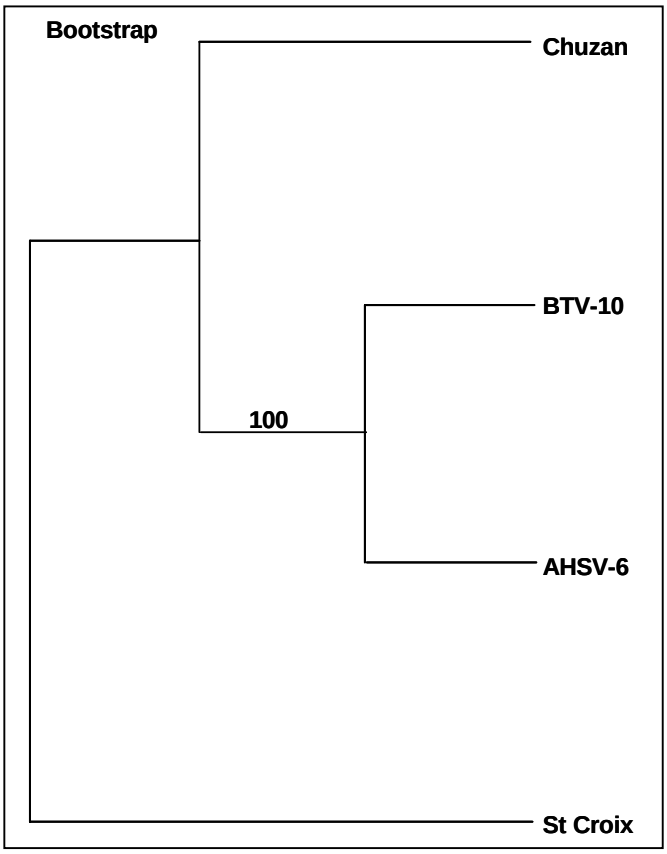

Figure 2.8 Nucleotide phylogenetic analysis of the genome segment encoding VP6 from four *Orbivirus* serogroups.

The phylogenetic tree was reconstructed using exhaustive maximum parsimony methods (PAUP ver 4.0b10).

A bootstrap value of 100 was obtained when each gap was treated as an individual character, when gaps were treated as individual events and when the gaps were treated as missing data (i.e. ignored).

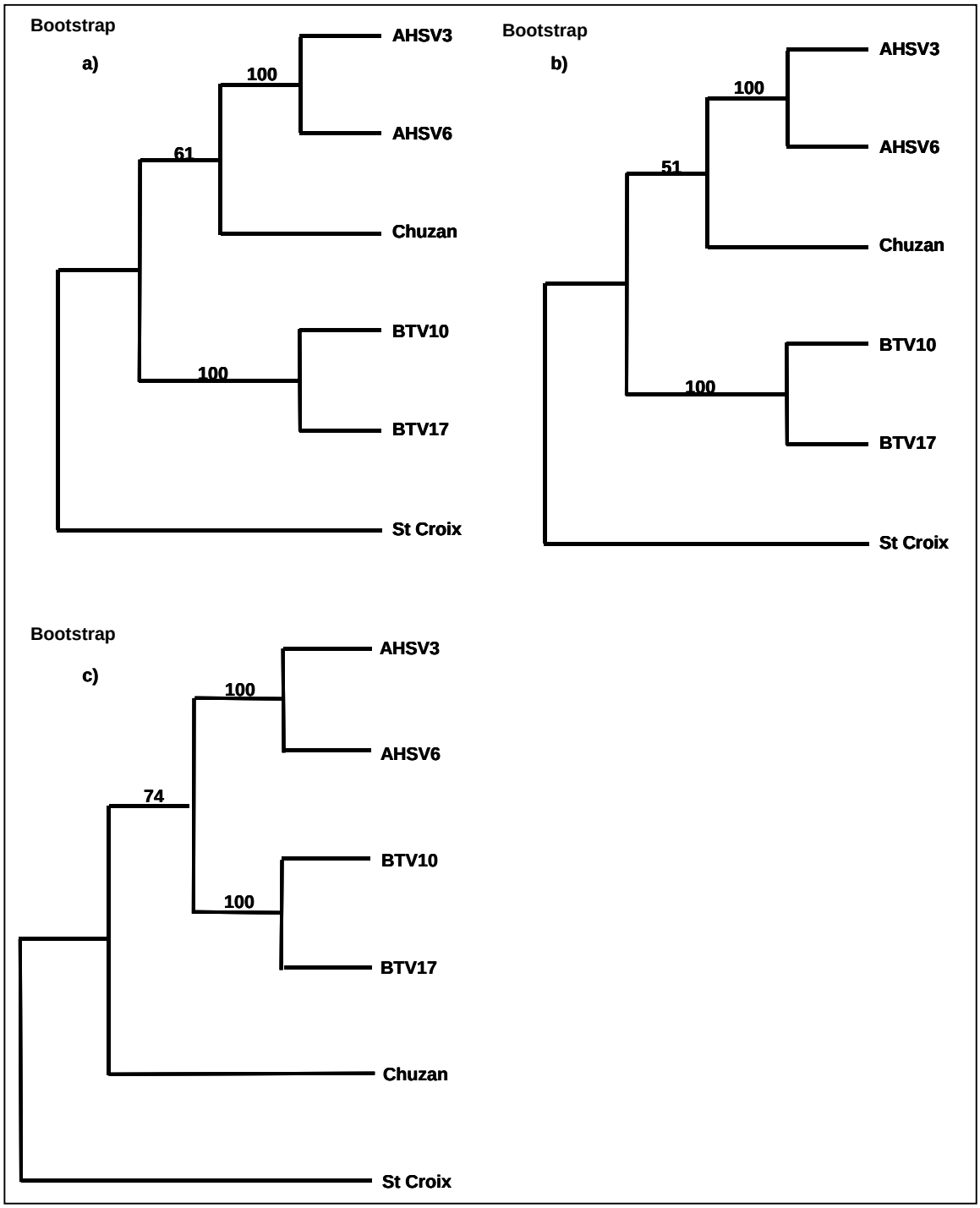

Figure 2.9 Functional constraint analyses of VP6 from four *Orbivirus* serogroups using parsimony methods (PAUP ver 4.0b5).

In a) gaps were treated as missing data (i.e. ignored).

In b) gaps were treated as single events and;

in c) each individual gap was treated as a character.

2.3.6 AMINO ACID SEQUENCE ANALYSIS

In order to identify RNA binding and possible enzymatic activities of AHSV VP6, a search was carried out to identify conserved motifs found in RNA/DNA helicases using a CLUSTALW alignment of VP6 of AHSV-3 and 6 (figure 2.10).

The results of the search are summarized in figures 2.11, 2.12 and table 2.4. In AHSV VP6, a glycine rich region with 36% identity and 45% similarity to glycine rich proteins was identified between amino acids 180 and 192. A region with 91% similarity to the Rep helicase of *E. coli* is also present. Numerous post-translational modification signals were identified. Amongst these are several protein kinase C and Casein kinase II phosphorylation sites and N-myristylation sites (not shown). A N-glycosylation site was also identified located at residues 48 – 51 (figure 2.10a).

VP6 of AHSV shows certain similarities to conserved helicase motifs. Both the A and B sites of the Walker box are present as well as motifs III and IV found in SF2 helicases (figure 2.10b). Molecular weights and isoelectric points were estimated using Calculate pI/Mw (Bjellqvist *et al*., 1993; Bjellqvist *et al*., 1994; Wilkins *et al*., 1998). The predicted molecular weights and isoelectric points of VP6 of AHSV-3 and AHSV-6 were 38 869.28 Da (pI 7.24) and 39 054.46 Da (pI 8.25) respectively.

With the objective of identifying motifs conserved among the VP6 proteins of various orbiviruses, amino acid sequences of AHSV-3 and AHSV-6, BTV-10 and BTV-17, Chuzan and St Croix River virus were aligned using CLUSTALW (Thompson *et al*., 1994). Four motifs were identified using MEME (Bailey and Elkan, 1994) and MAST (Bailey and Gribskov, 1998) (figure 2.11). Motifs were generated using the eMOTIF maker software (Nevill-Manning *et al*., 1998; Huang and Brutlag, 2001). eMOTIF search software and BLAST (Altschul *et al*., 1990; Altschul *et al*., 1997) were used to determine the significance of the identified motifs (table 2.4). Two of the motifs (i and iii) seem to be found in the *Orbivirus* VP6 proteins alone with no other significant similarities detected. Motif ii overlaps with a region with 91% similarity to the Rep helicase of *E. coli*. Similarity to ATP dependent helicases and DNA and RNA directed polymerases were determined. Motif iv showed similarity to ATP binding or dependent transporters and carrier proteins. The three putative RNA binding motifs identified by Hayama and Li (1994) are indicated in figure 2.11. So-called motif 71 overlaps with a motif identified in BTV reported to be important for RNA binding (Kar and Roy, 2003).

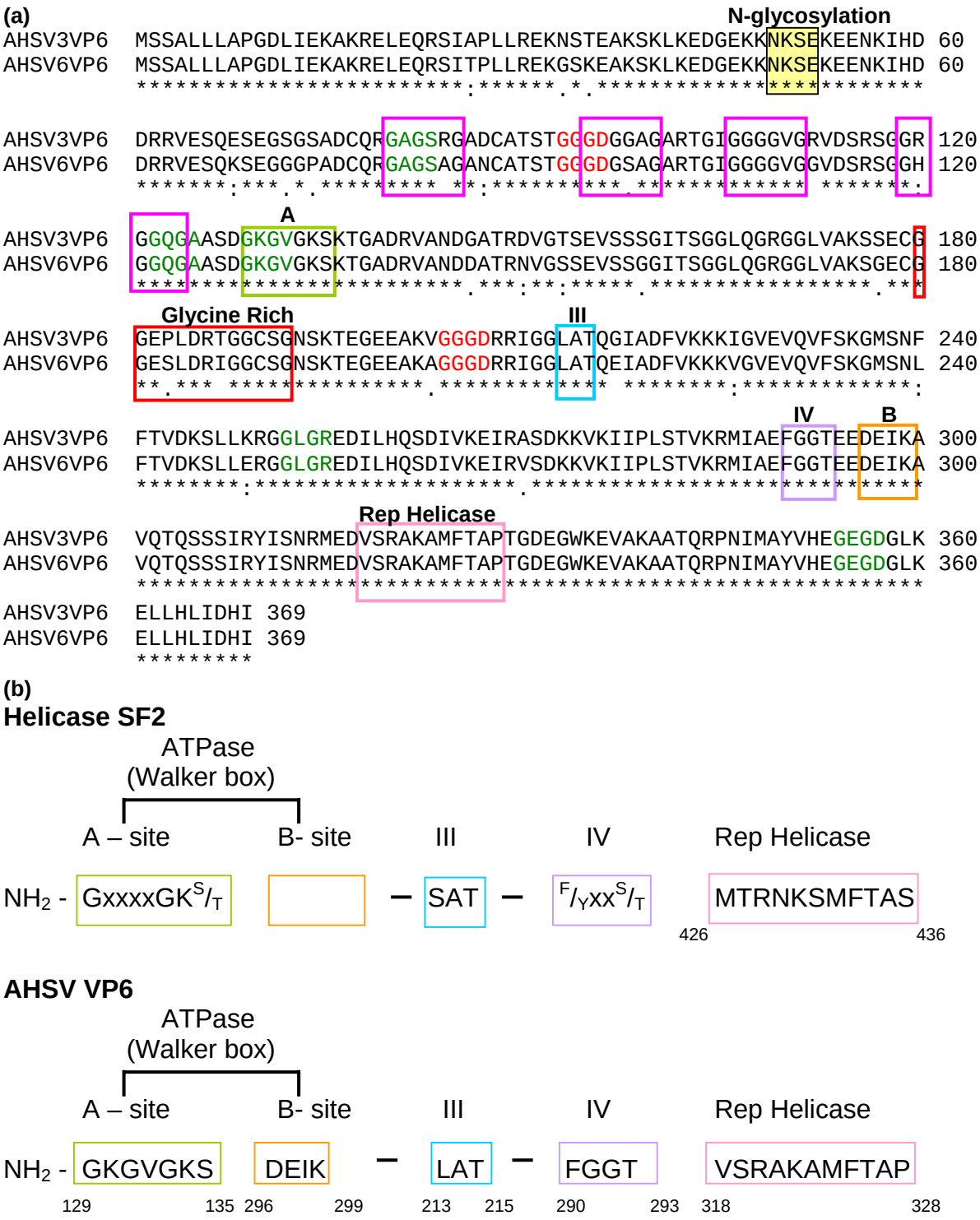

Figure 2.10 (a) CLUSTAL W alignment of VP6 of AHSV-3 and 6.

The yellow box indicates the N-glycosylation site. The red box shows a glycine rich region. (b) Schematic diagram indicating important helicase motifs (Bisaillon *et al*., 1997). The green and orange boxes show the A and B sites of the Walker box respectively. Conserved motifs III and IV found in SF2 helicases are represented by the blue and lavender boxes. The 11 amino acid region with 91% similarity to the Rep helicase of E. coli is indicated by the pink box. Various motifs found in glycine rich proteins are indicated as follows: GxGxxG (pink line box), GGGx (red font) and GxGx (green font).

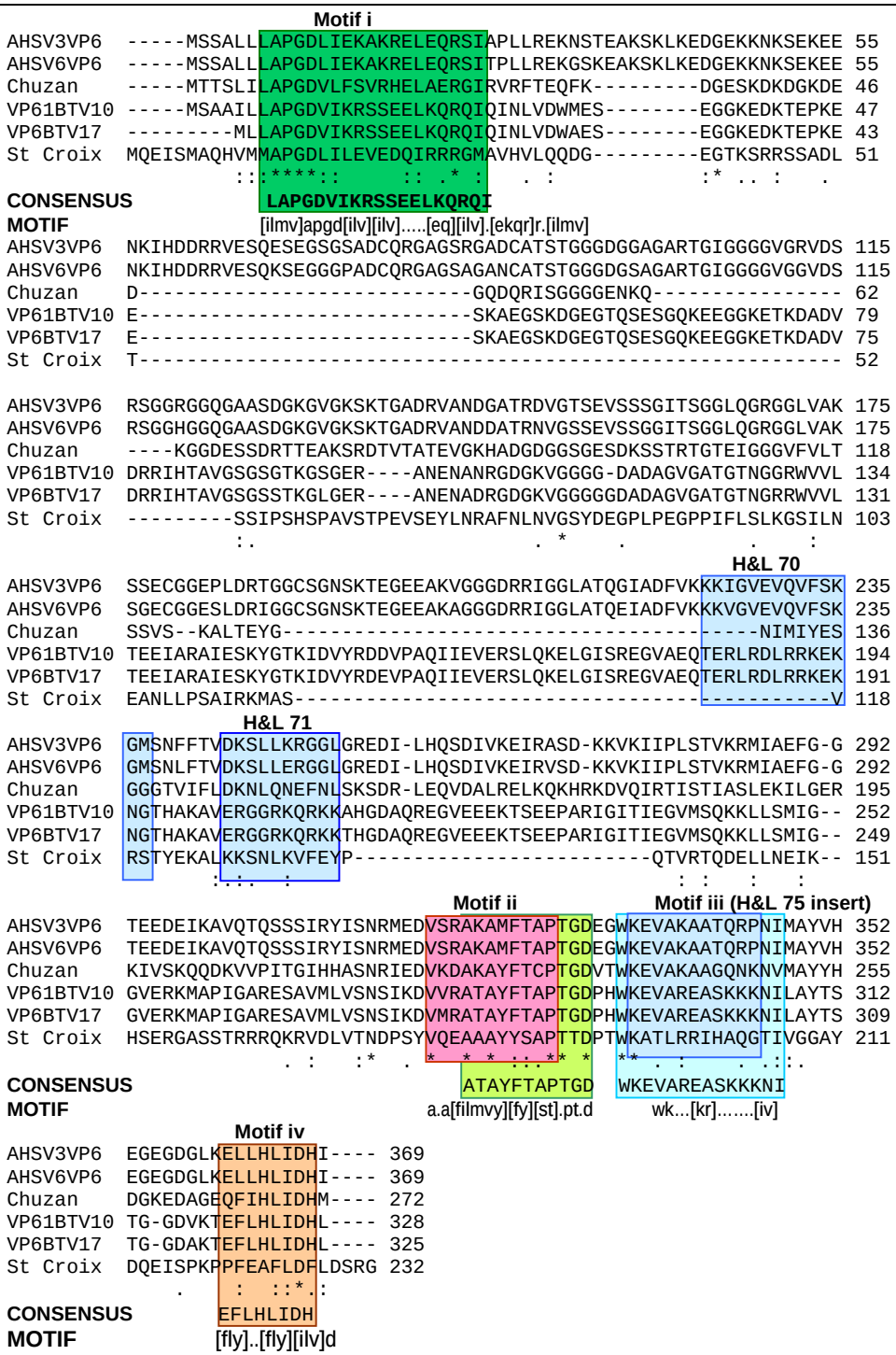

Figure 2.11 CLUSTAL W alignment of VP6 of AHSV-3 and 6; Chuzan virus; BTV-10 and 17 and St Croix River virus.

Four motifs identified using MEME (Bailey and Elkan, 1994) and MAST [motif alignment and search tool] (Bailey and Gribskov, 1998) and eMOTIF maker are indicated by the coloured blocks. Consensus sequences and motifs are indicated below motifs i – iv. The blue blocks are the three amino acid sequences identified by Hayama and Li (1994).

Table 2.4: Motif analysis from VP6 alignment of 6 orbiviruses.

A summary of the motif analysis using *e*MOTIF maker and search software (Nevill-Manning *et al*., 1998; Huang and Brutlag, 2001) and BLAST (Altschul *et al*., 1990; Altschul *et al*., 1997).

MOTIF	SIGNIFICANCE:
Motif i	Found in the *Orbivirus* VP6 proteins analyzed. 88% similarity to a coiled-coil protein. No other significant similarity detected.
Motif ii	Similar motif detected in the following proteins: calcium and potassium transporting ATPases. ATP dependent RNA helicases DDX19 (DEAD box protein), protein kinase P54, ATP dependent DNA helicase PCRA, zinc finger protein REP1, DNA and RNA directed RNA polymerases, Glycoprotein E precursor and RNA helicase MG425/255. The pink block indicates a region of AHSV VP6 with 91% similarity to the Rep helicase of E. coli.
Motif iii	Found in the *Orbivirus* VP6 proteins analyzed. No other significant similarity detected.
Motif iv	Similar motif detected in the following proteins: ATP binding cassette transporter, ATP dependent transporter, ADP/ATP carrier protein 1 precursor (maize), ADP/ATP translocase, potential phospholipid transporting ATPase IA (human) and probable calcium, copper, cation transporting ATPase.
H & L 70 H & L 71 H & L 75	3 motifs identified by Hayama and Li (1994) in BTV VP6, involved in nucleic acid binding activity. H&L 71 putative RNA binding site (Kar and Roy, 2003).

Kar and Roy (2003), proposed that BTV-10 VP6 has motifs consistent with superfamily 2 helicases. These include the Walker A site (common to all NTP hydrolysing proteins), the DEAD/DExH helicase motif and a RNA binding motif. The alignment was expanded to include the other *Orbivirus* VP6 sequences used in this study; to include the full motifs as published by Kadaré and Haenni (1997) for virus-encoded RNA helicases and to examine the importance of the Rep helicase motif (figure 2.12). The ATPase motif is conserved in all the proteins compared. The DEAD/DExH helicase motif (common to SF2 families) is well conserved in the known helicase proteins but not in the *Orbivirus* VP6 proteins. The Rep helicase motif is well conserved in the *Orbivirus* VP6 proteins (with the exception of St Croix River virus VP6). It is not well conserved in the known SF2 helicases. Although the RNA binding motif characteristic of SF2 helicases is well conserved in BTV-10 VP6 and the known SF2 helicases, it is not conserved in any of the other *Orbivirus* VP6 proteins (namely AHSV, Chuzan and St Croix River virus VP6).

Site	ATPase	Helicase	Rep helicase	RNA binding
Motif	**$Gx_{1-4}GK^{S}/_{T}$[1,2]**	**DEAD/DexH**	**[426] MTRNKSMFTAS[436]**	**Tx_5 QRxGRuGR[1] / RxGRxxR[2]**
AHSV-3 VP6	[129] **G**K**G**V**GKS**T[136]	[293] TEE**DE**IKAVQ[102]	[318] VS**RA**K**A**M**FTA**P[328]	[238] SNFFTVDKSLLKRGG[252]
AHSV-6 VP6	[129] **G**K**G**V**GKS**T[136]	[293] TEE**DE**IKAVQ[102]	[318] VS**RA**K**A**M**FTA**P[328]	[238] SNLFTVDKSLLERGG[252]
BTV-10 VP6	[105] NR**G**D**GK**VG[112]	[152] VYR**DE**VPAQI[161]	[278] VV**R**ATAY**FTA**P[288]	[198] **T**HAKAVE**RG**G**R**KQ**R**K[212]
BTV-17 VP6	[102] NR**G**D**GK**VG[109]	[149] VYR**DE**VPAQI[158]	[275] VM**R**ATAY**FTA**P[285]	[195] **T**HAKAVE**RG**G**R**KQ**R**K[209]
Chuzan VP6	[83] ATEV**GK**HA[90]	[42] DGK**DE**DGQDQ[51]	[221] VKDA**K**AY**FT**CP[231]	[140] **T**VIFLDKNLQNEFNL[154]
St Croix VP6	[37] QD**G**E**GT**K**S**[44]	[80] GSY**DE**GPLPE[89]	[177] VQEAAAYYS**A**P[187]	[121] **T**YEKALKKSNLKVFE[135]
HCV1 NS3	[1234]PT**G**S**GKS**T[1241]	[1313]IIC**DECH**SDA[1322]	[2379]DSDAE**S**YSSMP[2389]	[1480]DAVSRT**QR**R**GR**T**GR**G[1494]
DroMLE	[408] NT**G**C**GKT**T[415]	[504] IIV**DEIH**ERD[513]	[1089]IV**R**VDNWLNFD[1099]	[758] SKTNLE**QR**K**GR**A**GR**V[772]
PpvHel	[1255]AV**G**S**GKS**T[1262]	[1339]IIF**DECH**VHD[1348]	[2497]DDF**N**NQFYSLN[2507]	[1513]SYGERI**QR**L**GR**V**GR**[1527]
Reovirus λ1	[6] RKTK**GKS**S[13]	[97] EAK**DEAD**EAT[106]	[978] DLSDMLLEPLL[988]	[421] NFVASCV**R**N**R**V**GR**FD[435]
HuHelA	[412] AT**G**C**GKT**T[419]	[507] VIV**DEIH**ERD[517]	[1080]IVLVDDWIKLQ[1090]	[753] SKTNLE**QR**K**GR**A**GR**S[767]

Figure 2.12 Alignment of *Orbivirus* VP6 putative motifs important for helicase activity with known SF2 helicases.

At the top of the figure are the conserved helicase motifs with their function. Sequences were obtained from GenBank: AHSV-3, U19881; AHSV-6, U33000; BTV-10, D00509; BTV-17, U55798; Chuzan AB018088; St Croix, AF145406; HCV1 (Hepatitis C virus type 1), M62321; DroMLE (*Drosophila melanogaster* maleless protein), M74121; PpvHel (plum pox virus CI protein), M92280; Reovirus λ1 protein, NP-694679 and HuHelA (human RNA helicase), L13848. Motifs were obtained from [1]Kadaré and Haenni (1997) and [2]Kar and Roy (2003).

2.3.7 HYDROPHILICITY AND SECONDARY STRUCTURE

Hydrophilicity plots determined according to the method of Kyte and Doolittle (1982) on ANTHEPROT (Deléage *et al*., 1988) showed that VP6 of AHSV, as in the case of BTV and Chuzan virus, is a predominantly hydrophilic protein (figure 2.13). The hydrophilicity plots of AHSV-3 and AHSV-6 appear to be virtually identical, therefore, only the AHSV-6 hydrophilicity plot is presented here. When the hydrophilicity plots are compared, a large hydrophilic peak is present in all three VP6 proteins at the N-terminal from about redidue 25 – 75. There are several hydrophilic domains similarly distributed throughout the rest of the three VP6 proteins. These results suggest that the proteins are soluble which will facilitate further AHSV VP6 protein related studies.

Since the secondary structure of a protein may be important for the nucleic acid binding function, the secondary structure of VP6 of AHSV-6 was determined using the 3D-pssm server. This analysis allows the prediction of secondary structure and solvent accessibility of each residue of the VP6 proteins (Fischer *et al*., 1999; Kelley *et al*., 1999; Kelley *et al*., 2000) (figure 2.14). The results of this analysis did not reveal secondary structures characteristic of nucleic acid binding proteins. Secondary

structures such as βαββαβ (RNP-1 and RNP-2) and the β5 barrel (CSD) nucleic acid binding proteins were not found in AHSV VP6. Comparisons with two libraries on the 3D-pssm server predicted that AHSV VP6 is a left-handed beta helix. Comparison with another library predicted a beta-roll secondary structure.

Regions of significant solvent accessibility were not identified. This means that specific regions capable of direct contact with nucleic acids cannot be predicted from the amino acid sequence.

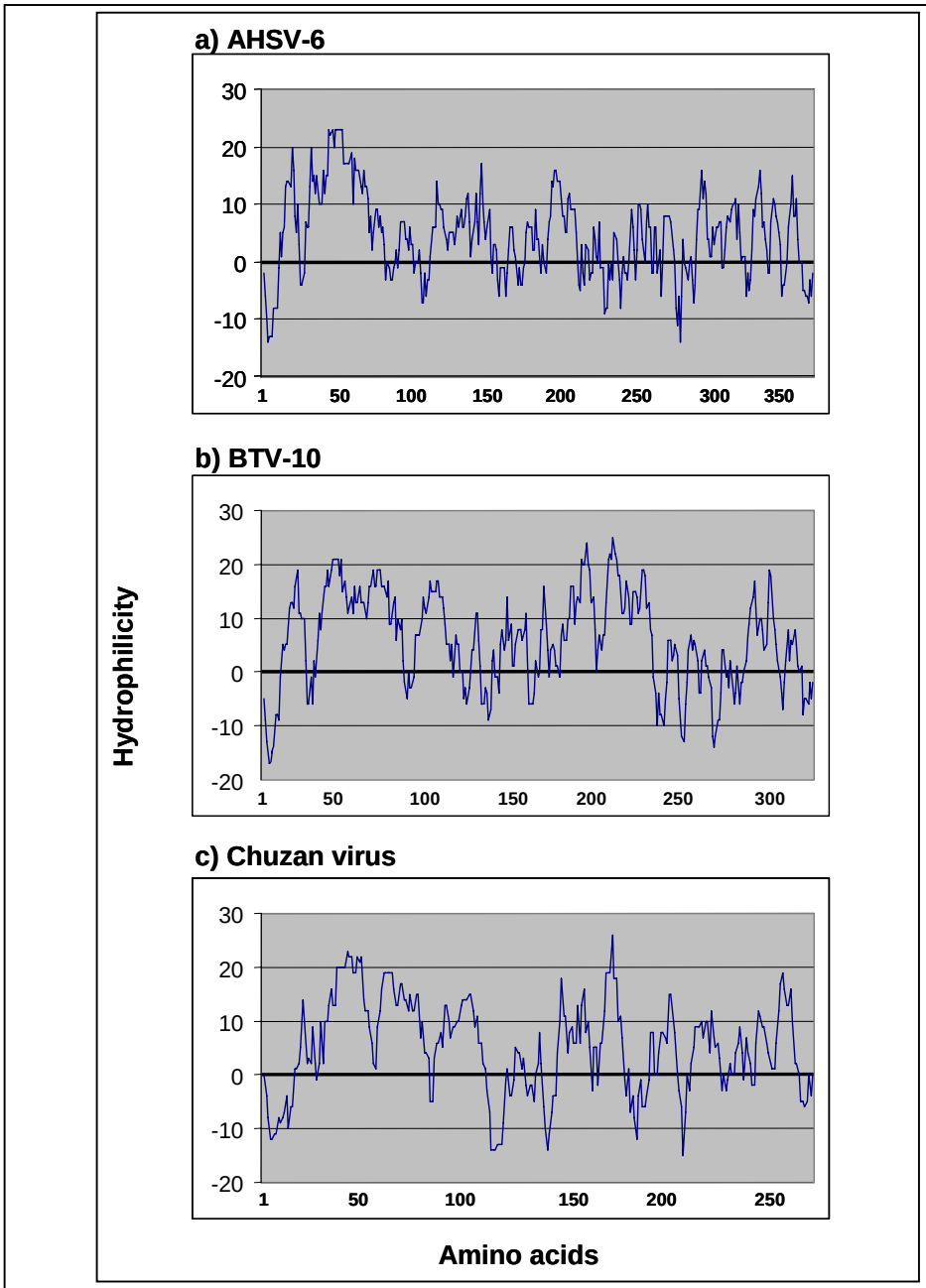

Figure 2.13 Hydrophilicity plots of the VP6 proteins of AHSV-6, BTV-10 and Chuzan virus.

Determined according to the method of Kyte and Doolittle (1982) plotted on ANTHEPROT (Deléage *et al*., 1988).

```
Conf: 95002006406888988866651022244105871343321212312574311013134 1
Pred: CCEEEECCCCHHHHHHHHHHHHHCCCHHHHHCCCHHHHHHHHHCCCCCCCCCCCCCCCC
  AA: MSSALLLAPGDLIEKAKRELEQRSITPLLREKGSKEAKSKLKEDGEKKNKSEKEENKIHD
            10        20        30        40        50        60

Conf: 00012122568881021015567677432245787775433346676662665445678 7
Pred: CCCCCCCCCCCCCCHHHHCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC
  AA: DRRVESQKSEGGGPADCQRGAGSAGANCATSTGGGDGSAGARTGIGGGGVGGVDSRSGGH
            70        80        90       100       110       120

Conf: 88764556664576656632023687542346331477753544423677546467555 5
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEECCCCCCCCCCCCCCCC
  AA: GGQGAASDGKGVGKSKTGADRVANDDATRNVGSSEVSSGGITSGGLQGRGGLVAKSGECG
            130       140       150       160       170       180

Conf: 88632136667888567632234788860122105889999986419635576078942 8
Pred: CCCCCCCCCCCCCCCCCCCCCCCCCCCCCCEEEECCHHHHHHHHHHHCCEEEEECCCCCEE
  AA: GESLDRIGGCSGNSKTEGEEAKAGGGDRRIGGLATQEIADFVKKKVGVEVQVFSKGMSNL
            190       200       210       220       230       240

Conf: 98702587620999799898999999998735750654210188988863687521235 4
Pred: EEECHHHHHHHCCCHHHHHHHHHHHHHHHHHCCCCEEEEEEHHHHHHHHHCCCCCCCCC
  AA: FTVDKSLLERGGLGREDILHQSDIVKEIRVSDKKVKIIPLSTVKRMIAEFGGTEEDEIKA
            250       260       270       280       290       300

Conf: 55310135761078878864011156788896478899885069988998103765047 8
Pred: CCCCEEEEEEEECCHHHHHHHHEEEECCCCCCCHHHHHHHHCCCCCEEEEECCCCCCHHH
  AA: VQTQSSSIRYISNRMEDVSRAKAMFTAPTGDEGWKEVAKAATQRPNIMAYVHEGEGDGLK
            310       320       330       340       350       360

Conf: 899887059
Pred: HHHHHHCCC
AA:   ELLHLIDHI
```

Figure 2.14 Secondary structure prediction of VP6 of AHSV-6.

This was performed on Psi-Pred (Jones, 1999) using the 3D-PSSM platform at the Biomolecular Modelling Laboratory hosted by the Imperial Cancer Research Fund, UK. The first line (designated Conf:) is the E-value which represents the confidence with which the match is made (0 = low; 9 = high). The second line (designated Pred:) gives the predicted secondary structure class for each residue. C=coil; E=β-strand/sheet and H=α-helix. The third strand gives the amino acid sequence.

2.4 DISCUSSION

The genes encoding the minor core components of AHSV, VP1, VP4 and VP6 have not been characterized in great detail. During the course of this investigation the full-sized VP6 genes of both AHSV-3 and AHSV-6 were cloned and fully sequenced for the first time. Subsequently, a method has been developed for sequence independent amplification of dsRNA segments with a single primer (Potgieter *et al*., 2002) which would have greatly facilitated the cloning of the genome segment encoding VP6.

It was found that both genes are 1169 nucleotides in length with an 1107 nucleotide open reading frame. The genes are flanked by the conserved 5' GTTAAA and 3' ACTTAC sequences (Mertens and Sangar, 1985; Roy, 1989) (figure 2.7). The start codon for both genes is at nucleotide 18 and there is a 42 nucleotide 3' non-coding region. Both polypeptides are 369 residues in length. The sequence results were submitted to GenBank (accession numbers: AHSV-3 U19881 and AHSV-6 U33000).

On a nucleotide level, there is 96% identity between the AHSV-3 and AHSV-6 VP6 genes. The changes are distributed throughout the gene with no one specific region where a cluster of changes occurs. The encoded proteins are highly conserved with 93.5% identity and 96% similarity as determined by CLUSTALW alignment. The predicted molecular weights are in the order of 38.8 - 39kDa. This is slightly larger than the predicted molecular weight for BTV-10 VP6 which is 35 458kDa (Fukusho *et al.*, 1989). Both VP6 of AHSV-3 and AHSV-6 are basic proteins with isoelectric points (pI) of 7.24 and 8.25 respectively.

In order to determine if there were characteristic features of VP6 that are conserved amongst all cognate genes of the orbiviruses, the amino acid content of AHSV-3 and AHSV-6 VP6 proteins was compared to BTV-10 VP6, Chuzan VP6 and St Croix River virus VP6. The number of cysteine residues for all the proteins compared was low. Aromatic residues (F + Y) were low for BTV (18 per 1000 residues) and AHSV (19 / 21). Chuzan and St Croix River virus were higher at 47 and 61 respectively. Glutamine and asparagine content are fairly low for all four orbiviruses. The content of charged amino acids (R + K + H; D + E) is high for all the proteins compared, the lowest being 267/1000 residues for St Croix and highest at 363/1000 residues for BTV. As in BTV VP6 (Roy, 1992), AHSV has a comparatively high number of basic residues. Possibly most significant is the glycine content which for AHSV is 162 – 168 per 1000 residues, 125 for BTV and 110 for Chuzan. For the previous three orbiviruses, glycine is the most abundant amino acid. In St Croix River virus VP6 there are fewer glycine residues (56), the most abundant residue being serine (103). In terms of residue content, AHSV, BTV and Chuzan are most similar and St Croix is somewhat different.

To summarize the relationship between the different VP6 proteins, the percentage similarity was calculated from CLUSTALW alignments for six *Orbivirus* VP6 proteins. Similar to BTV VP6 for which an identity of 97% and a similarity of 98% were observed between serotypes 10 and 17, AHSV VP6 is highly conserved within the serogroup.

Between BTV and AHSV the identity ranges from 26 – 27% and similarity 45 – 47%. When Chuzan VP6 is compared to VP6 of AHSV and BTV the identity falls within the same range (24 – 26%) with a similarity between 40 and 49%. St Croix River virus VP6 is the most distantly related to AHSV with an identity of 12 % (similarity 29%) and closest to Chuzan (19% identity and 37% similarity). It seems evident that the VP6 proteins of AHSV, BTV and Chuzan are most closely related while St Croix River virus is most distantly related. This may be as expected as the first three are biting midge borne orbiviruses while the latter is tick borne (Attoui *et al*., 2001).

Yamakawa *et al*. published the complete nucleotide sequence of the genome of Chuzan virus in 1999. They performed comparative sequence analysis for each segment with BTV and AHSV using UPGMA on amino acid sequence data. UPGMA (unweighted pairgroup method using arithmetic averages) assumes ultrametric data; an equal rate of evolution or an universal molecular clock (Swofford and Olsen, 1990; Page and Holmes, 1998). A broad correlation between the amount of molecular divergence and time is generally accepted. However, it has not been established that rates of evolution are constant. Evidence indicates enough heterogeneity in the rate of evolution that equal rates cannot be assumed (Moritz and Hillis, 1990). Genes do not diverge uniformly in all organisms (or organelles) and as a result systematic errors may well arise should equal rates of evolution be assumed. For each segment, Yamakawa *et al*. (1999) found that AHSV and Chuzan grouped together. The only exception was for the VP6 proteins where BTV and Chuzan grouped together. These results were problematic and as such an attempt was made to formulate a hypothesis of the evolutionary relationship and functional relatedness between the genome segments encoding VP6 of the four orbiviruses above by performing analyses on both nucleotide and amino acid data sets respectively.

Based on the data set of aligned VP6 genes for evolutionary analysis and VP6 amino acid sequences for functional analysis, trees were drawn using maximum parsimony methods performed using PAUP ver 4.0b10 (Swofford, 1999). Parsimony methods gave the best resolution as compared to neighbour joining methods. Characters were weighted to compensate for regions of poor alignment as specified by Swofford and Olsen (1990). All three methods of treating gaps are included in the trees. Bootstrap methods were used to test the robustness of the trees. According to the results with a nucleotide data set, AHSV and BTV grouped together with a bootstrap value of 100. Chuzan grouped on its own and St Croix River virus was used as an outgroup. In terms of inferred common ancestry (nucleotide data) the results presented here indicate that AHSV and BTV VP6 genes are more closely related to each other than to Chuzan virus VP6.

The protein is the most biologically relevant aspect of the gene and may, therefore, be used to infer functional relatedness or constraints (Swofford and Olsen, 1990). It is, however, not considered appropriate for evolutionary hypotheses. It is also implicit that amino acid sequences have only a third of the information of the nucleotide data and as

such fewer informative characters. Yamakawa *et al*. (1999), reported that VP6 of Chuzan virus grouped with BTV-10. Based on their analysis, VP6 of BTV-10 and Chuzan virus are functionally more similar to each other than to AHSV VP6. It was investigated whether parsimony methods would deliver the same results as those of Yamakawa *et al.* (1999) for functional constraints. Accordingly, a functional constraint tree was constructed using amino acid sequence alignments of the VP6 proteins of AHSV-3 and AHSV–6 VP6, BTV-10 and BTV–17, Chuzan and St Croix River virus (outgroup). Three data sets were generated differing in how the gaps were treated. When the gaps were treated as missing data (ignored), AHSV grouped with Chuzan virus with a bootstrap value of 61. In the second data set, the characters were weighted so that gaps in series of more than one were designated 1 = presence of a gap and 0 = absence of a gap i.e. single events, the same tree was obtained but with no support (bootstrap value = 51) and as such the relationship is unresolved. When each individual gap was treated as a character, AHSV grouped with BTV with a bootstrap value of 74. Indels (gaps) arise from biological events which may be as informative as nucleotide or amino acid changes. It would seem that to ignore the gaps would give a skewed picture. However, to treat each individual gap as a character may give inappropriate weight to what may be a single event. To clarify the functional relatedness, more sequence data are required which are currently not available for *Orbivirus* VP6 proteins.

Numerous phylogenetic studies have been performed to determine the relationships between the orbiviruses (De Sá *et al*., 1994; Williams *et al*., 1998; Yamakawa *et al*., 1999; Van Niekerk *et al*., 2001b). The most variable AHSV proteins are VP2 followed by NS3. VP2 is the type specific antigen and NS3 may be useful in differentiating between field isolates and live attenuated vaccine strains of the same serotype (van Niekerk *et al*., 2001b). VP6 proteins appear to be highly conserved within serogroups and very different in terms of total amino acid length between serogroups. VP6, the putative viral helicase, is therefore, not a particularly good choice for determining relationships within or between serogroups.

The amino acid sequences were aligned using CLUSTAL W. The regions of the different VP6 proteins that have the highest number of conserved residues are located in the amino terminal (first fifty residues) and the carboxyl terminal (last 126 residues). The difference in the number of amino acids in the VP6 proteins of the different orbiviruses (AHSV 369 amino acids to St Croix River virus 232 amino acids) is predominantly as a result of stretches of amino acids present in AHSV that are missing from some or all of the other VP6 genes used for comparison (figure 2.11).

RNA viruses evolve very quickly (Worobey and Holmes, 1999). Two types of genetic exchange are found in segmented RNA viruses namely, reassortment and recombination. In the case of recombination, a “donor” nucleotide sequence is inserted into an “acceptor” RNA molecule which results in a new RNA consisting of genetic information from more than one source (Worobey and Holmes, 1999). Homologous recombination has been described in rotaviruses (Suzuki *et al*., 1998). When strict

alignment is not maintained, genetic information may be lost or gained. From the protein alignment of the four *Orbivirus* VP6 proteins, it appears that the difference in length between AHSV, BTV, Chuzan and St Croix River virus VP6 proteins may be explained by a single indels. The amino acid sequences that have been inserted or deleted are most probably nonessential to the function of the protein.

Helicases enzymatically unwind double-stranded nucleic acid molecules by translocating along one strand. Mechanisms proposed for helicase action have two important features. Firstly, ATP-driven conformational rearrangements that provide energy and secondly, nucleic acid binding sites for binding single and double-stranded nucleic acids (Kadaré and Haenni, 1997). Five helicase families have been identified by means of a number of amino acid motifs. All the proteins grouped within these families have ATP binding properties and have the Walker A motif (phosphate-binding loop or P-loop) and the Walker B (Mg^{2+} binding aspartic acid) motifs (Caruthers and McKay, 2002). Superfamilies 1 and 2 have 7 characteristic motifs. SF1 includes alphavirus-like NSP2 proteins and SF2 includes the DexH/DEAD protein family. SF3 proteins have three conserved motifs and include the picornavirus 2C-like proteins (Kadaré and Haenni, 1997). The fourth family includes proteins which are related in sequence to the *E. coli* DnaB protein and describes hexameric ring-like helicases. The fifth family of proteins have sequences similar to the β subunit of proton-translocating ATPases (Caruthers and McKay, 2002). The seven conserved helicase motifs are presented in the table below with an example and functional interactions for each (**Table 2.5**).

Eight nucleic acid binding protein families have been identified on the basis of common conserved sequences. These motifs include the RRM (RNA recognition motif); the ARM (rich in arginine); RGG box (RGG repeats); KH motif; the dsRNA binding motif comprising 9 protein families characterized by a zinc-finger-knuckle; cold shock domain (CSD) consisting of conserved aromatic basic residues and GRPs (glycine rich proteins) characterized by GGGx/GGxxxGG or GxGx motifs.

Table 2.5: Conserved helicase motifs and functional interactions.

Abbreviations are as follows: b is a bulky aliphatic or aromatic hydrophobic amino acid; u is a bulky aliphatic amino acid and x is any amino acid. Functional interactions were determined by crystallographic structures with residues (in bold) involved in the interactions. The table was compiled using information from Kadaré and Haenni (1997)[1] and Caruthers and McKay (2002)[2].

Motif	Classical consensus sequence[1]	Conserved motifs of HCV, SF2 helicase[2] (alternative minimal motifs in parenthesis where applicable)	Functional interaction[2]
I	$b_2x_4GSGxGK^S/_Tx_3bP$	APTGSG**KT** ($GK^T/_S$)	The lysine (K) residue interacts with the phosphates of MgATP/MgADP and the Mg^{2+} ion is ligated by the threonine (T) or serine (S) residue.
Ia	$Rxbu_2xPTRxux_2Eb$	**P**SVAAT	Interacts with single-stranded nucleic acids. The backbone carbonyl of the P residue bonds with the oligonucleotide phosphodiester backbone.
II	b_4DExH	**DE**CH (/DEAD)	The aspartic acid (D) is thought to coordinate interactions with the Mg^{2+} ion of MgATP/MgADP while the glutamic acid (E) supposedly functions as a catalytic base (a Lewis base) during hydrolysis of ATP. H residue interacts with motif VI Q residue.
III	bxuTATPP	**T**AT (/ SAT)	T hydrogen bonds with H of motif II
IV	ub_2uPS	LIFCHS**K**K	K interacts with ssDNA
V	$BubxTDbxExGuxbx_4u_2$	ALMTGFT	No specific interactions documented
VI	$Tx_5QRxGRuGR$	**Q**RRGRTGR (QxxGRxxR)	Q interacts with H of motif II

The amino acid sequence of AHSV VP6 was analyzed to identify any motifs or amino acid signatures that could be used to predict the function of AHSV VP6. In the first instance, the amino acid sequences of the two AHSV serotypes were aligned. The Walker A and putative B motifs; similarity to the conserved helicase motifs III and IV of the SF1 and SF2 helicases and 91% similarity to a sequence found in the Rep helicase were identified. Several motifs characterizing glycine rich proteins were present as well as a N-glycosylation signal. Comparison of the AHSV VP6 amino acid sequence with those of BTV, Chuzan and St Croix River virus VP6 proteins showed four conserved amino acid regions. Two of these were found only in the amino acid sequences evaluated (namely motifs i and iii). Motif i showed similarity to a coiled-coil protein and motif iii was common to the *Orbivirus* VP6 proteins with no other significant similarity.

Motif ii showed marked similarity to ATP dependent helicases and overlapped with the region of high similarity to the Rep helicase. This region may be of functional importance in helicase activity. Motif iv showed similarity to ATP binding or dependent transporters and carrier proteins. No indication of conserved nuclear targeting signals reported for BTV VP6 (Yi *et al*., 1996) was determined from multiple *Orbivirus* protein alignments. Motif iii overlapped with a domain identified by Hayama and Li (1994) in BTV VP6 to be important in nucleic acid binding activity (designated 75). Hayama and Li (1994), describe three epitopes (70, 71 and 75) comprising two binding domains rich in basic residues (arginine and lysine). Motifs designated 70 and 71 were not well conserved in AHSV VP6.

In 2003, Kar and Roy proposed that BTV-10 VP6 has three motifs consistent with SF2 helicases. They presented a sequence alignment comparing BTV-10 VP6 to known SF2 helicases with respect to these motifs namely, the Walker A and B motifs and an RNA binding motif (similar to the conserved helicase motif VI, table 2.5). The alignment presented by Kar and Roy (2003) was expanded to include comparisons to AHSV serotypes 3 and 6 VP6; BTV-17 VP6; Chuzan VP6 and St Croix River virus VP6 proteins. The Rep helicase motif that is conserved within the *Orbivirus* VP6 proteins was included to assess its relevance with regard to other helicases.

The Walker A motif is the most conserved motif across the helicase families. Originally defined as GxxxxGKT, it requires a minimum of GK($^{T}/_{S}$) (Caruthers and McKay, 2002). It is characteristic of NTP dependent nucleic acid translocases (Soultanas and Wigley, 2001; Singleton and Wigley, 2002). Kar and Roy (2003), were able to demonstrate the importance of the lysine (GxG**K**T) in the Walker A motif for ATP binding, hydrolysis and RNA unwinding. This motif is conserved in all the amino acid sequences compared.

Kar and Roy (2003), evaluated the effect of modifying a putative helicase motif DEVP which resembles the DEAD / DExH motif of SF2 helicases. This represents the Walker B motif which was originally defined as a single aspartic acid (D) but most often takes the form of DExx (Caruthers and McKay, 2002). Kar and Roy (2003), changed the glutamic acid (E) to an asparagine (N). RNA binding and unwinding was only slightly affected by this change suggesting that this motif is not important for BTV VP6 enzyme activity. There is more than one DExx sequence in the proteins analysed, possibly this is not the correct one. Although several DExx motifs are present in all the *Orbivirus* VP6 proteins analyzed, the DEAD /DExH motif is not present.

The RNA binding motif identified in BTV VP6 corresponds to Hayama and Li's motif 71. By changing the **R**xGRxxR in the putative helicase RNA binding motif, RNA binding and unwinding as well as ATP hydrolysis activities were significantly reduced (Kar and Roy, 2003). These authors suggest that the functions are interlinked and require the same domains. This motif is highly positively charged and contains a RGG motif in AHSV and BTV VP6 proteins. Three basic amino acids are found in this region in AHSV and may perform a similar function even although the motif *per se* is not conserved.

The Rep helicase motif included in the alignment with known helicases (figure 2.12) is well conserved in AHSV and BTV VP6 proteins, less so in Chuzan VP6 and not conserved in St Croix River virus VP6. It is not conserved in the known SF2 helicases.

There are a number of short motifs characteristic of helicases in the AHSV VP6 protein. These include the Walker A motif, a putative Walker B motif and conserved motifs III and IV. Although not conserved in the SF2 helicases, the Rep helicase motif is conserved in the *Orbivirus* proteins and shows similarity to ATP dependent helicases. As such, it may play a role in *Orbivirus* helicases. However, based on the sequence analyses performed, there are no obvious target motifs for nucleic acid binding activity.

Hydrophilicity plots revealed that VP6 of AHSV-3, AHSV -6, BTV-10 and Chuzan are all predominantly hydrophilic proteins. This suggests that the proteins are soluble. Secondary structure plays and important role in some nucleic acid binding proteins. The RNP-1 and RNP-2 proteins have a βαββαβ secondary structure which forms a four-stranded antiparallel β sheet. As a result, the charged and aromatic side chains of the two sites are solvent exposed and can make direct contact with the bound RNA (Burd and Dreyfuss, 1994). Graumann and Marahiel (1996), report that the secondary structure of CSD proteins, which bind double-stranded and single-stranded DNA as well as RNA, have a β5 barrel secondary structure. GRPs which are proposed to be involved in protein-protein interactions, have a highly flexible secondary structure as a result of their glycine content (Sachetto-Martins *et al*., 2000). Secondary structure predictions were performed using the 3D-pssm server (Fischer *et al*., 1999; Kelley *et al*., 1999; Kelley *et al*., 2000).

The α-helix is a rod like, conformationally stable arrangement which involves a single polypeptide chain. The α-helix may be disrupted by charged side chains which repulse each other electrostatically or by a proline residue which causes a bend in the polypeptide backbone. The β-sheet is a two dimensional arrangement involving one or more polypeptide chains. Parallel or anti-parallel pleated sheets are formed with intrachain hydrogen bonds. The α-helix and β-sheet elements are combined in many ways as a polypeptide folds back on itself in a chain. For steric reasons, glycine is frequently encountered in reverse turns at which the polypeptide changes direction. The single hydrogen of the side chain prevents crowding. Combinations of α-helix and β-sheet elements produce a variety of supersecondary structures (Campbell, 1995). It can be predicted that the AHSV VP6 polypeptide has numerous reverse turns in its secondary structure as a result of the glycine rich residue content.

Fold recognition by comparisons with two libraries on the 3D-pssm server, predicted that VP6 of AHSV-6 is a single-stranded left-handed beta helix as found in the trimeric LpxA-like enzymes belonging to the UDP N-acetylglucosamine acetyltransferase and Xenobiotic acetyltransferase families. Comparison to another library predicted a beta-roll secondary structure as found in metalloprotease families. Solvent accessibility as

determined by DSSP (Kabsch and Sander, 1983) did not reveal regions of significant accessibility.

In a first step towards characterization of AHSV VP6, the primary structure of the genome segment encoding VP6 and encoded protein was analyzed. For further analysis of functional aspects of VP6, expression of the AHSV VP6 genes was investigated. This work is described in the next chapter.