# CHAPTER 2

# CHARACTERIZATION OF GENOME SEGMENT 10 OF EEV AND ITS ENCODED GENE PRODUCT NON-STRUCTURAL PROTEIN NS3

## 2.1. INTRODUCTION

A number of conserved characteristics or features have been identified within the nucleotide sequence of the NS3 gene of a number of orbiviruses such as AHSV, BTV, EHDV, Palyam virus and Broadhaven virus (e.g. Rao *et al.*, 1983; Mertens and Sangar, 1985; Van Staden and Huismans, 1991; Hwang *et al.*, 1992; Moss *et al.*, 1992; Jensen *et al.*, 1994; Jensen and Wilson, 1995; Van Staden *et al.*, 1995; Pierce *et al.*, 1998; Bonneau *et al.*, 1999; Yamakawa *et al.*, 1999; Attoui *et al.*, 2001; Van Niekerk *et al.*, 2001b; Potgieter *et al.*, 2002). These characteristics include the presence of two in-phase start codons, conserved terminal sequences, inverted repeats in the non-coding regions and the sizes of the non-coding and the coding regions. Characteristics of amino acid sequences of NS3 include glycosylation sites, myristylation sites, hydrophobic regions that form potential transmembrane helices, proline-rich regions, conserved regions, variable regions and conserved cysteine residues. A number of these regions within the genes or proteins have been found to have a functional role or to be related to the function of the protein.

Another important characteristic of both the S10 nucleotide sequence and the NS3 amino acid sequence is the amount of variation observed for this gene and gene product within serotypes and between different serotypes within a serogroup of orbiviruses. The BTV NS3 gene and gene product have been shown to be highly conserved between serotypes and sequence comparisons between BTV NS3 genes and between NS3 proteins have revealed a high percentage of similarity (e.g. Gould, 1988; Wade-Evans, 1990; De Mattos *et al.*, 1992a; Hwang *et al.*, 1992; Pierce *et al.*, 1998). In contrast, the AHSV NS3 gene and gene product have been shown to be highly variable between serotypes, and sequence comparisons between AHSV NS3 genes and between NS3 proteins have revealed a lower percentage of similarity (e.g. Van Staden and Huismans, 1991; De Sá *et al.*, 1994; Sailleau *et al.*, 1997; Martin *et al.*, 1998; Van Niekerk *et al.*, 2001b).

At present, no sequence information is available for S10 or NS3 of EEV, except that of the 5' and 3' terminal sequences (Potgieter *et al.*, 2002) and little or no information is available regarding the levels of variation within and between serotypes of the EEV serogroup, specifically regarding S10 and NS3. The limited S10/NS3 sequence data for orbiviruses other than AHSV or BTV also restrict a comprehensive comparison of the typical variation for this genome segment and non-structural protein of the orbivirus genus.

In order to characterize and compare genome segment 10 and the NS3 protein of EEV to other orbiviruses regarding conserved structural features and levels of variation, it will be necessary to determine the nucleotide sequences and the deduced amino acid sequences of a number of EEV isolates (including reference strains and field isolates), and to then compare genome segment 10 and the NS3 protein of EEV across the different serotypes, as well as to genome segment 10 and the NS3 protein of other orbiviruses using various bioinformatics tools and phylogenetic analysis.

The identification of conserved or unique structural features within the EEV S10 nucleotide sequence or NS3 protein sequence, and the determination of the level of variation seen for S10 and NS3 of EEV, compared to other orbivirus S10 or NS3 sequences, will allow the prediction of possible roles or functions for NS3 of EEV.

## 2.2. MATERIALS AND METHODS

### 2.2.1. Cells and viruses

The EEV Bryanston serotype (EEV-1) (a laboratory reference strain) (S1REF*) was obtained from the Equine Research Centre at the Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa. The additional EEV isolates used in this investigation were all South African isolates and all the isolates were serotyped at and received from the Equine Research Centre at the Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa. The EEV isolates included seven different serotypes (reference strains) and field isolates. All the available information for the EEV isolates are shown in Tables 2.1., 2.2. and 2.3. The information in Table 2.1. and 2.3. was provided with the samples when they were obtained from the Equine Research Centre. The Equine Research Centre however originally received the seven reference serotypes from the ARC-Onderstepoort Veterinary Institute and the information they were supplied with (origin, source and passage history of the seven serotypes) is shown in Table 2.2.

Other viruses used in experimental procedures include EHDV-1, BTV-10 and AHSV-9. In the case of EHDV-1 and BTV-10, purified dsRNA for each was obtained from Mr. M. K. Lombardi (Department of Genetics, University of Pretoria, South Africa). In the case of AHSV-9, dsRNA was obtained from Prof. H. Huismans (Department of Genetics, University of Pretoria, South Africa).

Other orbivirus S10 or NS3 sequences used for analysis were drawn from GenBank and all the relevant information, including the accession numbers are given in Appendix C, Table C1.

**Table 2.1.** The information provided with the EEV isolates when they were obtained from the Equine Research Centre, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa

| EEV isolate[a] | Isolate information[b] | Laboratory passage from[c] | Serotype identification using name and number | AHS ELISA | EEV ELISA |
|---|---|---|---|---|---|
| S1REF* | Bryanston (received in March 1999) | Unknown | EEV-1 | Information not available | Information not available |
| S1REF | Bryanston Ref. Antigen 7.8.00 | 19.7.99 | EEV-1 | Negative | Positive |
| S2REF | Cascara Ref. Antigen 7.8.00 | 14.8.98 | EEV-2 | Negative | Positive |
| S3REF | Gamil Ref. Antigen 7.8.00 | 6.8.98 | EEV-3 | Negative | Positive |
| S4REF | Kaalplaas Ref. Antigen 7.8.00 | 6.8.98 | EEV-4 | Negative | Positive |
| S5REF | Kyalami Ref. Antigen 7.8.00 | 22.10.98 | EEV-5 | Negative | Positive |
| S6REF | Potchefstroom Ref. Antigen 7.8.00 | 7.7.97 | EEV-6 | Negative | Positive |
| S7REF | E21/20 Ref. Antigen 7.8.00 | 17.4.00 | EEV-7 | Negative | Positive |
| S1FLD1 | E199/99 8 # 3 BHK 11.9.00 | # 2 BHK 5.7.00 | Bryanston; EEV-1 | Negative | Positive |
| S1FLD2 | E193/99 2 # 4 BHK 5.9.00 | # 3 BHK 18.7.00 | Bryanston; EEV-1 | Negative | Positive |
| S1FLD3 | E183/99 3 # 3 BHK 5.9.00 | #2 BHK 28.7.00 | Bryanston; EEV-1 | Negative | Positive |
| S1FLD4 | E300/99 # 3 BHK 5.9.00 | # 2 BHK 28.7.00 | Bryanston; EEV-1 | Negative | Positive |
| S3FLD1 | E131/2000 # 4 BHK 18.8.00 | # 3 BHK 9.5.00 | Gamil; EEV-3 | Negative | Positive |
| S6FLD1 | E35/2000 # 5 BHK 18.8.00 | # 4 BHK 22.3.00 | Potchefstroom; EEV-6 | Negative | Positive |
| S6FLD2 | E92/2000 # 3 BHK 18.8.00 | # 2 BHK 31.3.00 | Potchefstroom; EEV-6 | Negative | Positive |

[a] REF indicates laboratory reference strains, FLD indicates field isolates

[b] This is the information for each individual isolate. The name of the isolate is given, as well as its passage history up to the date when the isolate was received, e.g. S3FLD1. The isolate E131/2000, received on 18.8.00, had undergone 4 passages in BHK cells up to this date.

[c] This is the passage history of each individual isolate, e.g. S3FLD1. The sample had undergone 3 passages in BHK cells from 9.5.00.

**Table 2.2.** Origin and identification of the seven reference strains supplied by the Equine Research Centre, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa (adapted from Howell *et al.*, 2002)

| Reference strain | Identification and serotype | Source | Laboratory passage | Origin and year |
|---|---|---|---|---|
| S1REF | Bryanston, M8/76 (EEV-1) | Foetal liver/Spleen | 3 BHK[a] (28/06/76) | Colesberg, Northern Cape, SA (1976) |
| S2REF | Cascara (EEV-2) | Organ suspension | 1 SM[b], 2 BHK (19/03/80) | Kimberley, Northern Cape, SA (1967) |
| S3REF | Gamil, M9/71 (EEV-3) | Blood | 4 BHK, 1 Vero (24/04/93) | Naboomspruit, Limpopo, SA (1971) |
| S4REF | Kaalplaas, 7088 (7-2) (EEV-4) | Blood | 4 BHK (06/07/77) | Onderstepoort, Gauteng, SA (1974) |
| S5REF | Kyalami, 7084 (12-3) (EEV-5) | Blood | 4 BHK (03/07/85) | Johannesburg, Gauteng, SA (1974) |
| S6REF | Potchefstroom, Else EP8/91 (EEV-6) | Blood | 1 Vero[c], 1 CER[d] (26/08/91) | Potchefstroom, North West Province, SA (1991) |
| S7REF | E21/20 (EEV-7) | Blood | 5 BHK (17/04/2000) | St. Lucia, KwaZulu-Natal, SA (2000) |

[a] BHK – baby hamster kidney clone 13 cells; [b] SM – suckling mice; [c] Vero – African green monkey cells (ATCC CCL81); [d] CER – chicken/BHK hybridoma

**Table 2.3.** Origin and identification of the seven field isolates supplied by the Equine Research Centre, Faculty of Veterinary Science, University of Pretoria, Onderstepoort, South Africa (adapted from Van Niekerk *et al.*, 2003)

| Field isolate | Identification and serotype | Virus origin | Year of isolation |
|---|---|---|---|
| S1FLD1 | E199/99 8 (EEV-1) | Stellenbosch, Western Cape, South Africa | 1999 |
| S1FLD2 | E193/99 2 (EEV-1) | Stellenbosch, Western Cape, South Africa | 1999 |
| S1FLD3 | E183/99 3 (EEV-1) | Stellenbosch, Western Cape, South Africa | 1999 |
| S1FLD4 | E300/99 (EEV-1) | Stellenbosch, Western Cape, South Africa | 1999 |
| S3FLD1 | E131/2000 (EEV-3) | Kempton Park, Gauteng, South Africa | 2000 |
| S6FLD1 | E35/2000 (EEV-6) | Hoedspruit, Limpopo, South Africa | 2000 |
| S6FLD2 | E92/2000 (EEV-6) | Roodeplaat, Gauteng, South Africa | 2000 |

### 2.2.2. Isolation of dsRNA

Total RNA was isolated using the TRIZOL® reagent (Life Technologies) according to the manufacturer's instructions. Cells were collected by centrifugation at 1 252 x g for 10 min and lysed by the addition of 500 µl reagent. The cells were incubated for 5 min at room temperature, after which 100 µl chloroform was added, the sample shaken for 15 sec and incubated at room temperature for a further 2 - 3 min. The samples were centrifuged at 12 000 x g for 15 min at 2 - 8°C and the RNA in the aqueous phase precipitated by adding 250 µl isopropyl alcohol. The samples were incubated at room temperature for 10 min and centrifuged at 12 000 x g for 10 min at 2 - 8°C. The supernatant was removed and the RNA pellet washed once with 75% ethanol. The sample was mixed by vortexing and centrifuged at 7 500 x g for 5 min at 2 - 8°C. The RNA pellet was briefly air-dried for 5 - 10 min and dissolved in 50 µl RNase free water (DEPC treated $ddH_2O$) by passing the solution through a pipette tip and incubating for 10 min at 58°C. The samples were stored at -20°C until use

Double-stranded RNA was isolated from the total RNA by lithium chloride precipitation of ssRNA and ethanol precipitation of the dsRNA. LiCl was added to a final concentration of 2 M and ssRNA (rRNA) precipitated by leaving the samples overnight or longer at 4°C and centrifuging at 15 700 x g for 30 min. The dsRNA in the supernatant was ethanol precipitated by the addition of 2.5 x volume 96% ethanol and NaCl to a final concentration of 0.2 M. The samples were left at -20°C for at least 1 h. The dsRNA was collected by centrifuging at 15 700 x g for 15 min. The supernatant was discarded and the pellets washed with 96% ethanol by centrifuging at 15 700 x g for 20 min and discarding the supernatant. The pellets were air-dried for approximately 1 h and resuspended in DEPC treated $ddH_2O$. A small volume of each sample was analysed by 1% agarose gel electrophoresis (section 2.2.3.) and the remainder of the samples was stored at -20°C until use.

### 2.2.3. Agarose gel electrophoresis

DNA or dsRNA samples were analysed by horizontal electrophoresis in 1% (w/v) agarose horizontal slab gels (Seaplaque® agarose) using a Biorad Mini Sub™ agarose gel electrophoresis unit (7 x 10 cm) at 75 - 100 V for 30 - 90 min. Gels were prepared as described in Sambrook *et al.* (1989). Briefly, an agarose solution was prepared as required by adding 1 x TAE buffer (0.04 M Tris-acetate, 0.002 M EDTA, pH 8.5) to agarose powder. The solution was brought to boil, the volume adjusted with $dH_2O$ to compensate for any loss and EtBr added to a final concentration of 1 µg EtBr/ml gel solution just before casting. The sample loading buffer used was a 1 in 6 dilution of 6 x loading dye in 50% glycerol (Blue/Orange 6 x Loading Dye containing 15% Ficoll® 400, 0.03% bromophenol blue, 0.03% xylene cyanol FF, 0.4% orange G, 10 mM Tris-HCl pH 7.5, 50 mM EDTA). The bands were visualized on an UV transilluminator, and gels documented using the gel documentation system.

### 2.2.4. Reverse transcriptase polymerase chain reaction (RT-PCR)

The RT-PCR method used was based on the methods described by Wade-Evans (1990), De Sá *et al.* (1994), Zientara *et al.* (1998) and Van Niekerk *et al.* (2001b).

Briefly, 250 - 500 ng total RNA or dsRNA was denatured 10 min at room temperature using an equal volume of 10 mM methylmercuric hydroxide (MMOH). The methylmercuric hydroxide in the sample was reduced with addition of 2 µl of 0.7 M β-mercaptoethanol in the presence of either 159 U human placental RNase inhibitor (HPRI, Amersham Life Sciences) or 40 U Recombinant RNasin® Ribonuclease inhibitor (Promega), and left for a further 5 min at room temperature.

Complementary DNA was synthesized for each sample by adding the denatured RNA to a cDNA reaction mix containing 100 pmol of each primer (Table 2.4.), 2 µl of 2.5 mM dNTP mix (2.5 mM each of dATP, dCTP, dGTP, dTTP) (TaKaRa Biomedicals), 2.4 µl of AMV RT 5 x reaction buffer (250 mM Tris-HCl pH 8.3, 250 mM KCl, 50 mM $MgCl_2$, 50 mM DTT, 2.5 mM spermidine) (Promega) and 10 U AMV Reverse Transcriptase (Promega). The reaction was then incubated at 42°C for 90 min.

The S10 cDNA obtained was amplified using high fidelity PCR. Each reaction contained 5 µl 10 x PCR buffer (including 2 mM $MgCl_2$)(TaKaRa Biomedicals), 2.5 µl 2.5 mM dNTP mix (TaKaRa Biomedicals), 1 µl of each primer (100 pmol/µl), 0.5 µl TaKaRa Ex Taq™ polymerase (TaKaRa Biomedicals), cDNA and $ddH_2O$ to a final volume of 50 µl. The PCR program used for all the EEV isolates consisted of one cycle of 4 min at 94°C, followed by 5 cycles of 1 min at 94°C, 45 sec at 52°C and 1 min at 72°C, followed by another 25 cycles as above with the primer annealing temperature increased to 60°C, and a final cycle with the extension time increased to 5 min PCR was performed using a PTC-200 Peltier Thermal Cycler (MJ Research). A small volume of each reaction was analysed by 1% agarose gel electrophoresis (section 2.2.3.) and the remainder of the reactions stored at -20°C until use.

**Table 2.4.** Primers used for RT-PCR (supplied by GIBCO BRL Life Technologies)

| Name | Sequence[a] | Length | $T_m$ | | GC content |
|---|---|---|---|---|---|
| | | | (1 M $Na^+$) (Annealing) | (50mM $Na^+$) (PCR) | |
| EEVNS3BAM (forward) | 5'C**GGATCC**GTTAAGTTTCTGCGCC ATG3' | 26 | 78°C | 56°C | 53% |
| EEVNS3ECO (reverse) | 5'CG**GAATTC**GTAACACGTTTCCGC CACG3' | 27 | 79°C | 58°C | 55% |

[a] Restriction enzyme sites are indicated in bold type in the primer sequence and the first ATG of the coding region is underlined

### 2.2.5. Polymerase chain reaction product purification

PCR products were purified directly or from 1% agarose gels using the High Pure PCR Product Purification Kit (Boehringer Mannheim) according to the manufacturer's instructions.

The standard protocol for purifying the amplification products directly involved the following. Double-distilled $H_2O$ was added to the reaction to a final volume of 100 µl, after which 500 µl binding buffer (3 M guanidine-thiocyanate, 10 mM Tris-HCl, 5% ethanol, pH 6.6) was added. The High Pure filter tube and the collection tube

were combined, the sample pipetted into the upper reservoir and centrifuged for 30 - 60 sec at maximum speed in a standard table top centrifuge. The flow-through was discarded, 500 µl wash buffer (20 mM NaCl, 2 mM Tris-HCl, 80% ethanol, pH 7.5) was added and the sample centrifuged for 1 min at maximum speed. The wash step was repeated using 200 µl wash buffer. The filter tube was inserted into a 1.5 ml microfuge tube and the nucleic acids eluted by adding 50 µl $ddH_2O$ and centrifuging at maximum speed for 30 - 60 sec.

The standard protocol for purification of amplification products from a 1% agarose gel involved the following. The DNA band of interest was isolated by 1% agarose gel electrophoresis (section 2.2.3.). The desired band was cut from the gel and the gel slice placed in a 1.5 ml microfuge tube. The gel mass was determined and 300 µl of the binding buffer (3 M guanidine-thiocyanate, 10 mM Tris-HCl, 5% ethanol, pH 6.6) was added for every 100 mg agarose gel slice. The agarose gel slice was dissolved by vortexing 15 - 30 sec and incubating the suspension for 10 min at 56°C, vortexing every 2 - 3 min. After the agarose gel slice was completely dissolved, 150 µl isopropanol was added for every 100 mg agarose gel slice and vortexed. From here onwards, the standard protocol for purifying the amplification products directly, was followed, starting with the step where one High Pure filter tube was combined into one collection tube and the contents of the microfuge tube pipetted into the upper reservoir of the filter tube.

A small volume of each purified PCR product was analysed by 1% agarose gel electrophoresis (section 2.2.3.) and the remainder of the purified PCR product stored at -20°C until use

### 2.2.6. Preparation of competent *E. coli* cells – $CaCl_2$ method

The technique used was as described by Sambrook *et al.* (1989). Briefly, a 3 ml overnight culture of *E. coli* XL1-Blue cells was prepared in LB-medium (1% bacto-tryptone, 0.5% bacto-yeast extract, 1% NaCl, pH 7.4) containing the appropriate antibiotics or no antibiotics. Of this culture, 1 ml was used to inoculate 100 ml LB-medium and grown at 37°C with shaking to log phase. The $OD_{550}$ was determined using a Beckman Du®-64 Spectrophotometer. Readings were taken at 30 - 60 min intervals, and then shorter intervals as the OD approached 0.5. Upon reaching log phase ($OD_{550}$ = 0.45 - 0.5), the cells were transferred onto ice and 20 ml harvested by centrifugation at 1 830 x g for 5 min at 4°C in a Beckman J2-21 rotor. The cells were resuspended in half the original volume 50 mM $CaCl_2$ and centrifuged at 2 860 x g for 5 min at 4°C. The cells were resuspended in $1/20^{th}$ of the original volume 50 mM $CaCl_2$ and after 1 h on ice the cells were ready for transformation. The cells were used immediately or stored away in aliquots of 200 µl at -70°C until use.

### 2.2.7. Transformation of competent *E. coli* cells

The heat shock method (Sambrook *et al.*, 1989) was used to transform competent cells. A small volume of each of the plasmid vectors (1 µl) or half of each ligation reaction was added to 100 µl competent *E. coli* XL1-Blue cells and kept on ice for 30 min. After heat shock at 42°C for 90 sec, the cells were cooled on ice for 2 min. To this 0.8 - 1 ml LB-medium containing the appropriate antibiotics was added, and the cells were incubated at 37°C with shaking for 1 h. The cells were plated out in aliquots of 50 - 150 µl on 1.2% LB-agar plates containing

the appropriate antibiotic as well as 50 µl 2% (w/v) X-gal substrate and 10 µl 100 mM IPTG inducer to enable blue/white colour selection where appropriate. The plates were incubated overnight at 37°C.

### 2.2.8. Miniprep plasmid DNA isolation/Large scale plasmid purification

Plasmid DNA was isolated according to the alkaline lysis method originally described by Birnboim and Doly (1979). A standard mini-preparation procedure was followed as described in Sambrook *et al.* (1989). However, in cases where miniprep plasmid DNA was needed for sequencing, purification of plasmids or re-purification of previously purified plasmid DNA was performed using the High Pure Plasmid Isolation Kit (Boehringer Mannheim) according to standard protocols. Comments or values in brackets apply to the large scale plasmid purification procedure.

For a miniprep plasmid isolation, 5 ml LB-medium containing the appropriate antibiotic(s) was inoculated with a single bacterial colony containing a desired plasmid. For a large scale plasmid purification, 200 ml LB-medium containing the appropriate antibiotic(s) was inoculated with 0.2 - 1 ml of an overnight grown plasmid-containing culture. Cells were grown to saturation in a 37°C incubator with shaking for 16 - 24 h. Cells were harvested by centrifuging for 1 min in a microfuge (1 030 x g for 25 min (Beckman Model J2-21)). The cells were resuspended in 100 µl (10 ml) solution I (50 mM glucose, 10 mM EDTA-$Na_2.2H_20$, 25 mM Tris, pH 8). After 5 min at room temperature and 5 (15) min on ice, 200 µl (20 ml) solution II (0.2 N NaOH, 1% SDS) was added and the sample mixed. After another 5 min on ice, 150 µl (15 ml) 3 M NaAc, pH 4.8, was added. The sample was incubated on ice for 15 min (30 - 60 min) and centrifuged for 15 min at 13 400 x g (2 860 x g for 45 min at 4°C). The DNA in the supernatant was precipitated by adding 2 volumes of 96% ethanol and incubating at -20°C for at least 30 min (60 min or overnight). In the case of miniprep plasmid purification, the DNA precipitate was collected by centrifugation at high speed for 10 min and washed with 500 µl 70% ethanol. The sample was centrifuged again at high speed for 10 min, the pellet dried and resuspended in 25 µl $ddH_2O$. In the case of a large scale plasmid preparation, the DNA precipitate was collected by centrifugation (1 830 x g, 25 min) and the pellet was allowed to dry briefly. The pellet was resuspended in 8 ml $ddH_2O$, 4 ml 7.5 M ammonium acetate was added and the sample centrifuged at 2 860 x g for 25 min. The plasmid DNA in the supernatant was precipitated with 2 volumes of 96% ethanol and the precipitate collected by centrifugation at 4 110 x g for 30 min, washed with 70% ethanol, dried and resuspended in 200 µl $ddH_2O$. A small sample was analysed by 1% agarose gel electrophoresis (section 2.2.3.) and the remainder of the sample stored at -20°C until use.

### 2.2.9. Restriction enzyme digestion

Restriction enzyme digestion reactions were performed under the reaction conditions as recommended by the supplier. Restriction enzyme digestions consisted of 0.1 - 2 µg DNA, $1/10^{th}$ of the total volume 10 x buffer (Boehringer Mannheim), 5 - 10 U of enzyme/µg of DNA (Boehringer Mannheim) and $ddH_2O$ to a final volume of 15 - 20 µl. Reactions were incubated at 37°C (unless specified otherwise by the manufacturer) for 60 - 90 min.

### 2.2.10. Geneclean/Glassmilk procedure for isolating DNA fragments from agarose

DNA fragments were purified from agarose using the glassmilk procedure and the Geneclean™ II DNA Purification kit (BIO101, Inc.). The DNA bands were excised from the agarose gel, 2.5 volumes 6 M NaI stock solution was added, and the sample incubated at 45 - 55°C for 5 min. The glassmilk suspension was added (5 µl of glassmilk suspension per 5 µg or less DNA); the sample mixed and placed on ice for 5 - 15 min. The sample was centrifuged for 5 sec at high speed and the pellet washed three times by adding 500 µl NEW wash solution (NaCl, ethanol, water), resuspending, and spinning for 15 sec. The DNA was eluted by resuspending the pellet in $ddH_2O$ and incubating at 45 - 55°C for 2 - 3 min. The DNA in the supernatant was collected by centrifuging for 1 min. The elution step was repeated and a small sample was analysed by 1% agarose gel electrophoresis (section 2.2.3.) and the remainder of the sample stored at -20°C until use.

### 2.2.11. Ligation

The insert and vector fragments were added to the ligation reaction in a ratio of 2:1 or 3:1 as estimated by EtBr staining of the fragments in a 1% agarose gel (section 2.2.3). The reaction further contained $1/10^{th}$ of the final volume of 10 x ligation buffer (66 mM Tris-HCl, 50 mM $MgCl_2$, 10 mM DTT, 10 mM ATP, pH 7.5) (Boehringer Mannheim), 1 µl (5 U) T4 DNA ligase (Boehringer Mannheim) and $ddH_2O$ to a final volume of 10 - 20 µl. The reactions were incubated at 16°C for 16 h and stored at -20°C until use.

### 2.2.12. Automated sequencing

The sequencing method used is based on the dideoxynucleotide chain-termination method of Sanger *et al.* (1977). The ABI PRISM™ BigDye Terminator Cycle Sequencing Ready Reaction Kit (Perkin Elmer Applied Biosystems) was used for automated sequencing on the ABI PRISM 377 DNA Sequencer. The standard protocols provided for cycle sequencing and removal of excess dye terminators were followed with modifications where necessary. Sequence data obtained were analysed and evaluated through the use of computer software provided with the ABI 377 Sequencer on the Apple Macintosh computer, e.g. ABI PRISM Sequence Navigator and ABI PRISM Sequencing Analysis.

#### 2.2.12.1. Cycle sequencing

The concentration of the purified product (plasmid or PCR product) was determined by using a spectrophotometer (Beckman Du®-64 Spectrophotometer) as described in Sambrook *et al.* (1989), or by running an EtBr stained agarose gel, using a standard with known concentration and a dilution series of the sample with unknown concentration. The cycle sequencing reactions consisted of 30 - 90 ng DNA template (PCR product) or 250 ng DNA template (plasmid), forward primer and reverse primer (refer to Table 2.5.) (3.2 pmol of each), ready reaction terminator mix (contains dNTPs, ddNTPs, AmpliTaq® DNA polymerase FS) (4 µl), 5 x dilution buffer ($1/10^{th}$ of final volume)(if a 20 µl reaction is used) and $ddH_2O$ to a final volume of 10 - 20 µl. The cycle sequencing reactions were performed using the GeneAmp PCR System 9600 and the program used

consisted of 25 cycles of 10 sec at 96°C, 10 sec at 55°C and 4 min at 60°C. The cycle sequencing annealing temperature was lowered to 50°C for sequencing reactions with the internal primers.

**Table 2.5.** Primers used for cycle sequencing reactions (supplied by GIBCO BRL Life Technologies and BIOS)

| Name | Sequence[a,b] | Length | $T_m$ | | GC content |
|---|---|---|---|---|---|
| | | | (1 M $Na^+$) (Annealing) | (50 mM $Na^+$) (PCR) | |
| EEV Bryanston (EEV-1) (S1REF*) and all other EEV isolates (PCR products) | | | | | |
| EEVNS3BAM (forward) | 5'C**GGATCC**GTTAAGTTTCTGCGC CATG3' | 26 | 78°C | 56°C | 53% |
| EEVNS3ECO (reverse) | 5'CG**GAATTC**GTAACACGTTTCCG CCACG3' | 27 | 79°C | 58°C | 55% |
| All other EEV isolates (PCR products) | | | | | |
| EEVNS3INTF (internal forward) | 5'CGACSTTYTTGATGGCYATGAT GC3' | 24 | 74°C | 52°C | 50% |
| EEVNS3INR (internal reverse) | 5'CKTTTATCACAYTCRTCYGC3' | 20 | 66°C | 45°C | 45% |
| PBS.EEVB.S10 plasmid (EEV Bryanston (EEV-1) (S1REF*) S10 in pBS) | | | | | |
| M13-21 (forward) | 5'TGTAAAACGACGGCCAGT3' | 18 | NC[c] | 53.7°C | 50% |
| M13-RP (reverse) | 5'CAGGAAACAGCTATGACC3' | 18 | NC | 53.7°C | 50% |
| PFB.EEVB.S10 plasmid (EEV Bryanston (EEV-1) (S1REF*) S10 in pFASTBAC1) | | | | | |
| POLH (forward) | 5'TTCCGGATTATTCATACC3' | 18 | 59.9°C | 38.3°C | 39% |
| SV40 (reverse) | 5'CAGGTCGACTTCGGATCTCCTA GGCTC3' | 27 | 80.8°C | 59.2°C | 59% |

[a] Restriction enzyme sites are indicated in bold type in the primer sequence and the first ATG of the coding region is underlined

[b] Refer to Appendix A for description of single-letter codes (S, Y, and K)

[c] NC indicates that $T_m$ was not calculated

### 2.2.12.2. Precipitation of cycle sequencing products/Purifying extension products

Two methods were used namely the ethanol/sodium acetate precipitation method and purification using the DyeEx™ Spin Kit (Qiagen).

The standard protocol provided for the ethanol/sodium acetate method was followed and involved the following. Double-distilled $H_2O$ was added to the cycle sequencing reaction to a final volume of 20 µl, as well as. This was followed by the addition of 2 µl of 3 M NaAc, pH 4.6 and 50 µl 99.9% ethanol. The reaction was mixed and incubated on ice for 15 min. The reaction was centrifuged for 30 min at maximum speed, the supernatant removed and the pellet washed by adding 250 µl 70% ethanol. The sample was centrifuged for 20 - 30 min at maximum speed, the supernatant removed and the pellet air-dried for approximately 30 min, or dried under vacuum in a vacuum centrifuge (Savant) for 10 - 15 min. The precipitated product was stored in the dark at -20°C until use for up to one week.

The standard protocol provided with the DyeEx™ Spin Kit (Qiagen) recommended for sequence analysis using an ABI PRISM 377 sequencer was followed and it involved the following. The spin column was vortexed, the bottom closure snapped off, and placed in the provided collection tube. The column was centrifuged for 3 min at 16 100 x g and the spin column was transferred to a microfuge tube. Double-distilled $H_2O$ was added to the cycle sequencing reaction to a final volume of 20 µl and the sequencing reaction was applied to the gel bed. The column was centrifuged for 3 min at 16 100 x g and the spin column removed from the microfuge tube. The eluate contained the purified DNA. The sample was dried in a heat block or thermal cycler at 90°C for 1 min. The precipitated product was stored in the dark at -20°C until use for up to one week.

### 2.2.13. Phylogenetic analysis

The EEV S10 nucleotide sequences were translated into NS3 amino acid sequences using ExPASy (Appendix B) and both the S10 and NS3 sequences were aligned with either S10 or NS3 sequences of other orbiviruses retrieved from GenBank (refer to Appendix C for accession numbers) using ClustalX 1.81 (Thompson *et al.*, 1997). The aligned S10 and NS3 sequences were used to generate a table of pairwise distances (PAUP version 4.0b8) to evaluate the variation within EEV S10 and EEV NS3. In addition, a comparison was made between the variation found within EEV S10/NS3 and the variation found within the cognate genes/proteins of other orbiviruses. The phylogenies of EEV S10 and EEV NS3 were investigated with the construction of phylogenetic trees using the neighbour-joining method (PAUP version 4.0b8, Sinauer Associates) and the maximum parsimony method (MEGA version 2.1., Kumar *et al.*, 2001). Trees were rooted using Broadhaven virus S10 or NS3 sequences and bootstrap analysis was performed (1000 replicates for neighbour-joining, 500 replicates for maximum parsimony). Confidence scores are shown on the branches of the phylogenetic trees.

## 2.3. RESULTS

The results discussed in this chapter are focussed on obtaining information about genome segment 10 of EEV and its gene product NS3. To achieve these aims, segment 10 of a number of EEV isolates were sequenced. Different structural features were identified and the amount of variation seen in S10/NS3 of EEV was determined. Finally, the observed structural features and variation of EEV NS3 was compared to that of other orbiviruses.

### 2.3.1. Cloning of EEV Bryanston (EEV-1) (S1REF*) dsRNA of genome segment 10

In order to clone the NS3 gene of EEV, a RT-PCR copy of S10 of EEV Bryanston (EEV-1) (S1REF*) was cloned into two plasmid vectors, pFASTBAC1 (Gibco BRL Life Technologies) and pBS (Stratagene) (refer to Appendix A for plasmid maps). pFASTBAC1 is an expression vector that is used as the donor plasmid in the in the BAC-TO-BAC™ Baculovirus Expression System to construct a recombinant baculovirus for protein expression (refer to Chapter 3), while pBS is primarily used for determining the nucleotide sequence of cloned genes.

EEV Bryanston (EEV-1) (S1REF*) total RNA was isolated from infected BHK cells, using the phenol-SDS extraction method (Bremer, 1976) or the TRIZOL® reagent (Life Technologies) according to the manufacturer's instructions. Double-stranded RNA was then isolated from the total RNA (section 2.2.2.). The necessary precautions to prevent RNA degradation were followed at all times (Sambrook *et al.*, 1989; Farell 1993). The dsRNA was used for RT-PCR (section 2.2.4.) to obtain material for sequence determination and for cloning a cDNA copy of the S10 dsRNA segment. The results are shown in Fig. 2.1. The RT-PCR product has an estimated size of approximately 732 bp, which corresponds to what was previously reported for EEV Cascara (EEV-2) (Viljoen and Huismans, 1989) where S10 was estimated to be approximately 710 bp in size.

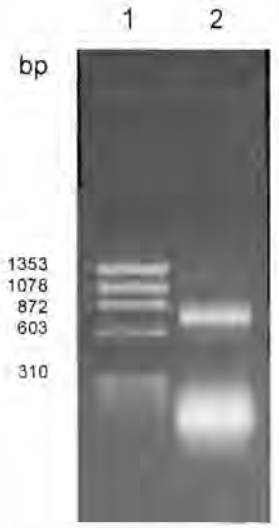


**Figure 2.1.** 1% Agarose gel electrophoresis of the RT-PCR product of EEV Bryanston (EEV-1) (S1REF*) (2) with a suitable size marker (φX 174) (1).

The oligonucleotide primers used for RT-PCR were designed using unpublished sequence data for 30 bases of the 5' and 3' terminal ends of the coding strand of EEV Kyalami (EEV-5) S10 as provided by Mr. A. C. Potgieter from the Onderstepoort Veterinary Institute (OVI), Onderstepoort, South Africa. The two synthetic 26- and 27-mer oligonucleotide primers, EEVNS3BAM and EEVNS3ECO (Table 2.4.), each consisted of 19 bp which were identical to the first 19 and complementary to the last 19 nucleotides of the provided terminal sequences respectively, preceded by a *Bam*HI site or *Eco*RI site

to facilitate the cloning of the RT-PCR product, and 1 or 2 additional C or G nucleotides to stabilize their 5' ends. The start codon of the coding region is found in EEVNS3BAM.

The RT-PCR product of EEV Bryanston (EEV-1) (S1REF*) S10 was purified (section 2.2.5.), sequenced (section 2.2.12.) and cloned into plasmid vectors pBS and pFASTBAC1, using the *Bam*HI and *Eco*RI restriction enzyme sites, as described in the methods (sections 2.2.7. and 2.2.9 to 2.2.11.). Numerous putative recombinant colonies were identified from each cloning based on their antibiotic resistance (pFASTBAC1 confers ampicillin resistance) or colour (putative recombinant pBS colonies are white) and miniprep plasmid DNA isolations were performed for selected colonies (section 2.2.8.). The putative recombinant plasmids were characterized by restriction enzyme digestion with *Bam*HI and *Eco*RI and successful ligation of the insert was confirmed by the excision of a band of the expected size. The results of the restriction enzyme digestion reactions can be seen in Fig. 2.2.

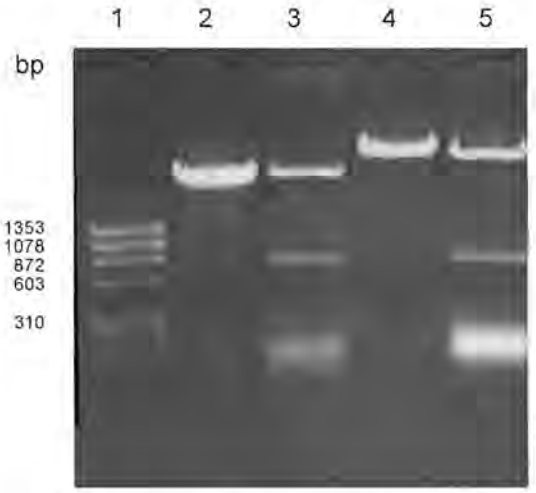


**Figure 2.2.** 1% Agarose gel electrophoresis of the putative recombinant plasmids after restriction enzyme digestion with *Bam*HI and *Eco*RI. Linearized pBS (2) and pFASTBAC1 (4) were included as controls to compare with the putative recombinant plasmids (3 and 5). φX 174 DNA size marker (1) was included as a control to estimate the sizes of the excised inserts.

A recombinant of each plasmid was selected and named: pFB.EEVB.S10 (pFASTBAC recombinant) and pBS.EEVB.S10 (pBS recombinant), and the remainder of the miniprep plasmid DNA samples purified for sequencing (section 2.2.5.). The plasmids were sequenced (section 2.2.12.) and the sequencing results of the recombinant plasmids were compared to the sequencing results of the RT-PCR product. The results obtained indicated that the recombinant plasmids contained the correct sequence, no errors had been introduced during cloning that would change the amino acid sequence of the final protein product.

Large scale plasmid DNA isolation was performed for each of the two recombinant plasmids using the NUCLEOBOND® kit (Machery-Nagel) according to the manufacturer's instructions in order to

generate a sufficient amount of each recombinant plasmid for further experiments such as the construction of a recombinant baculovirus (pFB.EEVB.S10).

### 2.3.2. Sequence analysis and evaluation of EEV Bryanston (EEV-1) (S1REF*)

The aim of the results in this section was to determine the nucleotide sequence of a RT-PCR copy of S10 of EEV Bryanston (EEV-1) (S1REF*), as well as the nucleotide sequence of the insert contained in the recombinant pFASTBAC1 and pBS plasmids.

Automated sequencing of the RT-PCR product of S10 of EEV Bryanston (EEV-1) (S1REF*), as well as pBS.EEVB.S10 and pFB.EEVB.S10 containing S10 of EEV Bryanston (EEV-1) (S1REF*), was performed, and the obtained sequence data analysed and evaluated, as described in section 2.2.12.

A consensus nucleotide sequence for EEV Bryanston (EEV-1) (S1REF*) S10 was compiled using the sequences obtained from the RT-PCR product and the plasmids and is shown in Fig. 2.3. In cases where there were differences between the sequences, the nucleotide that occurred the most frequently was taken as the correct nucleotide for that position. The consensus nucleotide sequence of EEV Bryanston (EEV-1) (S1REF*) S10 was translated using ExPASy (Appendix B) and the result is shown in Fig. 2.4.

gttaagtttctgcgcc**ATG**TATCCGGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGAGCGTGCGCTA**ATG**GTGTACCCGCCAACAGCC
CCAATGCCGCCTGTTATGACCCGGAATAATCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTGGACAAGAATA
TAACTAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTCGGTAGCCGAATCAGCTGCGGATAGCCCAATGGT
ACGCACTATTAAAATCCAGATATATAATAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTTGAGA
TTTATCTCGTACACCTTTATAACGTTAATGTTAACATCGACCTTCTTGATGGCTATGATGCAAACTCCACCAATAACACAGTATGTGGAGA
TGGCGTGTAATAAAACGGGGAACGCAGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTTGATAAT
GAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACACTTTCCACGAACGCGGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCG
TATATCGACGCAGCTAGATCAAATCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTTCGGGGATA
CGGCCCAT**TAG**ccgtggcggaaacgtgttac

**Figure 2.3.** Nucleotide sequence of EEV Bryanston (EEV-1) (S1REF*) S10. The sequence is a consensus sequence compiled using the sequences generated for both the RT-PCR product of EEV Bryanston (EEV-1) (S1REF*) S10 and the two recombinant plasmids, pBS.EEVB.S10 and pFB.EEVB.S10, containing a copy of EEV Bryanston (EEV-1) (S1REF*) S10. This consensus sequence was used in phylogenetic analysis and is the S1REF* sequence that was submitted to GenBank and received accession number AY115864. Non-coding sequences are shown in lowercase, start and stop codons in bold type and the primer sequences are underlined. (759 basepairs)

MYPVLSRTVVNNPEERALMVYPPTAPMPPVMTRNNLKIDSVDGMKDLALNILDKNITSTTGADECDKREKAMFASVAESAADSPMVRTIKI
QIYNRVLDDMEREKQKCEKRRAVLRFISYTFITLMLTSTFLMAMMQTPPITQYVEMACNKTGNAEKNDPCGLMRWSGAVQFLTLIMSGFLY
MCKRWITTLSTNADRISKNILKRRAYIDAARSNPDATVLTVTGGNTGDLPYQFGDTAH

**Figure 2.4.** Deduced amino acid sequence of EEV Bryanston (EEV-1) (S1REF*) NS3. The sequence is a translation of the consensus sequence of EEV Bryanston (EEV-1) (S1REF*) S10 compiled using the sequences generated for both the RT-PCR product of EEV Bryanston (EEV-1) (S1REF*) S10 and the two recombinant plasmids, pBS.EEVB.S10 and pFB.EEVB.S10, containing a copy of EEV Bryanston (EEV-1) (S1REF*) S10. (240 amino acids)

The amino acid sequence was used in a BLAST search of GenBank at NCBI (BLASTP version 2.0.14) (Altschul *et al.*, 1997) and sequences producing significant alignments included different serotypes of BTV non-structural protein P8 or non-structural protein NS3 (containing non-structural protein NS3A), different serotypes of AHSV non-structural protein NS3 (containing non-structural protein NS3A), Broadhaven virus non-structural protein NS3, EHDV non-structural protein NS3, as well as non-structural protein NS3 of a Palyam virus, Chuzan. Similar results were obtained using other tools (refer to Appendix B). This confirmed that the sequence that was amplified by RT-PCR of the dsRNA of EEV Bryanston (EEV-1) (S1REF*), and cloned into pBS and pFASTBAC1, was segment 10 encoding NS3. The next step is to analyse the complete nucleotide sequence of EEV Bryanston (EEV-1) (S1REF*) S10, as well as the deduced amino acid sequence of the gene product NS3.

### 2.3.3. Analysis of EEV Bryanston (EEV-1) (S1REF*) genome segment 10 and its encoded gene product, NS3

In this section analysis of the nucleotide sequence (S10) (Fig. 2.3.) and the derived amino acid sequence (NS3) (Fig. 2.4.) of EEV Bryanston (EEV-1) (S1REF*) was carried out in order to identify characteristic features of EEV NS3 for comparison to other orbivirus NS3 proteins.

Some of the basic characteristics of the EEV NS3 protein are summarized in Tables 2.6. and 2.7. The molecular weight of NS3 was also estimated using another suite of programs namely AnTheProt version 5 (Geourjon and Deléage, 1995) and the estimated value calculated using this program, corresponded well with the estimate of 27 kDa. The pI was also estimated using other tools at ExPASy and AnTheProt version 5, and in both cases the same value, 8.93, was found for NS3.

**Table 2.6.** Some basic characteristics of EEV Bryanston (EEV-1) (S1REF*) NS3 (and NS3A) determined using the ProtParam tool at Expasy (Appendix B)

| Characteristic | Value/Result |
|---|---|
| Number of amino acids | 240 amino acid residues (NS3); 222 amino acid residues (NS3A) |
| Molecular weight | 26 989 Da ≈ 27 kDa (NS3); 24 918 Da ≈ 25 kDa (NS3A) |
| pI | 8.93 (NS3); 8.94 (NS3A) |
| Total number of charged amino acid residues | total number of negatively charged residues (D + E) = 25 (NS3), 23 (NS3A); total number of positively charged residues (R + K) = 30 (NS3), 28 (NS3A) |

**Table 2.7.** Amino acid composition of EEV Bryanston (EEV-1) (S1REF*) NS3 protein determined using the ProtParam tool at Expasy (Appendix B)

| Amino acid | Number of amino acids in sequence | Percentage of amino acids of total | Amino acid | Number of amino acids in sequence | Percentage of amino acids of total |
|---|---|---|---|---|---|
| A (Alanine) | 20 | 8.3% | M (Methionine) | 16 | 6.7% |
| C (Cysteine) | 5 | 2.1% | N (Asparagine) | 14 | 5.8% |
| D (Aspartic acid) | 15 | 6.2% | P (Proline) | 13 | 5.4% |
| E (Glutamic acid) | 10 | 4.2% | Q (Glutamine) | 6 | 2.5% |
| F (Phenylalanine) | 7 | 2.9% | R (Arginine) | 16 | 6.7% |
| G (Glycine) | 10 | 4.2% | S (Serine) | 13 | 5.4% |
| H (Histidine) | 1 | 0.4% | T (Threonine) | 23 | 9.6% |
| I (Isoleucine) | 14 | 5.8% | V (Valine) | 14 | 5.8% |
| K (Lysine) | 14 | 5.8% | W (Tryptophan) | 2 | 0.8% |
| L (Leucine) | 19 | 7.5% | Y (Tyrosine) | 8 | 3.3% |

2.3.3.1. The identification of structural features of EEV Bryanston (EEV-1) (S1REF*) NS3

A number of functional features can be predicted directly from a protein sequence, these features include hydrophobicity profiles, transmembrane helices, and small sequence motifs which cells use to target proteins to particular cellular compartments.

The hydrophobicity profile (Kyte and Doolittle, 1982) of EEV Bryanston (EEV-1) (S1REF*) NS3 was determined and two hydrophobic regions were identified in the NS3 protein sequence at amino acid residues 115 - 136 (HDI) and 169 - 183 (HDII) (Fig. 2.5.). Within these two hydrophobic regions, two transmembrane helices were identified at amino acid residues 120 - 132 and 171 - 178 (Fig. 2.6.).

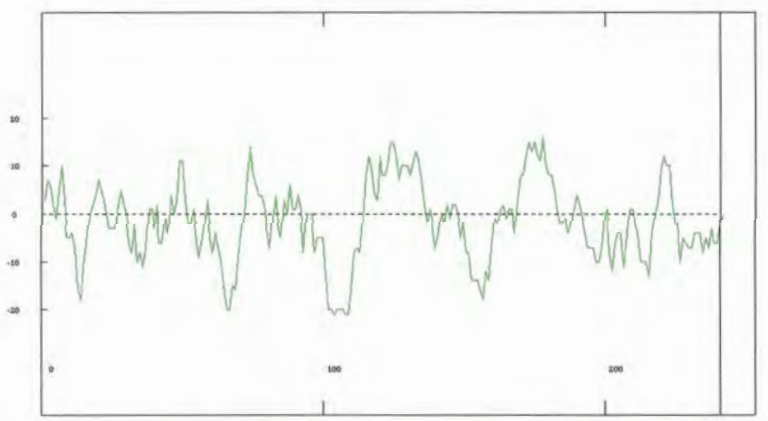

**Figure 2.5.** The hydrophobicity profile (Kyte and Doolittle, 1982) for EEV Bryanston (EEV-1) (S1REF*) NS3 protein determined using AnTheProt version 5 (Geourjon and Deléage, 1995).

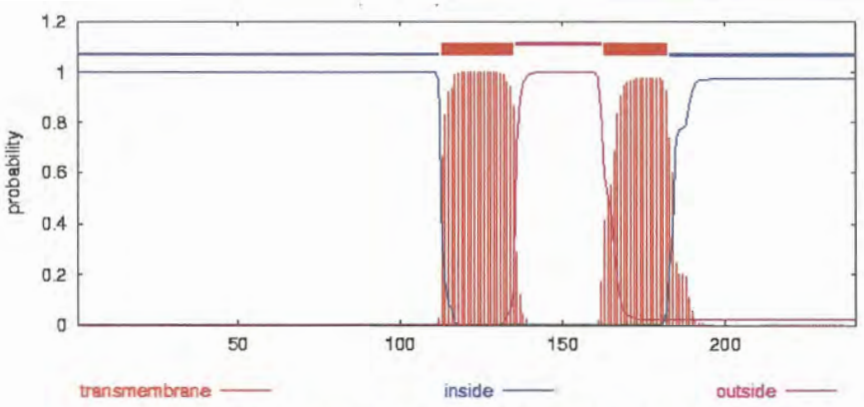


**Figure 2.6.** Transmembrane helices identified in the EEV Bryanston (EEV-1) (S1REF*) NS3 protein using TMHMM (Sonnhammer *et al.*, 1998). The figure shows the orientation of the protein in relation to the membrane.

The NS3 amino acid sequence of EEV Bryanston (EEV-1) (S1REF*) were also analysed for the presence of transmembrane regions or helices using a number of other programs (refer to Appendix B). All the different programs used identified two transmembrane helices in the NS3 protein and gave a similar location and orientation for the two transmembrane helices.

Two possible models for the transmembrane topology of the EEV Bryanston (EEV-1) (S1REF*) NS3 protein were determined, a strongly preferred model consisting of two possible inside to outside transmembrane helices, as well as an alternative model consisting of two possible outside to inside transmembrane helices. Inside is defined as the cytoplasmic face, while outside is defined as the lumenal face of the membrane depending on the organelle. The strongly preferred model is shown in Table 2.8.

**Table 2.8.** A model for the transmembrane topology of EEV Bryanston (EEV-1) (S1REF*) NS3 protein as predicted by using TMPRED (Appendix B)

Model 1: Strongly preferred model: N-terminus inside. Two strong transmembrane helices, total score 3578.

| From | To | Length | Score | Orientation |
|---|---|---|---|---|
| 1 | 116 | 116 | NA | inside region 1 (loop in cytoplasm) |
| 117 | 136 | 20 | 1865 | inside ⇨outside (transmembrane segment) |
| 137 | 165 | 29 | NA | outside region (external loop) |
| 166 | 184 | 19 | 1713 | outside ⇨ inside (transmembrane segment) |
| 185 | 240 | 56 | NA | inside region 2 (loop in cytoplasm) |

NA indicates information not available

The presence of a coiled-coil region consisting of 21 residues was identified within EEV Bryanston (EEV-1) (S1REF*) NS3 and the results are shown in Fig. 2.7.

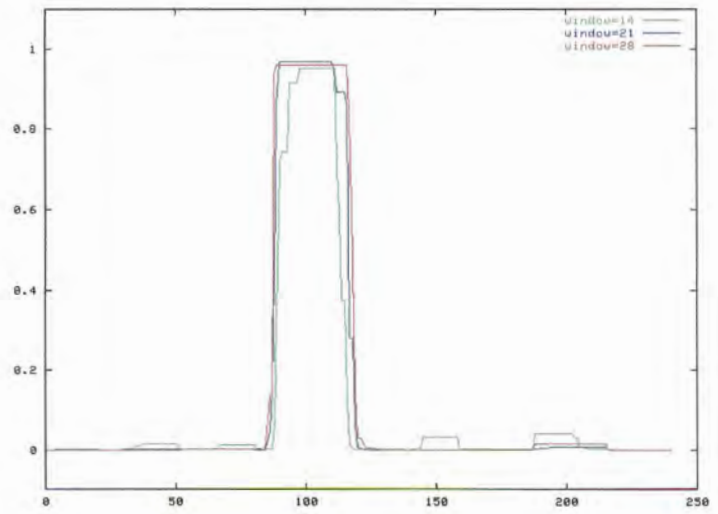


**Figure 2.7.** Coiled-coil regions identified in EEV Bryanston (EEV-1) (S1REF*) NS3 protein using COILS version 2.1. and 2.2. (Lupas, 1996). The default parameters, the MTK matrix and three window sizes namely 14, 21 and 28 were used to identify the coiled-coil domain.

Proteins can contain signals within their amino acid sequence, which assist in their processing within the cell, for e.g. leader sequences or signal peptide sequences, which target proteins to specific compartments within cells. The presence of leader sequences, their cleavage sites and protein localization can be predicted using various programs.

The presence of a leader sequence in EEV Bryanston (EEV-1) (S1REF*) NS3, with the most likely cleavage site between amino acid position 134 and 135 or amino acid position 133 and 134, was identified using SignalP (Nielson *et al.*, 1997.). The potential signal peptide sequence identified consisted of a positively charged n-region (position 110 to 114; lysine, arginine, arginine, alanine), followed by a hydrophobic h-region (position 115 to 136, a stretch of 22 amino acids) and a neutral but polar c-region (position 137 to 144, glutamine, threonine, proline, proline, isoleucine, threonine, glutamine, tyrosine). A potential cleavage site of a signal peptide was also identified at amino acid residue 134 in EEV Bryanston (EEV-1) (S1REF*) NS3 protein when using AnTheProt version 5 (Geourjon and Deléage, 1995). It is interesting to note that position 134 falls within hydrophobic region I (HDI, position 115 to 136) after the first identified transmembrane helix (position 120 to 132).

Protein localization was analysed using PSORT (Appendix B). It was predicted that EEV Bryanston (EEV-1) (S1REF*) NS3 is localized to the cytoplasm, lumen of the lysosome and the mitochondrial matrix space, it was however not clear whether it would be found in the membrane of the endoplasmic reticulum. The most likely subcellular location for EEV Bryanston (EEV-1) (S1REF*) NS3 as predicted by another program, TargetP version 1.0 (Emanuelsson *et al.*, 2000), was "other", i.e. the protein is not found in the mitochondria or secretory pathway.

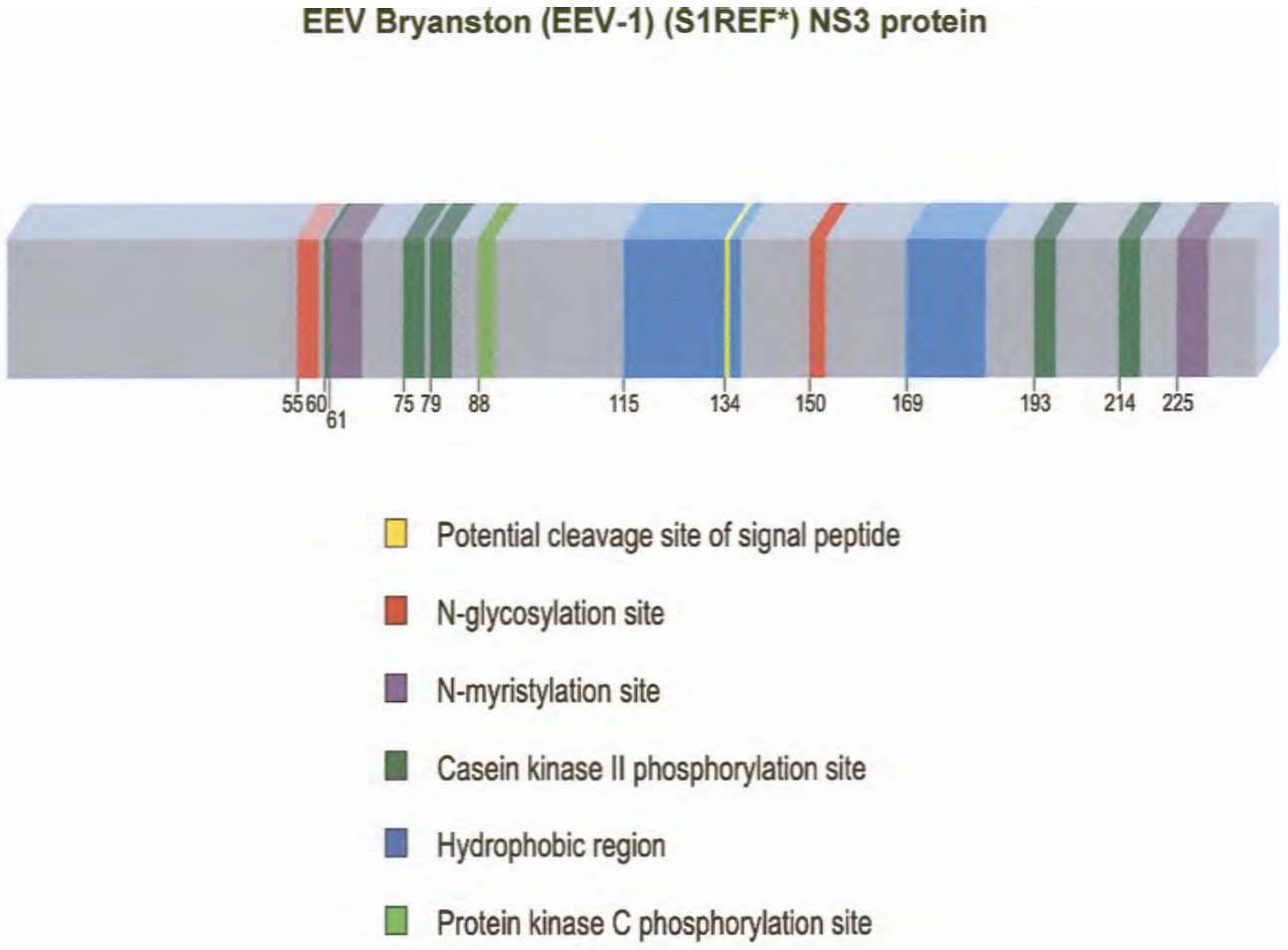


**Figure 2.8.** A summary of some of the different structural features or motifs that were identified for EEV Bryanston (EEV-1) (S1REF*) NS3 protein using different programs. The potential cleavage site of the signal peptide is indicated in yellow, the N-glycosylation sites are indicated in red, the N-myristylation sites are indicated in purple, the casein kinase II phosphorylation sites are indicated in dark green, the hydrophobic regions are indicated in blue and the protein kinase C phosphorylation site is indicated in light green. The first amino acid position of each of the sites or regions is indicated in the figure with the number of the first amino acid of the site or region.

In the preceding part of this section a number of structural features were identified in the NS3 protein sequence of EEV Bryanston (EEV-1) (S1REF*) and these features are summarized in Fig. 2.8. A number of structural features, including antigenic regions, proline-rich regions, conserved regions and the number and location of cysteine residues have however been identified previously for orbivirus NS3 proteins and the remainder of this section is therefore focused on identifying these structural features within the EEV NS3 protein.

The presence of charge clusters, charge runs and the number and spacing of cysteine residues in EEV Bryanston (EEV-1) (S1REF*) NS3 were analysed and the results are shown in Table 2.10.

**Table 2.10.** Results of a SAPS (Statistical Analysis of Protein Sequences) (version of April 11, 1996) (Brendel *et al.*, 1992) analysis of EEV Bryanston (EEV-1) (S1REF*) NS3 sequence

| Sequence property | Result |
|---|---|
| Charge clusters | Mixed charge cluster found from 87 – 116 (before transmembrane region), size 30, 10 positively charged residues (K and R), 5 negatively charged residues (E and D), 15 uncharged residues |
| Charge runs | Positive run found at 110 and 204 |
| Spacing of C residues | H2N-60-**C**-37-**C**-35-**C**-6-**C**-17-**C**-51-COOH; 5 cysteine residues |

A number of physico-chemical profiles (Fig. 2.9.) were determined for EEV Bryanston (EEV-1) (S1REF*) NS3 of which the first was the combined antigenicity profile (Parker *et al.*, 1986). Antigenic regions (high scoring residues) were found at amino acid position 51 to 67, 97 to 108 (before HDI), and 146 to 158 (between HDI and HDII). This profile and scores at the specific residues were compared to the other profiles seen in Fig.2.9. High scoring regions for the combined antigenicity profile generally corresponded to high scoring regions for the Welling antigenicity profile, the hydrophilicity profile (Hopp and Woods, 1981) and the solvent accessibility profile (Boger *et al.*, 1986), while low scoring regions for the combined antigenicity profile generally corresponded to high scoring regions for the hydrophobicity profile and helical transmembrane regions profile. High scoring regions in the hydrophobicity profile indicate regions probably embedded in the membrane which are not exposed to the cellular environment, while high scoring regions in the hydrophilicity profile probably identify regions outside the membrane that are exposed to the cytoplasm. The external or outside region identified in predictions of the membrane topology of EEV NS3 (Table 2.10, Fig. 2.8.) at position 137 to 165 (between HDI and HDII) generally corresponded to regions with high scores in the combined antigenicity profile (position 146 to 158) and hydrophilicity profile (position 148 to 160), indicating that this region is exposed extracellularly and may be involved in eliciting an immune response against the EEV NS3 protein.

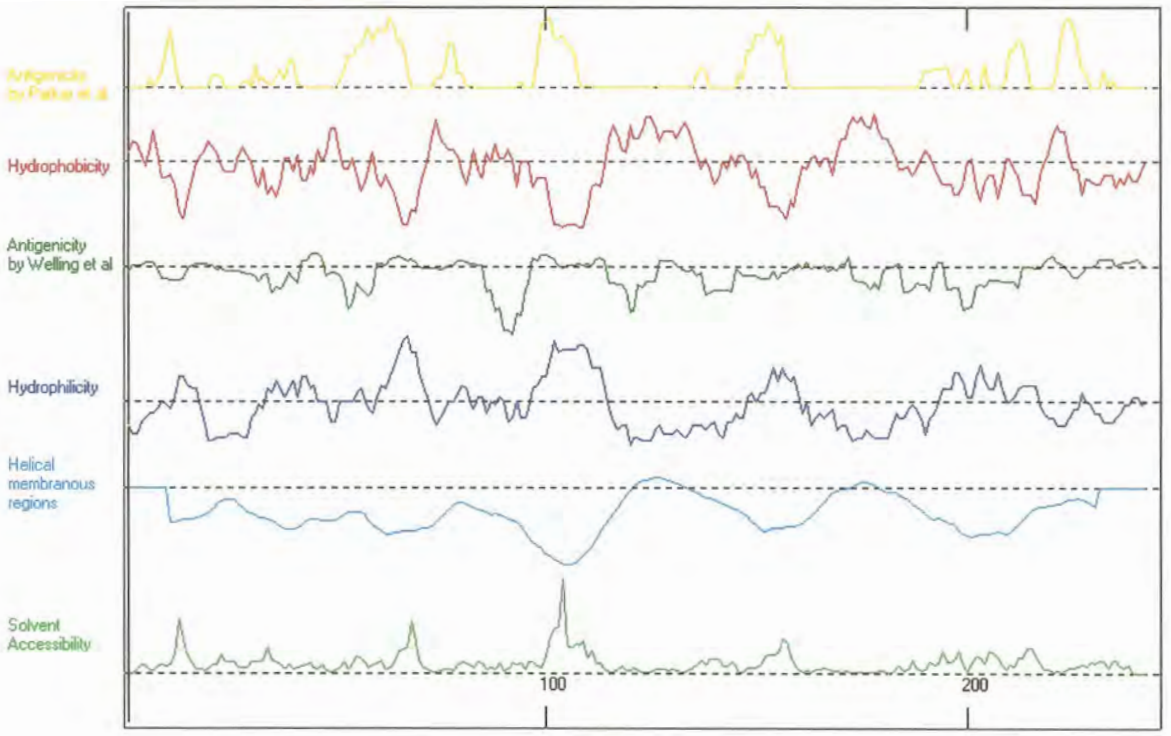


**Figure 2.9.** The different physico-chemical profiles that were determined for EEV Bryanston (EEV-1) (S1REF*) NS3 protein using AnTheProt version 5 (Geourjon and Deléage, 1995). The combined antigenicity profile is shown in yellow, the hydrophobicity profile is shown in red, the Welling antigenicity profile is shown in dark green, the hydrophilicity profile is shown in blue, the helical membranous regions profile is shown in turquoise and the solvent accessibility profile is shown in green.

The last structural feature that was identified in the EEV Bryanston (EEV-1) (S1REF*) NS3 protein was a proline-rich region. A total of 13 proline residues were identified in the protein sequence, 6 of these proline residues were located in the amino acid region from position 13 to 29 (a stretch of 17 residues), while 5 of these proline residues were located in the amino acid region from position 22 to 29 (a stretch of 8 residues). The proline-rich region (position 22 to 29) was located in a region of the EEV NS3 protein that forms a loop in the cytoplasm (region 1 to 116, Table 2.8., Fig. 2.6.). This region precedes the first identified transmembrane region at position 120 to 132. Similar proline-rich regions have been identified in other orbivirus NS3 proteins.

### 2.3.3.2. Protein structure prediction of EEV Bryanston (EEV-1) (S1REF*) NS3

Secondary structure predictions were performed by AnTheProt version 5 (Geourjon and Deléage, 1995) using the Double Prediction Method (Deléage and Roux, 1987). This method predicted that EEV Bryanston (EEV-1) (S1REF*) NS3 consists of 34% helix, 29% sheet, 29% coil and 7% turn. Overall the protein can be placed in a class where proteins contain a mixture of both $\alpha$-helices and $\beta$-sheets.

Secondary structure prediction was also performed using PHDsec (Rost *et al.*, 1994) and the results indicated that EEV NS3 was a mixed protein consisting of 53.8% helix, 8.8% extended (sheet) and 37.5% other (loop). It was a combination of α-helices and β-sheets, but did not conform to any of the 3 known classes of alpha/beta combinations. Protein secondary structure was also predicted using SSpro (Baldi *et al.*, 1999) and two long regions containing helix-forming residues were identified at amino acid position 85 to 139 and 171 to 210. These regions, although bigger, correspond to the hydrophobic or transmembrane regions predicted by other programs.

### 2.3.4. Isolation of dsRNA, RT-PCR and PCR product purification of the NS3 gene of a number of EEV reference strains and field isolates

In this section, the aim was to obtain a purified RT-PCR product of the EEV genome segment 10 of a number of reference strains, as well as a number of field isolates, for sequencing, and to then compare the nucleotide sequences and derived amino acid sequences of the different isolates to determine the levels of variation between the different serotypes on nucleotide and amino acid levels.

Double-stranded RNA was isolated from all 15 EEV isolates received (7 reference strains and 7 field isolates, Table 2.1. to 2.3.) as described in section 2.2.2. following the necessary precautions to prevent RNA degradation (Sambrook *et al.*, 1989; Farell 1993). The dsRNA profiles of the seven different EEV serotypes, as well as the dsRNA profiles of a number of other orbiviruses (AHSV-9, BTV-10 and EHDV-1) are shown in Fig. 2.10.

The seven different serotypes (reference strains) all showed similar dsRNA profiles. The same observation was made for the seven field isolates, and there were also no visible differences between the dsRNA profiles of a reference strain and field isolates of the same serotype (results not shown). There seemed to be at least 3 size classes of dsRNA segments namely large, medium and small segments. Only nine dsRNA segments were visible on a 1% agarose gel and it was assumed that genome segments 7 and 8 of the EEV isolates co-migrated.

In the case of the other orbivirus dsRNA profiles (Fig. 2.10B.), the dsRNA of the AHSV sample separated to form 8 bands where segment 7, 8, and 9 co-migrated, the dsRNA of the BTV-10 sample separated to form 8 bands where segment 2 and 3 co-migrated, as well as segment 7 and 8, and a total of 8 bands was observed for the EHDV-1 sample where genome segments 7, 8 and 9 co-migrated.

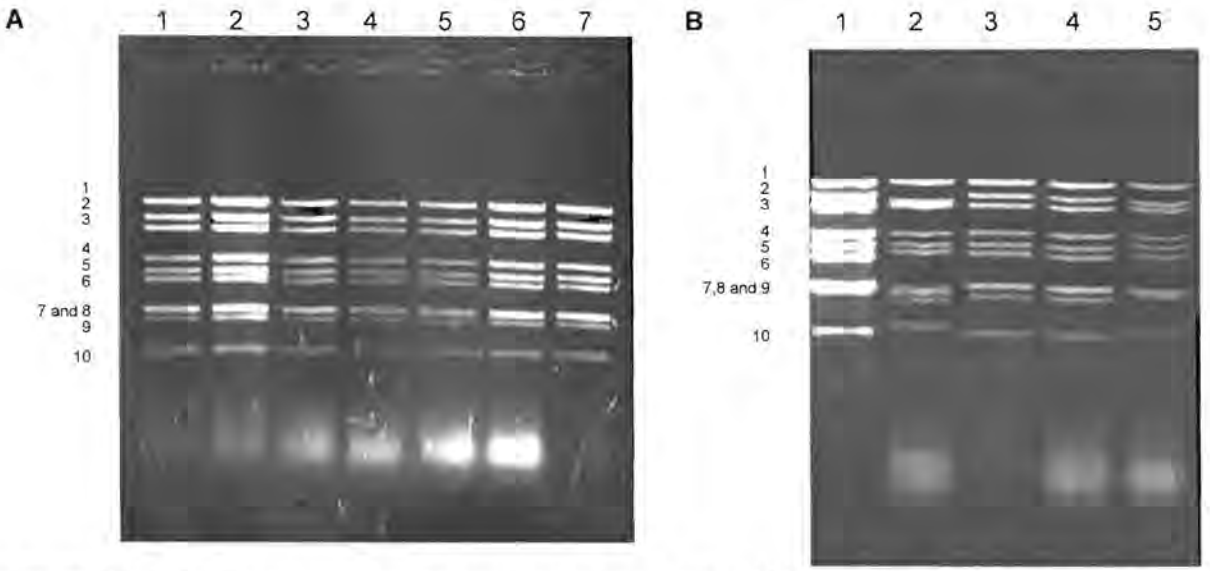


**Figure 2.10.** 1% Agarose gel electrophoresis to show the dsRNA profiles of the 7 different EEV serotypes (reference strains) as well as the dsRNA profiles of a number of orbiviruses compared to that of an EEV reference strain and an EEV field isolate. The seven EEV serotypes (A) are S1REF (1), S6REF (2), S4REF (3), S5REF (4), S7REF (5), S2REF (6) and S3REF (7). In B the dsRNA profiles of S1FLD3 (3) and S2REF (4) are compared to AHSV-9 (1), BTV-10 (2) and EHDV-1 (5). The genome segments were numbered 1 to 10 in order of decreasing size.

Although there are differences between the orbivirus profiles, the EEV dsRNA profiles seem to be more similar to that of the AHSV isolates than to the dsRNA profiles of EHDV or BTV. This observation correlates with what was previously observed (Viljoen and Huismans, 1989; Bremer *et al.*, 1990). The sizes of cognate AHSV and EEV genome segments seemed to be similar, with the exception of segments 7, 8 and 9 which co-migrated differently in the two profiles, another exception was segment 6. In EEV this segment migrated slightly slower than the cognate segment of AHSV, most likely indicating a size difference where the segment of EEV was slightly bigger than that seen for AHSV. Accurate comparisons of the sizes of the different segments can however only be made after the determination of the nucleotide sequences and number of basepairs of each segment.

From the results seen in Fig 2.10. it is clear that EEV has a specific dsRNA profile irrespective of the serotype involved, confirming what was previously seen by Viljoen and Huismans in 1989. The EEV dsRNA profile is also different from the dsRNA profiles observed for other orbiviruses such as BTV, AHSV and EHDV under the same electrophoretic conditions. These differences could possibly allow a preliminary identification of the orbivirus in question based on its dsRNA profile on a 1% agarose gel. Another important conclusion that can be drawn, based on the similarity of the EEV profiles, irrespective of serotype involved, is that no estimation of variation between or within serotypes is possible using the dsRNA profiles on an agarose gel as a guide. Another way in which variation can however be studied is by determining the nucleotide sequence of the individual dsRNA segments and to compare the sequences of the same segments between different serotypes and within the same serotype.

RT-PCR was performed on the reference strains and field isolates as described in section 2.2.4. All the reference strains and field isolates produced a RT-PCR product corresponding to the expected size of approximately 732 bp. After purification of the RT-PCR products (section 2.2.5.), the nucleotide sequences of the RT-PCR products of each of the reference strains and field isolates were determined (section 2.2.12.) and the sequencing results are shown in the next section.

### 2.3.5. Sequence evaluation and analysis of a number of EEV reference strains and field isolates

The aim of the results shown in this section was to determine the nucleotide sequence of EEV NS3 genes of 7 different serotypes and field isolates, as well as their deduced amino acid sequences, in order to compare the structural features of the segment 10 sequences and deduced amino acid sequences of EEV to that seen in other orbiviruses.

Automated sequencing of the RT-PCR products of the seven reference strains and the seven field isolates were performed as described in section 2.2.12. Forward and reverse primers, as well as an internal forward and an internal reverse primer (Table 2.6.) were used since the terminal ends of the S10 sequences could only be obtained when using internal primers sequencing towards the 5' and 3' ends. The degenerate internal primers were designed based on a sequence alignment of preliminary sequence information of S10 and NS3 of the 7 EEV reference strains to target conserved regions or conserved protein motifs within the sequence that were located relatively close to the 5' or 3' terminal end of the sequences. The primers targeted the myristylation site (nucleotide position 200 to 219, amino acid position 61 to 66) which partly overlapped a casein kinase II phosphorylation site (nucleotide position 194 to 205, amino acid position 60 to 63), as well as a conserved region seen in the nucleotide alignment between nucleotide position 402 and 425. The nucleotide sequence data obtained was analysed and evaluated as described in section 2.2.12. and consensus sequences were compiled for each of the seven reference strains and the seven field isolates. The complete sequences were submitted to GenBank and the accession numbers are as shown in Appendix C. The nucleotide sequences of the seven reference strains and seven field isolates were translated using ExPASy (Appendix B) to generate amino acid sequences.

#### 2.3.5.1. Nucleotide sequence analysis of a number of EEV reference strains and field isolates

The nucleotide sequences of the seven reference strains and the seven field isolates were compared to each other using sequence alignment. The aim of the sequence alignment was to attempt to align two or more nucleotide sequences in such a way that regions of structural or functional similarity between the molecules were highlighted. Another function of sequence alignment is however also to show which regions differ between sequences, giving an indication of the variation found within a set of sequences. ClustalX version 1.81 (Thompson *et al.*, 1997) was used to perform the sequence

alignments and the result of the comparison between the seven reference strains and the seven field isolates is shown in Fig. 2.11.

A number of interesting observations can be made from the nucleotide alignment in Fig. 2.11. Firstly, there are two conserved in-frame putative initiation codons at nucleotide position 17 to 19 and position 71 to 73, and the contexts in which they are found is shown in Table 2.11. The first is expected to initiate the synthesis of NS3 and the second the synthesis of NS3A as has been seen for other orbiviruses such as AHSV and BTV.

**Table 2.11.** Contexts surrounding the first and second AUGs of the seven different EEV reference strains

| Virus | First AUG codon[a] | | Second AUG codon | |
|---|---|---|---|---|
| | Position | Sequence | Position | Sequence |
| EEV[b] | 17-19 | **G**CCATG**T** | 71-73 | **C**TAATG**G** |

[a] Sequences representing the putative initiation codons are underlined, and nucleotides at positions –3 and +4 relative to the AUG are in bold type

[b] Context surrounding the first and second initiation codon was conserved for all seven EEV reference strains as well as for the seven field isolates and EEV Bryanston (EEV-1) (S1REF*)

Secondly, the conserved stop codon for both ORFs is located at nucleotide position 737 to 739, and the 5' untranslated region consists of 16 bases, while the 3' untranslated region consists of 20 bases. The 5' terminal sequence and the 3' terminal sequence found in all the sequences are similar due to the RT-PCR strategy used to obtain the S10 sequences. The terminal sequences correspond to the primer sequences, which are based on unpublished sequence data of EEV Kyalami (EEV-5) S10. This means that all the terminal sequences for all the isolates are EEV Kyalami (EEV-5) sequence data and no analysis of the terminal sequences of S10 of other EEV serotypes can be made in this investigation. A recent study (Potgieter *et al.*, 2002) has however shown that the 5' terminal sequence 5' GTTAAG 3' and the 3' terminal sequence 5' TGTTAC 3' found in all the sequences as part of the primer sequences, are the 5' and 3' terminal sequences of the complete S10 genome segment. This means that sequences obtained using the primers based on the received unpublished sequence data of EEV Kyalami (EEV-5) S10 will be full-length S10 sequences.

CLUSTAL X (1.81) multiple nucleotide sequence alignment of EEV serotypes (S1REF to S7REF), seven EEV field isolates and EEV Bryanston (EEV-1) (S1REF*)

```
                    10        20        30        40        50        60
S1REF*      gttaagtttctgcgccATGTATCCGGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S1REF       gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S1FLD4      gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S1FLD1      gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGACCAATCCAGAAGA
S1FLD2      gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S1FLD3      gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S7REF       gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S2REF       gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTGAACAATCCAGAAGA
S4REF       gttaagtttctgcgccATGTATCCAGTACTTTCGAGAACCGTGGTAAATAATCCAGAAGA
S6REF       gttaagtttctgcgccATGTATTCGGTACTATCAAGAACTGTTGTAAACAATCCTGAGGA
S6FLD2      gttaagtttctgcgccATGTATCCGGTACTATCAAGAACTGTTGTAAACAATCCTGAGGA
S6FLD1      gttaagtttctgcgccATGTATCCGGTACTATCAAGAACTGTTGTAAACAATCCTGAGGA
S3REF       gttaagtttctgcgccATGTATCCGGTACTATCAAGAACTGTTGTAAACAATCCTGAGGA
S3FLD1      gttaagtttctgcgccATGTATCCGGTACTATCAAGAACTGTTGTAAACAATCCTGAGGA
S5REF       gttaagtttctgcgccATGTATTCGGTACTATCAAGAACTGTTGTAAACAATCCTGAGGA
            ********************* * ***** ** ***** ** ** *  ***** ** **

                    70        80        90       100       110       120
S1REF*      GCGTGCGCTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCCGGAATAA
S1REF       GCGTGCGCTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCCGGAATAA
S1FLD4      GCGTGCACTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCTGGAATAA
S1FLD1      GCGTGCACTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCTGGAATAA
S1FLD2      GCGTGCACTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCTGGAATAA
S1FLD3      GCGTGCACTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCTGGAATAA
S7REF       GCGTGCACTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCTGGAATAA
S2REF       GCGTGCGCTAATGGTGTACCCGCCAACAGCCCCAATGCCGCCTGTTATGACCGGGAATAA
S4REF       GCGTGCGCTAATGGTGTACCCGCCGACGGCTCCGATACCGCCTGTTATGACTTGGAATAA
S6REF       ACGAGCACTAATGGTGTACCCGCCTACGGCACCTGTACCGCCCGTGATGACATGGAACAA
S6FLD2      ACGAGCACTAATGGTGTACCCGCCTACGGCACCTGTACCGCCCGTGATGACATGGAACAA
S6FLD1      ACGAGCACTAATGGTGTACCCGCCTACGGCACCTGTACCGCCCGTGATGACATGGAACAA
S3REF       ACGAGCACTAATGGTGTACCCGCCGACGGCACCCATGCCGCCCATGATGACACGGAACAA
S3FLD1      ACGAGCACTAATGGTGTACCCGCCGACGGCGCCCATGCCTCCCATGATGACATGGAACAA
S5REF       ACGAGCACTAATGGTGTACCCGCCGACGGCACCTGTGCCGCCCATGATGACATGGAATAA
             ** ** ****************** ** ** **  * ** **  * *****  **** **

                    130       140       150       160       170       180
S1REF*      TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTGGACAAGAA
S1REF       TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTGGACAAGAA
S1FLD4      TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTAGACAAGAA
S1FLD1      TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTAGACAAGAA
S1FLD2      TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTAGACAAGAA
S1FLD3      TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTAGACAAGAA
S7REF       TCTTAAAATTGACGGTGTTGACGGAATGAAAGACTTAGCTCTAAATATATTAGACAAGAA
S2REF       TCTTAAAATTGACAGTGTTGACGGAATGAAAGACTTAGCTTTAAATATATTGGACAAGAA
S4REF       TCTTAAAATCGATAGTGTAGACGGAATGAAAGACTTAGCTCTAAATATATTGGATAAAAA
S6REF       TCTGAAAATTGATAGTGTAGATGGAATGAGGGATCTAGCTTTGAACATTCTAGATAAAAA
S6FLD2      TCTGAAAATTGATAGTGTAGATGGAATGAGGGATCTAGCTTTGAACATTCTAGATAAAAA
S6FLD1      TCTGAAAATTGATAGTGTAGATGGAATGAGGGACCTAGCTTTGAACATTCTAGATAAAAA
S3REF       TCTCAAAATTGATAGTGTGGATGGAATGAGAGATTTAGCTTTGAACATTCTAGATAAAAA
S3FLD1      TCTGAAAATTGATAGTGTGGATGGAATGAGGGATTTAGCTTTGAACATTCTAGATAAAAA
S5REF       TCTGAAAATTGATAGTGTAGACGGAATGAGAGATCTAGCTTTGAACATTCTAGATAAAAA
            *** ***** **  **** ** ******* **  ***** * ** **  * ** ** **

                    190       200       210       220       230       240
S1REF*      TATAACTAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S1REF       TATAACTAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S1FLD4      TATAACCAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S1FLD1      TATAACCAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S1FLD2      TATAACCAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S1FLD3      TATAACCAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S7REF       TATAACCAGTACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S2REF       TATAACTAGCACAACAGGTGCGGATGAGTGTGATAAACGTGAGAAAGCGATGTTTGCCTC
S4REF       CATAACTAGTACAACAGGGGCGGATGAGTGTGATAAACGTGAAAAAGCAATGTTTGCCTC
S6REF       CATAACAAGTACGACCGGAGCAGATGAATGTGATAAAAGGGAGAAGGCGATGTTTGCGTC
S5FLD2      CATAACAAGTACGACCGGAGCAGATGAATGTGATAAAAGGGAGAAGGCGATGTTTGCGTC
S6FLD1      CATAACAAGTACGACCGGAGCAGATGAATGTGATAAAAGGGAGAAGGCGATGTTTGCGTC
S3REF       CATAACAAGTACGACTGGAGCAGATGAATGTGATAAAAGGGAGAAGGCGATGTTTGCATC
S3FLD1      CATAACAAGTACGACTGGAGCAGATGAATGTGATAAAAGGGAGAAGGCGATGTTTGCATC
S5REF       TATAACGAGTACGACCGGAGCGGACGAATGTGATAAAAGAGAGAAGGCGATGTTTGCGTC
             ***** ** ** ** ** ** ** ** ********* * ** ** ** ******** **

                    250       260       270       280       290       300
S1REF*      GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
S1REF       GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
S1FLD4      GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
S1FLD1      GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
```

```
S1FLD2          GGTAGCCGAATCAGCTGCGGATAGCCC                ATCCAGATATATAA
S1FLD3          GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
S7REF           GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
S2REF           GGTAGCCGAATCAGCTGCGGATAGCCCAATGGTACGCACTATTAAAATCCAGATATATAA
S4REF           AGTTGCCGAATCAGCTGCAGATAGCCCAATGGTGCGCACTATTAAAATTCAGATATATAA
S6REF           AGTGGCGGAGTCCGCGGCCGACAGCCCTGCAGTGAGGATGATCAAGATCCAGATTTACAA
S6FLD2          AGTGGCGGAGTCCGCGGCCGACAGCCCTGCAGTGAGGATGATCAAGATCCAGATTTACAA
S6FLD1          AGTGGCGGAGTCCGCGGCCGACAGCCCTGCAGTGAGGATGATCAAGATCCAGATTTACAA
S3REF           AGTGGCAGAGTCCGCAGCCGACAGCCCTGCAGTGAGGATGATCAAAATCCAGATCTACAA
S3FLD1          AGTGGCAGAGTCCGCAGCCGACAGCCCAGCAGTGAGGATGATCAAAATCCAGATCTACAA
S5REF           AGTAGCGGAGTCCGCAGCTGACAGCCCCGCAGTAAGAATGATCAAAATCCAGATCTACAA
                 ** ** ** ** ** ** ** *****    **  * *  ** ** ** ***** ** **

                       310        320        330        340        350        360
S1REF*          TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S1REF           TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S1FLD4          TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S1FLD1          TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S1FLD2          TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S1FLD3          TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S7REF           TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAAGAGACGTGCGGTGTT
S2REF           TAGAGTATTAGATGATATGGAGAGGGAGAAGCAAAAATGCGAGAGGAGACGTGCGGTGTT
S4REF           TAGAGTGTTAGATGATATGGAGAGGGAAAAGCAAAAATGTGAGAAGAGACGTGCGGTGTT
S6REF           TAGGGTGTTGGACGACATGGAAAAAGAGAAGGCCAAAGCGGAGAGGAGAAGAACGGTACT
S6FLD2          TAGGGTGTTGGACGACATGGAAAAAGAGAAGGCCAAAGCGGAGAGGAGAAGAACGGTACT
S6FLD1          TAGGGTGTTGGACGACATGGAAAAAGAGAAGGCCAAAGCGGAGAGGAGAAGAACGGTGCT
S3REF           TAGGGTGTTGGATGACATGGAAAAAGAGAAGGTCAAAGCGGAGAGGAGAAGAGCGGTACT
S3FLD1          TAGGGTGTTGGATGACATGGAAAAAGAGAAGGCCAAAGCGGAGAGGAGAAGAGCGGTACT
S5REF           TAGAGTGTTGGACGACATGGAAAAGGAAAAGGTTAAAGCGGAGAGGAGGAGAGCTGTACT
                *** ** ** ** ** ***** *  ** ***   ***   **** *** *  * ** *

                       370        380        390        400        410        420
S1REF*          GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACATCGACCTTCTTGATGGCTAT
S1REF           GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACATCGACCTTCTTGATGGCTAT
S1FLD4          GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACAGCGACCTTCTTGATGGCTAT
S1FLD1          GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACAGCGACCTTCTTGATGGCTAT
S1FLD2          GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACAGCGACCTTCTTGATGGCTAT
S1FLD3          GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACAGCGACCTTCTTGATGGCTAT
S7REF           GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACAGCGACCTTCTTGATGGCTAT
S2REF           GAGATTTATCTCGTACACCTTTATAACGTTAATGTTAACATCGACCTTCTTGATGGCTAT
S4REF           GAAATTCATTTCGTACACCTTTATAACGTTAATGCTAACATCGACCTTCTTGATGGCTAT
S6REF           CAGATTCGCGTCGTGTACAATCATAACAGTTATGCTTTTATCGACGTTCTTGATGGCCAT
S6FLD2          CAGATTCGCGTCGTGTACAATCATAACAGTTATGCTTTTATCGACGTTCTTGATGGCCAT
S6FLD1          CAGAATCGCGTCGTGTACAATCATAACAGTTATGCTTTTATCGACGTTCTTGATGGCCAT
S3REF           TAGATTCGCGTCGTATACAATCATAACAGTCATGCTTTTGTCGACGTTCTTGATGGCCAT
S3FLD1          TAGATTCGCGTCGTATACAATCATAACAGTTATGCTTTTGTCGACGTTCTTGATGGCCAT
S5REF           TAGATTGGCGTCGTATACAATCATAACAGTTATGCTTTTGTCGACGTTTTTGATGGCCAT
                 * * *    ****  **  * *****  * *** *     **** ** ******** **

                       430        440        450        460        470        480
S1REF*          GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAAACGGGGAACGC
S1REF           GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAAACGGGGAACGC
S1FLD4          GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAAACGGGGAGCGC
S1FLD1          GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAAACGGGGAGCGC
S1FLD2          GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAAACGGGGAGCGC
S1FLD3          GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAAACGGGGAGCGC
S7REF           GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAATACGGGGAGCGC
S2REF           GATGCAAACTCCACCAATAACACAGTATGTGGAGATGGCGTGTAATAAGACGGGGAACGC
S4REF           GATGCAGACGCCACCAATAACACAGTATGTGGAGAAGGCGTGTAATGGAACGGGGGAAGC
S6REF           GATGCAAACTCCACCTATAAATCAATACGTAGAGAGAGCCTGTAACGGTACGGGTCATAC
S6FLD2          GATGCAAACTCCACCTATAAATCAATACGTAGAGAGAGCCTGTAACGGTACGGGTCATAC
S6FLD1          GATGCAAACTCCACCTATAAATCAATACGTAGAGAGAGCCTGTAACGGTACGGGTCATAC
S3REF           GATGCAAACCCCACCTATAAATCAATACGTAGAGAGAGCCTGCAACGGTACGGGTCAAAC
S3FLD1          GATGCAAACCCCGCCTATAAATCAATACGTAGAGAAAGCCTGCAACGGTACGGGTCAAAC
S5REF           GATGCAGATCCCTCCGATAAATCAATACGTAGAAAGAGCCTGTAACGGTACGAACCCTAC
                ****** *  ** ** **** ** ** ** ** *  ** ** **    ***       *

                       490        500        510        520        530        540
S1REF*          AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S1REF           AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S1FLD4          AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S1FLD1          AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S1FLD2          AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S1FLD3          AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S7REF           AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S2REF           AGAGAAGAATGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S4REF           AGAGAAGACCGACCCATGCGGTTTGATGAGATGGAGTGGGGCTGTTCAGTTTCTGACGTT
S6REF           ACAGGGTGATGATCCTTGCGGTCTGATGAAATGGAGCGGAGCGCTGCAGTTCATAACGTT
S6FLD2          ACAGGGTGATGATCCTTGCGGTCTGATGAAATGGAGCGGAGCGCTGCAGTTCATAACGTT
S6FLD1          ACAGGGTGATGATCCTTGCGGTCTGATGAAATGGAGCGGAGCGCTGCAGTTCATAACGTT
S3REF           ACAGGGTGATGATCCTTGCGGTCTGATGAAATGGAGCGGAGCGCTGCAGTTCATAACGTT
```

```
S3FLD1          ACAGGGTGATGATCCTTGCGGTCTGATG                CAGTTCATAACGTT
S5REF           ACAAGGTGATGATCCTTGCGGTCTGATGAAATGGAGCGGAGCGTTGCAGTTCATAACGTT
                * *       ** ** ****** ******  ***** ** **  * *****  * ******

                       550       560       570       580       590       600
S1REF*          GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACACTTTCCACGAACGC
S1REF           GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACACTTTCCACGAACGC
S1FLD4          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S1FLD1          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S1FLD2          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S1FLD3          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S7REF           GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S2REF           GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACGCTTTCCACGAACGC
S4REF           GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATCGCCACGCTTTCAACGAATGT
S6REF           GATACTTAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
S6FLD2          GATACTTAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
S6FLD1          GATACTTAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
S3REF           GATACTCAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAATCGTTATCGCAGAACGC
S3FLD1          GATACTCAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAATCGTTATCGCAGAACGC
S5REF           GATACTCAGTGGATTTCTGTACATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
                **** * ** ** **  **** ***** *** * *****     *  * **   *** *

                       610       620       630       640       650       660
S1REF*          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1REF           GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD4          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD1          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD2          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD1          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S7REF           GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S2REF           GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S4REF           GGATAGAATTAGCAAAAATATTTTGAAGCGGCGAGCATATATTGACGCAGCTAGATCAAA
S6REF           GGATAGGATTAGAAAAAATATCTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S6FLD2          GGATAGGATTAGAAAAAATATTTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S6FLD1          GGATAGGATTAGAAAAAATATTTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S3REF           GGATAGGATTAGAAAAAATATCTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S3FLD1          GGATAGGATTAGAAAGAATATCTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S5REF           GGATAGAATCAGAAAGAATATCTTGAAGCGACGAGCTTACATCGATGCTGCAAAGACAAA
                ****** ** ** ** ** ** ***** ** ** ** ** ** ** ** ** * ****

                       670       680       690       700       710       720
S1REF*          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1REF           TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD4          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD1          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD2          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD3          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S7REF           TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S2REF           TCCAGATGCAACGGTTCTGAATGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S4REF           TCCAGATGCGACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S6REF           CCCAGATGCGACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAATTACCGTACCAGTT
S6FLD2          CCCAGATGCGACAGTTTTGACGGTAGCCGGAGGCAACACGGGGGAACTACCGTACCAGTT
S6FLD1          CCCAGATGCGACAGTTTTGACGGTAGCCGGAGGCAACACGGGGGAACTACCGTACCAGTT
S3REF           CCCAGATGCGACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAATTACCGTACCAGTT
S3FLD1          CCCAGATGCAACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAATTACCGTACCAGTT
S5REF           CCCAGATGCGACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAACTACCGTACCAGTT
                ******** ** **  ***  **  * ****************  **************

                       730       740       750       759
S1REF*          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD4          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD1          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD2          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD3          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S7REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S2REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S4REF           CGGGGATACGGCCCACTTAGccgtggcggaaacgtgttac
S6REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S6FLD2          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S6FLD1          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S3REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S3FLD1          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S5REF           CGGAGATGCGGCCCATTTAGccgtggcggaaacgtgttac
                *** *** ******* ************************
```

**Figure 2.11.** ClustalX 1.81 alignment of the nucleotide sequences of the seven reference strains, seven field isolates and EEV Bryanston (EEV-1) (S1REF*) to show similarities. Identical bases are indicated with "*", the first start codon in yellow, a putative second start codon in red, stop codons in grey, non-coding sequences in lowercase and primer sequences are underlined.

```
S3FLD1          ACAGGGTGATGATCCTTGCGGTCTGATG                CAGTTCATAACGTT
S5REF           ACAAGGTGATGATCCTTGCGGTCTGATGAAATGGAGCGGAGCGTTGCAGTTCATAACGTT
                * *       ** ** ****** ******  ***** ** **  * *****  * ******

                       550       560       570       580       590       600
S1REF*          GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACACTTTCCACGAACGC
S1REF           GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACACTTTCCACGAACGC
S1FLD4          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S1FLD1          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S1FLD2          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S1FLD3          GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S7REF           GATAATGAGCGGTTTCTTGTACATGTGCAAACGTTGGATTACCACACTTTCCTCGAACGC
S2REF           GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATTACCACGCTTTCCACGAACGC
S4REF           GATAATGAGCGGTTTCTTGTACATGTGTAAACGTTGGATCGCCACGCTTTCAACGAATGT
S6REF           GATACTTAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
S6FLD2          GATACTTAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
S6FLD1          GATACTTAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
S3REF           GATACTCAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAATCGTTATCGCAGAACGC
S3FLD1          GATACTCAGTGGATTCTTGTATATGTGTAAAAGGTGGATTCAATCGTTATCGCAGAACGC
S5REF           GATACTCAGTGGATTTCTGTACATGTGTAAAAGGTGGATTCAGTCGTTATCGCAGAACGC
                **** * ** ** **  **** ***** *** * *****     *  * **   *** *

                       610       620       630       640       650       660
S1REF*          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1REF           GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD4          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD1          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD2          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S1FLD1          GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S7REF           GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S2REF           GGATAGAATTAGCAAAAACATTTTGAAGCGGCGAGCGTATATCGACGCAGCTAGATCAAA
S4REF           GGATAGAATTAGCAAAAATATTTTGAAGCGGCGAGCATATATTGACGCAGCTAGATCAAA
S6REF           GGATAGGATTAGAAAAAATATCTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S6FLD2          GGATAGGATTAGAAAAAATATTTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S6FLD1          GGATAGGATTAGAAAAAATATTTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S3REF           GGATAGGATTAGAAAAAATATCTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S3FLD1          GGATAGGATTAGAAAGAATATCTTGAAACGACGGGCTTACATCGATGCTGCAAGAACAAA
S5REF           GGATAGAATCAGAAAGAATATCTTGAAGCGACGAGCTTACATCGATGCTGCAAAGACAAA
                ****** ** ** ** ** ** ***** ** ** ** ** ** ** ** ** * ****

                       670       680       690       700       710       720
S1REF*          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1REF           TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD4          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD1          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD2          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S1FLD3          TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S7REF           TCCAGATGCAACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S2REF           TCCAGATGCAACGGTTCTGAATGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S4REF           TCCAGATGCGACGGTTCTGACTGTGACTGGAGGCAACACGGGGGATCTACCGTACCAGTT
S6REF           CCCAGATGCGACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAATTACCGTACCAGTT
S6FLD2          CCCAGATGCGACAGTTTTGACGGTAGCCGGAGGCAACACGGGGGAACTACCGTACCAGTT
S6FLD1          CCCAGATGCGACAGTTTTGACGGTAGCCGGAGGCAACACGGGGGAACTACCGTACCAGTT
S3REF           CCCAGATGCGACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAATTACCGTACCAGTT
S3FLD1          CCCAGATGCAACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAATTACCGTACCAGTT
S5REF           CCCAGATGCGACAGTCTTGACGGTAGCCGGAGGCAACACGGGGGAACTACCGTACCAGTT
                ******** ** **  ***  **  * ****************  **************

                       730       740       750       759
S1REF*          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD4          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD1          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD2          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S1FLD3          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S7REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S2REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S4REF           CGGGGATACGGCCCACTTAGccgtggcggaaacgtgttac
S6REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S6FLD2          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S6FLD1          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S3REF           CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S3FLD1          CGGGGATACGGCCCATTTAGccgtggcggaaacgtgttac
S5REF           CGGAGATGCGGCCCATTTAGccgtggcggaaacgtgttac
                *** *** ******* ************************
```

**Figure 2.11.** ClustalX 1.81 alignment of the nucleotide sequences of the seven reference strains, seven field isolates and EEV Bryanston (EEV-1) (S1REF*) to show similarities. Identical bases are indicated with "*", the first start codon in yellow, a putative second start codon in red, stop codons in grey, non-coding sequences in lowercase and primer sequences are underlined.

2.3.5.2. Amino acid sequence analysis of a number of EEV reference strains and field isolates

Some basic characteristics of the different NS3 proteins of the EEV reference strains, such as the number of amino acids, the molecular weight and pI, were determined using the ProtParam tool at ExPASy (Appendix B). The amino acid sequences of the seven EEV reference strains were also analysed to identify conserved protein motifs as was done for EEV Bryanston (EEV-1) (S1REF*) NS3 (section 2.3.3.1.) by searching the PROSITE database. The results are shown in Table 2.12.

The amino acid sequences of the seven reference strains, as well as the amino acid sequences of the seven reference strains and the seven field isolates, were compared using sequence alignment. The aim of the sequence alignment was to attempt to align the amino acid sequences in such a way that regions of structural or functional similarity between the molecules were highlighted. Certain structural features were identified for the EEV Bryanston (EEV-1) (S1REF*) NS3 sequence (section 2.3.3.1.), and the amino acid sequence alignments will serve to identify which of these features are conserved within the EEV serogroup. The sequence alignments will however also show which regions differ between sequences, giving an indication of the variation found within a set of sequences.

ClustalX version 1.81 (Thompson *et al.*, 1997) was used to perform the alignment of the seven reference strains, including EEV Bryanston (EEV-1) (S1REF*) NS3 and the results are shown in Figure 2.12. The different protein motifs shown in Table 2.12. are also indicated in Fig. 2.12. The line below the S3REF sequence, containing the different symbols, is used to indicate conserved positions. Three characters ("*", ":" and ".") are used where "*" indicates positions which have a single, fully conserved residue, ":" indicates that one of the "strong" groups (STA, NEQK, NHQK, NDEQ, QHRK, MILV, MILF, HY, or FYW) is fully conserved and "." indicates that one of the "weaker" groups (CSA, ATV, SAG, STNK, STPA, SGND, SNDEQK, NDEQHK, NEQHRK, FVLIM or HFY) is fully conserved.

Two types of phosphorylation sites were identified for EEV Bryanston (EEV-1) (S1REF*) NS3 protein (Table 2.9.) namely a protein kinase C phosphorylation site and casein kinase II phosphorylation sites. The protein kinase C phosphorylation site was not conserved between the seven reference strains, while a number of the casein kinase II phosphorylation sites were conserved (Table 2.12., Fig. 2.12.). The EEV NS3 proteins contain a number of serine and threonine residues, which by chance could be in the correct context to be identified as a phosphorylation site when searching the PROSITE database, since many of the patterns contained in this database are small and have a reasonably high chance of appearing at random when performing a search. This implies that a significant number of the matches reported are potential false-positives or that PROSITE massively overpredicts certain motifs (Alphey, 1997). As phosphorylation has to date not been reported for any other orbivirus NS3 proteins, this result is only shown as a matter of interest but will not be pursued further in this investigation.

**Table 2.12.** Some basic characteristics of the seven EEV reference strain NS3 protein sequences and the results of ScanProsite (a protein against PROSITE search for protein motifs) (Appendix B)

| EEV reference strain | Number of amino acids, molecular weight and pI | Protein motif, number of matches, amino acid residues involved and amino acid sequence of motif | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | N-glycosylation site | | Protein kinase C phosphorylation site | | Casein kinase II phosphorylation site | | N-myristylation site | |
| S1REF<br><br>S2REF | 240 aa; ≈ 27 kDa; 8.93<br>240 aa; ≈ 27 kDa; 8.77 | 2 | 55 – 58 NITS<br>150 – 153 NKTG<br>222 – 225 NVTG (S2REF) | 1 | 88 – 90 TIK | 5 | 60 – 63 TGAD<br>75 – 78 SVAE<br>79 – 82 SAAD<br>193 – 196 TNAD<br>214 – 217 SNPD | 2 | 61 – 66 GADECD<br>225 – 230 GGNTGD |
| S3REF | 240 aa; ≈ 27 kDa; 9.03 | 2 | 55 – 58 NITS<br>150 – 153 NGTG | No sites identified | | 5 | 60 – 63 TGAD<br>75 – 78 SVAE<br>79 – 82 SAAD<br>155 – 158 TQGD<br>214 – 217 TNPD | 4 | 61 – 66 GADECD<br>151 – 156 GTGQTQ<br>153 – 158 GQTQGD<br>225 – 230 GGNTGE |
| S4REF | 240 aa; ≈ 27 kDa; 8.53 | 2 | 55 – 58 NITS<br>150 – 153 NGTG | 1 | 88 – 90 TIK | 5 | 60 – 63 TGAD<br>75 – 78 SVAE<br>79 – 82 SAAD<br>193 – 196 TNVD<br>214 – 217 SNPD | 3 | 61 – 66 GADECD<br>151 – 156 GTGEAE<br>225 – 230 GGNTGD |
| S5REF<br><br>S6REF | 240 aa; ≈ 27 kDa; 8.84<br>240 aa; ≈ 27 kDa; 8.79 | 2 | 55 – 58 NITS<br>150 – 153 NGTN (S5REF)<br>150 – 153 NGTG (S6REF) | No sites identified (S5REF)<br>1 cAMP- and cGMP-dependent protein kinase phosphorylation site<br>110 – 113 RRRT (S6REF) | | 5 | 60 – 63 TGAD<br>75 – 78 SVAE<br>79 – 82 SAAD<br>155 – 158 TQGD<br>214 – 217 TNPD | 3 | 61 – 66 GADECD<br>151 – 156 GTNPTQ (S5REF)<br>151 – 156 GTGHTQ (S6REF)<br>225 – 230 GGNTGE |
| S7REF | 240 aa; ≈ 27 kDa; 8.55 | 2 | 55 – 58 NITS<br>150 – 153 NNTG | 1 | 88 – 90 TIK | 5 | 60 – 63 TGAD<br>75 – 78 SVAE<br>79 – 82 SAAD<br>193 – 196 SNAD<br>214 – 217 SNPD | 2 | 61 – 66 GADECD<br>225 – 230 GGNTGD |

CLUSTAL X (1.81) multiple amino acid sequence alignment of the seven EEV reference strains and EEV Bryanston (EEV-1) (S1REF*)

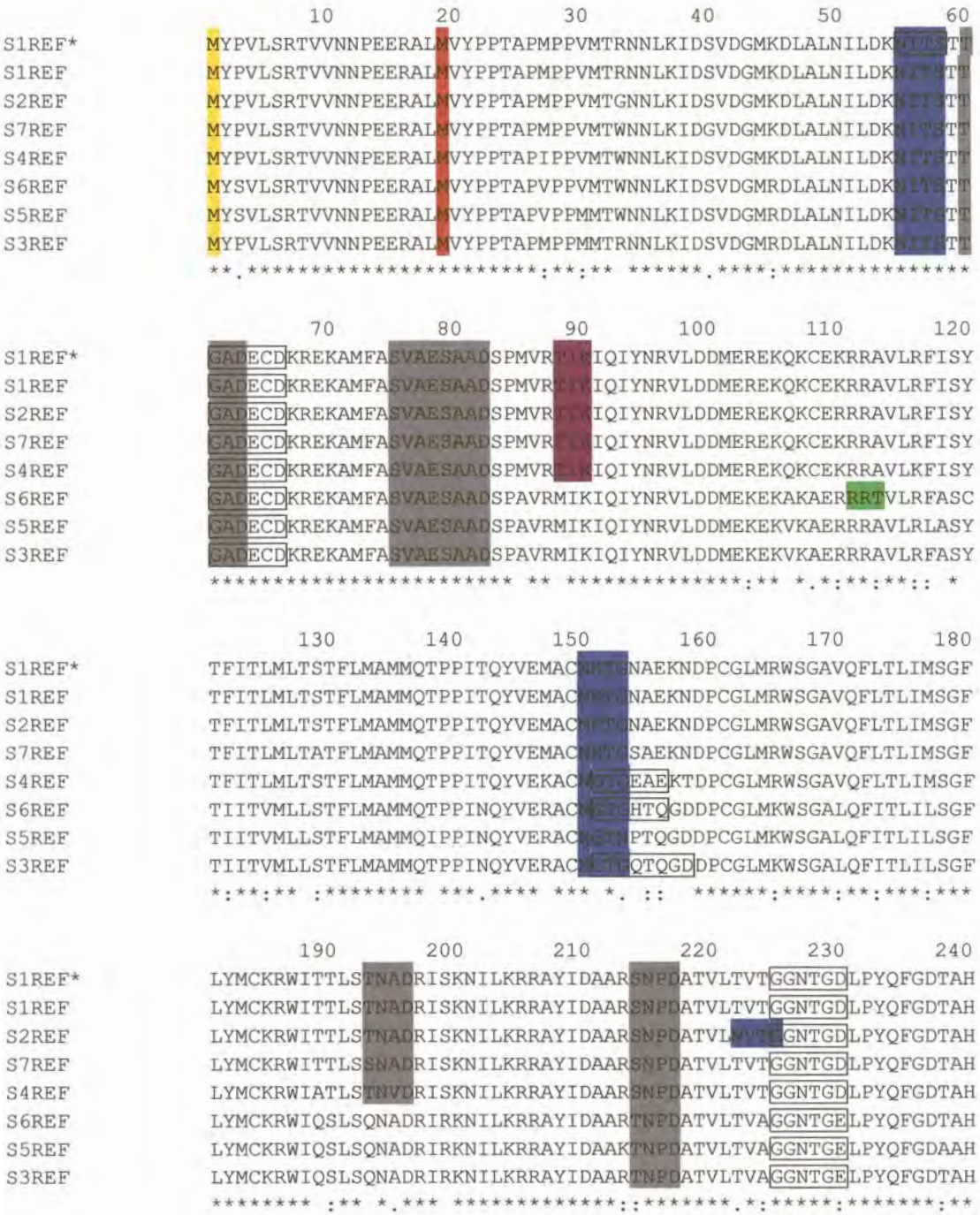


**Figure 2.12.** ClustalX 1.81 alignment of amino acid sequences of the seven EEV reference strains and EEV Bryanston (EEV-1) (S1REF*) to show similarities. Single, fully conserved, i.e. identical, residues are indicated with "*", "strong" groups that are conserved is indicated by ":" and "weak" groups that are conserved is indicated by "." The different protein motifs are also shown using different colours or blocks: N-glycosylation sites in blue, protein kinase C phosphorylation sites in purple, N-myristylation sites are blocked, cAMP- and cGMP-dependent protein kinase phosphorylation sites in green and casein kinase II phosphorylation sites in grey. The first start codon (first methionine residue) is indicated in yellow and a putative second start codon (second methionine residue) in red.

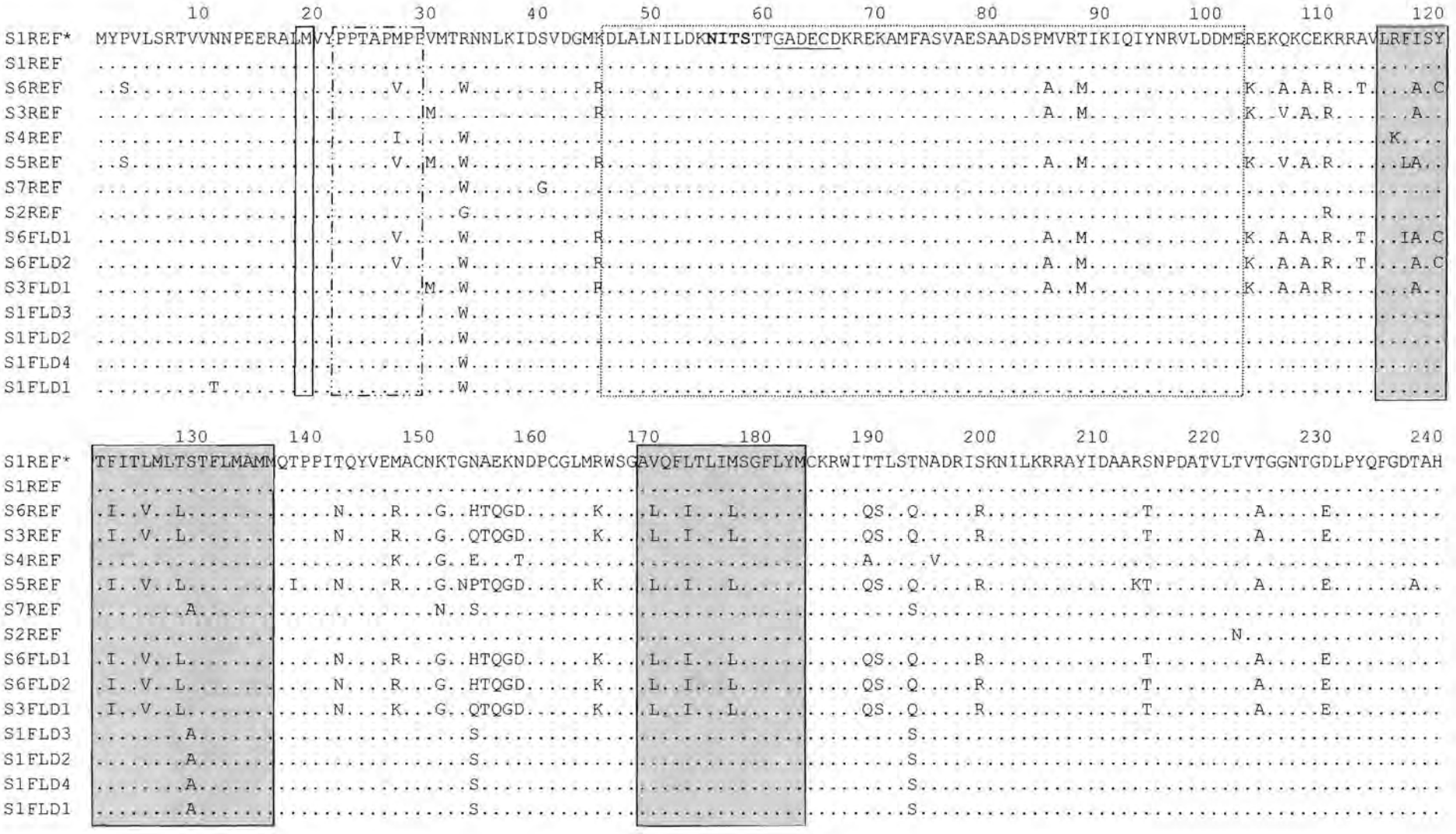


**Figure 2.13.** MEGA2 amino acid sequence alignment of all EEV isolates. Dots indicate identity to NS3 of the EEV Bryanston (EEV-1) laboratory reference strain, S1REF*. The NS3A start codon is blocked; the proline-rich region is blocked (amino acids 22 to 29). The highly conserved region is blocked (amino acids 46 to 102), the myristylation motif within this region is underlined and the N-linked glycosylation site (amino acids 55 to 58) is indicated by bold type. The two hydrophobic domains (amino acids 115 to 136 and 169 to 183) are blocked and shaded in grey. (adapted from Van Niekerk *et al.*, 2003).

**Table 2.13.** Characteristic structural features or protein motifs seen in EEV NS3

| Feature | Position(s) in amino acid sequence | Comment |
|---|---|---|
| Second in-phase methionine | Position 19 of full length NS3 protein | Second putative start codon for translation initiation of NS3A |
| Proline-rich region | Position 22 to 29 | Five proline residues in a stretch of 8 residues |
| Potential glycosylation motif | Position 55 to 58 and 150 to 153 | In case of S6REF, S3REF, S4REF, S6FLD2 and S3FLD1 the motif at position 150 to 153 is not conserved. A lysine (positively charged residue) is replaced by a glycine (non-polar residue) or an asparagine (polar uncharged residue). S2REF also has an additional site at position 222 to 225. |
| Putative myristylation site | Position 61 to 66 and 225 to 230 | |
| Positively charged region | Position 67 to 120 | Thirteen positively charged residues: R + K |
| Hydrophobic domains | HDI: position 115 to 136<br>HDII: position 169 to 183 | AnTheProt prediction of HDs, predicted TM helices (membrane spanning regions) found in these regions |
| Conserved region | Position 46 to 102 | Contains potential glycosylation motif at position 55 to 58 and putative myristylation site at position 61 to 66 |
| Cysteine residues: | Five residues seen at position 65, 108, 149, 161, 184 as well as an additional cysteine residue at position 120 for S6REF, S6FLD1 and S6FLD2 | Position 65 (conserved, in conserved region); position 108 (8 out of 15 sequences, others have alanine (non-polar residue) at this position); position 149 (conserved); position 161 (conserved); position 184 (conserved); additional cysteine at position 120 for S6REF, S6FLD1 and S6FLD2 (i.e. characteristic for EEV-6) (in proposed membrane spanning region, others tyrosine (polar, uncharged residue) at this position) |
| Variable region | Most variable region within NS3 protein seems to be located between the two hydrophobic domains i.e. between position 137 to 168 | Thirty-two residues of which residues differ at 11 positions between the different NS3 proteins, thus approximately 34% variation in this region |

The amino acid sequence alignment of the seven reference strains, the seven field isolates and EEV Bryanston (EEV-1) (S1REF*) NS3 was performed using MEGA2 (Kumar *et al.*, 2001) and the results are shown in Fig 2.13. Structural features or protein motifs were identified in the EEV Bryanston (EEV-1) NS3 protein sequence (section 2.3.3.1.); similar structural features or protein motifs were also identified for the seven reference strains in Table 2.12. In Fig. 2.13. the structural features or protein motifs that are conserved within the EEV serogroup are shown. The features shown in the figure were specifically chosen to allow comparison with other orbivirus NS3 proteins as similar features have been identified in these other orbivirus NS3 proteins. The different structural features or motifs that are seen in EEV NS3 proteins are also summarized in Table 2.13.

It is not unexpected to see similar structural features for NS3 of the different orbiviruses. This however, does not give an indication of how similar the viruses are regarding S10 or NS3 or what their relationship would be based on their S10 or NS3 sequences. In the next section the relationship of the EEV serogroup to other orbivirus serogroups, as well as the relationship of the different EEV serotypes to each other will be explored.

### 2.3.6. Phylogenetic analysis

The aim of the results in this section was to determine the relationships between EEV and other orbiviruses as well as the relationships between and within serotypes in the EEV serogroup based on S10 or NS3 sequences. These relationships can be visually represented as phylogenetic trees and the relationships can be given a specific numeric value through the use of distance matrixes.

Phylogenetic analysis was performed as described in section 2.2.13. Two nucleotide datasets and two amino acid datasets were used for analysis. The first nucleotide or amino acid dataset consisted of all the S10 or NS3 sequences shown in Appendix C, and the second nucleotide or amino acid dataset consisted of all the EEV S10 or NS3 sequences, as well as selected S10 or NS3 sequences from Appendix C. The S10 phylogenetic trees were generated using the neighbour-joining method (uncorrected "p"), while the NS3 phylogenetic trees were generated using the neighbour-joining method (mean character difference). Broadhaven S10 nucleotide or NS3 amino acid sequence was used as outgroup in the datasets. This virus differs from all the other viruses in the datasets in that it is a tick-transmitted orbivirus while the others are all transmitted by midges. Van Niekerk *et al.* (2001b) also found Broadhaven virus NS3 to be the most divergent NS3 protein in the orbivirus group. The results obtained from each type of analysis (neighbour-joining or maximum parsimony methods) for each of the two types of dataset (nucleotide or amino acid sequences) were evaluated by comparing the nucleotide results and by comparing the amino acid results. The neighbour-joining method seemed to be better suited to both the nucleotide and amino acid datasets, and the datamatrixes and phylogenies obtained using this method will be shown in section 2.3.6.1. and section 2.3.6.2.

Distance matrix showing the percentage nucleotide differences and amino differences seen between EEV S10/NS3 sequences and selected orbivirus S10 /NS3 sequences.

| | S1FLD2 | S1FLD4 | S1FLD3 | S1FLD1 | S7REF | S1REF* | S1REF | S2REF | S4REF | S6FLD1 | S6FLD2 | S6REF | S5REF | S3REF | S3FLD1 | AHSV4 | AHSV7 | AHSV2 | CHUZAN | BTV2 | BTV12 | EHDV1 | EHDV2 | BRDV | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| S1FLD2 | A | 0.0 | 0.0 | 0.1 | 0.3 | 1.2 | 1.1 | 1.8 | 7.4 | 23.7 | 23.8 | 24.4 | 25.0 | 23.8 | 24.1 | 60.9 | 62.0 | 61.9 | 64.2 | 61.2 | 61.4 | 61.2 | 62.2 | 64.0 | S1FLD2 |
| S1FLD4 | 0.0 | | 0.0 | 0.1 | 0.3 | 1.2 | 1.1 | 1.8 | 7.4 | 23.7 | 23.8 | 24.4 | 25.0 | 23.8 | 24.1 | 60.9 | 62.0 | 61.9 | 64.2 | 61.2 | 61.4 | 61.2 | 62.2 | 64.0 | S1FLD4 |
| S1FLD3 | 0.0 | 0.0 | | 0.1 | 0.3 | 1.2 | 1.1 | 1.8 | 7.4 | 23.7 | 23.8 | 24.4 | 25.0 | 23.8 | 24.1 | 60.9 | 62.0 | 61.9 | 64.2 | 61.2 | 61.4 | 61.2 | 62.2 | 64.0 | S1FLD3 |
| S1FLD1 | 0.4 | 0.4 | 0.4 | | 0.4 | 1.3 | 1.2 | 2.0 | 7.5 | 23.8 | 24.0 | 24.5 | 25.2 | 24.0 | 24.2 | 60.9 | 62.0 | 62.0 | 64.4 | 61.3 | 61.6 | 61.3 | 62.4 | 64.0 | S1FLD1 |
| S7REF | 0.8 | 0.8 | 0.8 | 1.3 | | 1.4 | 1.3 | 2.0 | 7.6 | 23.7 | 23.8 | 24.4 | 25.0 | 23.8 | 24.1 | 61.0 | 61.7 | 62.1 | 64.5 | 61.3 | 61.6 | 61.4 | 62.5 | 64.1 | S7REF |
| S1REF* | 1.7 | 1.7 | 1.7 | 2.1 | 2.5 | | 0.1 | 1.1 | 6.7 | 23.6 | 23.7 | 24.2 | 25.0 | 23.5 | 24.0 | 60.1 | 62.0 | 61.7 | 63.6 | 61.1 | 61.4 | 60.4 | 61.4 | 64.3 | S1REF* |
| S1REF | 1.7 | 1.7 | 1.7 | 2.1 | 2.5 | 0.0 | | 0.9 | 6.6 | 23.7 | 23.8 | 24.4 | 25.2 | 23.6 | 24.1 | 60.2 | 61.8 | 61.6 | 63.7 | 61.0 | 61.3 | 60.5 | 61.6 | 64.3 | S1REF |
| S2REF | 2.5 | 2.5 | 2.5 | 2.9 | 3.3 | 1.3 | 1.3 | | 7.1 | 23.6 | 23.7 | 24.2 | 25.0 | 23.6 | 24.0 | 60.6 | 62.4 | 62.3 | 64.1 | 60.9 | 61.1 | 60.9 | 62.0 | 64.3 | S2REF |
| S4REF | 4.2 | 4.2 | 4.2 | 4.6 | 4.6 | 3.8 | 3.8 | 4.6 | | 23.3 | 23.5 | 24.0 | 25.0 | 23.5 | 23.8 | 60.8 | 62.0 | 60.9 | 63.8 | 61.4 | 62.1 | 61.2 | 62.3 | 64.7 | S4REF |
| S6FLD1 | 14.6 | 14.6 | 14.6 | 15.0 | 15.0 | 14.6 | 14.6 | 14.6 | 14.6 | B | 0.4 | 0.9 | 8.3 | 4.5 | 4.9 | 62.7 | 62.5 | 62.4 | 64.7 | 63.7 | 62.7 | 62.9 | 62.6 | 63.7 | S6FLD1 |
| S6FLD2 | 14.2 | 14.2 | 14.2 | 14.6 | 14.6 | 14.2 | 14.2 | 14.2 | 14.2 | 0.4 | | 0.5 | 7.9 | 4.1 | 4.5 | 62.8 | 62.5 | 62.6 | 64.6 | 63.9 | 63.0 | 63.2 | 62.9 | 63.6 | S6FLD2 |
| S6REF | 14.6 | 14.6 | 14.6 | 15.0 | 15.0 | 14.6 | 14.6 | 14.6 | 14.6 | 0.8 | 0.4 | | 7.6 | 3.8 | 4.2 | 62.9 | 62.7 | 62.9 | 64.6 | 64.1 | 63.1 | 63.6 | 63.3 | 63.3 | S6REF |
| S5REF | 16.2 | 16.2 | 16.2 | 16.7 | 16.7 | 16.2 | 16.2 | 16.2 | 16.2 | 4.6 | 4.6 | 4.2 | | 7.8 | 8.2 | 63.9 | 64.4 | 64.1 | 63.6 | 62.5 | 62.1 | 63.7 | 63.8 | 63.1 | S5REF |
| S3REF | 13.8 | 13.8 | 13.8 | 14.2 | 14.2 | 12.9 | 12.9 | 13.3 | 14.2 | 3.3 | 2.9 | 3.3 | 3.8 | | 1.7 | 63.6 | 63.2 | 63.2 | 64.7 | 63.1 | 62.7 | 63.4 | 63.7 | 62.8 | S3REF |
| S3FLD1 | 13.3 | 13.3 | 13.3 | 13.8 | 13.8 | 13.3 | 13.3 | 13.3 | 13.3 | 2.9 | 2.5 | 2.9 | 4.2 | 1.3 | | 62.8 | 63.1 | 62.9 | 64.6 | 63.4 | 63.0 | 63.3 | 63.6 | 62.8 | S3FLD1 |
| AHSV4 | 78.6 | 78.6 | 78.6 | 78.6 | 79.1 | 78.6 | 78.6 | 78.1 | 78.6 | 77.7 | 78.1 | 78.1 | 78.6 | 78.6 | 78.1 | | 23.9 | 32.2 | 55.8 | 56.5 | 54.2 | 57.3 | 57.3 | 65.3 | AHSV4 |
| AHSV7 | 78.1 | 78.1 | 78.1 | 78.1 | 78.1 | 78.1 | 78.1 | 77.7 | 78.1 | 77.2 | 77.2 | 77.2 | 77.7 | 77.2 | 76.7 | 25.8 | | 31.7 | 55.1 | 55.4 | 54.0 | 57.7 | 57.0 | 65.8 | AHSV7 |
| AHSV2 | 77.8 | 77.8 | 77.8 | 77.8 | 78.2 | 77.8 | 77.8 | 77.8 | 76.9 | 77.8 | 78.2 | 77.8 | 77.8 | 78.2 | 77.8 | 34.9 | 34.0 | | 56.0 | 56.7 | 56.0 | 57.3 | 57.1 | 66.1 | AHSV2 |
| CHUZAN | 78.5 | 78.5 | 78.5 | 78.5 | 78.0 | 78.9 | 78.9 | 78.9 | 79.4 | 78.5 | 78.5 | 78.0 | 77.5 | 78.5 | 78.5 | 68.9 | 67.5 | 67.8 | | 58.1 | 58.8 | 58.6 | 58.1 | 63.1 | CHUZAN |
| BTV2 | 78.7 | 78.7 | 78.7 | 78.7 | 78.7 | 78.7 | 78.7 | 79.2 | 78.7 | 77.8 | 77.8 | 77.8 | 77.8 | 77.4 | 77.4 | 73.5 | 74.9 | 71.4 | 75.8 | | 7.3 | 38.4 | 39.2 | 65.3 | BTV2 |
| BTV12 | 78.7 | 78.7 | 78.7 | 78.7 | 78.7 | 78.7 | 78.7 | 79.2 | 78.7 | 77.8 | 77.8 | 77.8 | 77.8 | 77.4 | 77.4 | 73.0 | 74.9 | 71.0 | 74.9 | 2.2 | | 39.0 | 38.9 | 65.0 | BTV12 |
| EHDV1 | 78.5 | 78.5 | 78.5 | 78.5 | 78.5 | 78.5 | 78.5 | 78.1 | 78.5 | 76.7 | 76.7 | 76.7 | 76.7 | 76.3 | 76.3 | 72.8 | 71.8 | 72.6 | 69.7 | 43.1 | 42.7 | | 3.5 | 65.4 | EHDV1 |
| EHDV2 | 78.1 | 78.1 | 78.1 | 78.1 | 78.1 | 78.1 | 78.1 | 77.6 | 78.1 | 77.2 | 77.2 | 77.2 | 77.2 | 76.7 | 76.7 | 72.3 | 71.8 | 71.6 | 69.7 | 44.4 | 44.0 | 2.2 | | 65.1 | EHDV2 |
| BRDV | 76.6 | 76.6 | 76.6 | 76.6 | 76.6 | 76.6 | 76.6 | 76.6 | 77.1 | 76.6 | 76.6 | 76.6 | 78.1 | 76.6 | 76.6 | 76.3 | 77.9 | 76.1 | 76.4 | 81.3 | 81.3 | 79.0 | 79.5 | | BRDV |
| | S1FLD2 | S1FLD4 | S1FLD3 | S1FLD1 | S7REF | S1REF* | S1REF | S2REF | S4REF | S6FLD1 | S6FLD2 | S6REF | S5REF | S3REF | S3FLD1 | AHSV4 | AHSV7 | AHSV2 | CHUZAN | BTV2 | BTV12 | EHDV1 | EHDV2 | BRDV | |

**Figure 2.14.** Distance matrix showing the percentage nucleotide differences (upper triangle) and amino acid differences (lower triangle) between EEV S10 and NS3 sequences, and selected BTV S10/NS3 sequences, AHSV S10/NS3 sequences, EHDV S10/NS3 sequences, Chuzan virus S10/NS3 sequence and Broadhaven virus S10/NS3 sequence. The distance matrix was generated using the neighbour-joining method with the uncorrected "p" parameter (nucleotide sequence data) or the mean character difference parameter (amino acid sequence data) and Broadhaven S10/NS3 sequence as outgroup. Phylogenetic clusters A and B are indicated by large blocks and the percentages of S10/NS3 variation referred to in the text are blocked (adapted from Van Niekerk *et al.*, 2003).

2.3.6.1. Distance matrixes obtained after phylogenetic analysis of S10 nucleotide sequences and NS3 amino acid sequences

The distance matrixes give an indication of the amount of variation seen on nucleotide (S10) and amino acid (NS3) level within EEV serotypes, between EEV serotypes and between the EEV serogroup and the other orbivirus serogroups.

A single datamatrix is shown in Fig. 2.14. which combines the data from the second nucleotide dataset and the second amino acid dataset (both consisting of all EEV sequences and selected orbivirus sequences). The data from only one datamatrix of each type of dataset is shown, as the results obtained from both datamatrixes for each type of dataset were very similar. The distance matrixes were adjusted to show the values as the percentage difference and thus describe the percentage difference (variation) seen for S10 or NS3 between members of different serogroups, between members of the same serogroup (between serotypes) or within serotypes.

The maximum variation observed for S10 and NS3 between serotypes in the EEV serogroup and within serotypes of the EEV serogroup is shown in Table 2.14.

**Table 2.14.** Variation seen at nucleotide level and amino acid level within the EEV serogroup and within serotypes of the EEV serogroup

| Comparison[a] | Variation | |
|---|---|---|
| | Nucleotide level | Amino acid level |
| EEV serogroup (15) | 25.2% | 16.7% |
| Within EEV-1 serotype (6) | 1.3% | 2.1% |
| Within EEV-3 serotype (2) | 1.7% | 1.3% |
| Within EEV-6 serotype (3) | 0.9% | 0.8% |

[a] The number of sequences compared between serotypes and within serotypes in the EEV serogroup are indicated in brackets

The maximum variation of 25.2% at nucleotide level was observed between S1REF and S5REF, and the maximum variation of 16.7% at amino acid level was observed between S7REF and S5REF. The variation seen at nucleotide level or amino acid level within serotypes by comparing the reference strains and field isolates of a serotype to one another, are also shown in Table 2.14. The number of isolates for each serotype was however small (4 field isolates of EEV-1, one of EEV-3 and two of EEV-6), and more isolates should be sequenced for each serotype to confirm these preliminary values. The maximum variation seen at nucleotide and amino acid level between EEV and the other orbivirus serogroups in the first datamatrix (not shown) and the second datamatrix (Fig. 2.14.) varied and is shown in Table 2.15.

**Table 2.15.** Variation seen at nucleotide level and amino acid level between the EEV serogroup and the other orbivirus serogroups, and the maximum variation observed within serogroups for AHSV, BTV and EHDV

| Comparison[a] | Variation seen at nucleotide level | | Variation seen at amino acid level | |
|---|---|---|---|---|
| | First datamatrix | Second datamatrix | First datamatrix | Second datamatrix |
| EEV vs. AHSV | 60.2 – 65.0% | 60.1 – 64.4% | 76.7 – 79.5% | 76.8 – 79.1% |
| EEV vs. BTV | 60.1 – 65.6% | 60.9 – 64.1% | 76.5 – 78.3% | 77.4 – 79.2% |
| EEV vs. EHDV | 60.3 – 64.4% | 60.4 – 63.8% | 76.3 – 79.0% | 76.3 – 78.5% |
| EEV vs. Chuzan virus | 63.1 – 64.1% | 63.6 – 64.7% | 78.0 – 79.4% | 77.5 – 79.4% |
| EEV vs. BRDV | 62.3 – 64.9% | 62.8 – 64.7% | 78.0 – 79.1% | 76.6 – 78.1% |
| AHSV serogroup | 32.4% (9) | 32.2% (3) | 35.4% (9) | 34.9% (3) |
| BTV serogroup | 18.4% (11) | 7.3% (2) | 5.7% (11) | 2.2% (2) |
| EHDV serogroup | 3.5% (2) | 3.5% (2) | 2.2% (2) | 2.2% (2) |

[a] The number of serotypes compared within each serogroup is indicated in brackets

An interesting observation was the amount of variation seen at nucleotide level and amino acid level between S7REF and field isolates of EEV-1, only 0.3 - 0.4% (S10) and 0.8 - 1.3% (NS3) variation were seen. In contrast the variation seen at nucleotide and amino acid level between the EEV-1 reference strains, S1REF* and S1REF, and the EEV-1 field isolates varied between 1.1 - 1.4% (S10) and 1.7 - 2.1% (NS3). The maximum variation at amino acid level seen between EEV-1 field isolates was 0.4%. As will be seen later in Fig. 2.15., S7REF grouped closer to the EEV-1 field isolates than the two EEV-1 reference strains, although they are all grouped in the same subcluster.

S7REF is a new serotype identified in 2000 (Howell *et al.*, 2002) and its variation at nucleotide level (and amino acid level) with other reference strains varied: S1REF*, 1.4% (2.5%); S1REF, 1.3% (2.5%); S2REF, 2% (3.3%); S3REF 23.8% (14.2%); S4REF, 7.5% (4.6%); S5REF, 25.0% (16.7%); S6REF, 24.4% (15.0%).

When comparing the nucleotide sequence datamatrixes to the amino acid sequence datamatrixes, it is seen that the maximum level of variation within serotypes is slightly higher on a nucleotide level than on an amino acid level, this is however expected as generally more variation is seen on a nucleotide level due to the degeneracy of codons. It is interesting to note that when comparing EEV to other orbiviruses, the variation at nucleotide level is lower than that seen on the amino acid level, i.e. the different orbivirus sequences are more similar on a nucleotide level than on an amino acid level.

2.3.6.2. Phylogenetic trees obtained after phylogenetic analysis of S10 nucleotide sequences and NS3 amino acid sequences

The phylogenetic trees based on S10 or NS3 give an indication of the relationships seen between EEV serotypes and between the EEV serogroup and the other orbivirus serogroups when studying nucleotide or amino acid sequences of this genome segment or protein. Two phylogenies are shown. The trees in Fig. 2.15. is based on a S10 nucleotide sequence dataset and an NS3 amino acid sequence dataset, both containing EEV sequences and selected orbivirus sequences.

When comparing the phylogenies generated from nucleotide and amino acid sequences, it could be seen that overall the nucleotide based trees had enhanced resolution of the virus clusters as indicated by higher bootstrap values for the branching patterns. This was due to the greater level of sequence variation on the nucleotide level (25.2%) compared to that observed on the amino acid level (16.7%). However, the results of both the S10 and the NS3 trees will be given in this section.

The overall topology of the S10 tree and the NS3 tree is similar and generally there is a high level of confidence in the branching patterns as indicated by the bootstrap values at branch nodes of the trees. The results of the S10 tree and the NS3 tree will be reported together in this section, and for this reason the different groupings in the trees were given the same designation.

In general, the S10 and NS3 trees indicated that the EEV S10/NS3 sequences grouped together as a distinct, independent lineage with no close relation to the AHSV, Palyam virus (Chuzan virus), BTV or EHDV serogroups. The latter serogroups were of a closer relation to one another than to that of the EEV serogroup. Two big clusters are seen and were designated I and II for the purpose of further reference, the EEV serogroup in cluster II diverged to form two distinct subclusters, A and B.

Cluster I consists of AHSV sequences, the Chuzan sequence, the BTV sequences and the EHDV sequences. In this cluster there are two subclusters, IA and IB. Subcluster IA consists of the AHSV sequences and the Chuzan sequence. Chuzan is a member of the Palyam serogroup of the orbivirus genus, and previous reports have indicated that it is related to AHSV (Yamakawa *et al.*, 1999). The AHSV sequences are all South African isolates and cluster into three groups, referred to as phenogroups, according to serotype. AHSV-1, -8 and -2 form group $\gamma$, AHSV-3 and -7 form group $\beta$ and AHSV-5, -9, -4, and -6 form group $\alpha$.

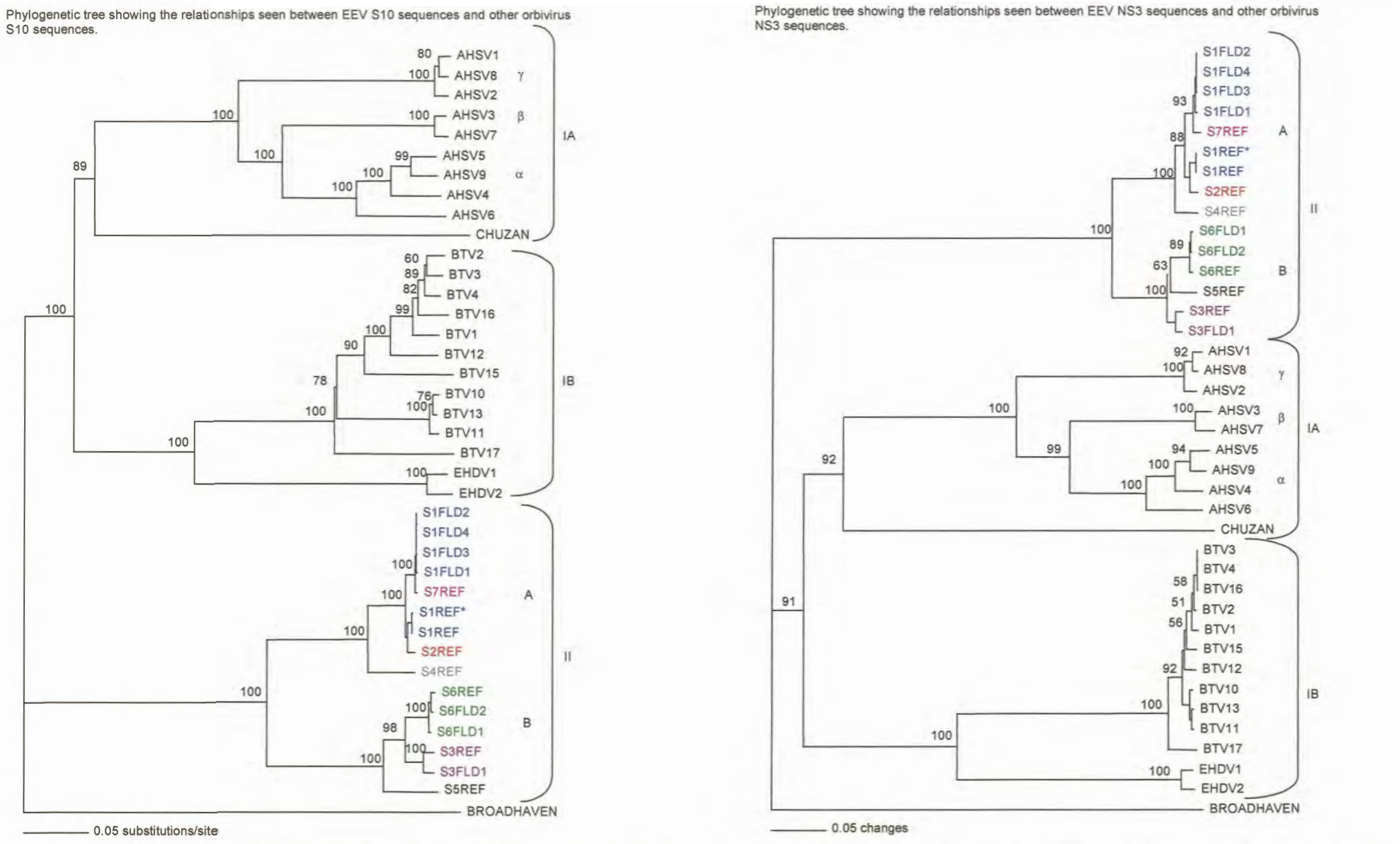


**Figure 2.15.** Phylogenetic trees of EEV S10/NS3 and other orbivirus S10/NS3 sequences. The horizontal branch lengths are indicative of the genetic distance between the sequences. The trees were constructed using the neighbour-joining method (uncorrected "p" for S10 and mean character difference for NS3) and bootstrapped with 1000 replicates. Confidence levels are as indicated. The outgroup was the Broadhaven S10/NS3 sequence.

Subcluster IB consists of the BTV sequences and the EHDV sequences. The BTV sequences in the S10 tree (Fig. 2.15.) seem to be clustered into 3 groups on one branch, BTV-2, -3, -4, -16, -1, -12, and -15 (Asian group), BTV-10, -13, and -11 (USA group 2) and BTV-17 (USA group 1), while the EHDV sequences cluster on the next branch in the same subcluster. Similarly, the BTV sequences of the NS3 tree (Fig. 2.15.) seem to be clustered into 3 groups on one branch where slight differences are observed for the clustering of individual BTV sequences within each group. The Asian group again consists of BTV -3, -4, -16, -2, -1, -15, and -12, the US group 2 consists of BTV -10, -13, and -11 and US group 1 consists of BTV-17, while the EHDV sequences cluster on the next branch in the same subcluster as for the S10 trees. The clustering of the BTV sequences into the 3 groups are based on the origin of the strains (refer to Appendix C.).

Cluster II consists of EEV sequences. In this cluster, 2 subclusters are seen. Subcluster IIA in the S10 and NS3 trees consists of S1FLD2, S1FLD4, S1FLD3, S1FLD1, S7REF, S1REF*, S1REF, S2REF with S4REF on its own branch, and subcluster IIB in the S10 trees consists of S6REF, S6FLD2, S6FLD1, S3REF and S3FLD1, and S5REF on its own branch, while subcluster IIB in the NS3 trees consists of S6FLD1, S6FLD2, S6REF, S5REF, and S3REF and S3FLD1. The trees generated for the nucleotide sequence datasets and the amino acid sequence datasets all showed a similar internal grouping of EEV S10/NS3, the only differences were seen for S5REF and S6REF. In the S10 trees S5REF is located on a separate branch in subcluster IIB, in the NS3 trees S5REF is located on a branch together with S6REF, S6FLD1 and S6FLD2, and S3REF and S3FLD1. In the S10 trees S6REF clusters closely with S6FLD2, in the NS3 trees, S6REF is replaced by S6FLD2 so that the two EEV-6 field strains cluster more closely together. These are slight differences, since the same serotypes are found within the same subclusters in both the S10 and NS3 trees. This only serves to show that the trees based on nucleotide sequences of a gene and the amino acid sequences of the same gene are not necessarily 100% identical. Subcluster IIA contains the bulk of the EEV sequences (9 of 15) and contains 4 of the seven known serotypes (EEV-1, EEV-7, EEV-2 and EEV-4). Subcluster IIB contains the rest of the sequences (6 of 15) and contains the remaining three serotypes (EEV-6, EEV-3 and EEV-5). The maximum variation seen in S10 trees and NS3 trees in subcluster IIA is 7.6% between S7REF and S4REF (S10) and 4.6% between S4REF and S1FLD1 or S7REF or S2REF (NS3). The maximum variation seen in S10 trees and NS3 trees in subcluster IIB is 8.3% between S5REF and S6FLD1 (S10) and 4.6% between S5REF and S6FLD1 or S6FLD2 (NS3). This indicates that on nucleotide level subcluster IIB is genetically more diverse than subcluster IIA, while on amino acid level the two subclusters are equally genetically diverse.

The field isolates of EEV-1 (S1FLD1 to S1FLD4) in subcluster IIA cluster closely together as can be expected, since their origin and year of isolation were similar (Table 2.3. and Appendix C). Field isolates of EEV-1 however group closer to EEV-7 (S7REF) than to the two EEV-1 reference strains (S1REF* and S1REF) (as indicated by the high bootstrap value, 100 (S10) or 93 (NS3)), although

they are all found in the same subcluster, IIA. The EEV-1 reference strains were sisters on a distal branch that were in turn closely related to the EEV-2 reference strain (S2REF). The tight grouping of the EEV-1 field isolates and S7REF was also implied by the amount of variation seen on both nucleotide and amino acid levels between S7REF and the S1FLD isolates. The grouping pattern and the small amount of variation observed could indicate a possible genome reassortment event. This type of reassortment event can also possibly explain the large variation seen within some orbivirus serotypes. However, the EEV reference strain (S1REF) was isolated in the Northern Cape in 1976, while the EEV-1 field isolates were isolated in the Western Cape in 1999. Over time the S10/NS3 sequence of the EEV-1 serotype could have changed through the process of genetic drift, so that 23 years later recent EEV-1 field isolates would not necessarily group closely with the EEV-1 reference strain, although the EEV-1 field isolates and the EEV-1 reference strains still group in the same subcluster (IIA). In the case of field isolates of EEV-6 in the S10 tree, the isolates were not members on the same distal node as found for the S1FLD isolates in subcluster IIA. The S6FLD2 strain was more closely related to the older S6REF (1991) strain than it was to the S6FLD1 of the same season. In the case of field isolates of EEV-6 in the NS3 tree, the isolates were members on the same distal node as found for the S1FLD isolates in subcluster IIA.

Cluster II thus appears to be divided into 2 subclusters according to serotype and possibly origin, EEV-1, EEV-7, EEV-2 and EEV-4 group together (W. Cape, N. Cape, KwaZulu-Natal) while EEV-6, EEV-3 and EEV-5 (Gauteng, N. West, Limpopo) form the second group. In subcluster IIB, the viruses group according to serotype into three subclusters. Field isolates generally grouped tightly with the corresponding reference strains, e.g. S6REF and S6FLD1 and S6FLD2, and S3REF and S3FLD1, the exception is the field isolates of EEV-1 that group with the EEV-7 reference strain.

The phylogenetic trees are based on the sequences of only 15 EEV isolates and more sequences are needed to confirm the analysis, specifically the groupings within the EEV serogroup. The datasets were also analysed using the neighbour-joining method (HKY85), as well as maximum parsimony and the results obtained showed the same overall topology as discussed above with the same grouping in the EEV serogroup, i.e. two subclusters containing the same strains in each subcluster. Overall, the bootstrap values obtained in these analyses were generally slightly lower than that seen for the neighbour-joining (uncorrected "p" or mean character difference) analysis indicating that the confidence in the branching order was lower.

## 2.4. DISCUSSION

The aim of this part of the study was to determine the nucleotide and deduced amino acid sequence of EEV NS3 genes of different serotypes and field isolates in order to identify structural features, to determine the level of variation seen in S10/NS3 of EEV, and to compare the observed structural

features and levels of variation with that observed for the cognate gene and gene product of other orbiviruses.

Previous studies with other orbiviruses (e.g. Gonzalez and Knudson, 1987, 1988; Brown *et al.*, 1988; Pedley *et al.*, 1988; Viljoen and Huismans, 1989; Bremer *et al.*, 1990; Brown *et al.*, 1991) suggest that the use of 1% agarose gels for the analysis of the genomic dsRNA give patterns of orbivirus genome segment migration which correlate with the serogroup rather than serotype classification of these viruses. This makes agarose gel electrophoresis useful for the initial characterization of orbivirus isolates prior to identification by serum neutralization tests (Pedley *et al.*, 1988). In this investigation, similar dsRNA profiles were observed for all the EEV reference strains and field isolates, and differences were observed between the dsRNA profiles of EEV and other orbivirus serogroups. These observations again confirm the use of agarose gel electrophoresis to distinguish between orbivirus serogroups and should allow the preliminary identification of EEV serotypes based on their dsRNA profile.

A summary of some basic characteristics of orbivirus S10 sequences, including the sizes of the nucleotide sequences and untranslated regions are shown in Table 2.16. The sizes seen for EEV generally correlate to that observed for the other orbiviruses where S10 varies between 702 nucleotides and 822 nucleotides in length and the size of the 5' untranslated region varies between 3 bp and 20 bp. The exception is the size of the 3' untranslated region, where EEV consists of 20 basepairs compared to between 70 bp and 113 bp. The reason for the reduction in length of the 3' untranslated region of EEV compared to the other orbivirus 3' untranslated regions is unknown and would be of interest in further studies, since it has been suggested that there might be a reason for the conservation of the length of this region in other orbiviruses. For example, in BTV it has been suggested that the reason for the conservation of this long (and G + C rich) 3' non-coding region involves some sequence that are conserved as either RNA replication signals and/or for encapsidation, while other areas may be involved in formation of "head-to-tail" concatemers of this RNA segment (Lee and Roy, 1986; Gould, 1988; Hwang *et al.*, 1992). There is also considerably sequence conservation at the 3' end of the AHSV S10 sequence (De Sá *et al.*, 1994), again indicating that within the *Orbivirus* genus the 3' non-coding region of S10 could have some unknown function.

As already indicated in Table 2.16., the EEV genome segment 10 sequences determined in this chapter potentially encode two proteins, NS3 and NS3A, using two conserved in-phase initiation codons. The contexts surrounding the first and second initiation codons of all the EEV isolates (Tables 2.11. and 2.17.) were investigated and did not conform to the Kozak consensus sequence (Kozak, 1981, 1984, 1989). The Kozak consensus sequence for the optimal context for initiation of translation in higher eukaryotes is GCC$^{A}/_{G}$CCAUGG and purines (preferably an A in position -3 and a G in position +4) have the strongest effects, modulating translation at least ten-fold (Kozak, 1989; Huismans *et al.*, 1992). The context of the first initiation codon of EEV corresponds well with that of

the consensus sequence, except that a pyrimidine (T) is seen at position +4 instead of a purine (G). The context of the second initiation codon of EEV does not correspond as well with the consensus sequence as the first initiation codon, the sequence preceding the ATG differs from the consensus sequence and a pyrimidine (C) is seen in position -3 instead of an purine (A or G). Of the two initiation codons, the first is probably the preferred initiation codon, although it is not in the optimal context. This conforms to the "leaky scanning model" (Kozak, 1989) which proposes that inefficient recognition of the first AUG by ribosomes scanning the 5' terminus of the RNA sense strand results in their migration to the second codon and thus initiation of translation at the second AUG codon, in the case of EEV S10 possibly leading to the synthesis of NS3 and NS3A with the amount of NS3 synthesized exceeding that of NS3A. The abovementioned situation however remains a hypothesis since to date no investigation has been performed regarding the synthesis of EEV NS3/NS3A. The expression of EEV NS3/NS3A will be however be investigated in Chapter 3.

Two in-phase initiation codons coding for two conserved methionine residues have also been identified in a number of other orbivirus S10 sequences. The location and context of the two initiation codons in the different orbiviruses are indicated in Table 2.17. In the case of AHSV, the potential initiation codons of the AHSV serotypes deviate from the optimum context for translation by not having a purine (G) at position +4 (Van Staden and Huismans, 1991; Huismans *et al.*, 1992; Sailleau *et al.*, 1997). Although neither AUG is in the optimal context for initiation of translation, the first codon is the major initiation site in eukaryotic cells because of its position relative to the 5' end of the mRNA (Kozak, 1991). In the case of BTV, the first AUG codon also does not conform to the Kozak consensus sequence owing to the presence of a pyrimidine (C) at position +4. The second AUG initiation codon, however, is closer to the consensus sequence, as it has a purine (A) at position +4. In the case of EHDV, the two in-frame initiation codons also do not conform to the optimum context described by Kozak, although the first AUG codon is closer to the optimum than the second (Jensen *et al.*, 1994). A similar observation is made for Chuzan virus where the flanking sequences of the initiation codons of segment 10 also deviate from the Kozak consensus sequence (Yamakawa *et al.*, 1999).

The sequence context of the two proposed initiation codons for BTV conform to the proposed "leaky scanning model" (Kozak, 1989), and it was suggested by Gould (1988) that the synthesis of NS3A and the resulting low level synthesis of NS3 *in vivo* may be due to this factor. A similar situation could apply to AHSV, and this could depress the amount of NS3 being synthesized at the level of translation initiation. The AHSV-3 NS3 gene also contains two G+C rich sequences at its 3' non-coding terminus at nucleotide position 688 to 712 and 737 to 758 (Van Staden and Huismans, 1991; Van Staden *et al.*, 1995) similar to S10 of BTV-1 (Gould, 1988). These sequences have been proposed to impede RNA transcription (Gould, 1988) and are therefore connected with the relative paucity of the NS3 protein and its mRNA in orbivirus infected cells.

**Table 2.16.** A summary of some basic characteristics of other orbivirus S10 and NS3 sequences

| Orbivirus | Number of basepairs | Size of untranslated regions | | Number of amino acids and pI | | Molecular weight in kDa | | References |
|---|---|---|---|---|---|---|---|---|
| | | 5' | 3' | NS3 | NS3A | NS3 | NS3A | |
| AHSV-1, -2, -and 8 | 764 bp | 19 bp | 91 bp | 218 aa | 207 aa | 24 K | 23 K | Van Staden and Huismans, 1991; Moss *et al.*, 1992; De Sá *et al.*, 1994; Van Staden *et al.*, 1995; Sailleau *et al.*, 1997; Martin *et al.*, 1998; Van Staden *et al.*, 1998; Van Niekerk *et al.*, 2001a; Van Niekerk *et al.*, 2001b |
| AHSV-3, and -4 | 758 bp | 19 bp | 88 bp | 217 aa | 206 aa | 24 K | 23 K | |
| AHSV-5 and -7 | 758 bp | 19 bp | 86 bp | 217 aa | 206 aa | 24 K | 23 K | |
| AHSV-6 | 755 bp (incomplete) | 17 bp | 84 bp | 217 aa | 206 aa | 24 K | 23 K | |
| AHSV-9 | 756 bp | 18 bp | 91 bp | 217 aa | 206 aa | 24 K | 23 K | |
| Broadhaven virus | 702 bp | 17 bp | 70 bp | 205 aa; +16.5 | 199 aa; +17.5 | 22 K | 21 K | Moss *et al.*, 1992 |
| BTV | 822 bp | 19 bp | 113 bp | 229 aa; + 4.5 or +5.5 or +6.6 | 216 aa; +5.5 | 20 K<br>28 K<br>25 K<br>26 K | 15 K<br>25 K<br>24 K<br>25 K | Mertens *et al.*, 1984; Lee and Roy, 1986; Gould, 1988; Van Dijk and Huismans, 1988; French *et al.*, 1989; Wade-Evans, 1990; Van Staden and Huismans, 1991; De Mattos *et al.*, 1992b; Hwang *et al.*, 1992; Pierce *et al.*, 1998; Bansal *et al.*,1998; Bonneau *et al.*, 1999 |
| Chuzan virus | 728 bp | 18 bp | 77 bp | 211 aa; +9.19 | 200 aa; +8.94 | 24 K | 22 K | Yamakawa *et al.*, 1999 |
| EEV-1 to -7 | 759 bp | 16 bp | 20 bp | 240 aa | 222 aa | 27 K | 25 K | This investigation |
| EHDV-1 | 809 bp | 20 bp | 103 bp | 228 aa | 214 aa | 26 K<br>25 K | 25 K<br>24 K | Van Staden and Huismans, 1991; Jensen *et al.*, 1994; Jensen and Wilson, 1995 |
| EHDV-2 | 809 bp | 20 bp | 105 bp | 228 aa; +4.42 | 214 aa; +4.42 | 25 K | 24 K | |
| Palyam virus | NA | 18 bp | NA | NA | NA | 26 K | 25 K | Van Staden and Huismans, 1991; Moss *et al.*, 1992 |
| St. Croix River virus | 764 bp | 3 bp | 86 bp | 224 aa | NA | 24 K | NA | Attoui *et al.*, 2001 |

NA indicates information not available

**Table 2.17.** The location and context surrounding the first and second initiation codons of different orbiviruses (adapted from Wu *et al.*, 1992)

| Virus | First AUG codon[a] | | Second AUG codon[b] | | Termination codon position[c] | Reference |
|---|---|---|---|---|---|---|
| | Position | Sequence | Position | Sequence | | |
| AHSV-1 | 20-22 | GUCAUGA | 50-52 | UAUAUGU | 674-676 | Van Staden and Huismans, 1991; De Sá *et al.*, 1994; Sailleau *et al.*, 1997 |
| AHSV-2 | 20-22 | GUCAUGA | 50-52 | UAUAUGU | 674-676 | |
| AHSV-3 | 20-22 | GUCAUGA | 53-55 or 50-52 | AUGAUGC | 671-673 | |
| AHSV-4 | 20-22 | GUCAUGA | 53-55 | AGUAUGC | 671-673 | |
| AHSV-5 | 20-22 | GUCAUGA | 53-55 | AGCAUGC | 658-660 | |
| AHSV-6 (incomplete) | 18-20 | GUCAUGA | 51-53 | AGUAUGC | 669-671 | |
| AHSV-7 | 20-22 | GUCAUGA | 50-52 | UAUAUGA | 670-672 | |
| AHSV-8 | 20-22 | GUCAUGA | 50-52 | UAUAUGU | 674-676 | |
| AHSV-9 | 19-21 | GUCAUGA | 52-54 | AGUAUGC | 670-672 | |
| Broadhaven virus | 18-20 | ACAAUGC | 36-38 | GAGAUGA | NA | Moss *et al.*, 1992 |
| BTV-1 | 20-22 | GCCAUGC | 59-61 | AAAAUGA | 707-709 (TGA or TAA) | Lee and Roy, 1986; Gould, 1988; Van Dijk and Huismans, 1988; Wade-Evans, 1990; De Mattos *et al.*, 1992b; Hwang *et al.*, 1992; Pierce *et al.*, 1998; Bonneau *et al.*, 1999 |
| BTV-10 | 20-22 | GCCAUGC | 59-61 | AAGAUGA | 707-709 (TGA or TAA) | |
| Chuzan virus | 19-21 | GAAAUGT | 52-54 | GCUAUGA | 652-654 | Yamakawa *et al.*, 1999 |
| EEV-1 to -7 | 17-19 | GCCAUGU | 71-73 | CUAAUGG | 737-739 | This investigation |
| EHDV-1 | 21-23 | AUCAUGC | 63-65 | GAAAUGA | 705-707 | Jensen and Wilson, 1995 |
| EHDV-2 | 21-23 | GGCAUCAUGC | 63-65 | AUCGAAAUGA | 705-707 | Jensen *et al.*, 1994 |
| Palyam virus | 19-21 | GACAUGU | 52-54 | GCGAUGA | NA | Van Staden and Huismans, 1991 |

[a] Sequences representing the putative initiation codons are underlined, and nucleotides at positions -3 and +4 relative to the AUG are in bold type

[b] Another in-phase initiation codon is present at nucleotides 50-52; it is not possible to predict which one would be used to initiate synthesis of NS3

[c] NA indicates information not available

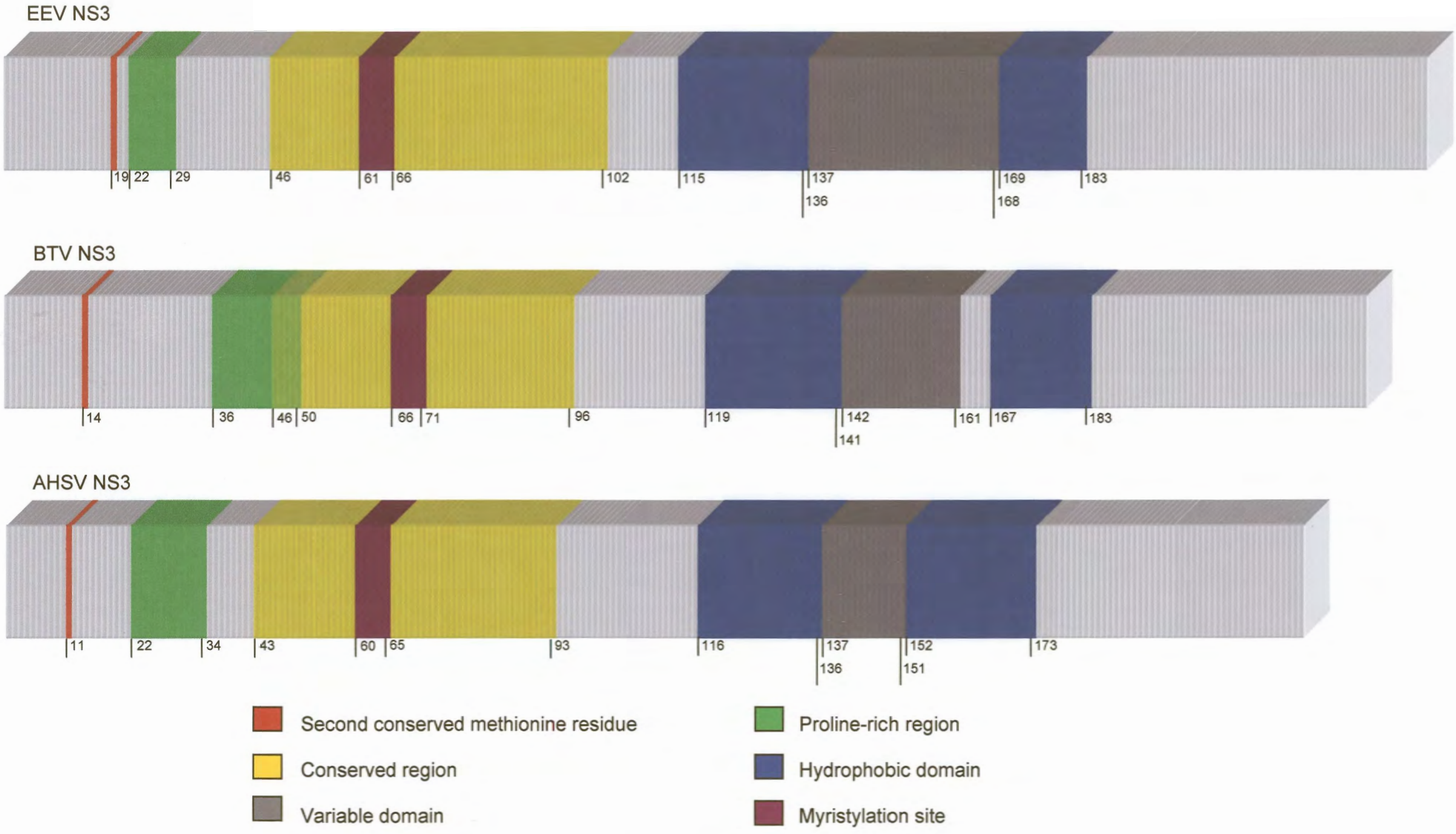


**Figure 2.16.** A comparison of conserved structural characteristics and protein motifs of NS3 proteins of selected orbiviruses.

A summary of some basic characteristics of orbivirus NS3 sequences, including the sizes of the deduced amino acid sequences and molecular weight of orbivirus NS3/NS3A proteins are also shown in Table 2.16. The sizes seen for EEV generally correlate to that observed for the other orbiviruses where NS3 varies between 205 and 229 amino acids in length and NS3A varies between 199 and 216 amino acids in length. It is interesting to note that the EEV NS3 protein is 11 amino acids longer than the biggest NS3 protein, which is that of BTV. The difference seen in size between EEV NS3 and NS3A was 18 amino acids, while for the other orbiviruses it varied between 6 and 14 amino acids. The variation seen in the estimates of molecular weight (e.g. the BTV NS3 protein) can be explained by differences in the experimental procedures used by the different research groups. In the case of BTV NS3 and NS3A (26 K and 24 K), glycosylated forms with higher molecular weights of 30 K and 29 K have been observed in the T7 transient expression system and BTV infected cells (Wu *et al.*, 1992). This could also explain the variation in size of the different estimates.

Several structural characteristics and different protein motifs were identified in the EEV Bryanston (EEV-1) (S1REF*) amino acid sequence (section 2.3.3.1.) which include hydrophobic regions, transmembrane helices, coiled-coil regions, a leader or signal sequence, glycosylation motifs, myristylation motifs, a putative virulence motif, antigenic regions and a proline-rich region. A number of these structural characteristics or protein motifs are conserved within the EEV serogroup (refer to Fig. 2.13. and Table 2.13.) and include features such as the presence of a second in-phase initiation codon or methionine residue, a proline-rich region, a glycosylation motif at amino acid position 55 to 58, a myristylation site at amino acid position 61 to 66, a positively charged region, two hydrophobic domains and a number of cysteine residues. A number of these structural characteristics and protein motifs are also conserved within the *Orbivirus* genus and a comparison of these structural characteristics and protein motifs or domains in three orbivirus serogroups are shown in Fig. 2.16.

The locations of the hydrophobic regions and transmembranous helices in the NS3 proteins of a number of different orbiviruses are shown in Table 2.18. AHSV contains 2 hydrophobic regions in the C-terminal half of the NS3 protein at position 116 to 137 (45% identity among 9 serotypes) and position 154 to 170 or 154 to 176 (80% identity among 9 serotypes for 154 to 176 region) (Van Staden and Huismans, 1991; De Sá *et al.*, 1994; Van Staden *et al.*, 1995; Sailleau *et al.*, 1997; Martin *et al.*, 1998; Van Staden *et al.*, 1998). EEV NS3 similarly contains two conserved hydrophobic regions at amino acid position 115 to 136 (63.63% identity among 7 serotypes, Fig. 2.13.) and 169 to 183 (80% identity among 7 serotypes, Fig. 2.13.). The conserved nature of these hydrophobic regions possibly indicates a functional significance. In the case of AHSV NS3, both hydrophobic domains have been implicated in the cytotoxic effect of AHSV NS3. When either one of these regions in AHSV NS3 was mutated to abolish its potential to form a transmembrane region, the cytotoxic effect of the protein on insect cells was significantly reduced compared to the wild-type AHSV NS3 protein. This was probably due to an alteration in the protein's structure which rendered it incapable of interacting with the cellular membrane components (Van Staden *et al.*, 1998; Van Niekerk *et al.*, 2001a).

**Table 2.18.** Hydrophobic regions and transmembrane helices identified in a number of different orbivirus NS3 proteins

| Virus | Amino acid residues[a] | | Reference |
|---|---|---|---|
| | Region 1 | Region 2 | |
| AHSV | 116 – 137 | 154 – 170 | Van Staden and Huismans, 1991; De Sá *et al.*, 1994; |
| | 116 – 146 | 154 – 176 | Van Staden *et al.*, 1995; Sailleau *et al.*, 1997; |
| | | 164 – 186 | Van Staden *et al.*, 1998; Martin *et al.*, 1998; |
| AHSV-3 | **114 – 136** | **157 – 170** | Van Niekerk *et al.*, 2001b |
| Broadhaven virus | **97 – 123** | **129 – 145** | Moss *et al.*, 1992 |
| BTV | 114 – 130 | 159 – 175 | Lee and Roy, 1986; Hyatt *et al.*, 1991; Hwang *et al.*, 1992; |
| | 118 – 148 | 156 – 181 | Wu *et al.*, 1992; Bansal *et al.*,1998; Pierce *et al.*, 1998; |
| | 115 – 145 | 160 – 185 | Bonneau *et al.*, 1999; |
| | 118 – 141 | 162 – 182 | |
| BTV-10 | **121 – 144** | **167 – 182** | |
| Chuzan virus | 114 – 137 | 146 – 167 | Yamakawa *et al.*, 1999 |
| EEV-1 to -7 | 115 – 136 | 169 – 183 | This investigation |
| | **120 – 132** | **171 – 178** | |
| EHDV-1 | **115 – 135** | **158 – 178** | Jensen and Wilson, 1995 |
| EHDV-2 | **114 – 130** | **159 – 175** | Jensen *et al.*, 1994 |
| Palyam virus | **113 – 136** | **151 – 164** | Van Staden and Huismans, 1991 |

[a] Transmembrane helices are indicated in bold type

The presence of transmembrane regions or transmembrane helices within the NS3 proteins of a number of orbiviruses indicates that these proteins interact with membranes in a specific way that is referred to as the membrane topology of the protein. The membrane topologies of different orbiviruses and rotavirus non-structural glycoprotein NS28 (NSP4) were discussed in sections 1.4.3.2. and 1.5.7. The most likely orientation of the different orbivirus NS3 proteins was with the N- and C-termini on the cytoplasmic side of the membrane, with each of the hydrophobic regions spanning the membrane (Fig. 1.3.). Only the localization of the first transmembrane region differed in the alternative model proposed for EHDV (Fig. 1.4.).

The membrane topology of the EEV Bryanston (EEV-1) (S1REF*) NS3 protein was investigated (section 2.3.3.1.) and it was predicted that each of the two transmembrane helices span the cell membrane so that the N- and C-termini are located on the inside (cytoplasmic side) of the membrane leaving a large part of the protein available for interactions with components of virus particles. The region between the two membrane spanning segments is exposed on the outside (luminal side) of the membrane. This model for the EEV NS3 membrane topology correlates well with the model predicted for AHSV NS3 (Van Staden *et al.*, 1995) and BTV NS3. However, the first hydrophobic region at position 115 to 136 of the EEV Bryanston (EEV-1) (S1REF*) NS3 protein contains a potential signal

peptide sequence with a potential cleavage site for this sequence (section 2.3.3.1., Table 2.19.) and an N-linked glycosylation site has been identified at position 55 to 58 (Table 2.19.), implying that this region may be located on the luminal side of the membrane. These two observations correspond to observations made for the EHDV-1 NS3 topology model. The membrane topology of EEV NS3 therefore contains elements of both membrane topology models as predicted for AHSV (Fig. 1.3.) and EHDV (Fig. 1.4.) and it is not possible to exclude any one of the two models, the specific model that applies to EEV NS3 will have to be determined experimentally. Van Niekerk *et al.* (2001a) re-evaluated the membrane topology of AHSV NS3, and as for EEV NS3, did not find any evidence to exclude any of the two models. It was however evident from their study that only a very specific membrane-associated conformation of AHSV NS3, dependent on the structure of both hydrophobic domains, made AHSV NS3 cytotoxic to host cells.

The region between the two membrane spanning segments of EEV NS3, predicted to be exposed, has also been predicted to have antigenic properties. Antigenic regions were identified at amino acid position 146 to 158 (between HDI and HDII). This region also showed a high degree of variability (61.54%) between 15 EEV NS3 amino acid sequences (Fig. 2.13.) and was located within the most variable region within the amino acid alignment, i.e. at amino acid position 137 to 168 (Table 2.13.) The corresponding region in AHSV NS3 (amino acid position 137 to 154, Van Staden *et al.*, 1995) is also predicted to be an antigenic determinant, and showed a high degree of variability (94%) between 13 AHSV NS3 amino acid sequences (Sailleau *et al.*, 1997). Antigenic determinants have also been identified for EEV NS3 at amino acid positions 51 to 67 and 97 to 108 (before HDI), and at amino acid positions 188 to 235 (after HDII). BTV similarly contains a region with antigenic qualities at amino acid position 29 to 50 (before HDI) (Hyatt *et al.*, 1992). In the case of AHSV, NS3/NS3A has been found to elicit an antibody response (Laviada *et al.*, 1993), and as such has been used as a marker to distinguish between horses vaccinated with a non-replicating vaccine and naturally infected animals or animals vaccinated with a live, attenuated AHSV vaccine (Laviada *et al.*, 1994, 1995; Bougrine *et al.*, 1998). If EEV NS3 elicits a similar immune response as suggested by the presence of antigenic regions, it is possible that EEV NS3 can similarly be used as a marker, specifically in cases where there are symptoms that could be ascribed to either AHS or EE. The presence of antigenic regions in EEV Bryanston (EEV-1) (S1REF*) NS3 also allows the preparation of monospecific antiserum directed at EEV NS3 that can be used in Western blotting or immunoprecipitation to study EEV NS3 expression or its subcellular localization.

A number of cysteine residues have been identified in the EEV NS3 protein sequences of which the residues at position 65 (in conserved region, part of myristylation site, refer to Table 2.13.), position 149 and 161 (between HDI and HDII, in exposed region), and position 184 (after HDII) are conserved. Cysteine residues are usually found to play a role in the stability, folding and maturation of viral proteins by forming intra-chain disulphide bonds (Doms *et al.*, 1993). These residues can however

also be involved in the formation of inter-chain disulphide bridges and can assist in the dimerization of proteins to form channel-like structures, specifically when a cysteine residue is found within or close to a region predicted to form a transmembrane helix. An example of this was seen in the membrane topology prediction of EHDV-1 NS3 (Jensen and Wilson, 1995). To date there is however no biochemical evidence for the dimerization of the NS3 protein of EHDV-1. Some of the programs used to predict the hydrophobic regions of EEV NS3 included the cysteine at position 184 as part of the second hydrophobic region or transmembrane helix, e.g. TMPRED (Appendix B), PHDhtm (Rost *et al.*, 1996) and TopPred2 (Cserzo *et al.*, 1997) and as such it is possible that in the case of EEV NS3 two helices could combine to form a channel, stabilized by a disulphide bridge as seen for EHDV-1 NS3. This possibility of potential dimerization is further strengthened by the fact that the cysteine residue at position 184 is conserved for all the EEV isolates investigated in this study (Table 2.13.).

Conserved cysteine residues have also been observed for other orbiviruses. BTV has 2 conserved cysteine residues at position 137 and 181, both located in one of the two hydrophobic regions of BTV (Table 2.18.) and the presence of only two conserved cysteine residues suggest the formation of a single disulphide bond within the protein (Lee and Roy, 1986; Hwang *et al.*, 1992; Jensen and Wilson, 1995; Pierce *et al.*, 1998; Bonneau *et al.*, 1999). The cysteine residue at position 137 seems to be conserved, as it is found within all BTV isolates studied to date, the cysteine residue at position 181 however, is not conserved. In a recent study (Van Niekerk *et al.*, 2003), it was common in 18 out of 21 BTV NS3 sequences; the exceptions had a positively charged lysine in this position. As already mentioned, the cysteine residue at position 181 is located within a proposed membrane-spanning region, making it unlikely that the residue would be available for disulphide bond formations within the protein (Van Niekerk *et al.*, 2003). In the case of AHSV, a conserved cysteine residue is found at position 162 in all serotypes, as well as one or two additional cysteine residues depending on the serotype. AHSV-1 to -7 and -9 have 2 cysteine residues, while AHSV-8 has 3 cysteine residues. AHSV-1, -2 and -8 share a cysteine at residue 120, AHSV-3, -4 and -9 share a residue at position 122, AHSV- 4, -5 and -7 share a residue at position 123 and AHSV-6 has a residue at position 215. The two cysteine residues were located in the hydrophobic region of the protein (De Sá *et al.*, 1994; Sailleau *et al.*, 1997; Martin *et al.*, 1998). In another study, a conserved cysteine residue was observed in position 123 (all serotypes except serotype 2), at position 120 for AHSV-2 and a cysteine at position 164 was conserved amongst all serotypes (Van Niekerk *et al.*, 2001b). In Palyam NS3, one cysteine residue is found at position 182 and in Broadhaven virus 2 residues are found at position 100 and 185 (Jensen and Wilson, 1995). In the case of EHDV, one cysteine residue was observed at position 168 of EHDV-1 and position 160 of EHDV-2 (Jensen *et al.*, 1994; Jensen and Wilson, 1995). The presence of cysteine residues within the various orbivirus NS3 proteins suggests the possible formation of disulphide bridges at least within the protein as no evidence of dimerization has been found to indicate the formation of disulphide bridges between two NS3 proteins.

Signal peptides (signal sequences) have been identified for orbivirus NS3 proteins and the nature and locations of these sequences are shown in Table 2.19. As already mentioned, a potential signal peptide sequence and a potential cleavage site at amino acid position 134 for this signal peptide has been identified in EEV Bryanston (EEV-1) (S1REF*) NS3 (section 2.3.3.1.). Signal peptides control the entry of virtually all proteins to the secretory pathway, both in eukaryotes and prokaryotes. They comprise the N-terminal part of the amino acid chain and are cleaved off while the protein is translocated through the membrane. The common structure of signal peptides (20 to 40 amino acids in length) from various proteins is described as a positively charged n-region, followed by a hydrophobic h-region (e.g. a stretch of 9 hydrophobic residues) and a neutral but polar c-region. Insertion signal peptides also often have a cleavage site for removal of the signal peptide after insertion of the protein in the membrane. An additional peptide signal is then needed to target the protein to the plasma membrane.

A putative membrane targeting signal of the orbivirus NS3 protein has also been identified provided cleavage occurs (Van Niekerk *et al.*, 2001b). It is a bipartite membrane targeting signal and is located in the conserved region between amino acid 43 and 92 of AHSV NS3. It consists of a myristylation motif that is followed by 25% to 28% positively charged residues (basic amino acids) arranged in a random order that are predicted to stabilize membrane interactions in analogy to what is seen for membrane associated proteins of other viruses such as Gag of HIV-1 (Zhou *et al.*, 1994) and Src of Rous sarcoma virus (Silverman and Resh, 1992). This motif is highly conserved in the amino terminal region of other orbivirus NS3 proteins including BTV, Chuzan virus (member of the Palyam serogroup), EHDV and Broadhaven virus (Van Niekerk *et al.*, 2001b). The location of the myristylation sites within the orbivirus NS3 proteins are indicated in Table 2.20.

**Table 2.19.** Potential signal peptide sequences identified in orbivirus NS3 proteins (adapted from Jensen and Wilson, 1995)

| Virus | Positively charged region | Hydrophobic stretch | Polar region |
|---|---|---|---|
| AHSV-3 | Position 100 – 111 | Position 112 – 131 | Position 132 – 137 |
| Broadhaven virus | Position 87 – 91 | Position 92 – 113 | Position 114 – 122 |
| BTV-10 | Position 114 – 118 | Position 119 – 133 | Position 134 – 141 |
| EEV-1 to -7[a] | Position 110 – 114 | Position 115 – 136 | Position 137 – 144 |
| EHDV-1 | Position 110 – 114 | Position 115 – 130 | Position 131 – 137 |
| Palyam virus | Position 103 – 110 | Position 111 – 125 | Position 126 – 131 |

[a] Data from this investigation

**Table 2.20.** Potential myristylation motifs identified in different orbivirus NS3 proteins (adapted from Van Niekerk *et al.*, 2001b)

| Virus | Amino acid residues |
|---|---|
| AHSV | 59 – 64 and 60 – 65 |
| Broadhaven virus | 37 – 42 |
| BTV | 66 – 71 |
| Chuzan virus | 47 – 52 |
| EEV-1 to –7[a] | 61 – 66 and 225 – 230 |
| EHDV | 62 – 67 |

[a] Data from this investigation

The myristylation motif is common to a vast number of proteins and therefore does not imply that a protein is myristylated. Myristylation involves the acylation of proteins by the covalent addition of myristate (a C14-saturated fatty acid) to their N-terminal residue via an amide linkage where the N-terminal residue is a glycine. The myristylation sites shown in Table 2.20. are located within the amino acid sequence, but subsequent proteolytic processing such as the cleavage of the signal peptide, could expose an internal glycine as the N-terminal of the mature protein. The major function of protein acylation is membrane targeting and association, myristylation alone does however not provide sufficient energy to attach the protein to the phospholipid bilayer (McLaughlin and Aderem, 1995).

Two putative myristylation sites have been identified for EEV Bryanston (EEV-1) (S1REF*) NS3 at position 61 to 66 and 225 to 230. The first is followed by a positively charged region between amino acids 67 and 120, and together the first myristylation motif and the positively charged region comprise a putative bipartite membrane targeting signal. Using this information, it is possible to predict that EEV NS3 will probably be targeted to the ER from where it will later be transported via the Golgi complex to the correct destination, which is most probably the plasma membrane of the host cell as discussed in section 1.5.3. At this stage, however, this prediction is based on theoretical analysis and will have to be tested experimentally.

A protein motif that often plays a role in the proper folding and transport of proteins (Doms *et al.*, 1993), as well as determining their orientation in a membrane, is the glycosylation motif where carbohydrate moieties are added to a specific protein at specific sites (sections 1.5.2. and 1.5.3.) Potential glycosylation motifs have been identified for a number of different orbivirus NS3 proteins and their location within the amino acid sequence is indicated in Table 2.21.

**Table 2.21.** Potential glycosylation motifs identified in different orbivirus NS3 proteins

| Virus | Amino acid residues | Type of glycosylation | Reference |
|---|---|---|---|
| AHSV-1, -2 and -8 | 122 –124 | N-linked | De Sá *et al.*, 1994;Van Staden *et al.*, 1995; Sailleau *et al.*, 1997 |
| AHSV-4, -5, -6 and -9 | 9 – 11 | N-linked | |
| AHSV-7 | 14 – 16 | N-linked | |
| Broadhaven virus | 11 | O-linked | Moss *et al.*, 1992; Jensen and Wilson, 1995 |
| BTV | 63 - 65<br>150 - 152 | N-linked | Hyatt *et al.*, 1991; Hwang *et al.*, 1992; Wu *et al.*, 1992; Jensen and Wilson, 1995; Bansal *et al.*,1998; Bonneau *et al.*, 1999; Pierce *et al.*, 1998 |
| | 46 – 50 | O-linked | |
| EEV-1 to -7 | 55 – 58<br>150 – 153 | N-linked | This investigation |
| EEV-2 | 222 – 225 | N-linked | |
| EHDV-1 | 33 – 37<br>42 – 46 | N-linked | Jensen and Wilson, 1995 |
| EHDV-2 | 59 – 61<br>146 – 148 | N-linked | Jensen *et al.*, 1994; Jensen and Wilson, 1995 |
| | 33 – 37<br>42 – 46 | O-linked | |
| Palyam virus | 54 | | Jensen and Wilson, 1995 |

BTV NS3 contains 2 potential N-linked glycosylation sites of which only the site at position 150 to 152 is glycosylated (Bansal *et al.*, 1998). Wu *et al.* (1992) showed that NS3 and NS3A of BTV are glycosylated in mammalian cells and modified by the addition of polylactosaminoglycan-like complex carbohydrate chains. The proteins are then probably processed in the Golgi apparatus after they have achieved a transport-competent form in the ER, and from the Golgi apparatus the proteins are most probably targeted towards the plasma membrane. Bansal *et al.* (1998) showed that glycosylation at position 150 was necessary for the correct processing of NS3 which involved the transport from the Golgi complex to the cell surface. The modification of BTV NS3/NS3A proteins by N-linked glycosylation, indicates that the products are synthesized by membrane-attached ribosomes and this makes it highly likely that the proteins are integral membrane glycoproteins (Wu *et al.*, 1992). NS3/NS3A of AHSV-1, -2 and AHSV-4 to -9 contain 1 potential glycosylation site at varying positions (Table 2.21.); AHSV-3 does not contain any sites. No experimental evidence of glycosylation has been found for AHSV-3, AHSV-4 or AHSV-9 when NS3 was expressed in the baculovirus system in the presence of tunicamycin, which inhibits glycosylation (Grubman and Lewis, 1992; Van Staden *et al.*, 1995). EEV Bryanston (EEV-1) (S1REF*) NS3 protein contains two potential N-linked glycosylation sites at amino acid position 55 to 58 (before HDI) and at position 150 to 153 (between the two hydrophobic domains). The glycosylation site at position 150 to 153 is located within the

predicted exposed region of the NS3 protein. Glycosylation may play a role in proper folding of the various NS3/NS3A proteins for transport or alternatively prevent proteolysis during transport, as the non-glycosylated proteins are not as stable as the glycosylated proteins. The presence of a putative glycosylation site is however not sufficient to conclude that glycosylation takes place, due to the fact that the folding of the protein plays an important role in the regulation of N-glycosylation. The requirement for carbohydrates for transport of membrane proteins is also not universal and is highly protein specific (Doms *et al.*, 1993). This means that differences in glycosylation of orbivirus NS3 proteins would not necessarily influence their possible similar function in viral morphogenesis, and it therefore remains to be experimentally determined which, if any, of the potential sites observed in the different orbivirus NS3 proteins are in fact glycosylated.

The NS3 proteins of other orbiviruses such as BTV and AHSV have been implicated in the final stages of viral morphogenesis by enabling the release of progeny viruses from infected cells (section 1.4.3.). In the case of AHSV NS3 a number of characteristic features have been identified in this protein that could allow the protein to perform this function. These characteristics include two hydrophobic domains followed by a cluster of basic amino acids (residues 182 to 190) and a region at amino acid 95 to 118 that is predicted to have a high potential for forming a coiled-coil structure based on the presence of a heptad-repeat motif (Van Niekerk *et al.*, 2001a). In general, such coiled-coil regions are implicated in protein stabilization and multimerization (Cohen and Parry, 1986). These characteristics are also found in a group of proteins referred to a viroporins which are small, integral membrane proteins with at least one hydrophobic membrane-spanning domain that affects an altered membrane permeability of the host cell. These proteins usually contain stretches of basic amino acids, and tend to oligomerize, forming hydrophilic pores in the membrane. The main function of these proteins is to help in the release of progeny viruses from infected cells (Van Niekerk *et al.*, 2001a). EEV Bryanston (EEV-1) (S1REF*) NS3 is similar to AHSV NS3 in that it also contains two hydrophobic domains followed by a cluster of basic amino acids (residues 204 to 206), as well as a predicted coiled-coil region around position 100, immediately adjacent to the first predicted transmembrane helix at position 120 to 132. The presence of these features suggests that EEV NS3 may also possibly function as a viroporin and may play a role in the release of progeny viruses from infected cells. This is further supported by the identification of a protein motif, TYPE3IMAPROT, in EEV Bryanston (EEV-1) (S1REF*) NS3 (section 2.3.3.1.). This motif is a 6-element fingerprint that provides a signature for the Type III secretion system inner membrane A protein family, a bacterial protein family that is involved in virulence. The type III protein subunits have a function that is reminiscent of the function of viroporins since they are located in the outer membrane and translocate secreted proteins through a channel-like structure. EEV NS3 also share a number of similarities with NSP4 (NS28), a non-structural membrane glycoprotein encoded by genomic segment 10 of rotaviruses (Au *et al.*, 1989; Meyer *et al.*, 1989) involved in viral morphogenesis by functioning as a virus-encoded receptor to which a structural protein can bind to mediate a membrane budding event

(Meyer *et al.*, 1989) (refer to section 1.5.7.). These shared features include the presence of hydrophobic domains, transmembrane helices, a similar membrane topology model and potential glycosylation sites in similar regions of the protein. These similarities again argue towards a role for EEV NS3 in the release of viral progeny from infected cells.

A proline-rich region has been identified for EEV NS3 between position 22 and 29, where 5 proline residues were observed in a stretch of 8 residues. Similar proline-rich regions have been observed for other orbivirus NS3 sequences. AHSV NS3 contains a cluster of 5 prolines within a region of 13 amino acids (position 22 to 34), a similar region is also conserved in BTV, Palyam, Broadhaven virus and EHDV (Hwang *et al.*, 1992; Moss *et al.*, 1992; Jensen *et al.*, 1994; Van Staden *et al.*, 1995; Van Staden *et al.*, 1998; Van Niekerk *et al.*, 2001b). In the case of EHDV, a cluster of 6 proline residues is observed at position 36 to 50 (Jensen *et al.*, 1994), while BTV NS3 contains 11 prolines, and 6 of the 11 prolines exist as a cluster near the amino terminus within a region of 15 amino acids (position 36 to 50). This cluster of proline residues should import structural constraints and effects on the NS3/NS3A proteins and should also have some role in the biological function of NS3/NS3A (Hwang *et al.*, 1992; Pierce *et al.*, 1998; Bonneau *et al.*, 1999). Proline has a cyclized side chain, this places tight structural restrictions on the protein conformation of proline-rich regions, and such regions frequently play an important part in binding proteins. Referring back to the predicted membrane topology model for EEV NS3, the proline-rich region falls in a part of the protein that is found in the cytoplasm. In the case of rotavirus NSP4, the part of the protein located in the cytoplasm has been implicated in binding other proteins in order to perform its function as a receptor (section 1.5.6.). The presence of a region frequently involved in binding in EEV and other orbivirus NS3 proteins indicate that this protein can possibly function as a receptor as in the case of NSP4, especially since AHSV and BTV NS3 have been shown to be involved in viral release. In addition, another region of the BTV NS3 protein, the 13 amino acids between the initiation codon for NS3 and the initiation codon for NS3A at the N-terminal of the NS3 protein, has recently been shown to have a binding function (Beaton *et al.*, 2002). This highly conserved region interacts with a cellular protein, p11, to mediate virus release (section 1.4.3.2.). In this investigation the region between the initiation codon for NS3 and NS3A of EEV is also highly conserved with only 2 amino acid differences within the 18 amino acid region in an alignment of 15 amino acid sequences (refer to Fig. 2.13.). This indicates a possible similar binding function for the 18 amino acid residues at the N-terminal of EEV NS3, again indicating that EEV NS3 could possibly function as a receptor and could be involved in viral release.

A conserved region was identified for EEV NS3 at position 46 to 102; this region also contained a conserved glycosylation motif, a conserved myristylation motif and a positively charged region. Similar conserved regions have been identified in other orbiviruses. AHSV NS3 contains a conserved amino acid region in the N-terminal half of the protein at position 43 to 92 or position 46 to 90 (79% identity, 96% similarity, conserved between 9 AHSV serotypes), a similar region is also found to be conserved

among NS3 proteins of BTV-10 (at position 49 to 98), Palyam virus, EHDV, Broadhaven virus and Chuzan virus (Van Staden and Huismans, 1991; Moss *et al.*, 1992; Jensen *et al.*, 1994; Sailleau *et al.*, 1997; Pierce *et al.*, 1998; Van Staden *et al.*, 1998; Yamakawa *et al.*, 1999; Van Niekerk *et al.*, 2001a, 2001b).

A variable region was also identified for EEV NS3; the region was located between amino acid position 137 and 168. This region falls between the two proposed hydrophobic domains and is also predicted to be exposed extracellularly and to have antigenic qualities. Other orbivirus NS3 proteins contain similar variable regions. In the case of AHSV NS3, most of the amino acid differences are grouped within three regions, the first 43 residues at the N-terminal, the region between residues 93 and 153, and the last 15 residues at the C-terminal. The most variable region (82.4% variation) is located between amino acid 136 to 153 and the NS3 membrane associated model proposed by Van Staden *et al.* (1995) maps this region to the exterior of the cell membrane. In BTV the most variable regions are located between the two hydrophobic domains (approximately 25% variation between position 142 to 161) and in the final 20 residues (position 211 to 226) (Van Niekerk *et al.*, 2003). As already mentioned, the variable regions are proposed to be exposed extracellularly and contain antigenic regions and as such could play a role in the virus-host immune interaction. These variable regions are also a reservoir of variation within the orbivirus NS3 protein, and as will be discussed in the following paragraphs, the levels of variation in the NS3 protein has certain important implications as well as possible uses.

The use of sequencing and phylogenetic analysis, have allowed researchers to determine the relatedness of viruses within a serogroup (between serotypes of the same virus) or viruses within a family (between serogroups of viruses) by determining the level of variation observed between members of the same serogroup and between members of different serogroups. In the case of the BTV serogroup the levels of S10/NS3 variation observed has been quantified in a number of studies (e.g. Gould, 1988; Wade-Evans, 1990; De Mattos *et al.*, 1992a; Huismans *et al.*, 1992; Hwang *et al.*, 1992; Pierce *et al.*, 1998) and the most recent results are shown in Table 2.22. and Fig.2.17. These studies have shown that BTV NS3 is highly conserved and it has been proposed that little variation is tolerated in the NS3/NS3A protein of BTV because of its crucial function of viral release in BTV morphogenesis, even though it is probably under selective pressure exerted by both the insect vector and ruminant host (Pierce *et al.*, 1998; Bonneau *et al.*, 1999). In the case of AHSV, the level of variation observed in S10 and NS3 has similarly been quantified in a number of investigations (e.g. Van Staden and Huismans, 1991; Huismans *et al.*, 1992; De Sá *et al.*, 1994; Sailleau *et al.*, 1997; Martin *et al.*, 1998; Zientara *et al.*, 1998; Sailleau *et al.*, 2000; Van Niekerk *et al.*, 2001b) and again the most recent results are shown in Table 2.22. and Fig. 2.17. The values obtained for AHSV S10/NS3 indicate a high level of variation between and within serotypes. NS3 is the second most variable AHSV protein with approximately 20% less variation than the observed 56% variation between the outer capsid protein VP2 of different serotypes in AHSV (Van Niekerk *et al.*, 2001b).

**Table 2.22.** The maximum level of variation seen on nucleotide level (S10) between different serotypes of orbivirus serogroups

| Serotypes analysed | Variation on nucleotide level | Reference |
|---|---|---|
| Between AHSV serotypes (AHSV-2 to -9) | 46% | Van Niekerk *et al.*, 2001b |
| Between BTV serotypes (BTV-10, -11, -13 and –17) | 18.2% | Pierce *et al.*, 1998 |
| Between EEV serotypes (EEV-1 to -7) | 25.2% | This investigation |
| Between EHDV serotypes (EHDV-1 and -2) | 3.7% | Jensen and Wilson, 1995 |

To date the variation found in EEV S10 and NS3 has only been studied using hybridization experiments (Viljoen and Huismans, 1989). In this investigation the variation found in S10 and NS3 was studied by determining the sequences of S10 and NS3 and performing a phylogenetic analysis (refer to section 2.3.6. and Table 2.22.). The level of variation observed for EEV NS3 is shown in comparison to that of AHSV, BTV, EHDV and rotavirus (Fig.2.17.).

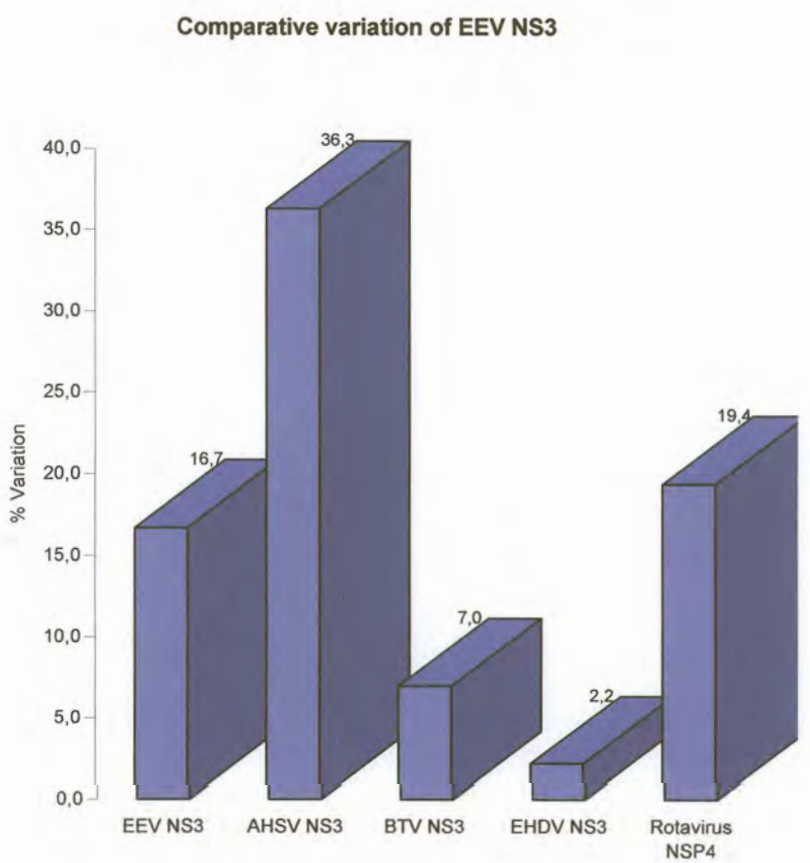


**Figure 2.17.** Comparative variation of EEV NS3. Total level of variation of EEV NS3 across serotypes is compared to that of AHSV NS3, BTV NS3, EHDV NS3 and rotavirus NSP4. Maximum variation percentages are indicated at the top of the bars (adapted from Van Niekerk *et al.*, 2001b).

The results obtained in this investigation indicate that EEV S10/NS3 is not highly conserved, but instead has a level of variation that is intermediate to that seen in BTV and AHSV. The limited NS3 sequence data of EEV and other orbiviruses such as EHDV, Palyam virus and Broadhaven virus however restricts a comprehensive comparison of the typical variation for this non-structural protein.

The level of variation seen for the NS3 protein between serotypes or within serotypes in the same serogroup has potential uses and significance. In the case of AHSV NS3, the level of variation observed between and within serotypes is relatively large, allowing it to be used as an epidemiological tool or marker. The large intra-serotype variation of AHSV NS3 has been found to be useful for distinguishing between sub-populations of the same specific serotype. In phylogenetic analysis performed, using sequences of NS3 proteins from recent field isolates, it has been found that the field isolates group according to their origin or geographic location (Van Niekerk *et al.*, 2001b). This could be of some advantage when outbreaks of the same serotype occur in different localities in the country and there is a request to link the outbreak to the transport of animals between these different regions. Despite the limited variation of BTV S10/NS3 it may still be possible to differentiate between vaccine and field strains of BTV and it could also have value as an epidemiological marker (Van Niekerk *et al.*, 2003). Sequencing of the comparatively small NS3 gene in outbreaks of the disease, together with serotyping, may therefore provide significant epidemiological information as to the origin of a virus involved in an outbreak, as well as to distinguish between field isolates and live attenuated vaccine strains. This has also been supported by a preliminary study and the unpublished results from an investigation of a recent outbreak of AHSV-7 in the Western Cape region of South Africa that has been free from AHSV for a number of years. Although the level of variation found in EEV S10/NS3 is not as high as that seen in AHSV S10/NS3, it can possibly be used in the same way as an epidemiological marker.

The origin of the observed variation in the NS3 protein is not clear and can include the contribution of many variables such as the interaction between the intermediate insect vector (*Culicoides* species) and the protein, founder effects, immunological pressures due to the fact that the protein is membrane associated, population size and mutation rate, errors incorporated by the reverse transcriptase during viral replication, natural selection, genetic drift (caused by mutations over time) and genetic shift (recombination caused by genomic reassortment events) (Gorman, 1983; Page and Holmes, 1998; Bonneau *et al.*, 2001; Van Niekerk *et al.*, 2001b). All of the mentioned processes could lead to diversification of S10/NS3 and can explain the sequence divergence between genome segments/proteins of viruses found in different regions as well as in the same region. The obvious question however remains that if all of the above processes play a role regarding S10/NS3 of all orbiviruses, why does BTV NS3 seem more conserved than AHSV and EEV NS3? It has been proposed that variation in BTV NS3 is limited by structural constraints important for its function (Pierce *et al.*, 1998), why these limitations do not apply to the same degree for AHSV NS3 remains uncertain

(Van Niekerk *et al.*, 2001b). EEV NS3 shows high variation but is also expected to be under functional constraint, these two conflicting factors can however both be accommodated when variation is observed on the protein sequence level, while the secondary structure and thus tertiary structure of the protein remains intact. Variation is found in or confined to regions of the protein that do not have specific structural or functional importance, other than being exposed extracellularly with a possible role in virus-host immune interactions. The large variation found in the region between the two hydrophobic membrane spanning domains illustrates that this area is able to tolerate a high level of variation (Van Niekerk *et al.*, 2001b).

The variation observed in orbivirus genome segments 10 and NS3 proteins can also be represented visually using phylogenetic trees. These trees also allow the visual representation of the relationships within serotypes and between serotypes in the same serogroup, as well as the relationships observed between different serogroups. Phylogenetic analysis has been performed for a number of different orbivirus S10/NS3 sequences, most notably AHSV and BTV.

In AHSV phylogenies based on S10/NS3, three groups (phenogroups or distinct phylogenetic lineages) can be defined in the AHSV serogroup. They are the alpha group (α) consisting of AHSV-4, -5, -6 and -9, the beta group (β) consisting of AHSV-3 and -7 and the gamma group (γ) consisting of AHSV-1, -2 and -8. Group β generally groups closer to group α than to group γ (De Sá *et al.*, 1994; Sailleau *et al.*, 1997; Martin *et al.*, 1998; Zientara *et al.*, 1998; Van Niekerk *et al.*, 2001b). In the AHSV S10/NS3 phylogenies, field isolates of the same season and same serotypes are generally closely related, those of different seasons are more distantly related. The close relatedness of the S10/NS3 sequences of viruses of the same serotype that were isolated from nearby geographic locations agrees with generally accepted epidemiological principles (Van Niekerk *et al.*, 2001b). Although 3 distinct S10/NS3 phylogenetic clusters are evident, the placement of specific AHSV S10/NS3 sequences is not exclusively defined by serotype, which indicates that variation in AHSV S10/NS3 sequences does not necessarily correlate with virus serotype. Van Niekerk *et al.* (2001b) observed that S10/NS3 of the same serotype could group in different lineages. Reassortment of AHSV genome segments may be able to explain some of this large inter-serotype variation. Barnard (1993) has observed that multiple AHSV serotypes can simultaneously be present in zebras, which creates the possibility for reassortment to occur.

In BTV phylogenies based on S10/NS3, 3 monophyletic lineages can be distinguished (Pierce *et al.*, 1998). These groups are most likely segregated based on geographic location, two of the groups represent BTV isolates from the USA, and the remaining group represents BTV isolates from Asian origin. The clustering into these groups is mostly independent of serotype, year of isolation and host species of isolation. It has been suggested that the viruses in the 3 different clusters were subjected to different evolutionary pathways leading to diversification (Bonneau *et al.*, 1999), as it is known for

BTV that considerable sequence variation can be found between viruses that evolved in isolation or in separate geographic regions (Gorman and Taylor, 1985). Evidence for reassortment of genomic segments has also been observed in BTV phylogenies based on S10/NS3, genomic reassortment is indicated by the presence of the same BTV serotypes in different groups (Pierce *et al.*, 1998).

Orbivirus NS3 sequences therefore have topotyping characteristics; i.e. the same sequence or a similar sequence is found for a gene from a region/geographic location, irrespective of the virus serotype. Distinct geographic groupings or topotypes seem to be a feature of orbiviruses and could be a consequence of co-evolution of the virus population and the specific insect vector that occur in the particular geographic region (Gould *et al.*, 1992; Bonneau *et al.*, 1999). This characteristic, together with the level of variation observed in different NS3 proteins is used in epidemiological studies to trace the origin of a virus as already described earlier in this section.

In this investigation, the EEV isolates, including reference strains and field isolates, grouped into two distinct clusters in a serotype dependant manner. This grouping may however change following a more extensive EEV S10/NS3 sequence analysis, including more sequences from different field isolates, as well as sequences from other countries such as Kenya or Botswana where EEV has also been identified. The data from this investigation indicated that the S10/NS3 of EEV circulated in two geographic pockets in SA, with an area of overlap in the Gauteng region. Viruses in cluster A were isolated from regions in SA stretching from Stellenbosch in the Western Cape (S1FLD strains) through to Gauteng (S4REF) and covers a vast southern region of the country. The most divergent strain of cluster A (S4REF) was isolated from the overlap region. The virus serotypes in cluster A are randomly distributed and the year of isolation of the viruses seemed to play more of a role in the final placing of the taxa. The recent 1999 and 2000 isolates were closely related and the older reference strains were members in another sublineage. The EEV strains in cluster B had a smaller northerly situated distribution with Gauteng forming the southern border (S5REF and S3FLD1) and Hoedspruit in Limpopo (S6FLD1) the northern border. The EEV isolates associated with the northern S10/NS3 geographic pocket grouped in a serotype dependent manner. There is no evidence of restricted geneflow within each of these pockets as virus isolates made from distant geographic locations in the pockets can harbour closely related genome segments 10 and NS3 proteins.

The occurrence of these pockets may be related to the distribution of the *Culicoides* midges that transmit EEV. To date, two species in the *Culicoides imicola* species complex are known to transmit EEV namely *C. imicola ss (senso stricto)* and *C. bolitinos* (Venter *et al.*, 1999). Sebastiani *et al.* (2001) revealed the presence of a high degree of polymorphism in the *C. imicola* species complex. *Culicoides imicola ss* is the most abundant and wide spread of the species in the complex and was highly variable within populations isolated from different geographic areas. *Culicoides bolitinos* is the second most abundant species but can be common in some areas in which *C. imicola ss* is scarce

(Sebastiani *et al.*, 2001). Interestingly, the geographic distribution of the EEV cluster A S10/NS3 correlates to some extent with the distribution of *C. bolitinos*, while the distribution of EEV cluster B S10/NS3 corresponds to that of *C. imicola ss* (Sebastiani *et al.*, 2001). The high degree of genetic diversity within these *Culicoides* species may effect efficacy of virus release from the midges. Certain S10/NS3 phenotypes of EEV may be better adapted to release virus from certain species in the *C. imicola* species complex. The two EEV S10/NS3 phenogroups, which differ by up to 25.2% in nucleotide sequence homology and up to 16.7% in amino acid sequence homology, occur in two regions of SA. These regions broadly correspond to the occurrence of *C. imicola* in the northern regions of the country while *C. bolitinos* was locally more abundant in the southern districts (Sebastiani *et al.*, 2001). It is unknown what the genetic variability and distribution of *C. imicola ss* is in the southern regions of the country and it is likely to differ from season to season dependent on the rainfall pattern. This subsequently may influence the dominance of certain EEV strains in specific outbreaks as well as their distribution in southern Africa. If this assumption is correct it would imply that EEV strains that harbour the cluster A phenotype of S10/NS3 are likely to occur in the southern regions of SA, while the strains that harbour the cluster B S10/NS3 phenotype will have a more northerly distribution. The overlap region can sustain either of the midge types therefore allowing for the occurrence of either S10/NS3 phenotype in this region.

The characteristics of EEV S10 and NS3 sequences have been studied in this chapter, with the focus on the structural characteristics of EEV NS3, and the results obtained were compared to genome segment 10 and the NS3 protein of other orbiviruses. EEV NS3 has been shown to possess a number of different characteristics, a number of which is conserved and of which most are found in other known membrane proteins (NSP4) or orbivirus NS3 proteins. The levels of variation observed for EEV S10 and NS3 were also determined using phylogenetic analysis, and the relationships between and within different EEV serotypes were visualized using phylogenetic trees. As already mentioned, the identified conserved characteristics or features of the NS3 protein, as well as the levels of observed variation for NS3, have been shown to have functional significance in other orbivirus NS3 proteins. To eventually determine the function of EEV NS3, and how the NS3 structure is related to the specific function, it will be necessary to establish a system to express EEV NS3. Once it is possible to selectively express EEV NS3 in a suitable system in sufficient quantities it will be possible to compare EEV NS3 expression to that seen for other orbivirus NS3 proteins, and the foundation will have been laid for future structure/function relationship studies of EEV NS3.