The molecular characterization of equine encephalosis virus non-structural protein NS3

by

Michelle Freeman

Submitted in partial fulfilment of the requirements for the degree M.Sc. (Agric.)
in the Faculty of Natural and Agricultural Sciences

University of Pretoria
March 2003

The Road Not Taken

By Robert Frost
(1874-1963)

Two roads diverged in a yellow wood,
And sorry I could not travel both
And be one traveler, long I stood
And looked down one as far as I could
To where it bent in the undergrowth;

Then took the other, as just as fair,
And having perhaps the better claim,
Because it was grassy and wanted wear;
Though as for that the passing there
Had worn them really about the same,

And both that morning equally lay
In leaves no step had trodden black.
Oh, I kept the first for another day!
Yet knowing how way leads on to way,
I doubted if I should ever come back.

I shall be telling this with a sigh
Somewhere ages and ages hence:
Two roads diverged in a wood, and I--
I took the one less travelled by,
And that has made all the difference.

Dedicated to my parents and to Theunis, who made all of this possible through their love, encouragement and patience

ACKNOWLEDGEMENTS

I wish to express my sincere thanks to the following people and institutions:

Prof. H. Huismans for allowing me the freedom to explore without restraints the ideas I had and providing the means necessary to do so

All the members of the Department of Genetics of the University of Pretoria for their encouragement and support, especially Mrs. Michelle van Niekerk and Mrs. Pamela de Waal for answering my numerous questions and giving good advice. As well as all my fellow Master's students, especially Miss. Joanne Riley, Mrs. Karen de Lange and Mr. Ruan van Rensburg, for their friendship, encouragement, ideas and criticism

Mr. Arrie Klopper for his friendship and all the equipment and computer crises he so often and so willingly solved over the years, also for his assistance with Figure 2.8.

Mrs. Karen de Lange for helping me to find all the books and articles I needed

Miss. Joanne Riley for keeping me informed, helping out when the tissue culture work became a crisis, and for understanding what it means to write up

Mr. Kimon Lombardi for the AHSV, BTV and EHDV dsRNA he provided, as well as for his ideas regarding some aspects of the phylogenetic analysis

Dr. P. Bloomer for providing some help with the phylogenetic analysis

Mrs. Trudie Goosen at the Department of Veterinary Tropical Diseases, Faculty of Veterinary Science, University of Pretoria for the propagation of the EEV Bryanston (EEV-1) laboratory reference strain (S1REF*)

Prof. A. J. Guthrie and Prof. P. G. Howell at the Equine Research Centre, Faculty of Veterinary Science, University of Pretoria for providing the additional EEV reference strains and field isolates

Mr. A. C. Potgieter at the Department Biochemistry, Onderstepoort Veterinary Institute for providing 30 basepairs of sequence information for both the 5' and 3' terminal ends of EEV Kyalami (EEV-5) NS3 gene for primer design

The National Research Foundation and the University of Pretoria for providing financial support

The Koekemoer family for their encouragement and support

Theunis for his love, support and encouragement, and most importantly, for putting up with me through all the ups and downs. As well as for his assistance with Figure 2.16. and other technical aspects of this thesis

My parents and family for supporting me, while I pursued what was important to me

Prof. V. Mizrahi for allowing me time off to work on my dissertation before starting and while working at the Molecular Mycobacteriology Research Unit at the NHLS, and my colleagues for their encouragement and interest

SUMMARY

The molecular characterization of equine encephalosis virus non-structural protein NS3

by

Michelle Freeman

Supervisor: Prof. H. Huismans
Department of Genetics
University of Pretoria

For the degree M.Sc. (Agric.)

Equine encephalosis virus (EEV) is a member of the *Orbivirus* genus within the family *Reoviridae*. The 10 dsRNA genome segments of these viruses encode 7 structural proteins and 3 non-structural proteins. It has been shown that the smallest of the non-structural proteins, NS3, is an integral membrane protein that plays a role in virulence and pathogenicity. The NS3 protein has two hydrophobic regions that can form transmembrane helices to allow the protein to incorporate itself into the cellular membrane. In the case of AHSV, NS3 is a highly variable protein with 36% variation within the serogroup. The variation of BTV NS3 is much less and in the case of EEV, NS3 has never been characterized.

The aim of this study was to investigate the EEV NS3 protein by looking at the structural characteristics of the protein; the variation found between different EEV NS3 proteins and its expression, using a suitable expression system.

The NS3 gene of EEV Bryanston (EEV-1) was cloned and sequenced and the deduced amino acid sequence determined. In addition, the nucleotide sequences of the NS3 gene of 7 different serotypes and 7 field isolates, as well as their deduced amino acid sequences, were also determined. These sequences represent the first sequence data for S10/NS3 of EEV.

A number of conserved structural features were identified in EEV S10/NS3, namely two putative initiation codons, two hydrophobic regions with the potential to form transmembrane helices, a proline-rich region at the N-terminal, a conserved region containing a conserved myristylation motif followed by a stretch of positively charged amino acids that together constitute a putative bipartite membrane targeting signal, glycosylation sites, conserved cysteine residues and a variable region.

These results suggested that EEV NS3 was an integral membrane protein, similar to BTV and AHSV NS3.

The level of variation observed in S10 and NS3 of EEV was determined using phylogenetic analysis, and the level of variation observed in the EEV NS3 protein was determined to be 16.7%, a value that is higher than that of BTV NS3, but lower than that of AHSV NS3. Based on S10/NS3 sequences, the EEV serogroup formed a lineage independent to that of the other orbiviruses and this lineage segregated into two clusters that corresponded to the northern and southern regions of South Africa. This geographic distribution of the EEV serotypes might be related to the distribution of the *Culicoides* vector subspecies in South Africa.

The NS3 gene of EEV Bryanston (EEV-1) was also studied by means of expression in an *in vitro* translation system using denatured EEV dsRNA, as well as by means of the baculovirus expression system. These expression studies indicated that approximately equal levels of EEV NS3 and NS3A were expressed during *in vitro* translation studies, whereas low levels of EEV NS3 expression was observed in the baculovirus expression system.

ABBREVIATIONS

A	adenine or adenosine
aa	amino acids
AcNPV/AcMNPV	*Autographa californica* nuclear polyhedrosis virus
AHS	African horsesickness
AHSV	African horsesickness virus
AMV	avian myeloblastosis virus
ATP	adenosine-5'- triphosphate
bp	basepairs
BHK	baby hamster kidney
BLAST	Basic Local Alignment Search Tool
BRDV	Broadhaven virus
BTV	bluetongue virus
°C	degrees Celsius
C	cytosine or cytidine
cDNA	complementary deoxyribonucleic acid
CLP	core-like particle
cm	centimetre
Da	Dalton
dATP	2'-deoxyadenosine-5'-triphosphate
dCTP	2'-deoxycytidine-5'-triphosphate
dGTP	2'-deoxyguanosine-5'-triphosphate
ddH_2O	double-distilled water (ultra high quality water)
dH_2O	distilled water
dNTP	deoxynucleoside-triphosphate
dTTP	2'-deoxythymidine-5'-triphosphate
DEPC	diethylpyrocarbonate
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
dsRNA	double-stranded RNA
DTT	1,4-dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
EE	equine encephalosis
EEV	equine encephalosis virus
e.g.	*exempli gratia* (for example)
EGTA	ethylene glycol bis(β-aminoethyl ether) N,N,N',N'-tetra-acetic acid

EHDV	epizootic haemorrhagic disease virus
ELISA	enzyme-linked immunosorbent assay
EMBL	European Molecular Biology Laboratory
ER	endoplasmic reticulum
et al.	*et alia* (and others)
EtBr	ethidium bromide
FCS	foetal calf serum
Fig.	figure
g	gram; centrifugal force
G	guanine or guanosine
GTP	guanosine-5'-triphosphate
h	hour(s)
i.e.	*id est* (that is to say)
IPTG	isopropyl-β-D-thiogalactopyranoside
IUB	International Union of Biochemistry
IUPAC	International Union of Pure and Applied Chemistry
kDa, K	kilodalton
LB	Luria-Bertani
M_r	relative molecular weight
M	molar
MCS	multiple cloning site
mg	milligram
min	minute(s)
ml	millilitre
mM	millimolar
MMOH	methylmercuric hydroxide
MOI	multiplicity of infection
mRNA	messenger ribonucleic acid
NCBI	National Center for Biotechnology Information
nm	nanometer
NS	non-structural
OD_{550}	optical density at 550 nm
OD_{600}	optical density at 600 nm
ORF	open reading frame
OVI	Onderstepoort Veterinary Institute
PAGE	polyacrylamide gel electrophoresis
pBS	plasmid Bluescribe
PBS	phosphate buffered saline

PCR	polymerase chain reaction
PDB	protein databank
PEG	polyethylene glycol
pI	isoelectric point
p.i.	post infection
pmol	picomole
PSB	protein solvent buffer
RER	rough endoplasmic reticulum
RNA	ribonucleic acid
RNase	ribonuclease
rpm	revolutions per minute
rRNA	ribosomal ribonucleic acid
RT	reverse transcriptase; reverse transcription
RT-PCR	reverse transcriptase polymerase chain reaction
S10	genome segment 10
SA	South Africa
SDS	sodium dodecyl sulphate
sec	second
Sf	*Spodoptera frugiperda*
ss	single-stranded
SV40	Simian virus 40
T_m	melting (or midpoint temperature); thermal denaturation
T	thymine or thymidine
TEMED	N,N,N',N'-tetramethylethylenediamine
Tris	tris-hydroxymethyl-aminomethane
tRNA	transfer ribonucleic acid
µCi	microcurie
µg	microgram
µl	microlitre
U	unit; uracil or uridine
USA, US	United States of America
UV	ultraviolet
V	Volts
VIB	viral inclusion body
v/v	volume per volume
VP	virus protein
w/v	weight per volume
X-gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

TABLE OF CONTENTS

CHAPTER 2

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

CHAPTER 1

LITERATURE REVIEW

1.1. INTRODUCTION

The concept of viruses as a natural phenomenon separate from other infectious organisms is less than 100 years old, and their nature began to be understood less than 50 years ago. The realization that many diseases of different organisms such as plants, animals and humans can be attributed to this newly recognized type of agent came even more recently, and the appreciation that practically all living species may have viruses associated with them defines the widespread prevalence of these agents and their potential importance in nature.

The concept of a biological disease agent, or pathogen, has its origins in the nineteenth century. The nature and history of viruses and virus diseases are reviewed in Levy *et al.*, 1994. Until about 1850, the intrinsic differences between various harmful agents, some of which were called by the Latin word for poison, *virus*, were poorly understood. The development of the concept of the virus as an agent entirely distinct from cellular disease agents such as bacteria and fungi was a slow process fraught with difficulties since viruses challenged the known principles and facts associated with bacteria and fungi. The first documented cases of viruses recognized as the disease-causing agents were in 1898 and involved tobacco mosaic disease of plants and foot-and-mouth disease of animals. The researchers in both cases implicated an infectious agent that was small enough to pass through the pores of filters and was thus much smaller than the smallest bacteria known. The agents were referred to as "filterable viruses", and it was understood that these "filterable viruses" were not only smaller than, but also inherently different from micro-organisms. At the beginning of the twentieth century, researchers in England showed that the human wart agent was also filterable, and there followed many other demonstrations of filterable viruses.

Today viruses are well known and can probably be best described as infectious agents and obligate intracellular parasites that use the genetic and metabolic apparatus in the cells of their hosts to reproduce themselves (Levy *et al.*, 1994). Viruses are immensely diverse, but each represents a definitive entity with shared properties and concepts different from all other forms in nature (Levy *et al.*, 1994). Viruses can be pathogenic and thus harmful to their hosts, while other viruses are symbiotic and in some cases infection with specific viruses can even provide an advantage to the host.

Viruses are however of interest not only as agents of disease but also as tools for investigating the cellular and molecular biology of the host organism. The study of viruses, or virology, plays an

important role in the field of molecular biology that can best be described as an amalgamation of scientific approaches involving the multidisciplinary effort of biochemists, physicists, chemists, geneticists, virologists, microbiologists and others (Levy *et al.*, 1994).

In this review of the literature, the present knowledge of the molecular nature of the orbiviruses, mainly BTV and AHSV, will be discussed. The focus will be on NS3, the non-structural protein encoded by segment 10 of the viral genome.

1.2. THE *ORBIVIRUS* GENUS OF THE *REOVIRIDAE* FAMILY

Viruses can be broadly classified based on the type of genomic nucleic acid and the types of molecules transcribed from the genomic nucleic acid as the virus directs the synthesis of messenger RNAs and new genomic nucleic acid (Levy *et al.*, 1994).

The genus *Orbivirus* in the family *Reoviridae* consists of viruses with similar morphological and physiochemical properties. The genus name, *Orbivirus*, was derived from the Latin *orbis*, meaning ring or circle (Borden *et al.*, 1971). Orbiviruses include pathogenic agents of man (e.g. Colorado tick fever), domestic animals (e.g. bluetongue of sheep and African horsesickness), native animals (e.g. epizootic haemorrhagic disease of deer) and many other viruses not yet linked with disease (e.g. St. Croix River virus) (Gorman, 1979).

The genus is divided into serogroups based on cross reactivities in complement fixation tests (Borden *et al.*, 1971) and serotypes within a serogroup are recognized by specific reactions in serum-neutralization tests (Gorman, 1979; Gorman and Taylor, 1985). Reassortment, a natural occurrence in the case of viruses with segmented genomes, has been shown to occur between some member viruses of a serogroup, but not between members from different serogroups (Holmes *et al.*, 1995). Examples for the AHSV serogroup include studies by Martin *et al.* (1998) and O'Hara *et al.* (1998); examples of studies where reassortment was observed for BTV S10/NS3 include Sugiyama *et al.* (1981), Collison and Roy (1983) and Pierce *et al.* (1998).

At present there are 11 genera in the family *Reoviridae*, and within the *Orbivirus* genus there are 19 serogroups of orbiviruses, each with a number of serotypes, as well as a group containing a number of unclassified viruses (Mertens *et al.*, 2000). Bluetongue virus is the prototype virus for this group (Murphy *et al.*, 1971). The present classification and nomenclature of the family *Reoviridae* are shown in Table 1.1.

Table 1.1. Classification and nomenclature of the *Reoviridae* family of viruses (adapted from Mertens *et al.*, 2000)

Family	*Reoviridae*	
Genera	*Orthoreovirus*	
	Orbivirus	
	Rotavirus	
	Coltivirus	
	Seadornavirus	
	Aquareovirus	
	Cypovirus	
	Entomoreovirus	
	Fijivirus	
	Phytoreovirus	
	Oryzavirus	
Serogroups in *Orbivirus* genus	African horsesickness group (AHSV-1 to 9)	*Culicoides*
	Bluetongue virus group (BTV-1 to 24)	*Culicoides*
	Changuinola virus group	phlebotomines
	Chenuda virus group	ticks
	Chobar Gorge virus group	ticks
	Corriparta virus group	*Culicine* mosquitoes
	Epizootic haemorrhagic disease virus group (EHDV-1 to 8)	*Culicoides*
	Equine encephalosis virus group (EEV-1 to 7)	*Culicoides*
	Eubenangee virus group	*Culicoides*, anopheline and *Culicine* mosquitoes
	Ieri virus group	mosquitoes
	Great Island virus group (e.g. Broadhaven virus)	*Argas*, *Ornithodoros*, *Ixodes* ticks
	Lebombo virus group	*Culicine* mosquitoes
	Orungo virus group	*Culicine* mosquitoes
	Palyam virus group (e.g. Chuzan virus)	*Culicoides*, *Culicine* mosquitoes
	Umatilla virus group	*Culicine* mosquitoes
	Wad Medani virus group	*Boophilus*, *Rhipicephalus*, *Hyaloma*, *Argas* ticks
	Wallal virus group	*Culicoides*
	Warrego virus group	*Culicoides*, anopheline and *Culicine* mosquitoes
	Wongor virus group	*Culicoides*, mosquitoes
	Unassigned virus group (e.g. St Croix River virus)	

The first EEV serotype was isolated in 1967, with the subsequent identification of another 3 serotypes in March 1971, December 1973 and April/May 1976 (Gorman and Taylor, 1985). Another serotype was also identified in 1993 (Gerdes and Pieterse, 1993). To date, seven serotypes namely Kyalami, Bryanston, Cascara, Gamil, Kaalplaas, Langeberg/Langebaan and Potchefstroom have been identified in southern Africa (Gerdes and Pieterse, 1993). Cross complement fixation tests have failed to demonstrate a serological relationship or relatedness between any of the EEV serotypes and other orbiviruses such as BTV, AHSV, EHDV or Corriparta virus (Erasmus *et al.*, 1970; Gorman and Taylor, 1985; Viljoen and Huismans, 1989).

The seven EEV serotypes are not equally prevalent under field conditions, probably indicating that some serotypes have adapted better to the specific vector and host environments than other serotypes (Barnard and Paweska, 1993; Venter *et al.*, 1999; Howell *et al.*, 2002; P. G. Howell, personal communication). Some of the seven identified serotypes, however, need to be validated (Venter *et al.*, 1999). In this regard samples of virus, representing what has been referred to as the Langeberg/Langebaan EEV serotype, have been investigated and have been shown to be AHSV-5 (Howell *et al.*, 2002). The same group however identified another isolate, E21/20, that was subsequently shown to be a previously undescribed member of the EEV serogroup, so that the number of known EEV serotypes remains at seven (Howell *et al.*, 2002). Howell *et al.* (2002) therefore concluded that to date seven valid serotypes of EEV have been identified which they arranged into alphabetical order, corresponding with a numerical serotype designation, in order to avoid confusion in the nomenclature (Table 1.2.).

Table 1.2. All known EEV serotypes with a proposed numerical serotype designation (Howell *et al.*, 2002)

Name of virus	Numerical serotype designation of virus
Bryanston	EEV-1
Cascara	EEV-2
Gamil	EEV-3
Kaalplaas	EEV-4
Kyalami	EEV-5
Potchefstroom	EEV-6
E21/20	EEV-7

Cascara (EEV-2) was the first serotype to be isolated (Erasmus *et al.*, 1970) and as such was viewed as the representative of the EEV group. Bryanston (EEV-1) has however superseded Cascara to become the dominant serotype and thus the present representative serotype of the EEV serogroup (Paweska, 1999; P.G. Howell, personal communication). For the purposes of this thesis, the EEV

serotypes will be referred to using the numerical serotype designation proposed by Howell *et al.* in 2002.

The diseases caused by some members of the *Orbivirus* genus, notably bluetongue virus (BTV) and African horsesickness virus (AHSV) are of great economical importance due to the livestock that are affected and the trade of these livestock. Trade often involves the import and export of the animals or products of the animals such as genetic material (semen), and there are many strict regulations regarding trade of the livestock or their products. AHSV is especially important, as it affects horses and thus plays an important role in the strictly regulated industries of horse racing and show jumping. In the case of EEV, which also infects horses, the importance of the disease lies not in the losses caused by it, as it is in many cases a subclinical infection, but rather in the nature of the affected animal and the potential impact it can have on the strictly regulated industry of which the animals are an integral part. The limited knowledge on the epidemiology, pathogenicity, and viral characteristics of EEV is seen as a potential barrier to the international movement of horses.

1.3. EPIDEMIOLOGY OF A NUMBER OF ORBIVIRUSES

1.3.1. Geographic distribution

Bluetongue seems to have originated on the African continent and for many decades it was believed to be confined to Africa. The first confirmed outbreak of the disease outside Africa occurred in Cyprus (Roy, 1989), subsequently several outbreaks of the disease have been reported in different parts of the world such as the Middle East (Erasmus, 1985; Roy, 1989), the United States of America, Portugal and in Spain. It is now known to occur in a broad belt around the world, encompassing the tropical and temperate zones in which a large percentage of the world ruminant population is also present (Ozawa, 1985). EHDV has also been isolated in countries in the same regions (Parsonson and Snowdon, 1985).

African horsesickness is endemic in eastern and central Africa, although AHS is also widely distributed in sub-Saharan Africa. Occasional outbreaks have occurred in North Africa, West Africa, the Middle East, parts of Asia and in Southern European countries (such as Spain and Portugal), but the virus has not been able to maintain itself outside the sub-Saharan enzootic zones (Williams *et al.*, 1993; Zientara *et al.*, 1993).

Equine encephalosis is endemic at least in southern Africa, including South Africa, Zimbabwe, Botswana and Kenya (Barnard, 1997; Venter *et al.*, 1999). EEV is found more widely throughout southern Africa than AHSV which is restricted to certain areas, this suggests an epidemiological pattern and vectors unlike that of AHSV (Barnard, 1997). Evidence of EEV infection has however been found in other regions of the world such as Australia (P.G. Howell, personal communication).

1.3.2. Transmission by insect vectors

The first breakthrough in identifying insect vectors of BTV came in 1943 when Du Toit (1944) succeeded in transmitting bluetongue to sheep with *Culicoides imicola* (*pallidipennis*) which had fed on a viraemic sheep 10 days earlier. Subsequently, *Culicoides variipennis* was incriminated as a vector of BTV in the US (Erasmus, 1985) and different species of *Culicoides* midges have been incriminated as vectors of BTV and other orbiviruses in various countries e.g. *Culicoides brevitarsis*, *Culicoides fulvus*, *Culicoides wadai*, *Culicoides acetoni* in Australia (Parsonson and Snowdon, 1985; St. George, 1985), and *Culicoides variipennis* (now known as *Culicoides sonorensis*) and *Culicoides insignis* in the US (Rao and Roy, 1983; Callis, 1985; Greiner *et al.*, 1985; Bonneau *et al.*, 2001). EHDV is also mainly transmitted by *Culicoides* species, e.g. *Culicoides brevitarsis* in Australia (Parsonson and Snowdon, 1985).

AHSV is also mainly transmitted by *Culicoides* species, specifically *Culicoides imicola* (Du Toit, 1944). In some cases, AHSV is also transmitted by mosquitoes of the *Aedes*, *Anopheles* and *Culex* species, as well as the brown dog tick *Rhipicephalus sanguines* (Lubroth, 1992; Tomori *et al.*, 1992).

EEV is also transmitted by *Culicoides* species, and specifically *Culicoides imicola* (Theodoridis *et al.*, 1979; Nevill *et al.*, 1992; Paweska, 1999; Venter *et al.*, 1999). The first isolation of a virus from the EEV serogroup from *Culicoides* species in South Africa was done by Theodoridis *et al.* (1979) during an unsuccessful attempt to isolate bovine ephemeral fever virus. Subsequently, *Culicoides bolitinos* has also been implicated in EEV transmission (Paweska, 1999).

Due to the nature of the insect vectors involved in the transmission of bluetongue, African horsesickness and equine encephalosis, the diseases all occur seasonally, and outbreaks (usually in the late summer and autumn) are influenced by conditions that favour the occurrence, breeding and spread of the vector (Theodoridis *et al.*, 1979; Venter *et al.*, 1999).

Culicoides imicola is widely distributed in sub-Saharan Africa, parts of North America and southern Europe, and southern Asia (Venter *et al.*, 1999). *Culicoides imicola* is a common vector for BTV, AHSV and EEV, and the potential thus exists for all these diseases to spread to these regions when conditions are favourable and even to persist within these regions since a suitable insect vector that can sustain the virus is present. Outbreaks in countries where AHSV and BTV are not endemic have occurred in the past due to the dissemination of the insect vectors (through the movement of the infected midges by wind, or climatic conditions that favour their growth) or the introduction of equidae or other livestock carrying the viruses. In this regard, the national and international movement of game or wild animals e.g. zebras, is an important factor relating to the spread and persistence of the diseases. Zebras can introduce disease into previously disease free regions, as zebras are potential

reservoirs of both AHSV and EEV. The status of previously disease free regions can therefore change permanently due to the introduction of livestock or game, as well as climate changes, that will allow the establishment and sustained growth of the vector species.

1.3.3. Host range

Bluetongue has traditionally been regarded as a disease of sheep and more particularly of European breeds. However, other domestic ruminants have also been shown to be susceptible, e.g. calves, cattle, goats, water buffaloes and camels (Ozawa, 1985; Roy, 1989). BTV and EHDV also share common hosts such as cattle, buffaloes, deer, antelope, white-tailed deer, mule deer, bighorn sheep, and in some cases EHDV can also infect zoological animals (Callis, 1985; Jessup, 1985; Parsonson and Snowdon, 1985).

In the case of AHSV, the host range may be limited mainly to equine animals such as horses, mules, donkeys and zebras (Barnard, 1993, 1997). The infection of man by AHSV has however also been reported. Humans exposed to freeze dried vaccine preparations of AHSV neurotropic strains 1 and 6 were shown to develop encephalitis and uveochorioretinitis (Swanepoel *et al.*, 1992).

EEV, like AHSV, infects horses and antibodies against EEV have also been found in zebras (Burchel's zebra and the Cape mountain zebra), in donkeys (Gerdes and Pieterse, 1993; Williams *et al.*, 1993) and African elephants (Barnard, 1997). It has been speculated that zebras function as a possible reservoir of the virus (Barnard and Paweska, 1993; Williams *et al.*, 1993; Barnard, 1997; Paweska, 1999).

AHSV and EEV thus share vectors and hosts as both infect equids. It has been found that the serum of a single animal can show antibodies to both viruses, this creates the possibility that an animal can be simultaneously infected with both viruses or rather that a single animal can be infected by both viruses in the same season, possibly even at the same time leading to confusion especially regarding symptoms of either disease (Williams *et al.*, 1993). This is further illustrated by the fact that during a recent outbreak of AHS in the Western Cape, EEV Bryanston (EEV-1) was isolated from horses showing disease symptoms (P.G. Howell, personal communication), confirming that EEV can indeed co-circulate with AHSV during the same season.

1.3.4. Epidemiology, clinical symptoms and pathogenesis of EEV

The history of the first known cases of EEV as well as the symptoms of the disease are discussed by Erasmus *et al.* (1970). EEV was first isolated in 1967 from a 13-year old thoroughbred mare (Cascara) at Mauritzfontein stud farm in the Kimberley district in South Africa. The horse first showed

symptoms 24 hours before death. The symptoms included listlessness and an anxious facial expression with tightening of the face, mouth and ocular commissures, and clinical examination revealed a slightly elevated temperature of 39.5°C and a pulse rate of 44 per minute. The autopsy revealed marked venous congestion particularly of the liver, kidneys and subcutaneous tissues. The liver showed signs of fatty degeneration and sharply demarcated areas of catarrhal enteritis were observed in the distal half of the small intestine. The brain was congested and oedematous with an excessive amount of cerebrospinal fluid.

In the next few days, two other adult mares became ill. One of them died the same day, whereas the other one recovered after about 14 days. Fatal cases occurred amongst horses at the Police College in Pretoria and were also reported from other parts of the country. All these animals showed essentially the same symptoms and lesions. It is interesting to note that all the clinical and fatal cases were invariably seen in animals older than seven years. Two young horses experimentally infected with liver and spleen suspensions from the horse Cascara, showed marked febrile responses that commenced 3 days p.i. and persisted for 4 days, attaining a maximum of 41°C, after which the horses recovered fully. Apart from the febrile responses, the horses showed no other clinical symptoms. This suggests that age seems to play a role in the pathogenesis of the disease.

Virus was isolated from organ specimens collected from case 1 (Cascara) as well as from other fatal cases, and from the blood of horses, which showed fever only. All these virus isolates appeared identical and consequently the Cascara isolate was selected as prototype strain for further characterization.

After isolation of the Cascara serotype, a serological survey was also conducted. The results indicated a widespread occurrence of infection with EEV during the summer (first three months) of 1967. An interesting point was that in the collected serum samples, unstabled horses were more likely to show antibodies against EEV than stabled horses, this suggested nocturnal biting insects possibly played a role in the transmission of the virus. This was confirmed by the first isolation of a virus from the EEV group from *Culicoides* species by Theodoridis *et al.* in 1979. Another interesting point was that antibodies against EEV were also found in horses that had shown no clinical signs, suggesting that the great majority of infections were subclinical in nature and that EEV exhibits a low pathogenicity for horses in endemic areas where annual exposure occurs (Gerdes and Pieterse, 1993). EEV thus causes a disease characterized by a high morbidity (60% to 70%) but a low mortality (5%) (Bremer and Viljoen, 1998). During this serological study, twenty serum samples collected during the ten years prior to 1967 were also examined and none of these were found to possess antibodies to EEV. This suggested the very recent appearance of the virus in South Africa.

Since 1967, equine encephalosis has taken on epidemic proportions in some years with more than 75% of horses tested having antibodies against EEV (e.g. 1994 to 1997) and 85% of donkeys tested (1995) (Venter *et al.*, 1999). Since the first outbreak in 1967, except for localized epizootics in 1976, 1978 and 1990 only sporadic isolations of the virus have been made (about 60 since 1983). A very high EEV seroprevalence in horses and donkeys in South Africa, but only a limited number of isolates made, thus again suggests that most EEV infections are subclinical in nature (Paweska, 1999). It has also been found that there is a continuous circulation of the virus between its host and vector in the case of zebras, which function as a possible reservoir of the virus and which ensures the persistence of the infection (Barnard and Paweska, 1993).

In a recent study, Howell *et al.* (2002), used serum samples from a ten year period (1990 to 2000) and found that 56.9% were positive for neutralizing antibody against one or more serotypes of the virus. The antibodies found were not necessarily against one type of virus, because it has been found that the host can be infected by more than one serotype, thus showing antibodies to more than one serotype at one point in time (Howell *et al.*, 2002). This allowed them to conclude that EE is a vector-borne infection of horses of a mild to subclinical character, which, when the diagnosis is confirmed, would in the vast majority of cases, closely resemble a mild attack of AHS (Howell *et al.*, 2002). In addition, it has been found that EEV Bryanston (EEV-1) can co-circulate with AHSV, this was illustrated in the Western Cape during a recent outbreak of AHS (P.G. Howell, personal communication).

In the Western Cape, a demarcated area of an estimated extent of 9 000 km^2, has been set aside as a surveillance zone to monitor the possible introduction or occurrence of AHS which would compromise the integrity of pre-export quarantine facilities for horses in Cape Town. Within this area serum samples were collected in May 1998 that showed the presence of antibodies against EEV. Subsequently, over a period of 18 months, serum samples were collected and assayed for neutralizing antibody against the serotypes of EEV. The results of these tests indicated the presence of EEV serotypes in the region, and it appeared that prior to the implementation of the surveillance program an outbreak of EE had taken place in the region. At the commencement of the survey, Bryanston was the predominant serotype, 18 months later Bryanston was still the most prominent serotype, followed by Potchefstroom and E20/21 (Howell *et al.*, 2002). The fact that EEV is present in this region is obviously of great concern and can have grave consequences for the export of horses.

1.3.5. Epidemiology of other orbiviruses

Other orbiviruses mentioned in this thesis include Orungo, Lebombo, Chuzan, Broadhaven and St. Croix River virus. Orungo and Lebombo viruses have both been isolated in Africa. Orungo virus has been isolated from *Aedes*, *Anopheles* and *Culex* mosquitoes, as well as from humans, where it has

been associated with febrile illness. Antibodies to this virus have been detected in sera from camels, cattle, goats, humans, monkeys and sheep. Lebombo virus has been isolated from *Aedes* and *Mansonia* mosquitoes, human plasma and rodents (Brown *et al.*, 1991). Chuzan virus, a member of the Palyam virus serogroup, was isolated from the biting midge *Culicoides oxystoma* and sentinel calves in Japan in 1985. The virus was subsequently implicated in an epizootic of congenital abnormalities with hydraencephaly-cerebellar hypoplasia syndrome of calves. Palyam serogroup viruses are usually associated with a variety of haematophagous arthropod vectors and large mammals, principally cattle, in many parts of the world including Asia, Australia and Africa. Although the Palyam serogroup of viruses seem to be involved in producing abortion and congenital malformations in cattle, the pathogenic importance of most of them remains unknown (Yamakawa *et al.*, 1999). Broadhaven virus is a member of the Great Island virus group, which is currently classified as a serogroup in the *Orbivirus* genus. This virus is transmitted by ticks and was first isolated from *Ixodes uriae* (Moss *et al.*, 1992; Attoui *et al.*, 2001). St. Croix River virus was isolated from a tick cell line established prior to 1994 from eggs of *Ixodes scapularis*. The eggs were collected from a tick obtained from a hunter-killed white-tailed deer in western Wisconsin, near the St. Croix River. This is a new orbivirus that has not been studied to a great extent, and although its genome has been sequenced, nothing is presently known about its replication characteristics and pathogenicity (Attoui *et al.*, 2001).

1.4. THE STRUCTURAL CHARACTERISTICS OF ORBIVIRUSES

Most of the work done on the viruses of the *Orbivirus* genus has concentrated on the prototype member, namely BTV. In recent years information has also become available regarding AHSV and EHDV. Very little is however known at this stage about the structural or functional characteristics of EEV.

1.4.1. Viral morphology and chemical composition

The prototype orbivirus, BTV, has a regular structure, is non-enveloped and the size of the virion is approximately 54 mμ. BTV consists of a dsRNA genome enclosed by a capsid of complex nature containing a single layer of 32 well-defined protein capsomeres with a hexamer-pentamer clustering of units arranged in accordance with icosahedral symmetry (5:3:2). The capsomeres are relatively widely spaced and have a flattened, hollow cylindrical or prismatic shape (Els and Verwoerd, 1969). This nucleocapsid structure (core particle) consists of 5 types of protein; two major (VP3 and VP7) and three minor components (VP1, VP4 and VP6) (Verwoerd *et al.*, 1970; Els, 1973; Huismans and Howell, 1973; Bremer, 1976; Mertens *et al.*, 1984; Roy, 1989; Huismans and Van Dijk, 1990; Roy, 1996). This icosahedral core particle is surrounded by an easily removed outer capsid structure (outer shell) composed of two proteins (VP2 and VP5). This outer capsid lacks clearly visible morphologic

subunits. None of the proteins is glycosylated or, as far as been determined, otherwise modified. (Roy, 1989). The core particle and outer capsid together is referred to as the virion. In protein synthesis studies at least 10 virus-induced polypeptides have been identified in infected cells. Therefore in addition to the seven structural proteins, there are at least three non-structural proteins, i.e. NS1 (P5a), NS2 (P6a) and NS3 (P8). Each of the major structural and non-structural proteins is encoded on a different genome segment (Mertens *et al.*, 1984; Pedley *et al.*, 1988; Van Dijk and Huismans, 1988; French *et al.*, 1989).

AHSV has a similar regular structure, also consisting of a protein capsid which consists of a single layer of 32 capsomeres arranged in a regular icosahedral symmetry to form a particle 55 nm in diameter (Verwoerd and Huismans, 1969).

EEV in turn possesses a similar structure to that seen for both BTV and AHSV. The virus particles are 73 nm in diameter (60 nm when negatively stained) and the particles do not appear to possess an envelope (Theodoridis *et al.*, 1979). Experimental observations suggested that the capsid is composed of 32 hexamer-pentamer morphological subunits with icosahedral symmetry or a dimer or trimer arrangement of the capsomeres (Erasmus *et al.*, 1970; Lecatsas *et al.*, 1973) and the capsid shell of the virion is composed of two layers (Holmes *et al.*, 1995).

Electron microscopic studies have shown that negatively stained particles of BTV, AHSV and EEV are indistinguishable, and as representative example, negatively stained BTV particles are shown in Fig. 1.1.

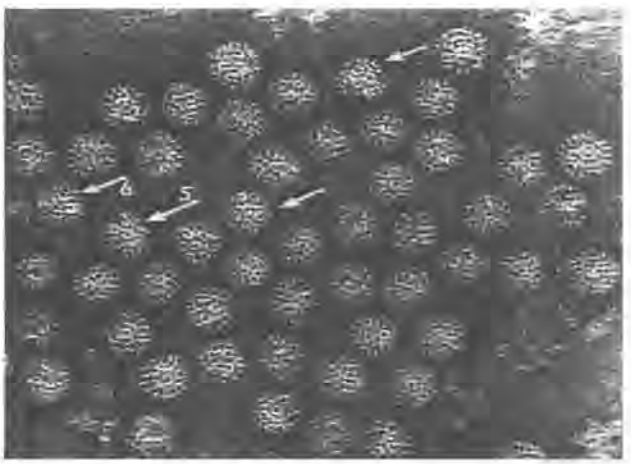


Figure 1.1. Negatively stained bluetongue virus particles. X 156 000. Particles are indicated by arrows (adapted from Els and Verwoerd, 1969).

Orbiviruses such as AHSV and BTV also share many similarities regarding their physical, chemical and biological properties. This includes their pH sensitivity, stability towards organic solvents and transmission by *Culicoides* species. They are thus mainly distinguished on serological grounds and because of their different host specificities (Verwoerd and Huismans, 1969).

BTV and AHSV, depending on the experimental conditions, can be either slightly sensitive or resistant to ether, chloroform and deoxycholate (Bowne and Ritchie, 1970; Borden *et al.*, 1971; Bremer, 1994). AHSV is resistant to the action of trypsin (Bremer, 1994), BTV is labile at pH 3.0 (Borden *et al.* 1971), while AHSV can be inactivated within 15 min at pH below 5.6 and above 10.9 (Bremer, 1994).

EEV, depending on the experimental conditions, can be either resistant or sensitive to chloroform, relatively sensitive to sodium deoxycholate and sensitive to trypsin. The virus is labile at pH 3.0 and 5.0, as well as temperature sensitive as it is inactivated at 56°C (stable at temperatures lower than 56°C). The virus is also susceptible to actinomycin D (Erasmus *et al.*, 1970; Theodoridis *et al.*, 1979).

1.4.2. The viral genome

Orbiviruses such as BTV, AHSV and EHDV, contain segmented genomes composed of 10 dsRNA segments ranging in size from 0.5 to 2.7 x 10^6 Daltons. A double layer protein coat surrounds these 10 segments (Bremer, 1976; Van Dijk and Huismans, 1988). The segments are numbered 1 to 10 in order of their migration in an agarose gel (Huismans and Van Dijk, 1990) and are divided into three size groups of segments (Els and Verwoerd, 1969; Verwoerd, 1969; Bowne and Ritchie, 1970; Verwoerd *et al.*, 1970), large segments (L, 1 to 3), medium segments (M, 4 to 6) and small segments (S, 7 to 10) (Roy, 1989). Each segment probably contains the information for the synthesis of one virus specific polypeptide; this is supported by a study that indicated that no secondary proteolytic cleavage of virus-induced polypeptides was observed to occur (Gorman *et al.*, 1981). As an example, the 10 dsRNA genome segments of AHSV encode 7 structural proteins (VP1 to VP7) and at least three distinct non-structural proteins (NS1 to NS3) (Bremer, 1976; Grubman and Lewis, 1992). Each segment codes for only one protein except the smallest segment, S10, which codes for two proteins, NS3 and NS3A, presumably via different in-frame translation initiation sites (Van Staden and Huismans, 1991; Van Staden *et al.*, 1991).

AHSV, similar to BTV, also contains double-stranded ribonucleic acid as its genetic material and can be fractionated into at least 10 components of similar but not identical size distribution after isolation (Verwoerd, 1969; Verwoerd and Huismans, 1969, Verwoerd *et al.*, 1970; Borden *et al.*, 1971; Bremer, 1976; Gorman *et al.*, 1981; Knudson *et al.*, 1982). A small and rather variable amount of hybridization has been found between bluetongue and African horsesickness viruses (Verwoerd and Huismans, 1969), which indicates that the genomic segments of AHSV and BTV show some similarity.

EEV, similar to both BTV and AHSV, contain RNA as its nucleic acid (Erasmus *et al.* 1970; Theodoridis *et al.* 1979) in the form of 10 segments of linear dsRNA (Viljoen and Huismans, 1989; Bremer *et al.*, 1990). The sizes of the ten genome segments for EEV Cascara (EEV-2) have been estimated by electrophoresis on a 1% agarose gel using BTV segments as a standard (Viljoen and Huismans, 1989) (Table 1.3.).

Table 1.3. EEV Cascara (EEV-2) genome segments and their estimated sizes (Viljoen and Huismans, 1989) compared to the estimated sizes of AHSV-3 (Bremer *et al.*, 1990) and BTV-10 (Roy, 1989)

Genome segment	EEV Cascara (EEV-2)	AHSV-3	BTV-10
	Segment length (bp)	Segment length (bp)	Segment length (bp)
1	3 900	3 314	3 954
2	3 220	3 038	2 926
3	2 750	2 663	2 772
4	2 020	2 033	2 011
5	1 750	1 894	1 638
6	1 570	1 639	1 769
7	1 080	1 137	1 156
8	1 080	1 137	1 124
9	1 080	1 137	1 046
10	710	693	822

Table 1.4. The conserved terminal sequences of the genome segments 10 of a number of different orbiviruses

Virus	Conserved terminal sequence at 5' end		Conserved terminal sequence at 3'end	Reference
	Position	Sequence[a]	Sequence	
AHSV	1-6 4	5' **GUU**UAA 3' A, U, or C	5' ACU**UAC** 3'	Mertens and Sangar, 1985; Van Staden and Huismans, 1991
BTV	1-6	5' **GUU**AAA 3'	5' ACU**UAC** 3'	Rao *et al.*, 1983; Mertens and Sangar, 1985; Gould, 1988; Roy, 1989; Van Staden and Huismans, 1991; De Mattos *et al.*, 1992b; Huismans *et al.*, 1992
BRDV	1-6	5' **GUA**AAA 3'	5' GGA**UAC** 3'	Moss *et al.*, 1992
EEV-1 and EEV-5	1-6	5' **GUU**WAD 3'	5' HSU**UAC** 3'	Potgieter *et al.*, 2002
Palyam virus	1-6	5' **GUU**AAA 3'	NA[b]	Van Staden and Huismans, 1991; Huismans *et al.*, 1992
St. Croix River virus	1-5	5' **GUA**AU 3'	5' CA**UAC** 3'	Attoui *et al.*, 2001

[a] Refer to Appendix A for description of single-letter codes

[b] Information not available is indicated by NA

Viruses belonging to the *Reoviridae* family have the interesting feature that all ten genomic dsRNA segments possess 5' and 3' terminal sequences which are conserved and genus specific. The

conserved terminal sequences of a number of different orbivirus genome segments 10 are shown in Table 1.4. The consensus terminal sequences of EEV Bryanston (EEV-1) and EEV Kyalami (EEV-5) have only recently been determined (Potgieter *et al.*, 2002). Orbiviruses additionally feature an inverted repeat which differs in sequence for each segment and which is present next to the conserved termini (Rao *et al.*, 1983; Roy, 1989; Nel *et al.*, 1990; Moss *et al.*, 1992). In AHSV it is a 7 bp inverted repeat adjacent to the terminal sequences (Van Staden and Huismans, 1991; Huismans *et al.*, 1992). No inverted repeats were however identified for EEV S10, although repeats have been identified in the genes that encode VP6 and VP7 (Potgieter *et al.*, 2002, refer to Table 1.5.). The conserved terminal features of the RNA segments of the viruses in the *Reoviridae* family are thought to play a role in transcription initiation, RNA replication, ribosome binding and translation of the mRNAs (Rao *et al.*, 1983), as well as in the determining of mRNA secondary structure, which could be important in the sorting and assembly of the genome during viral replication (Anzola *et al.*, 1987; Huismans *et al.*, 1992).

Table 1.5. Inverted terminal repeats of the cloned EEV Bryanston (EEV-1) gene encoding VP6 and EEV Kyalami (EEV-5) gene encoding VP7 (adapted from Potgieter *et al.*, 2002)

Gene	Inverted repeat	Position (bp)
VP6	AAATAACGTTC	4 and 1065
VP7	TTTGGCC	7 and 1161

1.4.3. The viral proteins

As seen in the previous section, each dsRNA segment encodes at least one viral polypeptide, with the exception of genome segment 10, which encodes two polypeptides. In this section the various viral proteins, structural proteins, as well as the non-structural proteins, will be discussed. The focus will however be on the non-structural proteins, specifically NS3.

In Table 1.6. the general coding assignment for the genome segments of BTV, as well as other information such as number of amino acids and predicted size of the encoded proteins, are indicated to serve as a framework for the discussion that will follow. The coding assignment of EEV has recently been studied (Potgieter *et al.*, 2002) and found to correspond to that of AHSV and BTV except for the fact that segments 7 and 8 are reversed encoding NS2 and VP7 respectively. The NS2 gene (segment 7) is also slightly larger than the VP7 gene (segment 8) (Potgieter *et al.*, 2002). In analogy to BTV, the protein(s) encoded by EEV genome segment 10 will be referred to as NS3/NS3A.

Table 1.6. Generalized coding assignments for the genome segments of BTV determined using 1% agarose gels. The number of amino acids, the predicted size (M_r) and function are as given by Roy, 1996.

Genome segment	Viral protein	Number of amino acids	Predicted size (Mr)	Type	Morphology/function
1	VP1	1302	149 588	Structural core protein	Part of core; RNA polymerase
2	VP2	956	111 112	Structural outer capsid protein	Part of outer capsid; serotype-specific antigen
3	VP3	901	103 344	Structural core protein	Part of core; structural protein
4	VP4	654	76 433	Structural core protein	Part of core; capping enzyme: guanyltransferase
5	NS1	552	64 445	Non-structural protein	Forms tubules
6	VP5	526	59 163	Structural outer capsid protein	Part of outer capsid; structural protein
7	VP7	349	38 548	Structural core protein	Part of core; group-specific structural protein
8	NS2	357	40 999	Non-structural protein	Phosphoprotein; binds ssRNA
9	VP6	328	35 750	Structural core protein	Part of core; binds ssRNA, dsRNA
10	NS3 NS3A	229	25 572	Non-structural protein	Glycoprotein; aids virus release

1.4.3.1. Structural proteins

Cryo-electron microscopy and computer image reconstruction techniques have been used as tools in the study of the architecturally complex virus particle of BTV and resulted in the proposal of a three dimensional model for BTV (reviewed by Roy, 1992). These studies indicated icosahedral symmetry with a triangulation number of 13 for the core. The core structure was shown to be divided into two concentric layers of protein, which enclose the inner core. The first layer, forming the outer surface of the core is made up of five or six-membered rings or clusters of VP7 trimers. The VP7 trimers form knob-like protrusions and are located at all the local and strict threefold axes. In between the VP7 structures are aqueous channels of which some penetrate the inner VP7 layer. The inner layer makes a smooth bed upon which VP7 trimers are located. The VP7 structures are the points for the deposition of the two surface proteins of this particle, VP2 and VP5. Under VP7 is a shell of VP3 molecules, which are arranged as 12 pentamers. The two layers enclose the inner core, which contains the three minor proteins, VP1, VP4 and VP6 and the viral genome.

Similar studies revealed that the structure of the outer capsid of the virus is unlike the fuzzy appearance revealed by earlier electron microscopic analysis of negatively stained virus particles, and that the proteins contained in the outer capsid have distinctive shapes. The VP5 protein, existing as a globular and almost spherical structure, sits in the channels formed by each of the six-membered rings of VP7 trimers. The VP2 protein exists as sail-shaped spikes, which project beyond the globular proteins and are located above the VP7 trimers. A cross-sectional diagram of the complete, assembled bluetongue virus particle, with the positions of the various structural proteins within the particle, is shown in Fig. 1.2.

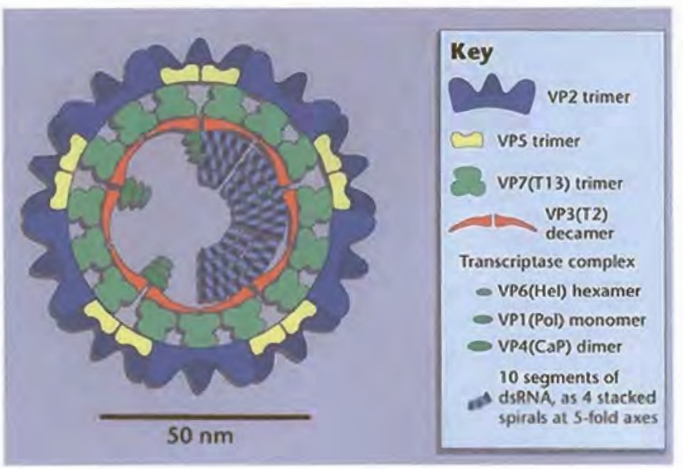


Figure 1.2. Cross-sectional diagram of the bluetongue virus particle showing the positions of the various proteins within the virus particle (Mertens, 2001)

The minor core proteins may be involved in virus replication and transcription processes. The largest of these proteins, VP1, is present in a low molar ratio in the core. Based on its size, and the presence of a motif common to all polymerases in the gene encoding this protein, it is the prime candidate for the virion RNA polymerase (Mertens *et al.*, 1984; Roy *et al.*, 1988; Huismans and Van Dijk, 1990; Roy, 1992; Vreede and Huismans, 1998). VP4 binds GTP and may act as a guanylyltransferase which catalyses the capping and methylation of the 5' end of viral mRNA (Roy, 1992; Roy *et al.*, 1994). The genome segment 4 of 5 different US BTVs have been analysed and a potential leucine zipper motif has been observed near the carboxy terminus of the deduced VP4 amino acid sequences (Huang *et al.*, 1993). A co-operative enzymatic function between VP1 and VP4 has been postulated involving RNA transcription and/or RNA replication (Huismans and Van Dijk, 1990). VP6, a highly basic protein and the smallest structural protein, is closely associated with the virus genome. This is based on the finding that it has a strong binding affinity for ss- and dsRNA species (Roy *et al.*, 1994). VP6 also binds ATP and exhibits RNA dependent ATPase activity and a helicase activity that catalyses the unwinding of dsRNA substrates prior to the synthesis of the mRNA (Roy, 1992). AHSV VP6 has a conserved motif in the primary sequence of the gene that is associated with a helicase (Turnbull *et al.*, 1996). VP6 may also aid or be responsible for the encapsidation of RNA (Roy, 1992).

VP3 and VP7 are the major constituents of the core particle which are associated in a highly ordered, three-dimensional manner resulting in icosahedral symmetry. Sequence data previously obtained for the VP3 gene, and deduced protein of BTV, AHSV and EHDV, demonstrates that of the four major capsid proteins, VP3 is the most conserved with between 57% to 58% amino acid identity (Iwata *et al.*, 1992). This high level of sequence conservation in the VP3 gene is probably due to the fact that the VP3 scaffold plays a major role in the structural integrity of the virus core. VP3 is hydrophobic and its size, as well as amino acid content, is similar for different orbiviruses e.g. AHSV-4, BTV-10 and EHDV (Iwata *et al.*, 1992). VP3 is also a group-specific antigen (Iwata *et al.*, 1992). VP3 has another function besides its structural function; it has been shown that it can bind ssRNA in a non-sequence specific manner, probably during the virus assembly process (Loudon and Roy, 1992).

The main component of the core, VP7, is highly conserved among serotypes and serogroups (Bremer *et al.*, 1990; Iwata *et al.*, 1992) and is the group specific antigen (Mertens *et al.*, 1984; Oldfield *et al.*, 1990; Chuma *et al.*, 1992). VP7 has been found to have limited exposure on the surface of the intact virion since a portion of the VP7 molecule is exposed to the surface (Roy, 1992). VP7 has at least two epitopes exposed on the virus surface (Lewis and Grubman, 1990) and Eaton *et al.* (1991) reported that the BTV serogroup reactive epitope in the amino terminal half of VP7 is accessible on the surface of BTV. VP7 exists as trimers in the virion, which appears as tripod-like structures with two distinct domains (Basak *et al.*, 1992; Grimes *et al.*, 1995, Basak *et al.*, 1996).

VP2 is the most surface exposed capsid protein of the orbiviruses and has been identified as the major serotype-specific neutralizing antigen (Huismans and Howell, 1973; Vreede and Huismans, 1994). VP2 is the most variable protein (Roy *et al.*, 1990) and as such exhibits low conservation between serotypes and serogroups (Iwata *et al.*, 1992; Vreede and Huismans, 1994). Regions involved in eliciting protective immune response and determination of specificity have not yet been fully identified. Studies with BTV have shown that VP2 is a haemagglutinin and is directly involved in attachment of virus to cells (Eaton and Crameri, 1989; Huismans and Van Dijk, 1990). There is also evidence that VP2 plays a role in the virulence of certain virus strains (Huismans and Howell, 1973; O'Hara *et al.*, 1998).

Little is known about VP5 relative to the other major capsid proteins. VP5 is more conserved than VP2 among serotypes, but not as conserved as the core proteins (Oldfield *et al.*, 1991). VP5 is also more conserved among serogroups including AHSV, BTV and EHDV, than VP2. It shows approximately the same level of conservation as the group specific antigen VP7 when compared between orbivirus serogroups (Iwata *et al.*, 1992). Although VP5 is located in the outer capsid, it is mostly unexposed or only partially exposed on the surface of the virion (Lewis and Grubman, 1990; Hewat *et al.*, 1992), and it does not appear to have any direct role in binding neutralizing antibodies. However, studies with BTV have demonstrated that VP5 enhances neutralization and protective immune response (Roy *et al*, 1990). The protein however, may also make some contribution to

virulence as was found in one case where the most obvious difference between a virulent and avirulent strain of BTV was in the mobility of the VP5 genome segment (Huismans and Howell, 1973). More recently, reassortment studies with BTV attributed increased neuro-invasiveness to VP5, an obvious virulence determinant (Carr *et al.*, 1994). Possibly an important role for the VP5 protein is to facilitate the interaction between the highly variable VP2 protein with the more highly constrained structure of the virus core, probably by imposing conformational constraints on VP2 (Cowley and Gorman, 1989).

The assembly of all seven structural proteins of BTV into virus-like particles has been demonstrated in the absence of viral genome or non-structural proteins, indicating that assembly is spontaneous requiring neither the genome nor the non-structural proteins (for review see Roy, 1992).

The structural proteins of EEV Cascara (EEV-2) have been characterized by comparison with BTV and AHSV using SDS-PAGE (Viljoen and Huismans, 1989). They identified seven structural proteins (VP1 to VP7 in analogy with BTV and AHSV) that ranged in molecular weight from 36 000 to 120 000. The protein fractionation pattern of EEV on polyacrylamide gels closely resembled that of AHSV and BTV, and showed a typical orbivirus profile with four major and three minor proteins. VP2 and VP5 of EEV was presumed to be the proteins in the outer capsid as they were lost or reduced in relative amount after centrifugation on caesium chloride density gradients, VP2 was also implicated as the possible serotype-specific antigen (Viljoen and Huismans, 1989). In a recent study by Potgieter *et al.* (2002) involving EEV Bryanston (EEV-1) and EEV Kyalami (EEV-5) the sizes of the genes encoding VP6 and VP7 were determined to be 1080 bp and 1175 bp respectively. This group also expressed proteins of EEV Kyalami (EEV-5) that were equivalent to the four major structural proteins of BTV namely VP2, VP3, VP5 and VP7, using the baculovirus expression system. Co-expression of the EEV Kyalami (EEV-5) VP3 and VP7 proteins resulted in the intracellular accumulation of particles they concluded to be EEV CLPs, since their size and morphology resembled that of native EEV cores and they had capsomeres typical of orbiviruses.

1.4.3.2. Non-structural proteins

At least three non-structural proteins (NS1 to NS3) have been identified (Gorman *et al.*, 1981; Mertens *et al.*, 1984). Van Dijk and Huismans (1988) and French *et al.* (1989) however reported that the NS3-encoding genome segment not only encodes NS3, but also a second, closely related polypeptide designated NS3A. Non-structural viral proteins are known to be involved in the replication and morphogenesis of viruses (specifically assembly and release) and are frequently responsible for the formation of characteristic structures in virus-infected cells. One of the non-structural proteins, NS3, has also been shown to influence virulence (O'Hara *et al.*, 1998). For BTV, the amino acid sequences of each of these NS proteins are highly conserved among different serotypes with an average homology of 96% (Huismans and Cloete, 1987; Huismans *et al.*, 1987a; Roy *et al.*, 1990).

Non-structural protein NS1:

Large quantities of NS1 are synthesized in orbivirus-infected (AHSV and BTV) cells, constituting approximately one quarter of the total virus-specified protein complement. Accumulation of NS1 in infected cells results in the formation of unique non-specific tubular structures (Mertens *et al.*, 1984; Nel *et al.*, 1990). The BTV NS1 protein is cysteine rich (Nel *et al.*, 1990) and NS1 has been found to be highly conserved within serogroups with strong regional conservation between serogroups (Huismans and Cloete, 1987; Nel and Huismans, 1991). Little is known about the function of NS1 tubules in BTV replication, it has been suggested that the tubules may be a repository of NS1 used in a prior stage of virus morphogenesis, and that NS1 may be involved in the translocation of virus particles to the host cell membrane (Hyatt *et al.*, 1993). This was supported by the demonstration of the association of NS1 with distinct areas of the VIBs especially at the periphery. These areas correlated with the presence of virus particles of all described size dimensions (Brookes *et al.*, 1993). This suggested that NS1 is involved in the early stages of BTV morphogenesis.

Non-structural protein NS2:

NS2 is also synthesized in large quantities and is responsible for the formation of virus inclusion bodies (VIBs) (Thomas *et al.*, 1990; Brookes *et al.*, 1993). NS2 has been shown to exist as multimers (Uitenweerde *et al.*, 1995) with ssRNA binding properties (Huismans *et al.*, 1987b; Theron *et al.*, 1994). This led to the proposal that NS2 may be involved in the selection and condensation of the RNA segments for encapsidation. The N-terminal half of NS2 is more conserved among BTV, AHSV and EHDV than the C-terminal half which has a hydrophilic character and is predicted to have a high content of α-helix conformation (Van Staden *et al.*, 1991). The N-terminal half of the three orbiviruses, as well as the ssRNA-binding proteins σ_{NS} of three reovirus serotypes and NS34 of bovine and simian rotavirus, contain a conserved 9 amino acid region. This, together with experiments where deletions of the N-terminal portion of the protein where found to abolish RNA binding activity (Zhao *et al.*, 1994), point to the importance of the N-terminus for NS2 function. NS2 has the unique property of being the only virus-specific phosphoprotein in BTV and AHSV-infected cells (Huismans *et al.*, 1987b; Theron *et al.*, 1994) The functional significance of NS2 phosphorylation is not yet known. Theron *et al.* (1994) reported that phosphorylation of NS2 of BTV, EHDV and AHSV affects ssRNA binding and may involve a ubiquitous cellular kinase.

Non-structural protein NS3/NS3A:

Unlike NS1 and NS2, the two closely related non-structural proteins NS3 and NS3A are synthesized in small amounts in orbivirus-infected cells (Huismans *et al.*, 1979, Van Dijk and Huismans, 1988; French *et al.*, 1989, Van Staden *et al.*, 1995). Both these proteins are encoded by genome segment 10. The two related protein products have been shown to be synthesized from two alternative in-phase initiation codons in the same open reading frame (Mertens *et al.*, 1984; Lee and Roy, 1986; Gould, 1988; Van Dijk and Huismans, 1988; French *et al*, 1989; Van Staden and Huismans, 1991;

Hwang *et al.*, 1992). The BTV counterparts have subsequently been identified in AHSV-infected cells; these proteins were designated NS3 and NS3A in accordance with BTV nomenclature by Van Staden and Huismans (1991), but they have also been described as NS4 and NS4a (Grubman and Lewis, 1992), and P21 and P20 (Laviada *et al.*, 1993). The nucleotide and amino acid sequences of the NS3 genes of all 9 AHSV serotypes have been determined (Sailleau *et al.*, 1997), following initial work by Van Staden and Huismans (1991) on S10 of AHSV-3 and AHSV-9, and De Sá *et al.* (1994) on S10 of AHSV-1, AHSV-4 and AHSV-8. Sequence data on the NS3 genes of a number of AHSV serotypes indicate that these genes are not as conserved as the cognate genes within the BTV serogroup (Hwang *et al.*, 1992; De Sá *et al.*, 1994). However, in all the BTV and AHSV NS3 genes investigated to date, two in-phase initiation codons are conserved in a suboptimal context for translation initiation, possibly explaining the low levels of expression of NS3 and NS3A (Gould, 1988; Van Staden *et al.*, 1995).

NS3 and NS3A proteins have characteristic features that are conserved amongst most orbiviruses (Lee and Roy 1986; Gould, 1988, Van Staden and Huismans, 1991; Hwang *et al.*, 1992; Moss *et al.*, 1992, Jensen *et al.*, 1994; Van Staden *et al.*, 1995; Yamakawa *et al.*, 1999). Analysis of the predicted amino acid sequence of the NS3 gene product of AHSV revealed two conserved hydrophobic regions in the C-terminal half of the proteins (residues 116 to 137 and 154 to 170), which may serve as transmembrane domains. Predicted protein sequence analysis of the NS3 gene products of the different BTV and AHSV serotypes has also revealed two N-linked glycosylation sites and a cluster of proline residues. In the case of AHSV, a cluster of five prolines is observed within the 13 amino acid region between residues 22 to 34, and in the case of BTV a cluster of six prolines within the 15 amino acid region between residues 36 to 50 has been observed. Alignment of the predicted amino acid sequences of NS3 of five different orbiviruses, BTV, AHSV, Palyam virus, Broadhaven virus and EHDV also revealed a 50 amino acid conserved region in the N-terminal half of the proteins between residues 43 to 92.

Models have been suggested for the membrane topology of a variety of orbivirus NS3 proteins. Van Staden *et al.* (1995) suggested a model for the membrane topology of AHSV NS3 that predicts that each of the two hydrophobic regions (HDI and HDII) of the protein spans the host cell membrane, resulting in the localization of both the N- and C-termini of NS3 on the cytoplasmic side of the membrane (Fig. 1.3.). This leaves a large part of the protein available for interactions with components of virus particles and the N-terminal domain may mediate the transport and egress of the virion from infected cells in the final stages of virus morphogenesis (Yamakawa *et al.*, 1999). Similar membrane topology models have also been predicted for the NS3 proteins of BTV, Chuzan virus, Palyam virus and Broadhaven virus (Van Staden and Huismans, 1991; Jensen and Wilson, 1995; Bansal *et al.*, 1998; Yamakawa *et al.*, 1999) confirming a conserved role for this protein amongst all orbiviruses.

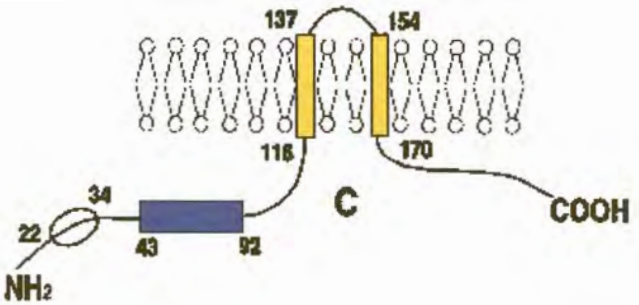


Figure 1.3. Model proposed for the membrane-associated topology of the AHSV NS3 protein. The yellow boxes represent membrane-spanning regions while the blue box depicts the 50 amino acid region, which is conserved amongst the different orbiviruses. The oval area illustrates a cluster of conserved proline residues (C = cytoplasm). The numbers refer to amino acid residues in AHSV-3 NS3 (adapted from Van Staden *et al.*, 1995)

A recent study by Beaton *et al.* (2002) has confirmed the role played by NS3 to mediate viral release. This group showed that NS3 interacts with a cellular protein p11, which is the light chain component of the annexin II complex, in which two heavy chains (p36) and two light chains (p11) interact to form the functional tetramer that is important for many cellular processes, including Ca^{2+}-dependant exocytosis and the correct trafficking of proteins out of the cell. They mapped the NS3 region of interaction with p11 to a 13-residue peptide found at the N-terminus of the protein and showed that it effectively competes with p36 for p11 ligand binding. This 13-residue peptide (the 13 residues between the initiation codon of NS3 and the initiation codon of NS3A) contains a putative amphiphatic α-helix and has been found to be highly conserved amongst various BTV serotypes, indicating its importance in the virus life cycle. They further showed that the C-terminal domain of NS3 interacts with VP2, the outermost protein of the fully assembled virus particle, thus suggesting that NS3 forms a bridging molecule that draws assembled virus into contact with the cellular export machinery, allowing orbivirus egress from infected cells. The exact method of virus release has however not been elucidated, but it is possible that the link between p11 and virus, via NS3, allows active transport of virus across the membrane. This suggests that NS3 may be analogous to rotavirus NSP4, which has been shown to bind cores and transport them across internal membranes to form mature enveloped virions (refer to section 1.5.7.). It can also be speculated, due to the fact that the region involved in the binding of p11, is found in NS3 and not in NS3A, that the synthesis of both types of proteins could be a form of regulation of the process of viral release.

A different model to that shown in Fig. 1.3. has however been suggested for the membrane topology of EHDV NS3. The first helix at position 115 to 135 (which is homologous to the 116 to 137 region of NS3 of AHSV) of the NS3 protein of EHDV-1 is predicted to be hydrophilic on one side and hydrophobic on the other, and would be expected to lie on the surface of the membrane, with the hydrophobic side touching the membrane. The second helix at position 158 to 178 is also predicted to be hydrophilic on the one side, but less so than the first, and there is a greater likelihood that this helix

inserts itself across the membrane rather than residing on top of it. This second helix was also predicted to be multimeric, i.e. associated with another transmembrane segment through use of a single cysteine residue at position 168 that forms a disulphide bridge with the cysteine residue of another transmembrane segment to form either a pair or a bundle. This would result in the combined helices forming a channel, stabilized by the disulphide bridge, where the non-polar sides were used for mutual association of the helices to each other (Jensen and Wilson, 1995). There is however no biochemical evidence for the dimerization of the NS3 protein of EHDV-1. The first membrane associated segment of EHDV-1 NS3 is consistent with the consensus sequence for a signal peptide. It has a highly positive charge (lysine, arginine, arginine, arginine) in the four residues preceding the hydrophobic region, followed by a slightly long hydrophobic region of 16 residues, and a strongly polar stretch of seven residues distally (serine, alanine, serine, threonine, leucine, threonine, serine). The first membrane associated region, although not likely to be stably integrated into the membrane, may function as an insertion signal peptide that is translocated into the ER lumen and lies buried in the lipid bilayer surface, while the second hydrophobic domain functions as a transmembrane segment, which then anchors NS3 as an integral membrane protein. Insertion signal peptides also often have a cleavage site for removal of the signal peptide after insertion of the protein in the membrane, the consensus sequence for a cleavage site was however not seen in the EHDV-1 NS3 sequence (Jensen and Wilson, 1995).

When a polypeptide is inserted into a membrane, either the amino or carboxy terminus is retained on the cytoplasmic side of the membrane. One indicator for determining this orientation is the presence of glycosylated residues, which are found on the luminal side of membranes. EHDV-1 NS3 has potential O- and N-linked glycosylation sites along its amino terminal end, this implies that the glycosylated amino terminus of EHDV-1 resides on the luminal side of the membrane while the carboxy terminus is free on the cytoplasmic side, having a complex extended-turn-helix conformation (Jensen and Wilson, 1995). All of the above information is used to predict the model of the membrane topology of EHDV-1 NS3 which is shown in Fig.1.4.

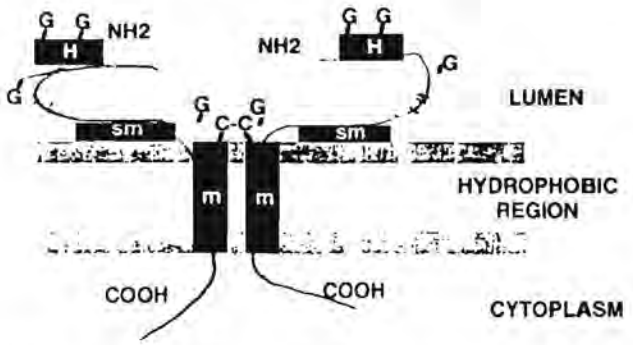


Figure 1.4. Model of the EHDV-1 NS3 protein inserted across the membrane. m = transmembrane segment; sm = surface membrane segment; H = hydrophobic region; G = potential glycosylation sites; C = cysteine residues; NH2 = amino terminus of protein; COOH = carboxy terminus of protein (adapted from Jensen and Wilson, 1995).

The 116 to 137 region in AHSV NS3 showed only 45% identity between the 9 AHSV serotypes (Sailleau *et al.*, 1997). The degree of variation in this region could be explained by the fact that this region would not be a complete transmembranous helix, as predicted by Van Staden *et al.* (1995), but would interact with the cellular membrane according to the membrane topology of the NS3 protein of EHDV (Wilson and Jensen, 1995; Sailleau *et al.*, 1997). Van Niekerk *et al.* (2001a) did not find any evidence to exclude any one of the two possible membrane topology models, it was however evident that only a very specific membrane-associated conformation of AHSV NS3, dependent on the structure of both hydrophobic domains, makes NS3 cytotoxic to host cells.

In BTV infected cells, the NS3 protein was found to be associated with intracellular smooth-surfaced vesicles (ER and Golgi) as well as with areas of the plasma membrane that had been disrupted by the egress of BTV (Hyatt *et al.*, 1989; Hyatt *et al.*, 1991). Similar results were obtained with a vaccinia virus expressing NS3, indicating that the localization of NS3 does not depend on the presence of BTV particles (Hyatt *et al.*, 1991). The exact role of NS3 and NS3A during virus replication is not clear, but co-expression of BTV structural proteins and NS3 by recombinant baculoviruses resulted in the release of BTV-like particles from infected insect cells (Hyatt *et al.*, 1993). These findings seem to indicate a role for BTV NS3 in the final stages of viral morphogenesis by enabling the release of progeny viruses from infected cells. Less information is available on the role of AHSV NS3 in the viral replication cycle, although results indicate a function similar to that of BTV NS3. AHSV NS3 has been shown to be membrane-associated and to be associated with areas of virus release involving both budding and extrusion (Van Staden *et al.*, 1995; Stoltz *et al.*, 1996). This membrane association of AHSV NS3 could result in a possible alteration of membrane permeability, caused by the insertion of the protein in the membrane, leading to osmotic disregulation and the eventual death and disruption of cells (Van Staden *et al.*, 1995; Stoltz *et al.*, 1996). This membrane destabilizing effect has also been detected in the disrupted plasma membranes of cells infected with an NS3 recombinant baculovirus (Stoltz *et al.*, 1996). The regulation of membrane permeability by NS3 could also lead to the control of the biochemical homeostasis of infected cells. NS3 may thus play a key role in both the viral pathogenesis and the viral release processes that regulate the disease progression of AHS (Van Niekerk *et al.*, 2001a).

The synthesis of AHSV NS3 in infected cells has been studied (Van Staden *et al.*, 1995). When the time and level of NS3 and NS3A expression in AHSV-3 infected Vero cells are monitored, both NS3 and NS3A are synthesized from 13 hours p.i. onward, and are present in approximately equimolar amounts, differing from the results obtained for BTV in which NS3 is the predominant protein product with nearly no NS3A being synthesized. This apparent aberration may find an explanation in the study of mRNA secondary structure, since two stable stem-loop structures, which can be formed on either side of the BTV NS3A initiation codon, may lead to disproportional amounts of NS3 and NS3A being expressed in BTV infected cells (Hwang *et al.*, 1992). The absence of similar structures in AHSV

mRNA possibly results in an equivalent frequency of initiation at either of the two AUGs, thereby producing equimolar amounts of NS3 and NS3A (Van Staden *et al.*, 1995).

When the AHSV NS3 gene is expressed using a recombinant baculovirus, low levels of a single product (NS3) of approximately 25 kDa are detected. No NS3A (24 kDa) is synthesized from the second initiation codon (Van Staden *et al.*, 1995) and it has been found in other studies that some membrane proteins are, for unknown reasons, not expressed to high levels by baculoviruses (King and Possee, 1992). It was further found that NS3 mRNA and protein were only synthesized during a relatively short period, after which time a drastic decline in the number of viable cells occurred (Van Staden *et al.*, 1995). Thus, the low levels of protein expression are due primarily to an apparent cytotoxic effect on the insect cells, and not because of a continued low level of expression during the viral infection cycle, although this may be a contributing factor. Electron microscopic examination of the morphology of cells expressing NS3 at late stages of infection revealed these cells to be completely permeabilized or disrupted (Van Staden *et al.*, 1995; Stoltz *et al.*, 1996), while immunofluorescence indicated a possible membrane association for NS3.

Immunofluorescence studies with a baculovirus-expressed BTV NS3 protein indicated that BTV NS3 and NS3A exist as N-linked glycoproteins and that they are transported to the cell membrane of infected cells via the endoplasmic reticulum (ER) and Golgi apparatus (Wu *et al.*, 1992). In another study, glycosylation was shown to be necessary for the correct processing of NS3 (Bansal *et al.*, 1998). NS3 and NS3A of six BTV serotypes all have two conserved glycosylation sites (Hwang *et al.*, 1992), while in contrast NS3 and NS3A of AHSV serotypes 4, 5, 6 and 7 (Sailleau *et al.*, 1997) have one potential glycosylation site but not in a conserved position, and the AHSV-3 NS3 protein does not contain such a site. Furthermore, no experimental evidence for the glycosylation of AHSV-3, 4 or AHSV-9 NS3 has been found (Grubman and Lewis, 1992; Van Staden *et al.*, 1995). The differences in the glycosylation of BTV and AHSV NS3 would not necessarily influence their predicted similar function in viral morphogenesis, as the requirement for carbohydrates for the transport of membrane proteins is not universal, and is protein specific (Doms *et al.*, 1993).

1.5. STRUCTURAL AND FUNCTIONAL CHARACTERISTICS OF MEMBRANE PROTEINS

Studies have indicated that BTV NS3, AHSV NS3 and Chuzan virus NS3 are all integral membrane proteins (Wu *et al.*, 1992; Yamakawa *et al.*, 1999; Van Niekerk *et al.*, 2001a). These proteins all have certain structural features in common, for e.g. two hydrophobic regions that are predicted to form transmembrane helices. For this reason, a brief discussion of membrane proteins, specifically integral membrane proteins, and their structures and functions will follow. The information in sections 1.5.1., 1.5.2. and 1.5.3. was obtained from Lodish *et al.* (1995); the information in subsequent sections was obtained from the references as indicated.

Membrane proteins are proteins that are associated with, or incorporated into, biological membranes. The ensemble of proteins associated with a membrane is a function of the cell and membrane type involved and the subcellular localization of the protein itself, which is determined by its function. Membrane proteins are usually found in only small amounts in cells, which makes the study of their structure and membrane interaction difficult. New technology and tools, notably bioinformatics, has made it possible to predict the structure of a protein from its amino acid sequence, which allows the prediction of its function, including its interaction with the membrane.

Membrane proteins can interact with membranes in various ways and can be classified into two broad categories based on the type of interaction: integral membrane proteins or peripheral membrane proteins.

1.5.1. Integral membrane proteins

Integral membrane proteins require the disruption of the lipid bilayer for release from the membrane. They contain one or more peptide segments that are embedded in the phospholipid bilayer itself, most integral proteins also contain residues with hydrophobic side chains that allow them to interact with the fatty acyl groups of the membrane phospholipids, thus effectively anchoring the protein within the membrane bilayer. Virtually all the integral membrane proteins characterized thus far have been transmembrane proteins, which span the phospholipid bilayer. These proteins may contain one or more membrane spanning regions (membrane-spanning α-helices), from a few residues up to several hundred residues in length, the bulk of which may extend into the aqueous medium bordering each side of the membrane. In some cases these multiple helices can interact with each other to form a coiled-coil structure stabilized by specific interactions between the amino acid side chains present at the interface between the two helices (Cohen and Parry, 1986). In contrast, some integral membrane proteins are anchored in only one of the two layers of the lipid bilayer, these proteins contain covalently bound hydrocarbon chains that function to anchor them in the bilayer. These membrane proteins are grouped into three classes according to the type of attached hydrocarbon chain. The first class is the glycosylphosphatidylinositol-anchored proteins, the second class is the farnesyl-anchored proteins and the third is the myristate-anchored proteins. The proteins of the third class are found within the cytoplasm, anchored to the cytosolic face of the plasma membrane by myristic acid, a 14-carbon saturated fatty acid. Myristate is always bound by an amide linkage to the glycine residue found at the N-terminus of such proteins.

1.5.2. Glycosylation

Carbohydrates are found associated with many of the known membrane proteins and the carbohydrate components of integral membrane proteins are almost always linked to the exoplasmic

domains of these proteins. When carbohydrates are covalently bound to proteins, the proteins become part of a class of proteins termed glycoproteins, which are especially abundant in the eukaryotic cell membrane and have also been adopted by virus systems to enhance their interaction with these cells. Carbohydrates have the function of increasing the hydrophilic character of proteins thereby increasing their solubility in aqueous medium, and may assist in the stabilization of protein conformations necessary for membrane association, protein folding and activity. The carbohydrate component of proteins are also necessary to ensure the correct integration of the protein in the membrane, as each type of membrane protein has a single, specific orientation with respect to the cytosolic and exoplasmic surfaces of the cell and is orientated accordingly during its synthesis.

Carbohydrate residues present in glycoproteins are linked to two classes of amino acid residues; sugars are O-linked if they are bound to the hydroxyl oxygen atom of serine, threonine or hydroxylysine, and N-linked when attached to the amide nitrogen of asparagine. The structures of O-linked and N-linked oligosaccharides are very different and most often contain different classes of sugars. N-linked oligosaccharides are generally longer and less varied than the O-linked components and often contain more sugar residues than is the case with the O-linked structures. N-linked glycans are often crucial to ensure the correct export of proteins from the endoplasmic reticulum and many proteins that are secreted from the cell are also glycoproteins. It is interesting to note that even oligosaccharides found at the same locus in a single protein type may frequently differ and the requirement for carbohydrates in the transport of membrane or secretory proteins is not universal and is highly protein specific.

1.5.3. Protein targeting and sorting

Protein targeting and sorting is of critical importance in determining the correct site of protein activity and function. Integral membrane proteins are part of the secretory pathway.

In order to be incorporated into a membraneous structure, the relevant proteins must be translated on membrane-bound ribosomes attached to the rough endoplasmic reticulum. Sequences up to 25 amino acids in length contained within the membrane-targeted protein called topogenic sequences, function to ensure that each transmembrane protein acquires its proper orientation during its insertion into the ER membrane during translation. These proteins then move, via transport vesicles, from the ER to their respective final destinations, which may be as diverse as the plasma membrane or the membranes of the lysosomal or Golgi apparatus. It is important to note that the subcellular targeting of viral (and cellular) membrane proteins are not exclusively determined by the sequences of the transmembrane regions. Retention sequences or transport sequences in the luminal and/or cytoplasmic domains of the protein, including glycosylation, myristylation or multimerization may further influence targeting. Specific mechanisms also operate to recognize and retain incorrectly folded proteins in the ER (Doms *et al.*, 1993).

It has also been found that plasma membrane glycoproteins follow the same maturation pathway as secretory proteins and it has been established that the newly synthesized glycoproteins are inserted into the rough ER membrane usually by way of a signal peptide translated early in the life of the targeted protein. The signal peptide is usually located at the N-terminal of the protein; it is 20 to 40 amino acids in length and contains a stretch of 9 hydrophobic amino acids flanked by basic residues on the N-terminal side. This signal peptide is recognized by a signal recognition particle, which directs the nascent protein to the rough ER. These proteins then move to the Golgi vesicles, targeting them to the plasma membrane. In addition to the signal peptide, an additional peptide signal consisting of a membrane spanning hydrophobic peptide is also needed to target a protein to the plasma membrane.

1.5.4. Integral viral membrane proteins

Virus particles enveloped in a membraneous structure usually contain one or more types of integral membrane proteins. These proteins have functions as diverse as recognizing receptors on the host cell plasma membrane, mediating penetration of host cells, co-ordinating viral morphogenesis at the budding site and serving as receptor-destroying enzymes necessary for viroid release. The majority of these proteins are also important antigens for the induction of neutralizing antibodies (Doms *et al.*, 1993).

Numerous viral integral membrane proteins have a single membrane-spanning domain. The largest parts of these proteins are usually found on the outside of the viral envelope. The cytoplasmic domains are usually less elaborate, but are nonetheless important in interactions with internal viral components. Nearly all viral membrane proteins examined to date have been found to be oligomeric structures, with some being homo-oligomers containing two to four identical subunits, while others are hetero-oligomers containing two or more different types of subunits. Usually, discrete genes encode the latter subunits, with each subunit consisting of either a single or double polypeptide chain, these being stabilized by non-covalent interactions or disulphide bonds. Interactions between adjacent subunits are extensive and extend throughout the length of the ectodomains (Doms *et al.*, 1993). The majority of these viral integral membrane proteins are also (just as is the case with normal cellular transmembrane proteins) post-translationally modified by the addition of N- or O-linked carbohydrate chains. The majority of the previously mentioned hydrophobic and disulphide bonds are generally hidden inside the folded subunits or subunit interfaces, while the N-linked oligosaccharides are found on the external surfaces (Doms *et al.*, 1993).

1.5.5. Folding of integral viral membrane proteins

Like most cellular glycoproteins, viral membrane proteins are typically translated on ribosomes bound to the membrane of the ER. These proteins are inserted co-translationally into the ER in an unfolded

form. Due to their being transmembrane proteins, their folding must also occur in three distinct environments: the ER lumen, the ER membrane and the cytosol. The large ectodomain which usually carries most of the protein's mass and all of the carbohydrate moieties and disulphide bonds, undergoes folding in the lumen of the ER, with the mechanisms resembling those associated with the secretory proteins and membrane glycoproteins of the cell. The transmembrane area(s) most likely adopt a α-helical structure, thereby anchoring the protein into the hydrophobic interior of the bilayer. The area, which is to become the cytoplasmic domain, generally follows the rules that apply for folding of cytosolic proteins (Doms *et al.*, 1993).

Folding of viral membrane proteins within the ER is not a spontaneous process, but is dependent on the presence of chaperone molecules and folding enzymes which are necessary to prevent the aggregation of proteins in the process of being translated, and to catalyse rate limiting steps such as the formation of disulphide bonds. This greatly increases the efficiency of the folding process and is especially important for the large, complex molecules, which sometimes constitute viral membrane proteins. However, the most important principle regarding protein folding is that the information required for a protein to attain its final three-dimensional structure resides in its primary amino acid sequence (Doms *et al.*, 1993). It is thus clear that folding of viral transmembrane proteins is a dynamic process involving a host of cellular folding enzymes and molecular chaperones, much the same as for cellular membrane proteins.

1.5.6. Integral viral membrane protein structures and functions

Various virus-encoded proteins have been identified which have transmembrane structures similar to those in eukaryotic transmembrane proteins. While showing structural similarity in the form of transmembrane spanning helices, coiled-coil structures and other elements found in membrane proteins of higher organisms, many of the models elucidated in the past few years indicate specific roles for these proteins in promoting successful virus attachment, disassembly, morphogenesis and release of virus particles. One of the more extensively characterized examples will be discussed, namely rotavirus NS28/NSP4 non-structural protein, with which BTV NS3 also share a number of similarities.

1.5.7. Rotavirus non-structural glycoprotein NSP4

Rotaviruses are non-enveloped and mature by budding through the rough (R) ER of infected cells prior to the final assembly of the outer capsid (Estes and Cohen, 1989). Subviral single-shelled particles (SSP) assemble in viral inclusion bodies bordering the RER membrane and subsequently bud into the lumen of the RER (Bellamy and Both, 1990). This event is mediated by the binding of VP6, a protein present on the SSPs, with a virus-encoded receptor, which is translocated prior to the

budding event to the RER membrane. Au *et al.* (1989) and Meyer *et al.* (1989) have shown that this receptor is the rotavirus non-structural protein NS28, which has subsequently been reclassified as non-structural protein NSP4.

NSP4 is a 28 K non-structural glycoprotein, classified as an integral membrane protein (specifically a transmembrane protein) in the ER membrane (Chan *et al.*, 1988; Au *et al.*, 1989; Estes and Cohen, 1989). This protein is encoded by genomic segment 10, is 175 amino acids in length (Chan *et al.*, 1988; Au *et al.*, 1989; Meyer *et al.*, 1989) and has three internal hydrophobic domains (H1, H2 and H3). The membrane topology of simian rotavirus non-structural glycoprotein NS28 (NSP4) has been determined by Chan *et al.* (1988) and is shown in Fig. 1.5. The amino terminus spans the membrane or is located on its lumenal side while the hydrophilic carboxyl half is exposed on the cytoplasmic side of the ER membrane. The hydrophobic region between position 7 to 21 (H1) was not predicted to be membrane-associated, the hydrophobic region between position 28 to 47 (H2) was suggested to be monomeric and a transmembrane region, and the third hydrophobic domain position 71 to 81 (H3) was predicted to be multimeric and a transmembrane region.

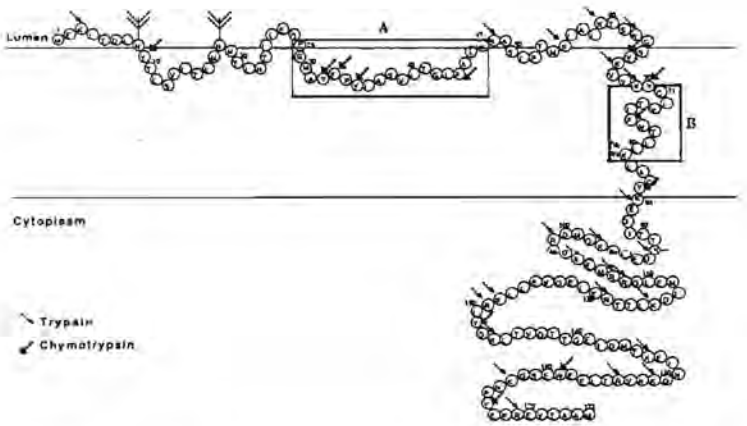


Figure 1.5. Proposed topology of NS28 (NSP4) in the ER membrane. Each amino acid is represented by one circle and the one-letter code. The carbohydrates on amino acids 8 and 18 are represented by Y. The small arrows pointing to the right indicate the possible trypsin cleavage sites, whereas the large arrows pointing to the left indicate the chymotrypsin cleavage sites. The boxed areas A and B show the regions predicted by analysis of hydrophobic moment plots to be transmembrane or membrane associated (adapted from Chan *et al.*, 1988).

Subsequent studies proposed that the second hydrophobic domain is the membrane spanning domain that anchors the protein, instead of the third domain as previously proposed (Estes and Cohen, 1989), a larger cytoplasmic domain is also suggested by this subsequent model. This means that the protein is orientated in the RER membrane with the amino terminus on the luminal side and approximately 131 amino acids accessible from the cytoplasmic side (Meyer *et al.*, 1989).

The three hydrophobic domains have specific functions within this membrane topology, based on their positions. The H2 domain contains a signal peptide and acts as the transmembrane domain leaving the C-terminal 131 residues exposed on the cytoplasmic side of the membrane and available to perform the receptor function which involves the binding of SSPs (Au *et al.*, 1989). H3 remains exposed in the C-terminal 131 residues on the cytoplasmic side of the membrane and allows the binding of the structural protein on the SSPs by orientating a C-terminal residue which is required to bind the structural protein (Taylor *et al.*, 1992). H1 is luminally orientated and includes two glycosylation sites at asparagine residues at amino acid positions 8 and 18 near the amino terminus, which are glycosylated with high mannose sugars (Chan *et al.*, 1988; Estes and Cohen, 1989).

The location of the glycosylation sites suggests that the signal peptide is not cleaved (Chan *et al.*, 1988). This NSP4 receptor protein is a homotetramer, its oligomerization is mediated by a coiled-coil domain that has been identified in the cytoplasmic tail, immediately adjacent to its membrane spanning region (Taylor *et al.*, 1996) and modification of its C-terminus has been shown to abolish SSP-binding (Au *et al.*, 1989; Taylor *et al.*, 1992). The budding event results in the inclusion of the virus in a transient envelope, but the mechanism whereby this is accomplished is poorly understood. Following translocation to the cell membrane, the membrane envelope containing NSP4 is removed and two structural proteins, VP4 and VP7, interact to form the outer virus capsid (Poruchynsky and Atkinson, 1991). It is not understood how the membrane envelope containing NSP4 is removed. However, it is likely that NSP4 itself is also involved in this process, and it has been shown that glycosylation of the protein is required for removal of the transient envelope from budding particles (Estes and Cohen, 1989). Non-enveloped particles are subsequently released from infected cells by cell lysis and cell death (Poruchynsky and Atkinson, 1991).

The region 114 to 135 of NSP4 has been predicted to fold as an amphipathic helix, to localize in the cytoplasmic domain and to be able to mobilize intracellular calcium in eukaryotic cells. A tyrosine residue at position 131 was subsequently shown to be important in the functionality of the protein (Ball *et al.*, 1996). This tyrosine-based motif in the cytoplasmic domain of membrane associated glycoproteins appears to be of critical value for other virus systems as well, such as the HIV-1 Env protein, as well as the alphavirus envelope protein (Zhao *et al.*, 1994). The 114 to 135 region also proved to be antigenic, NSP4 antibodies directed against the region showed the potential to block rotavirus-induced disease (Ball *et al.*, 1996).

Other studies indicated that NSP4 is cytotoxic to insect cells and it is hypothesized that the cytotoxic effect of this protein might be due to the sudden increase of intracellular calcium levels, leading to membrane destabilization and altered membrane permeability which causes the rupture of the cells and the subsequent release of rotavirus particles from the cell during late stages of the viral replication cycle (Tian *et al.*, 1994; Van Niekerk *et al.*, 2001a).

More recently, Newton *et al.* (1997) have delivered proof which indicates that NSP4 alters membrane permeability in cultured mammalian epithelial cells when expressed in a dual recombinant vaccinia virus system which mimics the approximate level of NSP4 synthesis as found in rotavirus-infected cells. Paradoxically, their work shows that the region of NSP4 necessary to effect this cytotoxic effect is not the region corresponding to residues 114 to 135 of the protein as was speculated by Ball *et al.* (1996) and Tian *et al.* (1996). Rather, they have shown by mutational analysis that it is an N-terminal region between amino acids 54 and 75 which causes the membrane destabilization activity as described. This region is predicted to form a protease accessible loop, which has the ability to form a helical structure with a pronounced amphipathic character. Cationic amphipathic helices are distinct motifs in lytic viral proteins, and cause permeabilization of cellular membranes. After having modified two clones with mutations overlapping the 54 to 75 residue sequence, Newton *et al.* (1997) found a significant reduction in the intrinsic membrane destabilizing activity of these mutants, indicating that this amphipathic α-helix represents a critical functional domain.

Zhang *et al.* (1998) recently sequenced two pairs of virulent and attenuated porcine rotaviruses. Comparisons made between the two pairs revealed that structural changes within a 10 amino acid region (amino acids 131 to 140) were important for pathogenesis. Their studies indicated that the 131 to 140 amino acid region, like the amino acid 54 to 75 region mentioned above, is also a domain critical to the functioning of NSP4.

1.6. AIMS OF THIS STUDY

From the review of the literature, it is clear that a vast amount of information has been gathered about the *Orbivirus* genus in the family *Reoviridae*. Most of the information however pertains to BTV and AHSV, while less is known about other viruses such as EHDV. Many of the details regarding structural characteristics, the process of viral replication, morphogenesis and assembly still need to be elucidated for other members of the family and this specifically includes the role of the viral non-structural proteins.

Over the past few years, equine encephalosis has emerged as an important disease of horses and other equids; a recent study by Lord *et al.* (2002) has indicated that in most parts of South Africa EEV has a higher transmission rate than AHSV (at least in donkeys), and from the literature review it is clear that little is known about the structural characteristics of EEV, specifically regarding its structural and non-structural proteins. To date, only two studies have been carried out regarding the molecular characterization of EEV. The first dates back to 1989 and involved the Cascara serotype (EEV-2) (Viljoen and Huismans, 1989). At that time Cascara (EEV-2) was still regarded as the predominant serotype and thus the representative of the EEV group; at present, the Bryanston serotype (EEV-1) has replaced Cascara as both the dominant serotype and as the representative of the group. The

second study is a very recent study and involves EEV Bryanston (EEV-1) and EEV Kyalami (EEV-5). Information regarding serotypes other than Cascara (EEV-2), EEV Bryanston (EEV-1) and EEV Kyalami (EEV-5), and information specifically regarding genome segment 10 and the gene product(s) it encodes, is lacking.

For these reasons it has become important to increase our knowledge of the virus. Investigations, such as that by Van Niekerk *et al.* (2001b), have shown that AHSV NS3, due to the large variation found in the protein, has potential for use as an epidemiological marker to subtype AHSV populations within an outbreak of AHS in order to trace the origin of the virus involved in the outbreak. NS3 of AHSV, BTV and EHDV has also been shown to play an important role in the release of virus particles from a cell and it also may be involved in virulence. Many other studies on the NS3 gene and its gene product(s) have also been undertaken in the case of BTV, AHSV, EHDV, and Palyam virus.

The long-term aim regarding these studies is to elucidate the structure/function relationship of NS3 and specifically its role in viral infection and morphogenesis, with emphasis on the mechanism responsible for viral virulence. Within that context, the aim of this study was to characterize the EEV NS3 gene and gene product(s) and to compare them to their orbivirus counterparts in an attempt to identify conserved and/or unique characteristics.

In order to achieve these aims, the following short-term aims were identified as part of this study:

1. To determine the NS3 sequences of the different EEV serotypes in order to identify unique structural features, and to compare the observed structural features and genetic variation within this serogroup to that observed for other orbiviruses (Chapter 2)

2. To determine if the expression of the EEV NS3 gene results in two overlapping gene products (NS3 and NS3A) and to establish a system to express NS3 of EEV in sufficient quantities for future functional studies (Chapter 3)

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.

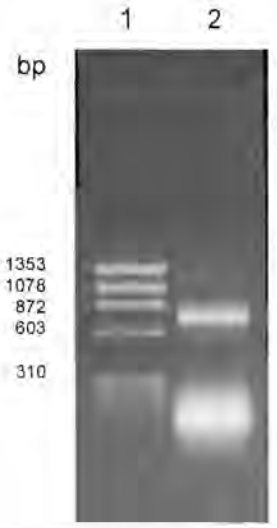


Figure 2.1. 1% Agarose gel electrophoresis of the RT-PCR product of EEV Bryanston (EEV-1) (S1REF*) (2) with a suitable size marker (φX 174) (1).

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.

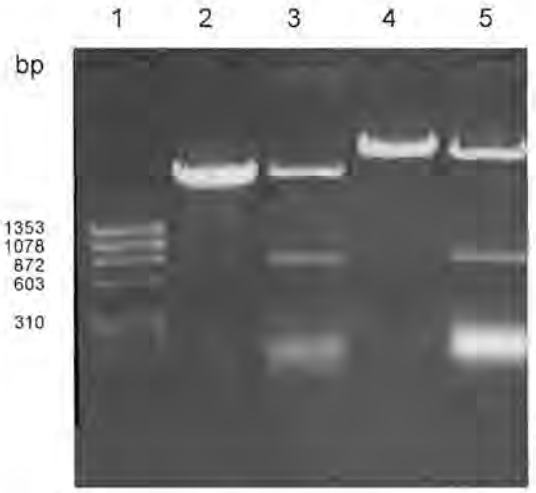


Figure 2.2. 1% Agarose gel electrophoresis of the putative recombinant plasmids after restriction enzyme digestion with *Bam*HI and *Eco*RI. Linearized pBS (2) and pFASTBAC1 (4) were included as controls to compare with the putative recombinant plasmids (3 and 5). φX 174 DNA size marker (1) was included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

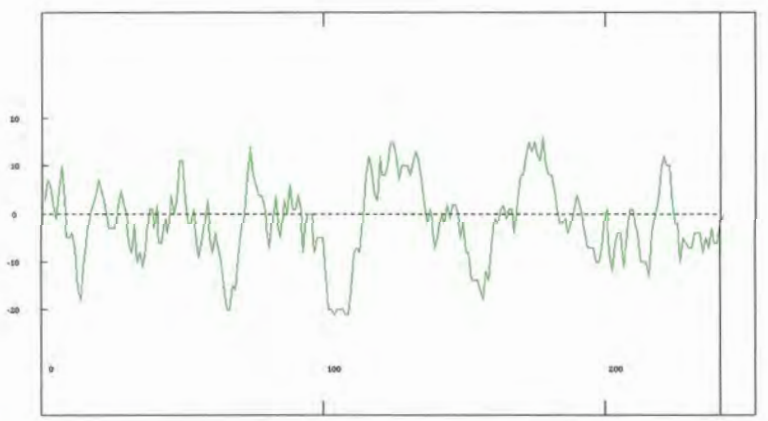


Figure 2.5. The hydrophobicity profile (Kyte and Doolittle, 1982) for EEV Bryanston (EEV-1) (S1REF*) NS3 protein determined using AnTheProt version 5 (Geourjon and Deléage, 1995).

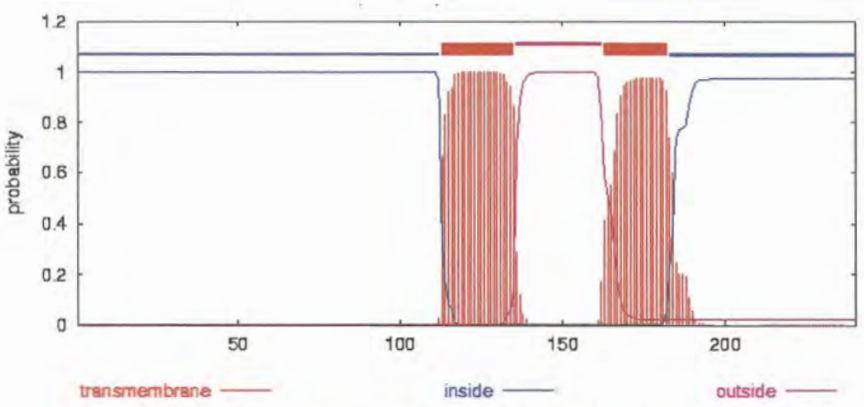


Figure 2.6. Transmembrane helices identified in the EEV Bryanston (EEV-1) (S1REF*) NS3 protein using TMHMM (Sonnhammer *et al.*, 1998). The figure shows the orientation of the protein in relation to the membrane.

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.

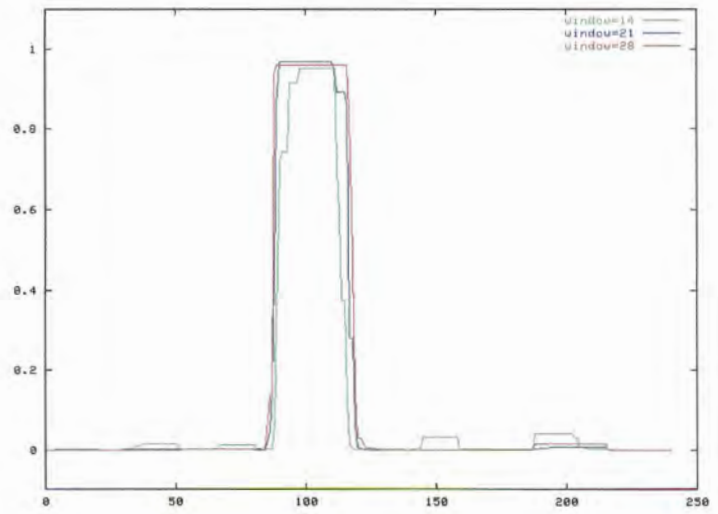


Figure 2.7. Coiled-coil regions identified in EEV Bryanston (EEV-1) (S1REF*) NS3 protein using COILS version 2.1. and 2.2. (Lupas, 1996). The default parameters, the MTK matrix and three window sizes namely 14, 21 and 28 were used to identify the coiled-coil domain.

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.

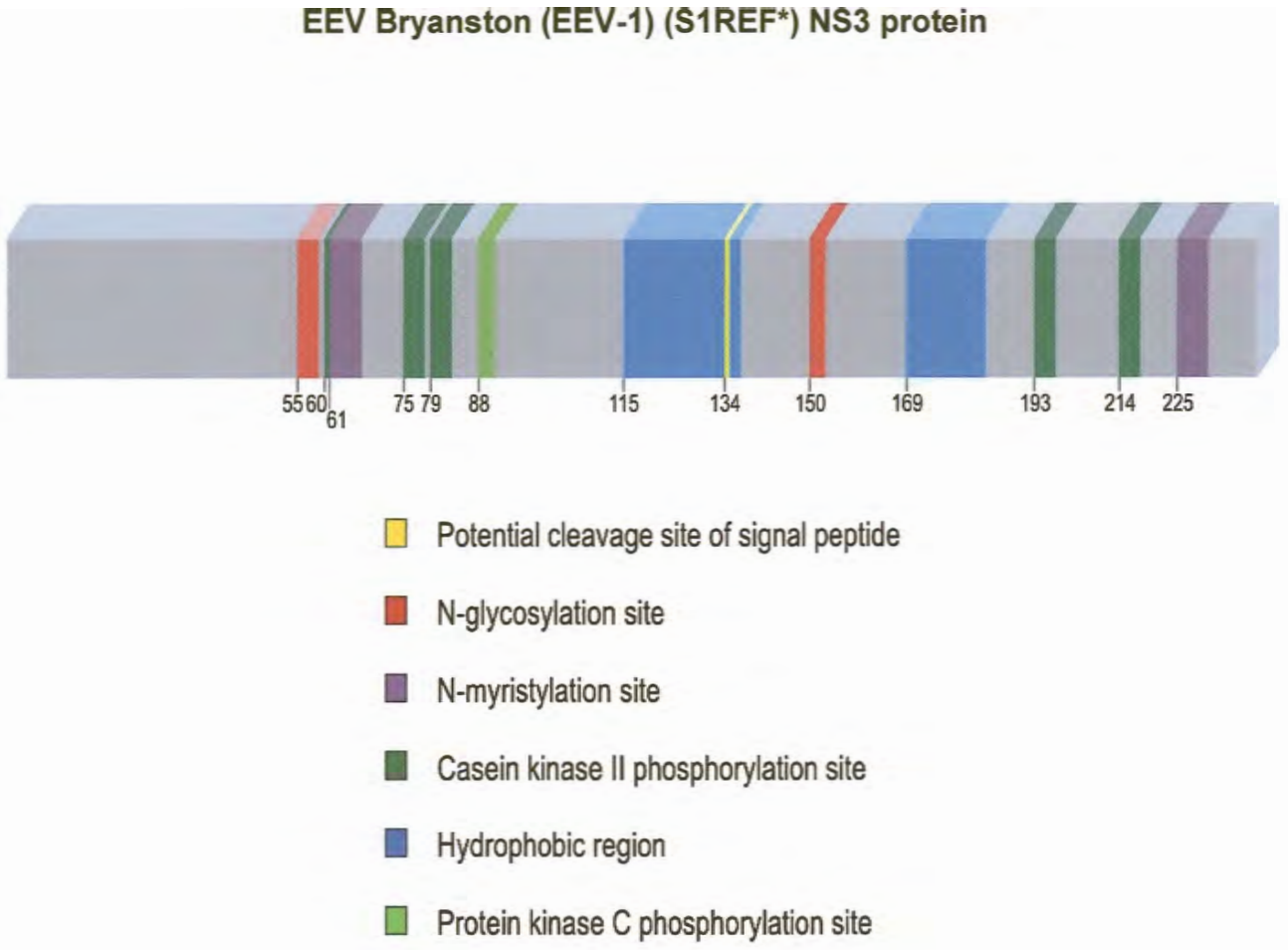


Figure 2.8. A summary of some of the different structural features or motifs that were identified for EEV Bryanston (EEV-1) (S1REF*) NS3 protein using different programs. The potential cleavage site of the signal peptide is indicated in yellow, the N-glycosylation sites are indicated in red, the N-myristylation 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.

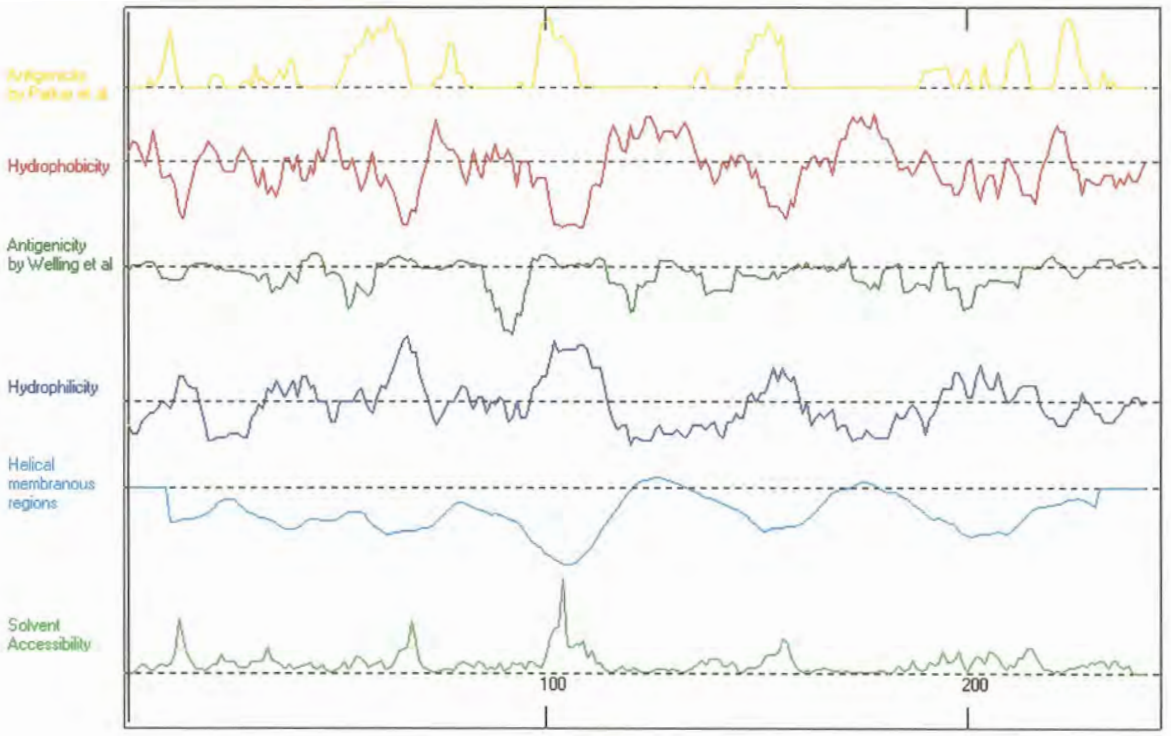


Figure 2.9. The different physico-chemical profiles that were determined for EEV Bryanston (EEV-1) (S1REF*) NS3 protein using AnTheProt version 5 (Geourjon and Deléage, 1995). The combined antigenicity profile is shown in yellow, the hydrophobicity profile is shown in red, the Welling antigenicity profile 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 α-helices and β-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.

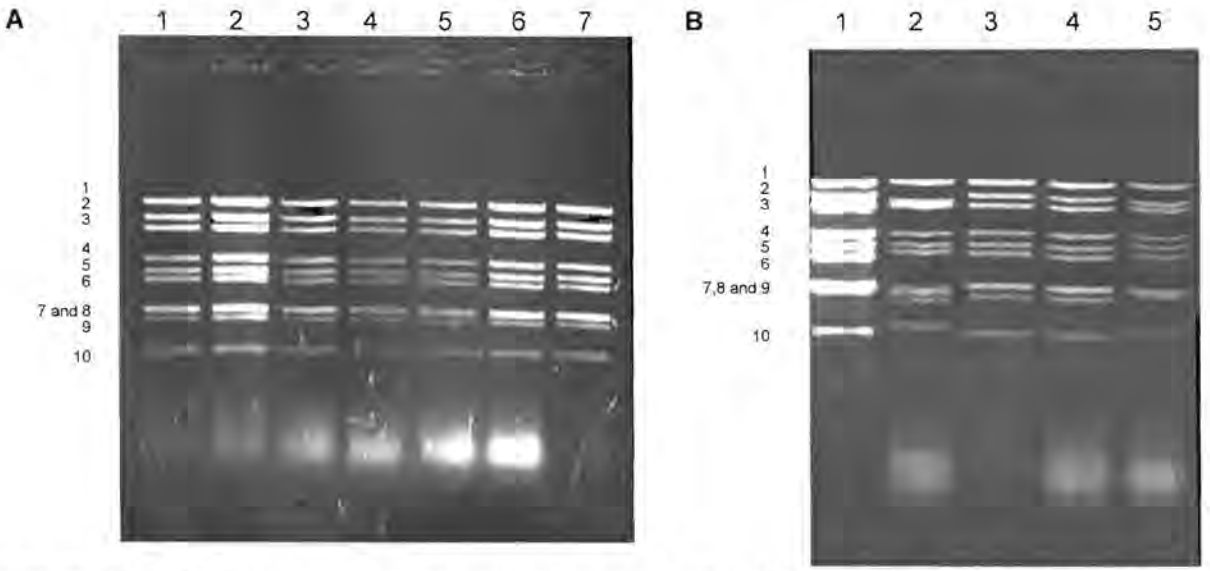


Figure 2.10. 1% Agarose gel electrophoresis to show the dsRNA profiles of the 7 different EEV serotypes (reference strains) as well as the dsRNA profiles of a number of orbiviruses compared to that of an EEV reference strain and an EEV field isolate. The seven EEV serotypes (A) are S1REF (1), S6REF (2), 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 S2REF	240 aa; ≈ 27 kDa; 8.93 240 aa; ≈ 27 kDa; 8.77	2	55 – 58 NITS 150 – 153 NKTG 222 – 225 NVTG (S2REF)	1	88 – 90 TIK	5	60 – 63 TGAD 75 – 78 SVAE 79 – 82 SAAD 193 – 196 TNAD 214 – 217 SNPD	2	61 – 66 GADECD 225 – 230 GGNTGD
S3REF	240 aa; ≈ 27 kDa; 9.03	2	55 – 58 NITS 150 – 153 NGTG	No sites identified		5	60 – 63 TGAD 75 – 78 SVAE 79 – 82 SAAD 155 – 158 TQGD 214 – 217 TNPD	4	61 – 66 GADECD 151 – 156 GTGQTQ 153 – 158 GQTQGD 225 – 230 GGNTGE
S4REF	240 aa; ≈ 27 kDa; 8.53	2	55 – 58 NITS 150 – 153 NGTG	1	88 – 90 TIK	5	60 – 63 TGAD 75 – 78 SVAE 79 – 82 SAAD 193 – 196 TNVD 214 – 217 SNPD	3	61 – 66 GADECD 151 – 156 GTGEAE 225 – 230 GGNTGD
S5REF S6REF	240 aa; ≈ 27 kDa; 8.84 240 aa; ≈ 27 kDa; 8.79	2	55 – 58 NITS 150 – 153 NGTN (S5REF) 150 – 153 NGTG (S6REF)	No sites identified (S5REF) 1 cAMP- and cGMP-dependent protein kinase phosphorylation site 110 – 113 RRRT (S6REF)		5	60 – 63 TGAD 75 – 78 SVAE 79 – 82 SAAD 155 – 158 TQGD 214 – 217 TNPD	3	61 – 66 GADECD 151 – 156 GTNPTQ (S5REF) 151 – 156 GTGHTQ (S6REF) 225 – 230 GGNTGE
S7REF	240 aa; ≈ 27 kDa; 8.55	2	55 – 58 NITS 150 – 153 NNTG	1	88 – 90 TIK	5	60 – 63 TGAD 75 – 78 SVAE 79 – 82 SAAD 193 – 196 SNAD 214 – 217 SNPD	2	61 – 66 GADECD 225 – 230 GGNTGD

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

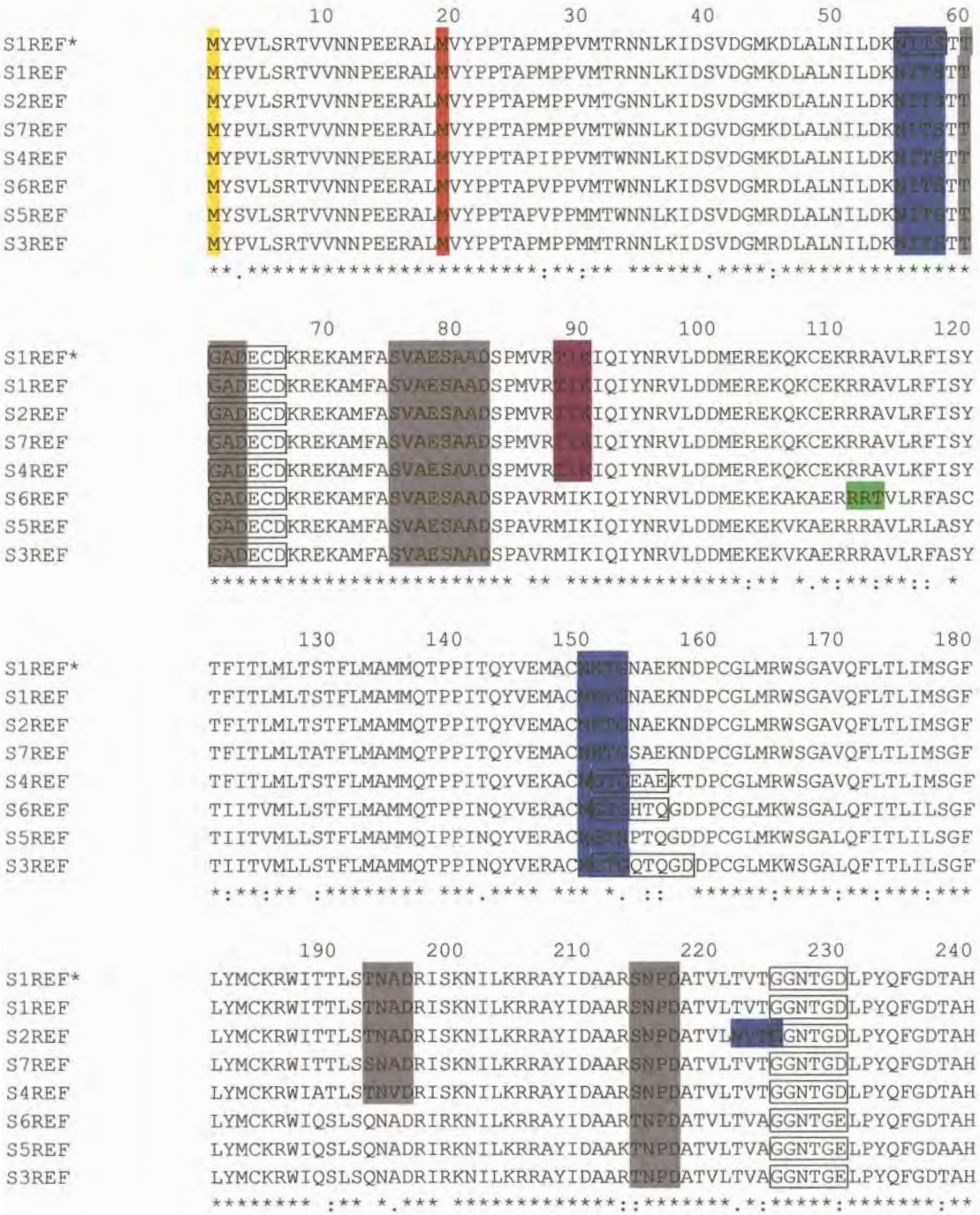


Figure 2.12. ClustalX 1.81 alignment of amino acid sequences of the seven EEV reference strains and EEV Bryanston (EEV-1) (S1REF*) to show similarities. Single, fully conserved, i.e. identical, residues are indicated with "*", "strong" groups that are conserved is indicated by ":" and "weak" groups that 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.

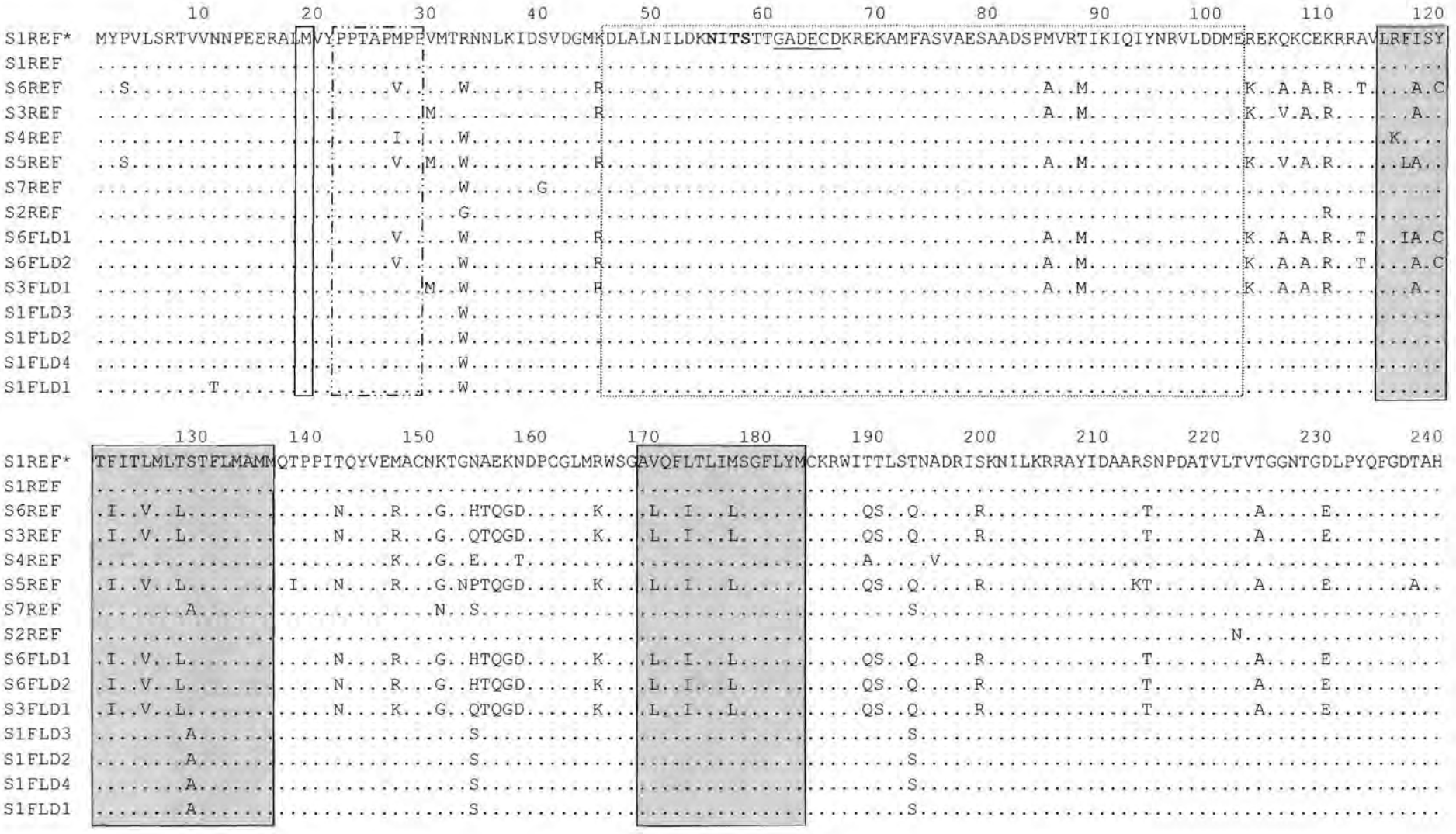


Figure 2.13. MEGA2 amino acid sequence alignment of all EEV isolates. Dots indicate identity to NS3 of the EEV Bryanston (EEV-1) laboratory reference strain, S1REF*. The NS3A start codon is blocked; the proline-rich region is blocked (amino acids 22 to 29). The highly conserved region is blocked (amino 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 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 γ, AHSV-3 and -7 form group β and AHSV-5, -9, -4, and -6 form group α.

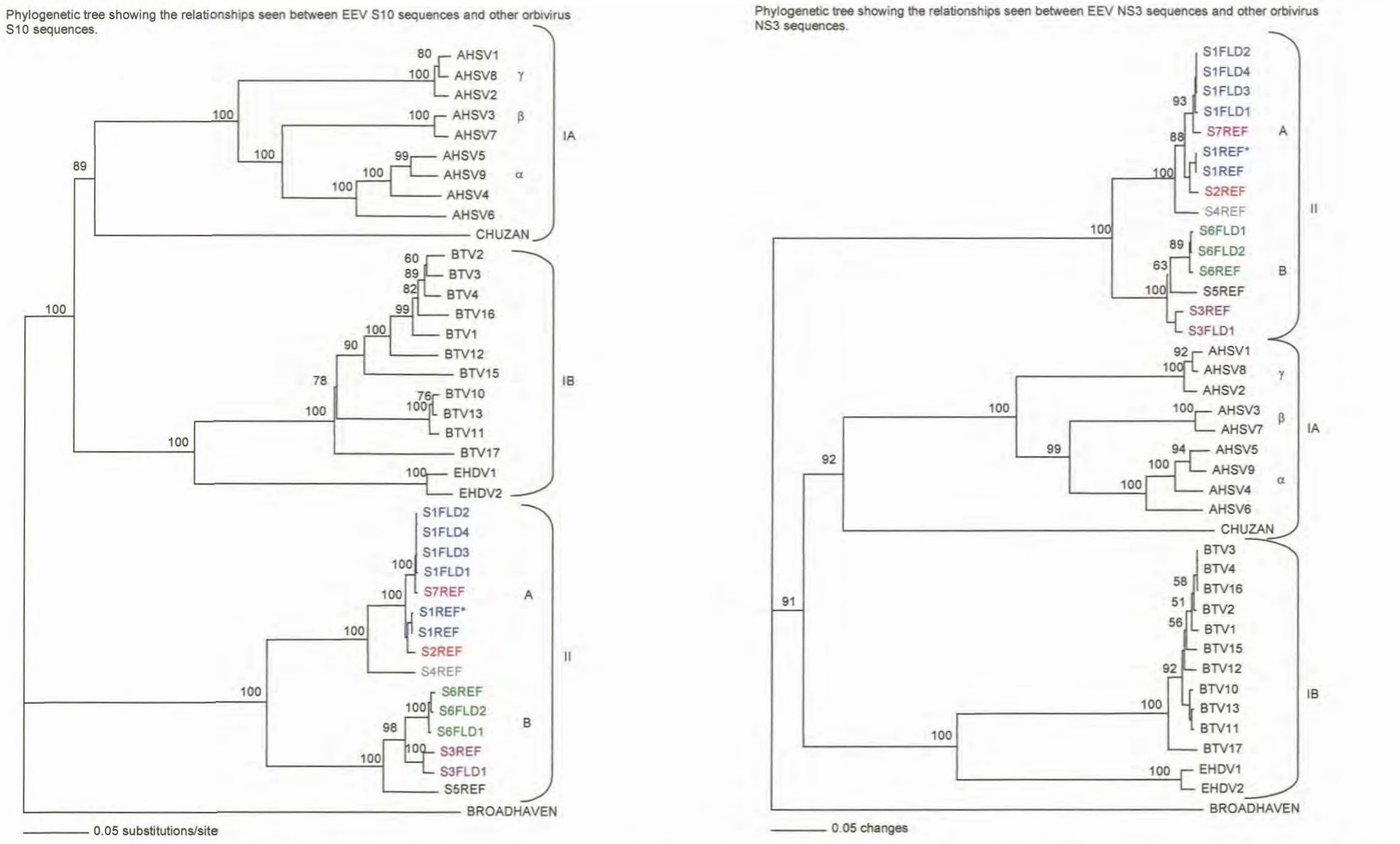


Figure 2.15. Phylogenetic trees of EEV S10/NS3 and other orbivirus S10/NS3 sequences. The horizontal branch lengths are indicative of the genetic distance between the sequences. The trees were constructed using the neighbour-joining method (uncorrected "p" for S10 and mean character difference for NS3) 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 28 K 25 K 26 K	15 K 25 K 24 K 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 25 K	25 K 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

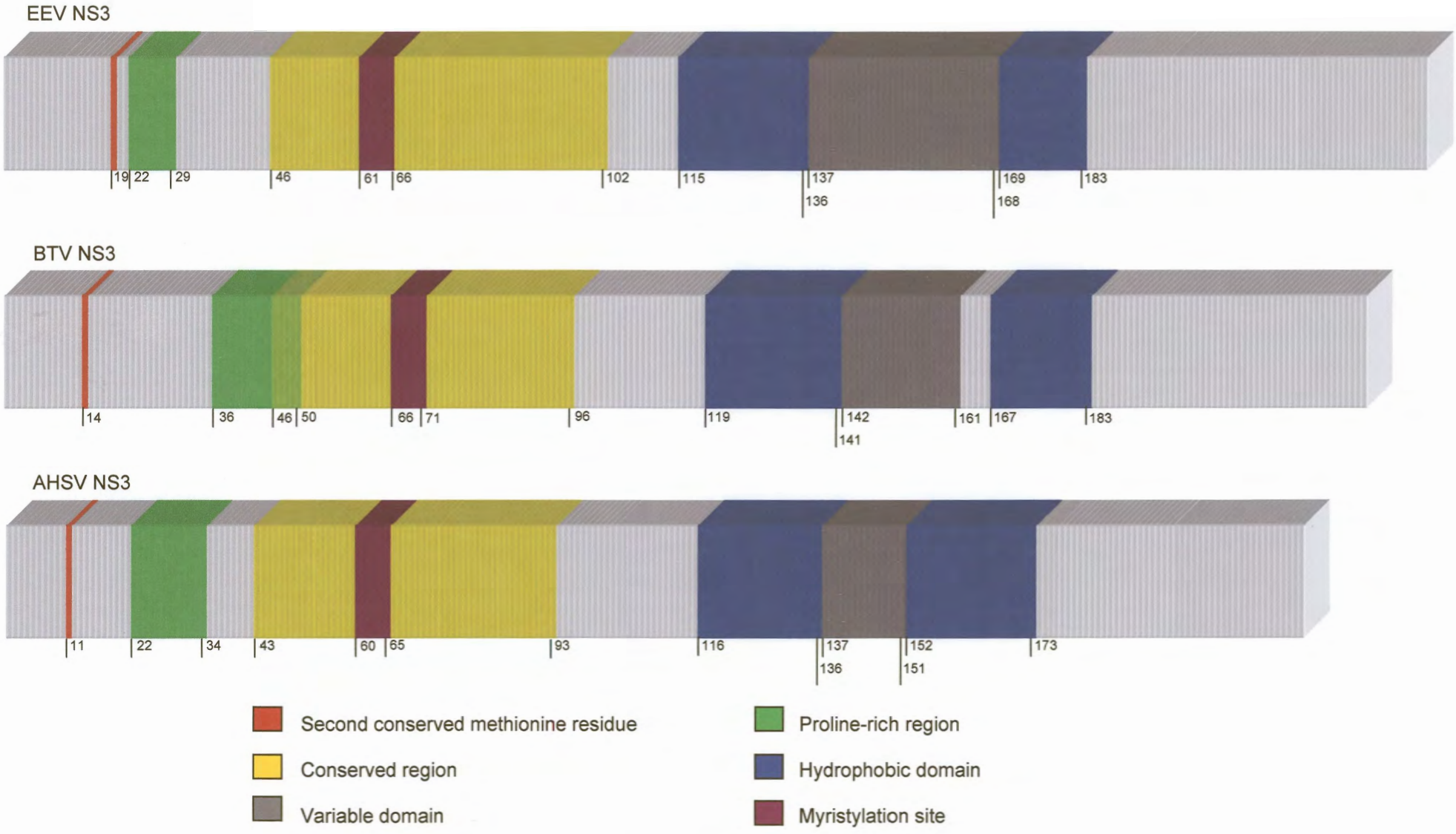


Figure 2.16. A comparison of conserved structural characteristics and protein motifs of NS3 proteins of selected orbiviruses.

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 - 65 150 - 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 – 58 150 – 153	N-linked	This investigation
EEV-2	222 – 225	N-linked	
EHDV-1	33 – 37 42 – 46	N-linked	Jensen and Wilson, 1995
EHDV-2	59 – 61 146 – 148	N-linked	Jensen *et al.*, 1994; Jensen and Wilson, 1995
	33 – 37 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.).

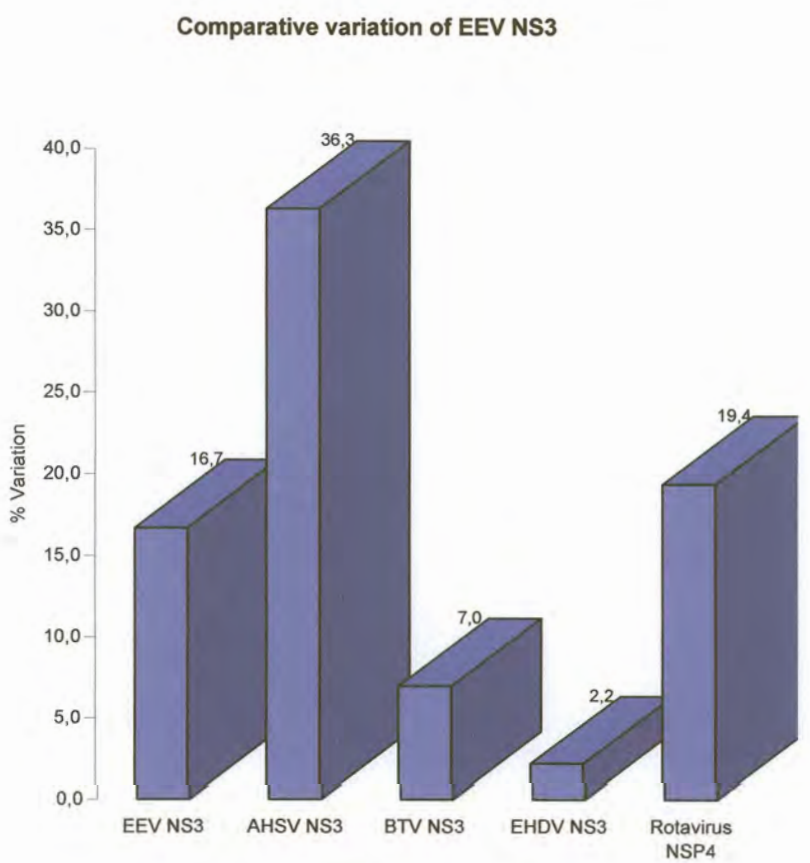


Figure 2.17. Comparative variation of EEV NS3. Total level of variation of EEV NS3 across serotypes is compared to that of AHSV NS3, BTV NS3, EHDV NS3 and rotavirus NSP4. Maximum variation percentages are indicated at the top of the bars (adapted from Van Niekerk *et al.*, 2001b).

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.

CHAPTER 3

EXPRESSION OF THE GENE ENCODING NON-STRUCTURAL PROTEIN NS3 OF EEV BRYANSTON (EEV-1)

3.1. INTRODUCTION

A number of conserved features or characteristics have been identified in the EEV NS3 protein in the preceding chapter, and similar features have previously been identified in other orbivirus NS3 proteins. In order to further study and characterize EEV NS3, specifically regarding the relationship between structure and function, it is therefore necessary to express the protein and to study its expression.

The NS3 genes of BTV, AHSV and EHDV have been demonstrated *in vitro* to code for two gene products, NS3 and NS3A (Van Dijk and Huismans, 1988; Van Staden and Huismans, 1991; Jensen *et al.*, 1994). Cloning and sequencing of these genes revealed the presence of two in-phase initiation codons, which could be responsible for the independent translation initiation of these two proteins. In the case of BTV and AHSV the NS3 and NS3A proteins are synthesized late in the infection cycle and are then only present at low concentrations *in vivo*, making it time-consuming or impossible to purify large amounts of protein via this route. More detailed information on the structure and function of NS3 and NS3A could only be obtained after expression of the proteins in yeast, insect and mammalian cells (e.g. French *et al.*, 1989; Hwang *et al.*, 1992; Hyatt *et al.*, 1993; Van Staden *et al.*, 1995; Stoltz *et al.*, 1996).

At present no data is available regarding the expression of EEV NS3 using *in vitro* or *in vivo* methods of expression. The first part of this chapter is therefore dedicated to the *in vitro* translation of the dsRNA of EEV Bryanston (EEV-1) alone and in comparison with AHSV, BTV and EHDV, while the second part of the chapter focuses on the expression of the EEV Bryanston (EEV-1) NS3 gene in an eukaryotic expression system. This should allow us to determine the nature of the segment 10 translation products and specifically whether the ability of BTV and AHSV segment 10 to express two proteins are conserved in EEV segment 10. The expression of the EEV Bryanston (EEV-1) NS3 gene product in an expression system will further allow the comparison of the mode of expression and expression levels to what has previously been observed for AHSV and BTV, and will establish a system to express and possibly purify NS3 of EEV in sufficient quantities to enable future functional studies (e.g. the study of post-translational modifications, and intracellular transport of proteins, and analysis of the properties of normal versus mutant proteins).

Various expression systems, prokaryotic and eukaryotic, are available for this purpose. Prokaryotic expression systems have the advantage that production can easily be scaled up to large volumes and

their host cells can be cultured continuously in relatively cheap medium, but in general such systems do not perform co-translational and post-translational modifications which are essential when immunogenic material suitable for production of antibodies is required, or if structure-function relationships are to be studied (King and Possee, 1992). Eukaryotic expression systems are capable of performing a number of post-translational modifications (e.g. glycosylation; fatty acid acylation (palmitylation and myristylation); nuclear transport; phosphorylation; C-terminus amidation; disulphide bond formation; proteolytic processing (including signal peptide cleavage, cellular targeting and secretion) and the formation of tertiary and quaternary structures), which would be found if the proteins were synthesized in the normal eukaryotic cellular environment (King and Possee, 1992).

The expression system of choice was the baculovirus expression system, an eukaryotic system, for which the techniques for cloning and expressing genes is well established in our laboratory. This expression system utilizes the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcMNPV) and insect cells. It has been used to express a wide range of proteins from many sources, including virus structural and non-structural proteins. The proteins are expressed to high levels and are accurately processed and biologically active, and in most cases the proteins have also proved to be antigenic (King and Possee, 1992).

Other advantages of the baculovirus expression system also include the fact that the foreign genes are under control of a strong late promoter, and are expressed after maturation of infectious virus particles (King and Possee, 1992). Consequently, a cytotoxic protein will not adversely affect virus replication. Baculovirus genomes can also accommodate large amounts of foreign DNA without affecting normal replication and DNA packaging, and can be propagated in a variety of insect cell lines (King and Possee, 1992). Baculoviruses are safe to work with as they have a restricted host range, and insect cells are amenable to large scale volume production in fermenter systems which allows the scale up off protein production (King and Possee, 1992).

3.2. MATERIALS AND METHODS

3.2.1. Cells and viruses

The origin of the EEV Bryanston serotype (EEV-1) (a laboratory reference strain) (S1REF*), the EEV Bryanston (EEV-1) (S1REF) reference strain and the EEV Bryanston (EEV-1) (S1FLD3) field isolate was described in section 2.2.1. All the available information for the laboratory strain, reference strain and field isolate was shown in Tables 2.1., 2.2. and 2.3.

Double-stranded RNA isolated from EEV Bryanston (EEV-1) (S1REF*) was used to synthesize a cDNA copy of genome segment 10 used for cloning into pFASTBAC1 (pFB.EEVB.S10).

Other viruses used in experimental procedures include AHSV M322/97 (AHSV-3), EHDV-1, and BTV-10. In the case of AHSV M322/97, EHDV-1 and BTV-10, purified dsRNA for each was obtained from Mr. M. K. Lombardi (Department of Genetics, University of Pretoria, South Africa). The origin and history of the AHSV M322/97 isolate has been described in Van Niekerk *et al.* (2001b). AHSV M322/97 was a serotype 3 field isolate, originally obtained from the OIE Reference laboratory at the Onderstepoort Veterinary Institute (OVI), South Africa and was isolated from the spleen of dead dog in Gauteng, South Africa in 1997.

3.2.2. *In vitro* translation

In vitro translations were performed using the rabbit reticulocyte lysate system (Amersham Pharmacia Biotech) according to the manufacturer's instructions. Generally single-stranded RNA (mRNA) is used for *in vitro* translation, but double-stranded RNA can also be used as a template (McCrae and Joklik, 1978).

Samples containing 0.5 - 2 µg of dsRNA were denatured by adding an equal volume of 10 mM MMOH and incubating at room temperature for 15 min (Mertens *et al.*, 1984). The *in vitro* translation reaction mixtures contained 0.5 - 2.0 µg denatured RNA, 8% (v/v) 12.5 x translation mix (25 mM DTT, 250 mM HEPES, pH 7.6, 100 mM creatine phosphate, 19 amino acids (312.5 µM each) minus the appropriate amino acid) (4µl of translation mix minus methionine), 100 mM KAc (2µl of 2.5M KAc solution), 0.5 mM MgAc (1 µl of 25mM MgAc solution), 40% (v/v) rabbit reticulocyte lysate (supplemented by calf liver tRNA, EGTA, creatine phosphokinase and hemin) (20 µl), RNase free water to the desired final volume of 25 or 50 µl and 10 - 15 mCi/ml (4 µl) [^{35}S]-methionine (EASYTAG™ METHIONINE L – [^{35}S]) (NEW™ Boston, MA, USA). The reactions were incubated at 30°C for 75 min and samples were stored at -20°C until use. Protein expression was analysed on 12% or 15% denaturing polyacrylamide gels (section 3.2.3.).

3.2.3. Polyacrylamide gel electrophoresis

In vitro translation protein samples were prepared for SDS-PAGE by removing 5 µl of the translation reaction and diluting with 50 µl 2 x PSB (0.125 M Tris, pH 6.8, 4% SDS, 20% glycerol, 10% 2-mercaptoethanol, 0.002% bromophenol blue), while protein samples from baculovirus infected *Sf*9 cells were treated with an equal volume of 2 x PSB. All protein samples were denatured by boiling at 95°C for 3 - 4 min before use.

SDS-PAGE was carried out as described by Laemmli (1970). The stacking gels contained 5% acrylamide, 0.125 M Tris, pH 6.8 and 0.1% SDS, and the separating gels 12% or 15% acrylamide, 0.375 M Tris, pH 8.8 and 0.1% SDS. The gels were prepared from a 30% acrylamide/0.8% bisacrylamide stock solution and polymerized by addition of TEMED and 10% ammonium persulphate. Electrophoresis was performed in TGS electrophoresis buffer (0.025 M Tris, pH 8.3, 0.192 M glycine, 0.1% SDS) using either the Mighty Small™ II SE 250 system (Hoefer Scientific Instruments) for 3.5 h at 120V, or the Sturdier SE 400 vertical slab gel units (Hoefer Scientific Instruments) for 16 h at 80V. Gels were either stained in 0.125% Coomassie blue, 50% methanol and 10% acetic acid for 20 min and destained in 5% methanol, 5% acetic acid overnight (or at 45 - 50°C for 10 - 15 min), or fixed in 7% acetic acid for 1 h at room temperature. Gels were then either stored in ddH_2O, or dried under vacuum on a piece of Whatman filterpaper in a slab gel-drier (Hoefer Scientific Instruments) for 2 h at 60 - 80°C.

3.2.4. Fluorography and autoradiography

Radiolabelled proteins separated by polyacrylamide gel electrophoresis (section 3.2.3.) were detected as follows. The proteins were fixed by staining/destaining or by using a fixing solution as described in section 3.2.3. The gel was rinsed using ddH_2O and incubated in Amplify™ (NAMP100) (Amersham Life Sciences) at room temperature for 15 - 30 min. The gel was rinsed briefly in ddH_2O, and dried under vacuum in a slab gel-drier (Hoefer Scientific Instruments) for 2 h at 70°C. The dried gel was exposed to X-ray film (Konica Medical Film AX) at -70°C and the X-ray film developed by agitating in developing solution (Ilford Multigrade Paper Developer) for 3 min or until image became visible, rinsing in H_2O, and fixing for 1 min in fixing solution (Ilfospeed Multigrade Paper Fixer). Finally, the gel was rinsed in H_2O, and allowed to dry.

3.2.5. Construction of a recombinant baculovirus for protein expression

The BAC-TO-BAC™ Baculovirus Expression System (GIBCO BRL Life Technologies) was used to construct a recombinant baculovirus for protein expression.

Insect cell culture handling techniques were essentially as described in the instruction manual provided for the BAC-TO-BAC™ Baculovirus Expression System (GIBCO BRL Life Technologies) with additional information obtained from King and Possee (1992). *Spodoptera frugiperda* (*Sf*9) cells (Highveld Biological) were grown as suspension cultures in spinner flasks at 27°C in Grace's insect cell culture medium containing 3.3 g/l yeastolate and 3.3 g/l lactalbumin hydrolysate (Highveld Biological), modified by the addition of pluronic (0.1% PLURONIC F-68 solution) (Sigma Cell Culture), containing 10% foetal calf serum (FCS) (Highveld Biological) and antibiotics. For suspension cultures the cell density was determined using a haemocytometer; trypan blue (0.4% in PBS) was added to check cell viability. Cultures were seeded at an initial density of 0.2 - 0.5 x 10^6 cells/ml, and subcultured when they reached 1 - 2 x 10^6 cells/ml.

3.2.5.1. Preparation of the donor plasmid

The donor plasmid, pFASTBAC1 expression vector, contained the polyhedrin promoter followed by an extensive MCS. The gene of interest was cloned into the donor vector in the correct orientation with respect to the polyhedrin promoter; i.e. the 5' end of the gene was inserted into the first selected site of the MCS. The integrity of recombinant clones was confirmed by automated sequencing (section 2.2.12.) before being used to transform competent DH10BAC cells.

3.2.5.2. Preparation of competent DH10BAC cells

MAX EFFICIENCY DH10BAC™ cells (GIBCO BRL Life Technologies) are provided as one of the components of the Baculovirus Expression System. If competent DH10BAC cells are not available, they can be prepared using the DMSO method (Chung and Miller, 1988).

Five ml of the *E. coli* strain was grown overnight in LB-broth at 37°C with shaking, and 1 ml of this overnight culture was used to inoculate 100 ml LB-broth. The cells were grown to early log phase (OD_{600} = 0.3 - 0.6) and

collected by centrifugation (2 860 x g for 5 min) at 4°C. The cells were resuspended in 1/10th of the culture volume TSB (10% (w/v) PEG, 5% (v/v) DMSO, 10 mM $MgCl_2$, 10 mM $MgSO_4$, in LB-broth) and incubated for 10 - 20 min on ice. Cells were used directly or were frozen away at -70°C in 400 µl aliquots containing 15% glycerol.

3.2.5.3. Transformation of competent DH10BAC cells with pFASTBAC donor plasmid

Competent DH10BAC cells (200 µl) were mixed with up to 100 ng of recombinant pFASTBAC1 donor plasmid DNA and incubated on ice for 30 min. The cells were then heat shocked at 42°C for 45 sec and chilled on ice for 2 min. This was followed by the addition of 0.9 ml TSBG (TSB plus 20mM glucose) and incubation with agitation at 37°C for 4 h before plating out 200 µl per LB-agar plate. The LB-agar plates contained 50 µg/ml kanamycin sulphate, 10 µg/ml tetracycline and 7 µg/ml gentamycin, as well as 50 µl 2% X-gal substrate and 10 µl 100 mM IPTG inducer which were added to each plate before plating. The plates were incubated at 37°C for two days.

3.2.5.4. Isolation of recombinant bacmid DNA

The following protocol, as provided by GIBCO BRL Life technologies, was adapted for the isolation of high molecular weight bacmid DNA (Luckow *et al.*, 1993), and it involved the following. Two to five ml LB-medium (containing 50 µg/ml kanamycin, 7 µg/ml gentamycin, 10 µg/ml tetracycline) was inoculated with each of the selected colonies and incubated for 16 h with shaking at 37°C. When well grown, 1.5 ml culture was transferred to an eppendorf tube and centrifuged at maximum speed (14 000 x g) in a microfuge for 1 min. The supernatant was removed, centrifugation repeated and the remaining supernatant removed. The cells were resuspended in 0.3 ml of a solution containing 15 mM Tris-HCl, pH 8 and 10 mM EDTA, after which .0.3 ml of a solution containing 0.2 N NaOH and 1% SDS was added and the sample mixed. The sample was incubated at room temperature for 5 min, after which 0.3 ml 2.5 M KAc, pH 5.5 was added, the sample mixed and left on ice for 5 - 10 min. The sample was centrifuged for 10 min at 14 000 x g. A new microfuge tube was labelled and 0.8 ml isopropanol added to it. The supernatant was transferred to the tube containing isopropanol, mixed and placed on ice for 5 - 10 min. The sample was stored overnight at -20°C or carried on directly and centrifuged for 15 min at 14 000 x g. The supernatant was removed, 0.5 ml 70% ethanol added and the sample centrifuged for 5 min at 14 000 x g. If desired, this wash step was repeated. The supernatant was removed, the pellet briefly air-dried for 5 - 10 min at room temperature and resuspended in 30 µl ddH_2O. The DNA was stored at -20°C until use.

3.2.5.5. PCR analysis of bacmid DNA to confirm the presence of the gene of interest

A standard PCR reaction was performed and the reaction contained 1.5 µl bacmid DNA, 5 µl thermophilic DNA polymerase 10 x buffer (10 mM Tris-HCl pH 9.0, 50 mM KCl and 0.1% Triton® X-100) (Promega), 3 µl 2.5 mM dNTP mix (TaKaRa Biomedicals), 2.5 µl of each primer (POLH and M13-RP, each 100pmol/µl) (refer to Table 2.5.), 3 µl 25 mM $MgCl_2$, 0.5 µl Taq polymerase (5U/µl) (Promega), and ddH_2O to a final volume of 50 µl. The PCR program used consisted of one cycle of 2 min at 94°C, 30 cycles of 45 sec at 94°C, 45 sec at 55°C and 5 min at 72°C, followed by one cycle of 7 min at 72°C. PCR was performed using a GeneAmp PCR System 9600 (Perkin Elmer). A small volume of each reaction was analysed by 1% agarose gel electrophoresis (section

2.2.3.) using suitable DNA molecular weight size markers to estimate the size of the PCR products, and the remainder of the reactions was stored at -20°C until use.

3.2.5.6. Transfection of recombinant bacmid DNA into *Spodoptera frugiperda* (*Sf*9) cells

Cells were seeded (1 x 10^6 cells/cm^2 for 6-well plates (Nunclon™)) in Grace's medium (Highveld Biological) without supplements (antibiotics, pluronic and FCS) and allowed to attach at 27°C for at least an h. Two solutions were prepared, solution A contained 5 µl bacmid miniprep plasmid DNA diluted into 100 µl Grace's medium without supplements, and solution B contained 6 µl CELLFECTIN™ reagent (GIBCO BRL Life Technologies) diluted into 100 µl of Grace's medium without supplements. The two solutions were combined, mixed and incubated at room temperature for 45 min. The cells were washed once or twice with 2 ml Grace's medium without supplements. Grace's medium without supplements (800 µl) were added to each tube containing DNA-lipid complexes and mixed gently. The wash media were removed from the cells and the washed cells overlayed with 1 ml of the diluted DNA-lipid complexes. The cells were incubated at 27°C for 5 h. The transfection mixture was removed, 2 ml Grace's medium containing antibiotics, pluronic and FCS added, and the cells incubated at 27°C for a further 96 h.

3.2.5.7. Harvest and storage of recombinant and wild type baculoviruses

Baculoviruses were harvested from transfection or p.i. supernatants by transfer of the supernatant to a sterile, capped tube. If so desired, the supernatant could be clarified through the use of centrifugation (5 min at 500 x g). The baculovirus was then contained in the supernatant, which was again transferred to a fresh, sterile tube. The baculovirus stock was stored at 4°C, protected from light. For long term storage of baculovirus at -70°C, fetal bovine serum was added to a final concentration of 2%.

3.2.6. Infection of insect cells with a baculovirus to analyse protein expression

Analysis of recombinant or wild type baculovirus expression was carried out in 24-well plates using virus stocks harvested 96 h p.i. (section 3.2.5.7.). The protocol used was as provided by GIBCO BRL Life Technologies, and was adapted from a protocol described by Luckow and Summers (1988).

Cells were seeded (0.4 x 10^6 cells/cm^2 in 24-well plates) in Grace's medium (Highveld Biological) containing antibiotics, pluronic and FCS, and allowed to attach for at least 60 min at 27°C before washing the cells once with Grace's medium containing antibiotics, pluronic and FCS. New medium was added to each well and the appropriate volume of recombinant or wild type baculovirus stock so that the total volume per well was 500 µl. The plates were then incubated at 27°C for 72 h. The cells were harvested by removing the medium from each well and using it to rinse the cells from the well. The medium containing the cells, was put into an Eppendorf tube and the cells collected by centrifugation at 800 x g for 5 min. The supernatant was discarded and the remaining cells washed from the wells with 500 µl 1 x PBS (137 mM NaCl, 2.7 mM KCl, 4.3 mM $Na_2HPO_4.2H_2O$, 1.4 mM KH_2PO_4, pH 7.3), which was added to the cell pellet and the cells resuspended. Again,

the cells were collected by centrifugation at 800 x g for 5 min. The cell pellet was resuspended in 40 µl 1 x PBS and stored at -20°C until use. Protein expression was analyzed by SDS-PAGE (section 3.2.3.).

3.2.7. Amplification of recombinant baculovirus and wild type baculovirus stocks

Cells were seeded (1 x 10^6 cells/cm^2 for 6-well plates) in Grace's medium (Highveld Biological) with antibiotics, pluronic and FCS, and allowed to attach at 27°C for at least an h. The media were removed from the cells and 1.8 ml medium containing antibiotics, pluronic and FCS added per well, followed by 50 - 100 µl of the virus stock that needed to be amplified. The cells were incubated at 27°C for 48 h. After 48 h, the virus was harvested as described in section 3.2.5.7.

3.2.8. ^{35}S-methionine labelling of proteins

Baculovirus-infected insect cells can be labeled with ^{35}S-methionine in methionine deficient media to monitor protein expression (Luckow and Summers, 1988).

Cells were seeded (1 x 10^6 cells/cm^2 for 6-well plates) and incubated at 27°C for at least 1 h. New Grace's medium (Highveld Biological) containing antibiotics, pluronic and FCS, as well as the appropriate volume of recombinant baculovirus stock, were added to each well so that the total volume per well was 1 ml. After incubation for 1 h at 27°C, 2 ml Grace's medium containing antibiotics, pluronic and FCS was added to each well, and the cells were incubated for 18 - 24 h at 27°C. After incubation, the medium was removed and the cells washed twice with 2 ml methionine-free Eagle's medium (minimum essential medium without L-methionine with non-essential amino acids, with L-glutamine and 2g/ml $NaHCO_3$) (Highveld Biological). The wash medium was replaced with 1 ml fresh methionine-free Eagle's medium and the cells starved for 1 h at 27°C to deplete the intracellular pools of methionine. The medium was replaced with fresh methionine-free Eagle's medium (1 ml per well) containing 30 µCi [^{35}S] methionine/ml (EASYTAG™ METHIONINE L – [^{35}S] (NEW™ Boston, MA, USA) and labelling was allowed for at least 3 h. The cells were harvested as described in section 3.2.6. and protein expression was analysed using SDS-PAGE (section 3.2.3) followed by fluorography and autoradiography (section 3.2.4.)

3.3. RESULTS

In order to determine if EEV segment 10 expressed two related proteins, as has been observed for segment 10 of other orbiviruses such as BTV, EHDV, Palyam virus, and AHSV, *in vitro* translation experiments using purified dsRNA were performed and the protein products of the NS3 gene of EEV and the other orbiviruses compared. In addition, to study the expression of EEV NS3 and to establish a system to express NS3 of EEV in sufficient quantities for future functional studies, a recombinant baculovirus was constructed to express the EEV Bryanston (EEV-1) (S1REF*) NS3 gene in the BAC-TO-BAC™ Baculovirus Expression System. The results obtained using *in vitro* translations were also compared to the expression of the EEV NS3 gene in the baculovirus expression system.

3.3.1. *In vitro* translation of dsRNA of EEV Bryanston (EEV-1) (S1REF) and other orbiviruses

In vitro translation reactions were performed using a commercially available kit of the rabbit reticulocyte lysate system (Amersham Pharmacia Biotech) according to the manufacturer's instructions (section 3.2.2.). The proteins in the translation reaction mixtures were analysed by 12% SDS-PAGE (section 3.2.4.), followed by fluorography and autoradiography (section 3.2.4.) and the results are shown in Fig. 3.1. and 3.2.

At least 10 proteins were synthesized during the *in vitro* translation reaction of dsRNA of EEV Bryanston (EEV-1) (S1REF) (Fig. 3.1.). The proteins were named using the nomenclature described for other orbiviruses by comparing the obtained EEV protein profile with that of other orbiviruses such as BTV (refer to Table 1.6.). At this stage the coding assignment of EEV is unpublished and the proteins as they are observed in Fig. 3.1. and the assignments they have been given, have not been verified, but are based only on analogy to BTV and AHSV proteins.

The *in vitro* translation experiment using dsRNA of EEV was repeated, and dsRNA of AHSV, BTV and EHDV isolates were included in order to compare the *in vitro* translation of these orbiviruses to that of EEV, specifically regarding genome segment 10. The results can be seen in Fig. 3.2. The *in vitro* translation reactions were not equally successful, e.g. the lanes containing the BTV or EHDV samples, this could be due to a variety of factors, the most important factor probably relating to the concentration of denatured dsRNA within the reaction. When the concentration of ssRNA is low, fewer templates are available for translation, which in turn leads to less protein being made. Generally, the smaller RNA segments are translated more frequently in the *in vitro* system than the larger RNA segments, which would explain the fact that the NS3 protein was observed even when all the other proteins were not.

The sizes of the NS3 proteins seen on the 12% SDS-polyacrylamide gels (Fig. 3.1. and 3.2.) were estimated using the Rainbow™ protein molecular weight marker (Amersham) as a standard and plotting $\log_{10}$ of molecular weight as a function of the distance migrated by the various proteins. The estimated size of AHSV, BTV and EHDV NS3 was approximately 24 K; the estimated size of EEV NS3 was approximately 27 K, which is similar to the predicted molecular weight (section 2.3.3.).

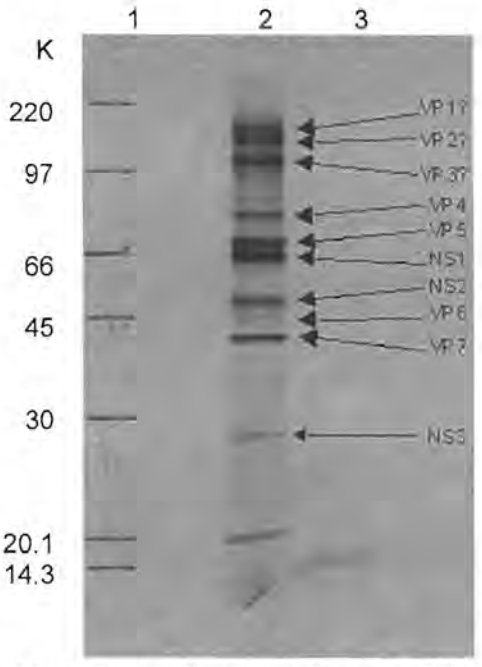


Figure 3.1. Proteins synthesized from the *in vitro* translation of dsRNA of EEV Bryanston (EEV-1) (S1REF) (2) analysed by 12% SDS-PAGE followed by fluorography and autoradiography. A negative control sample, in which no RNA was added to the *in vitro* translation reaction (3), was also included. Molecular weight standards (1) are Rainbow protein markers (non-radioactive, positions redrawn from gel). The arrows indicate the positions of the different protein products. The identities of the proteins have not been verified and the assignments are based on analogy to AHSV and BTV proteins.

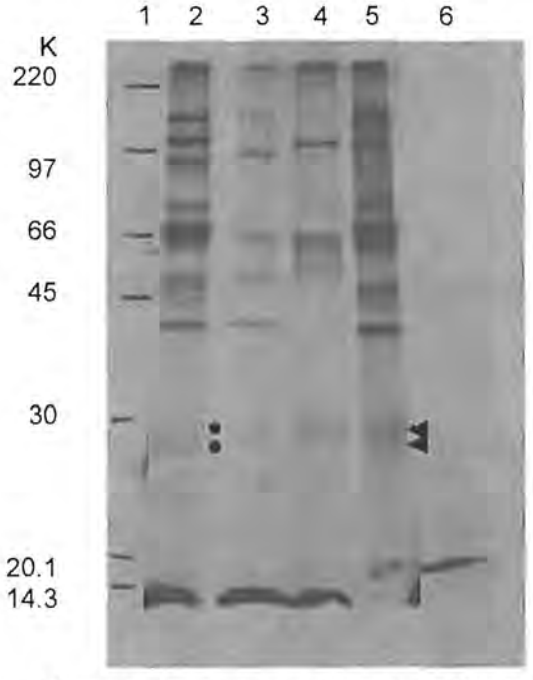


Figure 3.2. Proteins synthesized from the *in vitro* translation of dsRNA of AHSV-3 (AHSV M322/97) (3), BTV-10 (4), EHDV-1 (5) and EEV Bryanston (EEV-1) (S1REF) (6) analysed by 12% SDS-PAGE, followed by fluorography and autoradiography. A negative control sample, in which no RNA was added to the *in vitro* translation reaction (6), was also included. Molecular weight standards (1) are Rainbow protein markers (non-radioactive, positions redrawn from gel). The arrows indicate the position of NS3 and NS3A of EEV Bryanston (EEV-1) (SREF), and dots indicate NS3 and NS3A of AHSV-3 (AHSV M322/97).

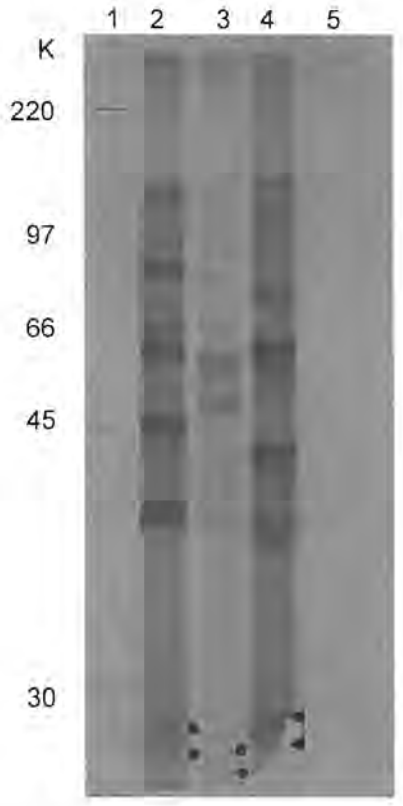


Figure 3.3. Proteins synthesized from the *in vitro* translation of dsRNA of BTV-10 (2), EHDV-1 (3) and EEV Bryanston (EEV-1) (S1REF) (4) analysed by 12% SDS-PAGE followed by fluorography and autoradiography. A negative control sample, in which no RNA was added to the *in vitro* translation reaction (5), was also included. Molecular weight standards (1) are Rainbow protein markers (non-radioactive, positions redrawn from gel). The arrows indicate the position of NS3 and NS3A of EEV Bryanston (EEV-1) (S1REF) and the dots indicate NS3 and NS3A of BTV-10 and EHDV-1.

As can be seen from the results in Fig. 3.1. and Fig. 3.2., it is not possible to distinguish NS3 and NS3A clearly. In Fig. 3.1. only one band is clearly visible, while in Fig. 3.2. the same band is fuzzy and seems to consist of two individual bands. The poor resolution can probably be attributed to the fact that small 12% gels were used that were run for a few hours at high voltage. This prevented the proper separation of the NS3 and NS3A proteins that are closely related to each other regarding their size; NS3 is estimated to be approximately 27 K, while NS3A is estimated to be approximately 25 K (refer to Table 2.6.). For this reason, the same *in vitro* translation samples used in the previous experiments (Fig. 3.1. and 3.2.) were again analysed. First, by performing 12% SDS-PAGE, allowing the gel to run for a longer time until the 20.1 K and 14.3 K markers ran off the gel (Fig.3.3.), or secondly by performing 15% SDS-PAGE overnight at a low voltage (refer to Fig.3.4.).

In Fig. 3.3., a broad fuzzy band can be seen in the region below the 30 K marker for lanes containing BTV, EHDV and EEV samples. This band seems to consist of two individual bands which are identified as NS3 and NS3A (also refer to Fig. 3.4., section 3.3.2., here two bands are again seen and their sizes estimated).

The results obtained in this section seem to indicate that EEV S10 encodes two proteins, NS3 and NS3A, as has been seen for other orbiviruses such as AHSV, BTV and EHDV. The two proteins are also produced in approximately equal amounts as can be estimated from the intensity of the bands representing the two proteins on the various autoradiographs. This result confirms the expectation created by the two in-phase initiation codons observed for EEV S10 (section 2.3.5.), namely that two proteins are encoded by S10 of EEV. The *in vitro* translation system however useful has the disadvantage that it is not a truly accurate representation of the expression of a protein or proteins in living cells. For this reason EEV S10 was also expressed using the baculovirus expression system.

3.3.2. Construction of a recombinant baculovirus for expression of EEV Bryanston (EEV-1) (S1REF*) NS3 and the expression of this protein in insect cells

The BAC-TO-BAC™ Baculovirus Expression System (GIBCO BRL Life Technologies) makes use of a baculovirus (an insect virus) which infects cultured insect cells. A rapid and efficient method has been developed by Luckow and co-workers (1993) to generate recombinant baculoviruses containing the gene of interest under the transcriptional control of the strong polyhedrin promoter of the *Autographa californica* nuclear polyhedrosis virus (AcNPV). This method is based on site-specific transposition of an expression cassette into a baculovirus shuttle vector (bacmid) which is propagated in *E. coli*.

Genome segment 10 of EEV Bryanston (EEV-1) (S1REF*) that encodes NS3, was cloned into the donor plasmid, pFASTBAC1 (section 3.2.5.1.), and the recombinant plasmid FB.EEVB.NS3 (pFASTBAC1 containing S10 of EEV Bryanston (EEV-1) (S1REF*)) was used to transform competent DH10BAC cells (sections 3.2.5.2. and 3.2.6.3). After transformation, colonies were selected on the basis of their antibiotic resistance (pFASTBAC1 confers gentamycin resistance to the cells) and colour (white colonies contain recombinant bacmid DNA that has undergone transposition). Bacmid DNA was isolated (section 3.2.5.4.) from selected colonies and was analysed by PCR (section 3.2.5.5.) to ascertain if the correct size insert had been transposed into the bacmid genome.

The PCR analysis (section 3.2.5.5.) is based on the observation that a reaction containing non-recombinant or wild type bacmid DNA generates a PCR product of approximately 723 bp in size. A reaction containing recombinant bacmid DNA should thus generate a PCR product of 723 bp plus the size of the gene of interest, in this case 759 bp, which produces a PCR product of approximately 1482 bp. PCR analysis was performed for a number of selected colonies (results not shown). In most cases, PCR products were obtained that conformed to the expected size of approximately 1482 bp and a single bacmid DNA sample was used to transfect insect cells (section 3.2.5.6.) After transfection, the putative recombinant baculovirus was harvested (section 3.2.5.7.) and stored at 4°C until use, or subjected to a round of amplification (section 3.2.7.) in order to increase the viral titre, and then stored at 4°C until use.

The next step was to test the expression of the putative recombinant baculovirus by infecting *Sf*9 cells with the virus, harvesting the virus-infected cells and performing SDS-PAGE analysis (section 3.2.7.) to determine if the EEV NS3 protein was being expressed. A number of these infection experiments (section 3.2.6.) were performed using different experimental conditions such as using different amounts of harvested putative recombinant baculovirus, using different amounts of harvested putative recombinant baculovirus after a round of amplification (section 3.2.8.), and harvesting the infected cells at different time points (24, 48, 72, and 96 hours) post infection. In each case, when comparing the cells infected with the putative recombinant baculovirus to wild type infected or mock infected cells, no unique protein band representing EEV NS3 was seen on the SDS-PAGE gels after Coomassie staining (results not shown).

The fact that no expressed NS3 protein bands were observed using Coomassie blue staining did not necessarily mean that EEV NS3 was not being expressed, but possibly that the detection technique used was not suitable or sensitive enough to detect expression; this would be the case if EEV NS3 was expressed at low levels. It was thus decided to test expression of the recombinant baculovirus using a more sensitive detection method such as labelling with radioactivity which would allow the incorporation of a radioisotope into proteins (cellular and viral) as they are synthesized in the insect cells.

Labelling of proteins with ^{35}S-methionine was performed as described in section 3.2.8. The labelling experiment was first performed using a time point of 18 hours p.i., no unique protein bands were however observed when comparing the proteins from recombinant baculovirus infected cells to the cellular proteins or the proteins generated by a wild type baculovirus infection (results not shown). The labelling experiment was repeated; the time period post infection was however increased to 24 hours, and labelling was allowed to take place for approximately 46 hours. The result of this experiment is shown in Fig. 3.4.

In Fig. 3.4. a single unique band, designated as NS3, is observed in all the samples of a recombinant baculovirus expressing EEV Bryanston (EEV-1) (S1REF*) NS3. In contrast, two bands, designated as NS3 and NS3A, are observed for the sample of an *in vitro* translation sample of an EEV Bryanston (EEV-1) field isolate (S1FLD3). This result confirms that in the baculovirus expression system, only NS3 is observed, while for *in vitro* translation reactions both NS3 and NS3A are observed. The size of the EEV Bryanston (EEV-1) (S1REF*) NS3 proteins seen on the 15% SDS-polyacrylamide gels (Fig.3.4.) were estimated as before by using the Rainbow molecular marker as a standard and plotting $\log_{10}$ of molecular weight as a function of the distance migrated by the various proteins (refer to section 3.3.1.).

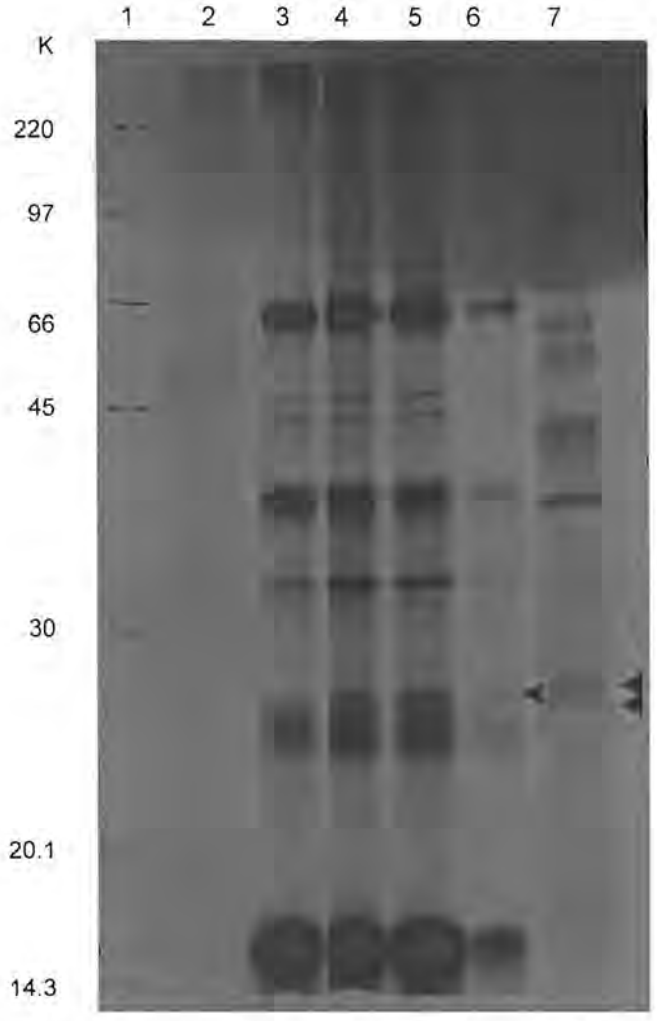


Figure 3.4. Labelling with ^{35}S-methionine of proteins synthesized in *Sf*9 cells using the baculovirus expression system. Recombinant baculovirus expressing EEV Bryanston (EEV-1) (S1REF*) NS3 (4, 5 and 6) analysed by 15% SDS-PAGE, fluorography and autoradiography. Three controls, mock infected cells (2), cells infected with wild type baculovirus (3) and an aliquot of an *in vitro* translation sample of an EEV Bryanston (EEV-1) field isolate (S1FLD3) (7), were also included. Molecular weight standards (1) are Rainbow protein markers (non-radioactive, positions redrawn from gel). The arrows indicate the position of NS3 and NS3A of EEV.

The estimated size of the single NS3 band observed in Fig.3.4. (lane 4, 5 and 6) was approximately 27 K. The sizes of the two proteins, NS3 and NS3A, seen in Fig. 3.4. (lane 7) were estimated to be approximately 28 K and 26 K respectively. The estimates for both the single NS3 band and the two bands representing NS3 and NS3A are very similar to the theoretical estimate of 27 K for NS3 and 25 K for NS3A (section 2.3.3).

Another method to test for expression of NS3 would be through the use of EEV antiserum, preferably a monospecific antiserum directed at NS3, in a Western blot procedure. An honour's project was launched in 2000 to obtain serum directed at EEV Bryanston (EEV-1) (S1REF*) NS3, but unfortunately due to technical difficulties, the project failed and no serum was obtained and made available. As will be mentioned in Chapter 4, the preparation of serum directed against EEV NS3 is

one of the priorities for future study of EEV NS3. The availability of antiserum will allow further studies of the nature of EEV NS3 expression, specifically its subcellular localization as well as variation as indicated by cross-reactivity of antiserum directed at one serotype with NS3 of other serotypes.

3.4. DISCUSSION

The aim of this part of the study was firstly to determine if both of the in-phase initiation codons identified in EEV NS3 (section 2.3.5.1.) are used when the denatured NS3 dsRNA gene is translated in an *in vitro* rabbit reticulocyte system, and secondly to construct a recombinant baculovirus that expresses NS3 of EEV. The successful expression of EEV NS3 in insect cells using the baculovirus expression system would also allow the establishment of a system to express NS3 of EEV to make future functional studies possible.

In the *in vitro* translation experiments performed using the rabbit reticulocyte lysate system, and dsRNA from EEV Bryanston (EEV-1) (S1REF) and other orbiviruses (AHSV, BTV and EHDV), ten protein bands were identified for EEV Bryanston (EEV-1) (S1REF) that were thought to represent the 7 structural proteins and the 3 non-structural proteins. The identities of the proteins were assigned using known data for other orbiviruses as a guide, since no coding assignment has yet been published for EEV; although a recent study indicates that the assignment is similar to that of AHSV and BTV (Potgieter *et al.*, 2002) (refer to section 1.4.3.). The question thus arises of which genome segment encodes which protein(s) when using all the dsRNA segments in *in vitro* translation experiments. Isolating the individual dsRNA segments before *in vitro* translation could solve this problem, however, in the case of genome segment 10, as it is the smallest genome segment, it is expected to encode the smallest protein(s). The focus on S10 in this investigation thus eliminates the need to isolate the individual dsRNA genome segments. EEV NS3/NS3A was therefore identified using both the theoretical molecular weight estimate of 27 K (section 2.3.3.) and the known molecular weights of other orbivirus NS3 proteins as guidelines (refer to Table 2.15.). In the case of EEV NS3 and NS3A, the two proteins were synthesized in approximately equal amounts, and could only be separated by means of 15% SDS-PAGE analysis. The sizes of EEV NS3 and NS3A were estimated to be approximately 28 K and 26 K respectively, which falls in the same size range as the theoretical estimate of 27 K for NS3 and 25 K for NS3A (section 2.3.3., Fig. 3.4.).

The molecular weight of the EEV Bryanston (EEV-1) (S1REF*) baculovirus expressed NS3 protein was estimated to be approximately 27 K, an estimate that again is similar to the theoretical estimate. Size differences for the same proteins can also be observed on SDS-PAGE gels in some cases. A factor to take into account when estimating molecular weight, specifically in eukaryotic expression systems such as the baculovirus system, are post-translational modifications.

Glycosylation of BTV NS3/NS3A was first detected by Wu *et al.* (1992). It has been shown that heavily glycosylated proteins can be difficult to detect, even when synthesized efficiently, as the protein can appear as a diffuse rather than a tight band on gels. Also, the glycosylation of a protein can vary, giving rise to multiple bands on a gel or to differently sized forms of the same protein under different circumstances. When NS3/NS3A proteins are glycosylated, their migration patterns on a polyacrylamide gel can therefore change giving rise to differently sized bands on a gel, which could also explain the variation seen when estimating the size of proteins. This situation is especially important when the proteins isolated were *in vivo* expressed in mammalian cells, since the glycosylation of proteins are generally more efficient in mammalian cells than in insect or yeast cells (Guirakhoo *et al.*, 1995). King and Possee (1992) has also made the observation that the glycosylation of proteins in insect cells are different from that seen in mammalian cells, this can also cause diffuse bands of varying molecular weight. The recombinant proteins in insect cells can also have a smaller apparent molecular weight, after analysis by PAGE, than their authentic counterparts in the normal host cell. In the case of EEV Bryanston (EEV-1) (S1REF*) NS3, the protein has two potential N-linked glycosylation sites (section 2.3.3.), this could potentially lead to different molecular weight estimations for NS3, especially in cases where a single diffuse band is visible on gels.

The expression of EEV NS3 in the baculovirus expression system could not be detected using Coomassie blue staining, indicating that EEV NS3 was probably expressed at low levels. Low levels of EEV NS3 expression were however not unexpected. Although large amounts of NS3 has been observed for BTV NS3 expressed in the baculovirus expression system (French *et al.*, 1989), low levels of expression has been seen for AHSV NS3 in the same expression system (Van Staden *et al.*, 1995) and the expression of AHSV NS3 was not sufficient to be detected by Coomassie blue staining (Van Niekerk *et al.*, 2001a) as was the case in this investigation for EEV NS3. EEV and AHSV have similar hosts; EEV NS3 and AHSV NS3 also share a number of structural characteristics and protein motifs (Chapter 2), it could therefore be postulated that the expression of NS3 of EEV could be more similar to that of AHSV than that of BTV and that low levels of EEV NS3 could be expected when expressed in the baculovirus expression system. In order to detect the proposed low level of expression of EEV NS3 it was decided to radiolabel the synthesized proteins with ^{35}S-methionine.

A few problems were experienced with the radiolabelling procedure (section 3.2.8.), which were most probably due to a low viral titre of the recombinant baculovirus as it is known that if the infectivity of a viral stock is too low, the insect cells will be inadequately infected and virus gene expression will be poor (King and Possee, 1992). When the time before labelling was increased from 18 hours to 24 hours and the labelling period increased from 3 hours to approximately 46 hours to compensate for the suspected low viral titre, expression of NS3 could be seen clearly, NS3A was however not observed (Fig. 3.4., lane 4, 5 and 6). This problem can however be avoided in further studies using this recombinant viral stock by performing virus titrations to determine the viral titre and by viral

amplification. It should also be noted that when using the baculovirus expression system, the foreign protein is placed under the control of the polyhedrin promoter, which is a very late promoter (active from about 18 to 72 hours p.i.). King and Possee (1992) suggest that when this promoter is used to drive expression of a foreign gene, protein synthesis should initially be examined at 24 or 48 hours p. i., since a rapid accumulation of material is observed between 24 and 30 hours p.i., followed by a more gradual increase over the final part of the infection cycle. Thereafter a time experiment can be performed to determine the point of maximum synthesis of the foreign protein. It has been found for AHSV that NS3 expression is maximal between 24 to 30 hours p.i. (Van Staden *et al.*, 1995; Van Niekerk *et al.*, 2001a). A similar finding has been made for BTV NS3. French *et al.* (1989) found that when using the baculovirus expression system, high level expression of BTV NS3 is observed in the insect cells when the cells are infected at a high MOI and harvested 24 to 48 hours p.i.

As already seen, EEV NS3 was expressed at low levels in the baculovirus expression system, while no EEV NS3A was observed. In contrast, equimolar levels of expression were observed for EEV NS3 and NS3A in *in vitro* translation experiments, and the level of NS3 seen exceeded that observed in the baculovirus system. There are a number of possible explanations for this observation. The secondary structure of ssRNA (mRNA) in *in vivo* expression systems may impede transcription/translation leading to low levels of the NS3/NS3A proteins. In the case of BTV, stable stem-loop structures have been identified in the S10 RNA sequence of BTV between the first and second AUG translation initiation codons, as well as about 40 nucleotides downstream from the second AUG triplet (Hwang *et al.*, 1992). The initiation codon for BTV NS3A might therefore be flanked by two stable stem-loop structures and this might explain the existence of disproportional amounts of NS3 and NS3A proteins found in BTV-infected cells. However, in AHSV-infected Vero cells, both NS3 and NS3A were shown to be synthesized in approximately equimolar amounts (Van Staden *et al.*, 1995) and a possible explanation for the difference in translation efficiency could again be on the level of secondary structure. AHSV S10 lacks similar structures to the two stable stem-loop structures formed on either side of the BTV NS3A initiation codon, and this could result in a similar frequency of initiation at either of the two AUGs, with equimolar amounts of NS3 and NS3A being synthesized. The use of denatured dsRNA segment 10 transcripts in *in vitro* translation experiments therefore prevents the problem of secondary structure and equal amounts of NS3 and NS3A are produced. It has however also been observed that the AHSV-3 NS3 gene 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). BTV-1 S10 also contains similar sequences (Gould, 1988). These sequences have also been proposed to impede RNA transcription (Gould *et al.*, 1988; Van Staden and Huismans, 1991) and are therefore connected with the relative paucity of the NS3 protein and its mRNA in orbivirus infected cells. Similar sequences, which could explain the low levels of EEV NS3, have however not been identified in the 3' non-coding terminus of the NS3 gene of EEV, and it begs the question of the involvement of other factors specifically relating to EEV NS3. Another

factor that could determine the levels of expression of EEV NS3 and NS3A is the context in which the two initiation codons reside (sections 2.3.5.1. and 2.4.). King and Possee (1992) concluded that the requirement for a Kozak consensus sequence in the baculovirus expression system was debatable. In relation to the AUG start codon, the normal AcMNPV polyhedrin gene has an A at position -3; however, foreign gene coding sequences lacking this arrangement have been successfully expressed to high levels, while others with an ideal Kozak sequence have been poorly expressed. Low levels of expression could however also be due to other factors within the baculovirus expression system or the nature of the expressed protein (e.g. if it is cytotoxic protein or a membrane-bound glycoprotein, lower levels of expression can be expected) (King and Possee, 1992).

Mertens *et al.* (1984) observed a similar discrepancy in the amount of BTV NS3/NS3A synthesized in *in vitro* and *in vivo* systems. When performing *in vitro* translation experiments, using the rabbit reticulocyte lysate system, they found in the majority of cases that the relative amount of each protein synthesized *in vitro*, was similar to that of the analogous protein synthesized *in vivo* (BTV infected BHK 21 cells), except in the case of proteins 8 and 8A (NS3 and NS3A). They were synthesized in approximately equal amounts and in relatively large amounts *in vitro* compared to *in vivo*. This was confirmed by French *et al.* (1989), they reported for BTV that NS3 was predominant *in vivo*, but approximately equal amounts of NS3 and NS3A were observed *in vitro*. This may be a reflection of the relatively greater efficiency of the translation of a small RNA species *in vitro*, but may also indicate that some regulatory mechanism (or characteristic of the protein itself) reduces the relative level of expression of this genome segment *in vivo*. It has also been found in various studies of orbivirus transcription and translation that S10 is transcribed at less than half its predicted frequency (Huismans *et al.*, 1979; Van Dijk and Huismans, 1988). This explains the low level of expression of NS3 in cells, although EHDV NS3/NS3A is expressed at higher amounts in virus-infected cells than BTV NS3/NS3A, with NS3 being the primary product (Jensen *et al.*, 1994).

The expression of other orbivirus NS3 and NS3A proteins has also been studied using the baculovirus expression system. BTV NS3/NS3A are produced in very small amounts in BTV infected cells with NS3 as the primary product (Van Dijk and Huismans, 1988; French *et al.*, 1989). When using the baculovirus expression system, a high level of expression of BTV NS3 is observed in the insect cells and the synthesis of NS3A is substantially less than NS3, NS3A is only observed after pulse labelling. The identity of the NS3A protein was also confirmed by using anti BTV-10 serum (Van Dijk and Huismans, 1988; French *et al.*, 1989). AHSV NS3 and NS3A are expressed at similar low levels in AHSV-infected Vero cells. AHSV NS3 is however also synthesized in small amounts in baculovirus infected cells, and co-migrates with a cellular protein when visualized by Coomassie staining, but it can easily be identified by radiolabelling or an immunoblot. AHSV NS3A was not detected in recombinant baculovirus infected cells, and the NS3 protein remained cell associated and was not released into the tissue culture medium (Van Staden *et al.*, 1995). The relatively low level of

expression of AHSV NS3 in the baculovirus system suggests a possible cytotoxic effect of NS3 and this was supported by studies of the insect cells using the vital exclusion dye trypan blue. The number of cells unstained by the dye declined rapidly after 24 hours p. i., at 48 hours p.i. only 6% of the cells were still unstained, indicating an increase in cell death. This decline in the ability of cells in which the recombinant baculovirus was being expressed to exclude trypan blue, was found to follow immediately after the period of maximal NS3 mRNA and protein synthesis thereby showing the involvement of NS3 in the cellular death due to the disruption of the plasma membrane (Van Staden *et al.*, 1995). The cytotoxic effect seen for AHSV NS3 also raises the question of the involvement of NS3 in other virus attributes such as pathogenicity and virulence. O'Hara *et al.* (1998) used reassortment of S10 to study the difference between a virulent strain of AHSV and an avirulent strain, their experimental data confirmed the role of NS3 in virus release from cells and supported the hypothesis that S10/NS3 may influence virulence phenotypes by altering the timing of virus release (Martin *et al.*, 1998). NS3 of AHSV can also significantly influence vector competence and therefore its transmission (Riegler *et al.*, 2000). The two identified hydrophobic regions within the AHSV NS3 protein have been implicated in the cytotoxicity and virulence of the NS3 protein. Studies have indicated that these domains are necessary to disrupt the plasma membrane, as well as for a stable association with the membrane (Van Staden *et al.*, 1995; Van Staden *et al.*, 1998; Van Niekerk *et al.*, 2001a). It has also been suggested that the high level of variation seen within the AHSV NS3 protein has some functional significance regarding virulence. The level of variation in the NS3 protein may also influence the efficacy of dissemination (refer to section 2.4.). In the case of EEV NS3/NS3A, NS3 seems to be the predominant gene product *in vivo*, although approximately equal amounts of EEV NS3 and NS3A are observed *in vitro*. A similar role to that of AHSV NS3 can also be proposed for EEV NS3 as it shows low levels of expression in the same expression system, EEV NS3 and AHSV NS3 also share a number of structural similarities, including the two hydrophobic domains. This proposed hypothesis will however need to be tested experimentally before further conclusions about the role of EEV NS3 can be drawn.

In order to study the expression of EEV NS3, it will be necessary to have some means of expressing the NS3 protein and a way to specifically detect NS3/NS3A. The first has been achieved by constructing a recombinant baculovirus containing S10 of EEV Bryanston (EEV-1) (S1REF*); the second objective can be achieved by preparing monospecific antiserum directed against NS3. This will allow a number of different studies to be performed in the future as has been done for other orbivirus NS3 proteins. In the case of BTV the NS3 and NS3A proteins are synthesized late in the infection cycle, and then only in small amounts (Van Dijk and Huismans, 1988) and elucidation of the possible function of NS3 came from immunoelectron microscopic and immunofluorescence studies on BTV infected cells and NS3 expressed in mammalian and insect cells, using antibodies raised against NS3. From these studies it was seen that NS3 and NS3A are associated with intracellular smooth surfaced vesicles and are located on the surface of infected cells in association with BTV particles

which are in process of extrusion (Hyatt *et al.*, 1993). Monospecific polyclonal antibodies have also been raised against AHSV-3 NS3 (Van Staden *et al.*, 1995) in order to study the subcellular localization of AHSV NS3.

Antibodies directed against EEV NS3 can be utilized in a number of ways. Western blotting will allow the detection of denatured forms of the protein (as an alternative to labelling with ^{35}S-methionine and to determine beyond any doubt the identity of a novel band or bands on a gel as NS3/NS3A), immunoprecipitation will recognize native forms of the protein, while immunofluorescence and electron microscopy studies will identify the subcellular localization of NS3 in mammalian and insect cells. It is therefore abundantly clear that the study of EEV NS3 is in its infancy, many questions still have to be answered in order for our knowledge of EEV NS3 to be on par with that of other orbivirus NS3 proteins such as AHSV and BTV.

CHAPTER 4

CONCLUDING REMARKS

The long-term aim regarding all studies concerning NS3 is to elucidate the structure/function relationship of NS3 and specifically its role in viral infection and morphogenesis, with emphasis on the mechanism responsible for the release of viral particles from infected cells, as well as the possible role of NS3 in virulence and disease. Within that context, the focus of this investigation was on genome segment 10 of EEV and the resulting gene product(s) with the aim to characterize the EEV genome segment 10 and gene product(s) and to compare them to their orbivirus counterparts in an attempt to identify conserved and/or unique characteristics. In order to achieve this aim, a number of short term aims relating to EEV NS3 were investigated and the details of the results that were obtained in the course of achieving these objectives were discussed in the individual chapters. In this conclusion, the main results which could contribute to a better understanding of the molecular nature of the NS3 protein of EEV will be summarized and some suggestions regarding future research in this field will be made.

A number of conserved structural features or characteristics have been identified for EEV NS3 within the EEV serogroup. These features include two putative initiation codons, two hydrophobic domains with the potential to form transmembrane helices, a coiled-coil domain, a proline-rich region, glycosylation sites, myristylation sites, a conserved region, a variable region and conserved cysteine residues. Similar features or characteristics have been found in other orbivirus NS3 proteins such as AHSV NS3, BTV NS3 and EHDV NS3.

The hydrophobic regions suggest that the EEV NS3 protein is a membrane protein and as such the membrane topology of the protein was investigated. The membrane topology of EEV NS3 was shown to be similar to both the AHSV model and the EHDV model, not excluding either one of the models. The identified coiled-coil domain and the conserved cysteine residues could play a role in dimerization as was proposed for EHDV NS3 where NS3 protein dimers form channels or viroporins. These viroporins play a role in viral release and cell damage by disrupting the cellular membrane by changing membrane permeability and interfering with cellular ionic gradients, most likely intracellular calcium levels. EEV NS3 also contains a putative bipartite membrane targeting signal, a signal peptide and a potential cleavage site for the signal peptide. The bipartite signal consists of a conserved myristylation site and a stretch of positively charged amino acids and is found within the identified conserved region. Using the information available, 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.

The predicted membrane topology of EEV NS3 also correlated with the identified variable region. This region is located between the two hydrophobic domains and as such is predicted to be exposed extracellularly. This region is also predicted to possess antigenic properties, which could play a role in the virus-host immunological interaction. The identified proline-rich region is found in the N-terminal portion of the protein that is predicted to be located in the cytoplasm, this region could function by binding other viral proteins in order to perform a function as a receptor in analogy with rotavirus NSP4.

The variation found in the NS3 protein is not confined to the variable region and has certain useful applications, including epidemiological studies where the outbreak of disease within a region is traced to the source. The maximum variation observed within the EEV serogroup between serotypes is 25.2% on a nucleotide level and 16.7% on an amino acid level. These values indicate that EEV NS3 is more variable than BTV NS3, but less variable than AHSV NS3. This level of variability within EEV NS3 was also observed between the two subclusters (A and B) in the phylogenetic trees and indicated that EEV NS3 could be grouped into two groups according to firstly, serotype and secondly, geographic location. A similar observation has been made for BTV where 3 groups (2 USA and 1 Asian) were identified based only on geographic origin. This implies that EEV NS3 has topotyping characteristics and that based on S10 or NS3 sequences the possible origins of a viral isolate can be predicted. As already discussed, there also seems to be a link between vector distribution and the distribution of the EEV serotypes within the two regions as indicated by the grouping of EEV in the two subclusters (A and B). A question now arises regarding the evolution of the EEV serogroup. Is there evidence for a co-speciating event where a specific vector is better adapted to disseminate a specific EEV serotype in a specific region? These types of questions, as well as the specific grouping pattern within the EEV serogroup will have to be investigated by sequencing more different isolates and more recent field isolates. The phylogeny of both S10 and NS3 also indicated that the EEV serogroup was a distinct lineage separate from the other orbivirus groups. In this investigation it was also found that a single set of primers based on EEV Kyalami (EEV-5) S10 could amplify all seven known different serotypes of EEV S10 by using RT-PCR, but no amplification was observed for EEV S10 when using primers directed at AHSV S10. This provides another way, by using RT-PCR, in addition to the agarose dsRNA profiles, to distinguish between AHSV and EEV during infection.

The expression of EEV NS3/NS3A was investigated using both *in vitro* and *in vivo* systems. *In vitro* translation studies were performed using EEV dsRNA and two proteins, NS3 and NS3A, were observed that were expressed in approximately equimolar amounts. In the baculovirus expression system, a recombinant baculovirus expressing EEV S10 was constructed, and expression of one protein, NS3, was observed at low levels and only after labelling with ^{35}S-methionine. The low levels of expression in the *in vivo* system prohibit the purification of the protein in its native form, and can possibly be explained by the context in which the initiation codons are found. A comparison with the

Kozak consensus sequence of initiation indicated that the two putative initiation codons are both found in suboptimal contexts for the initiation of translation. In addition, in *in vivo* systems RNA secondary structure can play a role in impeding translation by forming hairpin loops in the vicinity of the initiation codons.

The identified conserved structural features, the variation within NS3 of the EEV serogroup and the expression of EEV NS3, all hold clues as to the role EEV NS3 will play in the viral lifecycle. Some of these potential roles have been mentioned in the preceding chapters and in the preceding paragraphs. The work done and described in this investigation will therefore form the basis of future studies of EEV NS3.

The relationship between structure and function has to be proven experimentally and in this regard there are a large number of experiments that will have to be conducted in future research projects in order to elucidate the relationship. It is important to link the features that are conserved amongst the NS3 proteins of different orbiviruses to specific functions. The study of the functional significance/importance of these domains will be important not only for the EEV serogroup but also for the *Orbivirus* genus as a whole.

Future research should involve investigations relating to the subcellular localization of EEV NS3, cytotoxicity studies which could provide a link to virulence, and further sequencing of EEV S10/NS3, including more recent field isolates, to confirm the levels of variation observed as well as the grouping within the EEV serogroup.

In order to continue with these future projects a number of molecular tools are required. Firstly, monospecific antiserum directed against EEV Bryanston (EEV-1) (S1REF*) NS3 is needed. The use of antiserum will allow the detection of expression of EEV NS3/NS3A in *in vitro* and *in vivo* systems without the use of radiolabelled amino acids by Western blotting, as well as the identification of its subcellular localization in insect and mammalian cells in immunofluorescence studies as has been done for AHSV NS3 (Stoltz *et al.*, 1996). The identification of the location of a viral protein in an infected cell can give an indication of the function of the protein and the association of a particular protein with specific cellular and viral structures or events may provide insight in its role in virus morphogenesis.

The second tool that is needed is a way to express sufficient quantities of the protein under investigation. As already mentioned, the low levels of expression of EEV NS3/NS3A in the baculovirus system prohibits the purification of sufficient quantities of these proteins for functional studies. The low levels of expression obtained in the baculovirus expression system when using *Sf*9 cells can however be increased by using an alternative cell line derived from *Mamestra brassicae*

(Mb) as it appears to provide an increase in both polyhedrin and foreign protein levels that are about two- to three-fold higher than are given with *Sf*9 cells. Unfortunately, these cells are not as easy to use as the *Sf* cell lines. The production of recombinant protein in insect cells can also be scaled up (for methods refer to King and Possee, 1992); the nature of EEV NS3 (predicted to be a membrane glycoprotein and possibly cytotoxic) could however limit the amount of protein that can be produced. Therefore, the use of a bacterial expression system will be necessary to produce a sufficient amount of the protein for the preparation of monospecific antiserum or to experimentally study and confirm the predicted structural characteristics. An interesting study to perform, for which sufficient amounts of protein are needed, would be peptide mapping to experimentally identify the second initiation codon and to confirm the NS3/NS3A relationship for EEV suggested by the amino acid sequences.

The first future research project mentioned involves cytotoxicity. As already discussed, low levels of expression of a specific protein can be linked to cytotoxicity, this has been proven in the case of AHSV NS3 (Van Staden *et al.*, 1995; Van Staden *et al.*, 1998; Van Niekerk *et al.*, 2001a). The low levels of expression of AHSV NS3 in the baculovirus system were linked to cytotoxicity and the regions involved in the cytotoxic function of AHSV NS3 were also identified namely the two hydrophobic regions. As EEV NS3 exhibits a similar low level of expression in the same system, and also contains similar structural features such as the hydrophobic regions, cytotoxicity studies should be performed for EEV NS3. The information gathered in this investigation predicts that EEV NS3 could possibly show a similar cytotoxic effect as that observed for AHSV NS3. When cytotoxicity has been shown for EEV NS3 it will also be necessary to identify the regions involved in this function. The availability of mutant forms of the EEV NS3 protein would provide the opportunity to analyse the relationship between the structure and function of the protein by enabling identification of domains or individual amino acid residues that are essential for the structural integrity and/or function of the protein. These mutant forms of the EEV NS3 protein can be obtained by using site directed mutagenesis to construct recombinant baculoviruses expressing the mutant proteins as has been done for AHSV NS3 (Van Staden *et al.*, 1998; Van Niekerk *et al.*, 2001a).

A number of structural features or protein motifs were predicted for EEV NS3, e.g. glycosylation sites and myristylation sites. These sites involve post-translational modification of the EEV NS3 protein. The analysis of these post-translational processing events can be performed in insect cells and a number of methods to do this have been described (King and Possee, 1992).

In mammalian and insect cells, treating the cells with tunicamycin may inhibit N-linked glycosylation; this will allow the researcher to determine if EEV NS3 is glycosylated. Comparison of mammalian cell- and insect cell-derived glycoproteins, however, has revealed several differences in the nature of the added oligosaccharide side chains. However, where assays of biological activity have been carried out, glycoproteins synthesized in insect cells have been shown to have similar, if not identical,

activities to their authentic counterparts synthesized in mammalian cells and in most cases the glycoproteins synthesized were antigenic, giving rise to high-titre antibody preparations after injection into animals. Glycosylation in insect cells, however, might be slightly less efficient than in mammalian cells. There are thus differences in glycosylation between insect and other cell types. Whether this has a bearing on the biological activity or antigenicity of any given glycoprotein will probably have to be tested empirically for each example. This however implies that any analysis of glycosylation should be repeated in mammalian cells. Recombinant glycoproteins synthesized in insect cells may be readily detected by radiolabelling with [^{3}H] or [^{14}C] mannose or by binding to mannose-specific lectins such as concanavalin A. In addition, the baculovirus expression system can also be used to analyse phosphorylation, palmitylation or myristylation. Transcription can also be analysed, i.e.; the amount of RNA produced by the recombinant virus, if yields of the foreign protein as assessed by protein gels are low. This type of analysis should provide insight in whether the low levels of expression are due to the RNA secondary structure (transcriptional level) or the nature of the protein (translational level).

This investigation has provided important information regarding the nature of genome segment 10 and its encoded gene product NS3 of EEV. The molecular characterization of equine encephalosis virus non-structural protein NS3 is by no means complete, a solid foundation has however been laid for future research. This investigation provides structural information, as well as information about the expression of EEV NS3, that will form the basis of future functional studies of EEV NS3.

Parts of these results are included in the following paper:

Van Niekerk, M., Freeman, M., Paweska, J. T., Howell, P. G., Guthrie, A. J., Potgieter, A. C., Van Staden, V. and Huismans, H. (2003). Variation in the NS3 gene and protein in South African isolates of bluetongue and equine encephalosis viruses. *Journal of General Virology* **84**(3), 581-590.

Parts of these results have been submitted to GenBank as nucleotide sequence entries and obtained the following accession numbers:

EEV isolate	Accession number
S1REF*	AY115864
S1REF	AY115865
S2REF	AY115871
S3REF	AY115867
S4REF	AY115868
S5REF	AY115869
S6REF	AY115866
S7REF	AY115870
S1FLD1	AY115878
S1FLD2	AY115876
S1FLD3	AY115875
S1FLD4	AY115877
S3FLD1	AY115874
S6FLD1	AY115872
S6FLD2	AY115873

APPENDIX A

IUB/IUPAC CODES, ONE- AND THREE-LETTER CODES FOR THE COMMON AMINO ACIDS, PREDOMINANT STATE OF AMINO ACID SIDE CHAINS AT pH 7 AND PLASMID MAPS OF THE TWO VECTORS USED

Table A1. IUB/IUPAC codes

Code[a]	Bases	Derivation	Complements
A, C ,G ,T	A, C ,G ,T		T, G, C, A
R	A or G	puRine	Y
Y	T or C	pYrimidine	R
W	A or T	Weak[b]	W
S	C or G	Strong[b]	S
M	A or C	aMino	K
K	G or T	Keto	M
B	C, G or T	not A	V
D	A, G or T	not C	H
H	A, C or T	not G	D
V	A, C or G	not T	B
N	A, G, C or T	aNy	N

[a] These single-letter codes are used to represent mixed or ambiguous bases. They are derived from shared features of the bases (refer to the third column).

[b] Strong and weak refer to the strength of the hydrogen bonds between these pairs.

Table A2. One- and three-letter codes for the common amino acids

Amino acid	Abbreviation		Amino acid	Abbreviation	
	3-letter	1-letter		3-letter	1-letter
Alanine	Ala	A	Leucine	Leu	L
Arginine	Arg	R	Lysine	Lys	K
Asparagine	Asn	N	Methionine	Met	M
Aspartic acid	Asp	D	Phenylalanine	Phe	F
Cysteine	Cys	C	Proline	Pro	P
Glutamine	Gln	Q	Serine	Ser	S
Glutamic acid	Glu	E	Threonine	Thr	T
Glycine	Gly	G	Tryptophan	Trp	W
Histidine	His	H	Tyrosine	Tyr	Y
Isoleucine	Ile	I	Valine	Val	V

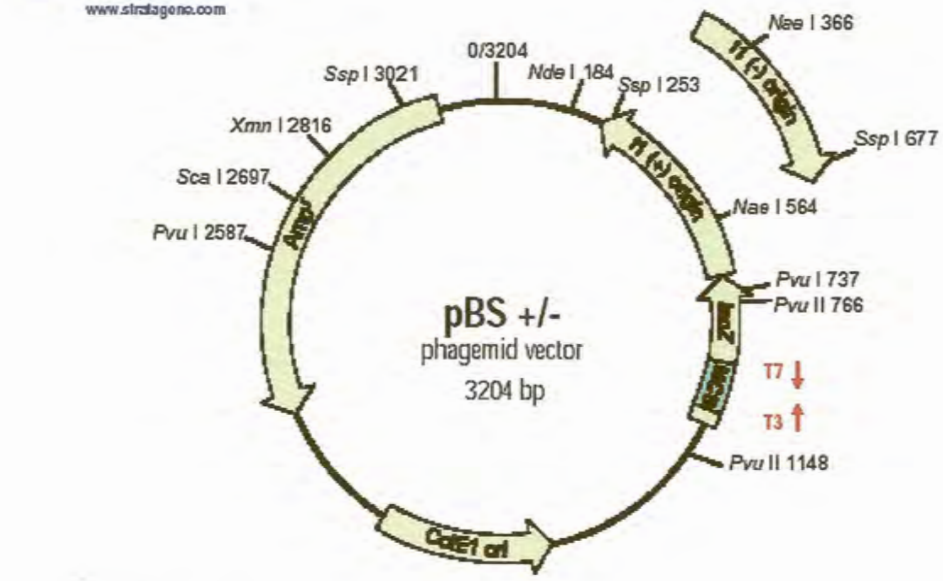
Table A3. The predominant state of amino acid side chains at pH 7 (adapted from Campbell, 1995)

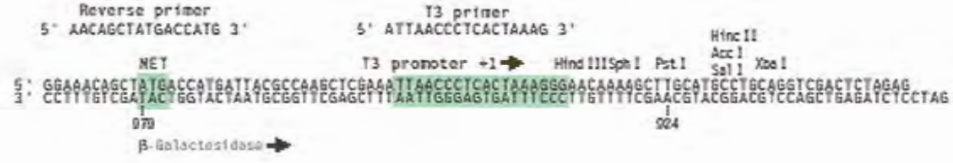
One-letter code	Amino acid	Predominant state at pH 7
A	Alanine	non-polar side chain
C	Cysteine	polar, uncharged side chain
D	Aspartic Acid	acidic side chain
E	Glutamic Acid	acidic side chain
F	Phenylalanine	non-polar side chain
G	Glycine	polar, uncharged side chain
H	Histidine	basic side chain
I	Isoleucine	non-polar side chain
K	Lysine	basic side chain
L	Leucine	non-polar side chain
M	Methionine	non-polar side chain
N	Asparagine	polar, uncharged side chain
P	Proline	non-polar side chain
Q	Glutamine	polar, uncharged side chain
R	Arginine	basic side chain
S	Serine	polar, uncharged side chain
T	Threonine	polar, uncharged side chain
V	Valine	non-polar side chain
W	Tryptophan	non-polar side chain
Y	Tyrosine	polar, uncharged side chain

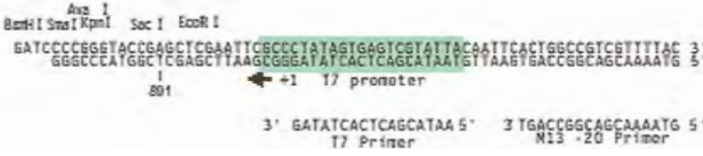
Diagram Courtesy of STRATAGENE

11011 North Torrey Pines Road
La Jolla, CA 92037
www.stratagene.com

Corporate Headquarters (619) 536-5400
Fax (619) 535-0071
Toll-Free Order & Technical Services (800) 424-5444
Toll Free Fax (Canada) (888) 267-4010





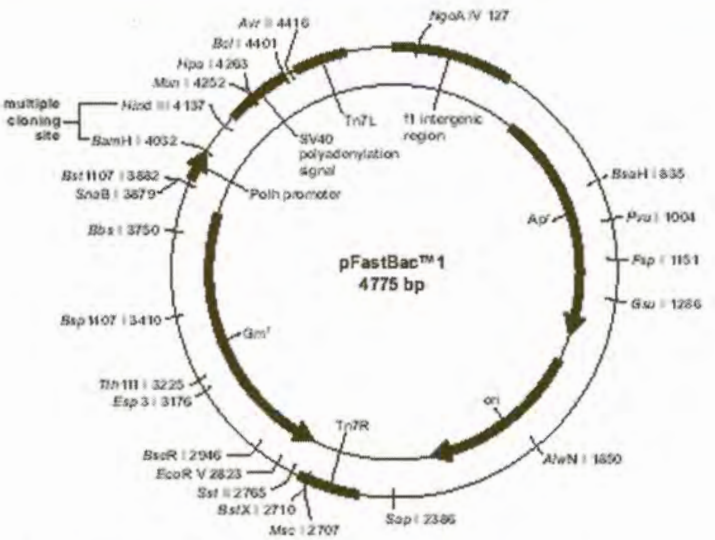


3' GATATCACTCAGCATAA 5'
T7 Primer

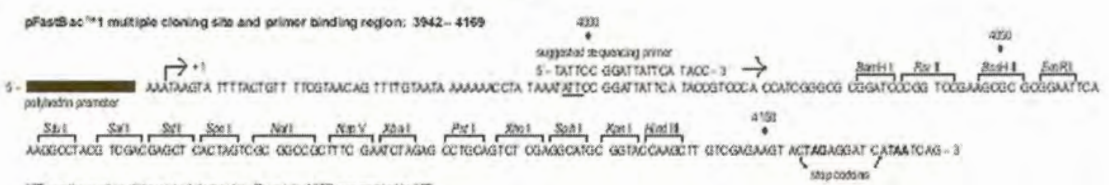
3 TGACCGGCAGCAAAATG 5'
N13 -20 Primer

Appendix A:
Maps and Restriction Endonuclease Sites for pFastBac™ Expression Vectors

8



pFastBac™1 multiple cloning site and primer binding region: 3942–4169



Restriction endonucleases that do not cleave pFastBac™1 DNA:

Aat II	*Cla* I	*Nco* I	*Pme* I	*Sse* I
Afl II	*Cvn* I	*Nde* I	*Psp5* II	*Sse8387* I
Apa I	*Eco47* III	*Nhe* I	*Pvu* II	*Sun* I
Asc I	*Eco72* I	*Nru* I	*SexA* I	*Swa* I
Bpu1102 I	*EcoN* I	*Nsi* I	*Sfi* I	*Xcm* I
Bsg I	*EcoO109* I	*Pac* I	*SgrA* I	
BspM I	*Mfu* I	*PflM* I	*Sma* I	
BstE II	*Nar* I	*PinA* I	*Srf* I	

Restriction endonucleases that cleave pFastBac™1 DNA twice:

Acc I	3882	4070	*Bsm* I	4247	4346	*Rca* I	536	1544
Afl III	2264	3246	*BsmF* I	4015	4775	*Sca* I	893	4150
Ban II	157	4076	*Dra* III	230	3578	*Tfi* I	2290	4101
Bgl II	2547	3017	*Eam1105* I	1371	4651	*Xmn* I	772	3797
Bsa I	1304	3661	*Kpn2* I	3888	3998			

Figure 8. Map and restriction endonuclease sites for pFastBac™1. Restriction endonucleases that cleave pFastBac™1 once are shown on the outer circle. The nucleotide position refers to the 5' base of the recognition sequence.

The sequence has not been confirmed by sequence analysis. It was assembled from the known sequence of fragments used to construct the vector. Vector sequences, restriction information, and maps can be found in the Vector Data area of our web site, www.invitrogen.com.

31

APPENDIX B

LIST OF BIOINFORMATICS TOOLS AND RESOURCES USED IN THIS STUDY

Table B1. List of bioinformatics tools and resources used in this study, compiled and adapted from information contained in Alphey, 1997

Name	Description	Where available	Reference (where applicable)
*AnTheProt (ANalyse THE PROTeins) version 5.0	A suite of programs that allows the user to perform different analyses such as secondary structure prediction, protein profile analyses, detection of known protein domains or patterns, similarity searches, detection of potential cleavage site of a signal peptide and determination of titration curves using the same interface. (The main idea was to integrate into a single package most of the methods designed for protein sequence analysis).	http://pbil.ibcp.fr/ANTHEPROT	Deléage *et al.*, 1988; Deléage *et al.*, 1989; Geourjon *et al.*, 1991; Geourjon and Deléage, 1995
#BLAST (Basic Local Alignment Search Tool) (BLASTP version 2.2.1.)	A program for comparing an unknown sequence against a sequence database. Uses an approach based on matching short sequence fragments, and a powerful statistical model to find the best local alignments between the unknown sequence and the database. Will only match continuous sequence – an alignment with insertions and deletions will be displayed as a number of separate fragment matches.	http://www.ncbi.nlm.nih.gov/BLAST/	Altschul *et al.*, 1990; Altschul *et al.*, 1997
#BLASTP + BEAUTY search	A combination search consisting of a BLAST search (BLASTP 2.0.4 Feb 24 1998) of the National Center for Biotechnology Information's NR Protein Database and BEAUTY post-processing provided by the Human Genome Sequencing Center, Baylor College of Medicine.		BEAUTY: Worley *et al.*, 1995 BLASTP: Altschul *et al.*, 1997

Name	Description	Where available	Reference (where applicable)
*CBS (Center for Biological Sequence Analysis)	A prediction server to analyse nucleotide and amino acid sequences, contains links to SignalP, TargetP and TMHMM.	http://www.cbs.dtu.dk/services/	
%CDD (Conserved Domain Database) (version 1.54) at NCBI	A database of conserved domains present in protein sequences. Contains domains derived from two collections namely Smart and Pfam.	http://www.ncbi.nlm.nih.gov/Structure/cdd	
ClustalX1.81	A program used to perform multiple sequence alignments of nucleotide and amino acid sequences. Details of algorithms, implementation and useful tips on usage of Clustal programs can be found in the mentioned references.	ftp://ftp-igbmc.u-strasbg.fr/pub/ClustalX/ or ftp.embl-heidelberg.de or ftp.ebi.ac.uk or http://www-igbmc.u-strasbg.fr/BioInfo/	Higgins and Sharp, 1988; Higgins and Sharp, 1989; Higgins *et al.*, 1996; Thompson *et al.*, 1997; Jeanmougin *et al.*, 1998
COILS version 1999_2.2	A program to predict regions of coiled-coil arrangements of α-helices in proteins. It compares a sequence to a database of known parallel two-stranded coiled-coils and derives a similarity score. By comparing this score to the distribution of scores in globular and coiled-coil proteins, the program then calculates the probability that the sequence will adopt a coiled-coil conformation.	http://www.ch.embnet.org/cgi-bin/COILS	Lupas *et al.*, 1991; Lupas, 1996
%DART (Domain Architecture Retrieval Tool)	A program at NCBI used to search for proteins with similar domain architectures than your query protein.		

Name	Description	Where available	Reference (where applicable)
*$DAS	A server to predict transmembrane segments.	http://www.biokemi.su.se/~server/DAS/	Cserzo *et al.*, 1997
*@EASY	A program that simultaneously searches a number of databases (BLOCKS, Pfam, PRINTS-BLOCKS, PRINTS, PROFILES, PROSITE) using BLASTP and FASTA3.	http://proline.sbc.man.ac.uk/cgibin/dbbrowser/easy	
*ExPASy (Expert Protein Analysis System)	A molecular biology server dedicated to the analysis of protein sequences and structures. Includes links to databases, tools and software packages, education and services, documentation, lists of molecular biology resources, major molecular biology servers, mirror sites, miscellaneous, announcements and new features, and local links. Proteomics tools listed include tools for protein identification and characterization, DNA translation to protein, similarity searches, pattern and profile searches, post-translational modification prediction, primary structure analysis, secondary structure prediction, tertiary structure prediction, transmembrane regions detection, and sequence alignment.	http://www.expasy.ch/	
#FASTA	A program for comparing an unknown sequence against a sequence database. The program makes approximations to try and concentrate only on alignments that are likely to be significant. Sometimes misses weak but significant scores. Attempts to find global alignments.	http://www.ebi.ac.uk/htbin/fasta.py?request	Lipman and Pearson, 1985
#GenBank	A DNA database based at the National Center for Biotechnology Information (NCBI), a division of the National Library of Medicine, located on the National Institutes for Health (NIH) campus in the US.	www.ncbi.nlm.nih.gov/web/search/index.html	

Name	Description	Where available	Reference (where applicable)
#GeneFIND	A gene family identification system which aims at high-throughput full-scale gene family identification by taking advantages of the strengths of various search methods using multi-level filters, and incorporating ProClass family information.	http://pir.georgetown.edu/gf-cgi/genefind.pl	Wu *et al.*, 1996; Wu *et al.*, 1998
#MaxHom version 1.99.04	A multiple sequence alignment program which finds similar sequences in a database.		Sander and Schneider, 1991
*MEGA (Molecular Evolutionary Genetic Analysis) version 2.1	A program that carries out parsimony, distance matrix and likelihood methods for molecular data. It can also be used calculate basic sequence statistics.	MEGA2 is available as Windows95/98/NT/2000 executables, and MEGA is available as a DOS executable, at http://www.megasoftware.net. The MEGA manual is also available on-line in HTML at http://evolgen.biol.metrou.ac.jp/MEGA/manual/default.html.	Kumar *et al.*, 2001
*META PredictProtein	A service for sequence analysis and prediction. It includes database searches such as prediction-based threading, also incorporating purely sequence-based database searches (FRSVR) and hidden Markov model method (SAM-T98) for finding remote homologs of protein sequences. It also includes predictions of presence and location of signal peptide cleavage sites in amino acid sequences from different organisms (SignalP), location and orientation of transmembrane helices (TMHMM, TOPPRED), and location of transmembrane helices (DAS).	http://www.embl-heidelberg.de/predictprotein	

Name	Description	Where available	Reference (where applicable)
*PHD (Profile fed neural network systems from HeiDelberg) version 1.96	A suite of programs consisting of three different programs that predicts one-dimensional structure from multiple sequence alignments. The three programs are PHDsec that predicts secondary structure, PHDacc that predicts solvent accessibility and PHDhtm for prediction of helical transmembrane regions.		Rost, 1996a
$PHDhtm version 1.96	A program that predicts the location and topology of transmembrane helices from multiple sequence alignments.	www.emblheidelberg.de/Services/sander/predictprotein/predictprotein.html	Rost *et al.*,1996
PHDsec	A program that predicts the secondary structure of a protein from multiple sequence alignments.	http://www.embl-heidelberg .de/predictprotein/predictprotein.html	Rost and Sander, 1993; Rost and Sander, 1994; Rost *et al.*,1994
#PIR (Protein Information Resource)	A community resource that provides protein databases and analysis tools to support research on molecular evolution, functional genomics and computational biology. Includes PIR-PSD (PIR-International Protein Sequence Database) which is a protein sequence database in which entries are annotated and classified into family groups. This database is used in a comparison of an amino acid sequence to the other amino acid sequences in the database using a BLAST search.	http://pir.georgetown.edu/ or http://pir.georgetown.edu/cgi-bin/pirwww/blast.pl	Barker *et al.*, 2001
*PredictProtein version 1.99.08	A service for sequence analysis and structure prediction. It includes database searches such as PROSITE to detect functional motifs, ProDom to detect protein domains, MaxHom to generate multiple sequence alignments and Topits for fold recognition by prediction-based threading. It also includes predictions of	http://maple.bioc.columbia.edu/pp/ or http://www.embl-heidelberg.de/predictprotein	Rost, 1996b

Name	Description	Where available	Reference (where applicable)
	secondary structure (PHDsec, PROFsec), residue solvent accessibility (PHDacc, PROFacc), transmembrane helix location and topology (PHDhtm, PHDtopology), coiled-coil regions (COILS), and cysteine bonds (CYSPRED).		
@PRINTS	A profile database to identify motifs in a protein sequence.	http://www.biochem.ucl.ac.uk/bsm/ dbbrowser/ PRINTS/PRINTS.html	Attwood *et al.*, 1994
%ProDom release 2001.2 Aug 2001	A profile database to identify domains in a protein sequence.	http://protein.toulouse.inra.fr/prodom/ prodom.html or http://www.toulouse.inra.fr/prodom.html	Sonnhammer and Kahn, 1994; Corpet *et al.*, 1998; Corpet *et al.*, 2000
@PROSITE (release 16.51 of 25 Nov 2001)	A database that consists of biologically significant patterns and profiles formulated in such a way that with appropriate computational tools it can help to determine to which known family of protein (if any) a new sequence belongs, or which known domain(s) it contains.	http://www.expasy.ch/sprot/prosite.html	Bairoch *et al.*, 1997; Hofmann *et al.*, 1999
*ProtParam tool	A tool that allows the computation of various physical and chemical parameters for a given protein sequence. The computed parameters include the molecular weight, theoretical pI, amino acid composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY).	http://ca.expasy.org/cgi-bun/protparam	

Name	Description	Where available	Reference (where applicable)
PSORT (Prediction of Protein Localization Sites) version 6.4	A program that analyses prokaryotic or eukaryotic sequences and searches for protein sorting signals. The program reports back a probability for the protein being localized to different compartments within the cell.	http://psort.nibb.ac.jp/form.html	
#SAM-T99	A program consisting of three functional sections covering alignment, database search results and secondary structure prediction. The database that it searches is the PDB database.	http://www.cse.ucsc.edu/~farmer/target99-query	
*SAPS (Statistical Analysis of Protein Sequences) version of April 11, 1996	A program that uses statistical criteria to evaluate a wide range of protein sequence properties.	http://www.isrec.isb-sib.ch/cgi-bin/SAPS	Brendel *et al.*, 1992
SignalP version 1.1	A program that predicts the presence and location of leader sequences and cleavage sites within both prokaryotes and eukaryotes.	http://www.cbs.dtu.dk/services/SignalP	Nielson *et al.*, 1997
&SSpro version 2.0	A server for protein secondary structure prediction based on an ensemble of 11 bi-directional recurrent neural networks.	http://promoter.ics.uci.edu/BRNN-PRED/	Baldi *et al.*, 1999
&SSpro8 version 2.0	An experimental extension to SSpro using the full DSSP 8 classes instead of three classes to classify the secondary structure of a protein.	http://promoter.ics.uci.edu/BRNN-PRED/	Baldi *et al.*, 1999

Name	Description	Where available	Reference (where applicable)
#SWISS-PROT Release 39.0 (5/00)	A protein database consisting of properly checked and annotated translations of the EMBL database. It distinguishes itself from other protein sequence databases by three distinct criteria namely annotations, minimal redundancy and integration with other databases.	http://www.expasy.ch/sprot/sprot/ or http://www.ebi.ac.uk/swissprot/	Bairoch and Apweiler, 2000
TargetP (version 1.0)	A program that predicts the subcellular localization of proteins, specifically whether the protein is localized within the mitochondria, the chloroplast, is part of the secretory pathway or other which does not include the previously mentioned locations.	http://www.cbs.dtu.dk/services/	Emanuelsson *et al.*, 2000
$TMbase Release 25	A database of naturally occurring TM helices based on SWISS-PROT release 25.		Hofmann and Stoffel, 1993
$TMHMM version 2.0.	A program to predict the location and topology of transmembrane helices.	http://www.cbs.dtu.dk/services/TMHMM-1.0/ or http://www.cbs.dtu.dk/	Sonnhammer *et al.*, 1998
$TMPRED	A program that predicts the location and topology of transmembrane helices.	http://www.ch.embnet.org/software/TMPRED_form.html or http://www.ch.embnet.org/cgi-bin/TMPRED_form_parser	
$TopPred2	A program to predict the location and orientation of transmembrane helices.	http://bioweb.pasteur.fr/seqanal/tmp/toppred or http://www.biokemi.su.se/~server/toppred2/	Von Heijne, 1992; Claros and Von Heijne, 1994; Cserzo *et al.*, 1997

Molecular Biology Servers, services or suites of programs which are available on the internet that can serve as a platform from which links are available to most of the regularly used bioinformatics tools or software on the world wide web are indicated with an asterisk (*), tools used in similarity searches are indicated by #, tools used to identify transmembrane regions or helices are indicated by $, tools used to identify protein domains or profiles are indicated by %, tools used to identify protein motifs are indicated by @ and tools used for secondary structure prediction are indicated by &.

APPENDIX C

LIST OF NUCLEOTIDE AND AMINO ACID SEQUENCES OBTAINED FROM GENBANK USED IN THIS STUDY FOR PHYLOGENETIC ANALYSES

Table C1. Relevant information regarding all orbivirus nucleotide and amino acid sequences obtained from GenBank used in this study

Serogroup and serotype	Accession number	Origin of virus and year of isolation	Number of basepairs	Reference
AHSV-1	U02711	South Africa (UN)	764 bp	De Sá *et al.*, 1994
*AHSV-2	U59279	South Africa (UN)	764 bp	Sailleau *et al.*, 1997
AHSV-3	AJ007303	South Africa (UN)	758 bp	Martin *et al.*, 1998
*AHSV-4	AJ007305	South Africa (UN)	758 bp	Martin *et al.*, 1998
AHSV-5	AJ007309	South Africa (UN)	758 bp	Martin *et al.*, 1998
AHSV-6	U60189	South Africa (UN)	755 bp	Sailleau *et al.*, 1997
*AHSV-7	AJ007306	South Africa (UN)	758 bp	Martin *et al.*, 1998
AHSV-8	AJ007307	South Africa (UN)	764 bp	Martin *et al.*, 1998
AHSV-9	AJ007308	South Africa (UN)	756 bp	Martin *et al.*, 1998
*Broadhaven NS3	M83197	Scotland, St. Abb's Head (UN)	702 bp	Moss *et al.*, 1992
BTV-1	AF135223	China, Shizong (1986)	785 bp	Bonneau *et al.*, 1999
*BTV-2	AF135224	Jinghong, China (1996)	785 bp	Bonneau *et al.*, 1999
BTV-3	AF135225	Ershang, China (1996)	785 bp	Bonneau *et al.*, 1999
BTV-4	AF135226	Dehong, China (1996)	785 bp	Bonneau *et al.*, 1999
BTV-10	AF044372	United States, California (1953)	822 bp	Pierce *et al.*, 1998
BTV-11	AF044373	United States, Texas (1962)	822 bp	Pierce *et al.*, 1998
*BTV-12	AF135227	China, Jinghong (1996)	785 bp	Bonneau *et al.*, 1999
BTV-13	AF044374	United States, Idaho (1967)	822 bp	Pierce *et al.*, 1998
BTV-15	AF135228	China, Ershang (1996)	785 bp	Bonneau *et al.*, 1999

Serogroup and serotype	Accession number	Origin of virus and year of isolation	Number of basepairs	Reference
BTV-16	AF135229	China, Wuxi (1988)	785 bp	Bonneau *et al.*, 1999
BTV-17	AF044707	United States, California, Kern (1981)	822 bp	Pierce *et al.*, 1998
*Chuzan NS3	AB018091	Japan (1985)	728 bp	Yamakawa *et al.*, 1999
EEV-1 (S1FLD1)	AY1158Y8	South Africa, W. Cape, Stellenbosch (1999)	759 bp	Van Niekerk *et al.*, 2003
EEV-1 (S1FLD2)	AY115876	South Africa, W. Cape, Stellenbosch (1999)	759 bp	Van Niekerk *et al.*, 2003
EEV-1 (S1FLD3)	AY115875	South Africa, W. Cape, Stellenbosch (1999)	759 bp	Van Niekerk *et al.*, 2003
EEV-1 (S1FLD4)	AY115877	South Africa, W. Cape, Stellenbosch (1999)	759 bp	Van Niekerk *et al.*, 2003
EEV-1 (S1REF*)	AY115864	UN	759 bp	Van Niekerk *et al.*, 2003
EEV-1 (S1REF)	AY115865	South Africa, N. Cape, Colesberg (1976).	759 bp	Van Niekerk *et al.*, 2003
EEV-2 (S2REF)	AY115871	South Africa, N. Cape, Kimberley (1967).	759 bp	Van Niekerk *et al.*, 2003
EEV-3 (S3FLD1)	AY115874	South Africa, Gauteng, Kempton Park (2000)	759 bp	Van Niekerk *et al.*, 2003
EEV-3 (S3REF)	AY115867	South Africa, Limpopo, Naboomspruit (1971)	759 bp	Van Niekerk *et al.*, 2003
EEV-4 (S4REF)	AY115868	South Africa, Gauteng, Onderstepoort (1974)	759 bp	Van Niekerk *et al.*, 2003
EEV-5 (S5REF)	AY115869	South Africa, Gauteng, Johannesburg (1974)	759 bp	Van Niekerk *et al.*, 2003
EEV-6 (S6FLD1)	AY115872	South Africa, Limpopo, Hoedspruit (2000)	759 bp	Van Niekerk *et al.*, 2003
EEV-6 (S6FLD2)	AY115873	South Africa, Gauteng, Roodeplaat (2000)	759 bp	Van Niekerk *et al.*, 2003
EEV-6 (S6REF)	AY115866	South Africa, N. West, Potchefstroom (1991)	759 bp	Van Niekerk *et al.*, 2003
EEV-7 (S7REF)	AY115870	South Africa, KwaZulu-Natal, St. Lucia (2000)	759 bp	Van Niekerk *et al.*, 2003
*EHDV-1	L29023	United States, New Jersey (UN)	809 bp	Jensen and Wilson, 1995
*EHDV-2	L29022	United States, South Carolina (UN)	809 bp	Jensen *et al.*, 1994

REF denotes reference strain; FLD denotes field strain; (UN) denotes year of virus isolation unknown; UN denotes information unknown; N. Cape denotes Northern Cape province, S.A.; W. Cape denotes Western Cape province, S.A.; N. West denotes North West province, S.A.; Limpopo denotes the province previously known as the Northern Province, S.A. and the asterisk (*) indicates selected sequences for use in the second nucleotide or amino acid dataset

REFERENCES

Alphey, L. (1997). *Introduction to Biotechniques. DNA Sequencing. From experimental methods to bioinformatics.* First Edition. Bios Scientific Publishers Limited, New York, USA.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology* **215**, 403-410.

Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research* **25**, 3389-3402.

Anzola, J. V., Xu, Z., Asamizu, T. and Nuss, D. L. (1987). Segment-specific inverted repeats found adjacent to conserved terminal sequences in wound tumor virus genome and defective interfering RNAs. *Proceedings of the National Academy of Sciences USA* **84**, 8301-8305.

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

Attwood, T. K., Beck, M. E., Bleasby, A. J. and Parry-Smith, D. J. (1994). PRINTS – a database of protein motif fingerprints. *Nucleic Acids Research* **22**, 3590-3596.

Au, K. -S., Chan, W. -K., Burns, J. W. and Estes, M. K. (1989). Receptor activity of rotavirus nonstructural glycoprotein NS28. *Journal of Virology* **63**(11), 4553-4562.

Bairoch, A. and Apweiler, R. (2000). The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Research* **28**(1), 45-48.

Bairoch, A., Bucher, P. and Hofmann, K. (1997). The PROSITE database, its status in 1997. *Nucleic Acids Research* **25**, 217-221.

Baldi, P., Brunak, S., Frasconi, P., Pollastri, G. and Soda, G. (1999). Exploiting the past and future in protein secondary structure prediction. *Bioinformatics* **15**, 937-946.

Ball, J. M., Tian, P., Zeng, C. Q. Y., Morris, A. P. and Estes, M. K. (1996). Age dependent diarrhea induced by rotaviral nonstructural glycoprotein. *Science* **272**, 101-104.

Bansal, O. B., Stokes, A., Bansal, A., Bishop, D. and Roy, P. (1998). Membrane organization of bluetongue virus nonstructural glycoprotein NS3. *Journal of Virology* **72**(4), 3362-3369.

Barker, W. C., Garavelli, J. S., Hou, Z., Huang, H., Ledley, R. S., McGarvey, P. B., Mewes, H-W., Orcutt, B. C., Pfeiffer, F., Tsugita, A., Vinayaka, C. R., Xiao, C., Yeh, L-S. L. and Wu, C. (2001). Protein Information Resource: a community resource for expert annotation of protein data. *Nucleic Acids Research* **29**(1), 29-32.

Barnard, B. J. H. (1993). Circulation of African horsesickness virus in zebra (*Equus burchelli*) in the Kruger National Park, South Africa, as measured by the prevalence of type specific antibodies. *Onderstepoort Journal of Veterinary Research* **60**,111-117.

Barnard, B. J. H. (1997). Antibodies against some viruses of domestic animals in southern African wild animals. *Onderstepoort Journal of Veterinary Research* **64**, 95-110.

Barnard, B. J. H. and Paweska, J. T. (1993). Prevalence of antibodies against some equine viruses in zebra (*Zebra burchelli*) in the Kruger National Park, 1991-1992. *Onderstepoort Journal of Veterinary Research* **60**, 175-179.

Basak, A. K., Stuart, D. I., Roy, P. (1992). Preliminary crystallographic study on bluetongue virus capsid protein VP7. *Journal of Molecular Biology* **288,** 687-689.

Basak, A.K., Gouet, P., Grimes, J., Roy, P. and Stuart, D. (1996). Crystal structure of the top domain of African horsesickness virus VP7: comparisons with bluetongue virus VP7. *Journal of Virology* **70,** 3797-3806.

Beaton, A. R., Rodriguez, J., Reddy, Y. K. and Roy, P. (2002). The membrane trafficking protein calpactin forms a complex with bluetongue virus protein NS3 and mediates virus release. *Proceedings of the National Academy of Sciences USA* **99** (20), 13154-13159.

Bellamy, A. R. and Both, G. W. (1990). Molecular biology of rotaviruses. *Advances in Virus Research* **38**, 1-43.

Birnboim, H. G. C. and Doly, J. (1979). A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513-1522.

Boger, J., Emini, E. A. and Schmidt, A. (1986). Surface probability profile. A heuristic approach to the selection of synthetic peptide antigens. *Reports on the Sixth International Congress in Immunology*, Toronto, p 250.

Bonneau, K. R., Mullens, B. A. and MacLachlan, N. J. (2001). Occurrence of genetic drift and founder effect during quasispecies evolution of the VP2 and NS3/NS3A genes of bluetongue virus upon passage between sheep, cattle, and *Culicoides sonorensis*. *Journal of Virology* **75**(17), 8298-8305.

Bonneau, K. R., Zhang, N., Zhu, J., Zhang, F., Li, Z., Zhang, K., Xiao, L., Xiang, W. and MacLachlan, N. J. (1999). Sequence comparison of the L2 and S10 genes of bluetongue viruses from the United States and the People's Republic of China. *Virus Research* **61**(2), 153-160.

Borden, E. C., Shope, R. E. and Murray, F. A. (1971). Physicochemical and morphological relationships of some arthropod-borne viruses to bluetongue virus – a new taxonomic group. Physiochemical and serological studies. *Journal of General Virology* **13**(2), 261-271.

Bougrine, S. I., Fihri, O. F. and Fehri, M. M. F. (1998). Western immunoblotting as a method for the detection of African horse sickness virus protein-specific antibodies: differentiation between infected and vaccinated horses. *Archives of Virology Supplement* **14**, 329-336.

Bowne, J. G. and Ritchie, A. E. (1970). Some morphological features of bluetongue virus. *Virology* **40**, 903-911.

Bremer, C. W. (1976). A gel electrophoretic study of the protein and nucleic acid components of African horsesickness virus. *Onderstepoort Journal of Veterinary Research* **43**(4), 193-199.

Bremer, C. W. (1994). Characterisation of the African horsesickness virus genome and the development of new methods for virus detection. PhD thesis, Faculty of Veterinary Science, University of Pretoria, 208 pp.

Bremer, C. W. and Viljoen, G. J. (1998). Detection of African horsesickness virus and discrimination between two equine orbivirus serogroups by reverse transcription polymerase chain reaction. *Onderstepoort Journal of Veterinary Research* **65**, 1-8.

Bremer, C. W., Huismans, H. and Van Dijk, A. A. (1990). Characterization and cloning of the African horsesickness virus genome. *Journal of General Virology* **71**, 793-799.

Brendel, V., Bucher, P., Nourbakhsh, I., Blaisdell, B. E., Karlin, S. (1992). Methods and algorithms for statistical analysis of protein sequences. *Proceedings of the National Academy of Sciences USA* **89**, 2002-2006.

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

Brown, S. E., Gonzalez, H. A., Bodkin, D. K., Tesh, R. B. and Knudson, D. L. (1988). Intra- and inter-serogroup genetic relatedness of orbiviruses. II. Blot hybridization and reassortment in vitro of epizootic haemorrhagic disease serogroup, bluetongue type 10 and Pata viruses. *Journal of General Virology* **69**(1), 135-147.

Brown, S. E., Morrison, H. G., Karabatsos, N. and Knudson, D. L. (1991). Genetic relatedness of two new *Orbivirus* serogroups: Orungo and Lebombo. *Journal of General Virology* **72**, 1065-1072.

Callis, J. (1985). Bluetongue in the United States. In: *Bluetongue and Related Orbiviruses*. Proceedings of an International Symposium. Barber, T. L., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 37-42.

Campbell, M. K. (1995). *Biochemistry*. Second Edition. Saunders College Publishing, USA.

Carr, M. A., De Mattos, C. C., De Mattos, C. A. and Osburn, B. I. (1994). Association of bluetongue virus gene segment 5 with neuro-invasiveness. *Journal of Virology*, **68,** 1255-1257.

Chan, W. -K., Au, K. -S. and Estes, M. K. (1988). Topography of the simian rotavirus nonstructural glycoprotein (NS28) in the endoplasmic reticulum membrane. *Virology* **164**, 435-442.

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

Chung, C. T. and Miller, R. H. (1988). A rapid and convenient method for the preparation and storage of competent bacterial cells. *Nucleic Acids Research* **16**(8), 3580.

Claros, M. G. and Von Heijne, G. (1994). TopPred II: an improved software for membrane protein structure predictions. *CABIOS* **10**, 685-686.

Cohen, C. and Parry, D. A. (1986). α-Helical coiled coils – a widespread motif in proteins. *Trends in Biochemical Sciences* **11**, 245-248.

Collison, E. and Roy, P. (1983). Analysis of the genomes of bluetongue virus serotype 10 vaccines and a recent BTV-10 isolate from Washington. *American Journal of Veterinary Research* **44**(2), 235-237.

Corpet, F., Gouzy, J. and Kahn, D. (1998). The ProDom database of protein domain families. *Nucleic Acids Research* **26**, 323-326.

Corpet, F., Servant, F., Gouzy, J. and Kahn, D. (2000). ProDom and ProDom-CG: tools for protein domain analysis and whole genome comparisons. *Nucleic Acids Research* **28**, 267-269.

Cowley, J. A. and Gorman, B. M. (1989). Cross-neutralization of genetic reassortants of bluetongue virus serotypes 20 and 21. *Veterinary Microbiology* **19**, 37-51.

Cserzo, M., Wallin, E., Simon, I., Von Heijne, G. and Elofsen, A. (1997). Prediction of transmembrane alpha-helices in prokaryotic membrane proteins: the dense alignment surface method. *Protein Engineering* **10**, 673-676.

De Mattos, C. A., De Mattos, C. C. P., Dangler, C. A., Osburn, B. I., Ianconescu, M. and Kaufmann, R. (1992a). Hybridization relatedness of Israeli and U.S. bluetongue (BLU) serotypes using cDNA probes from BLU virus strain 11-UC8. *Archives of Virology* **122**(1-2), 23-33.

De Mattos, C. A., De Mattos, C. C., Muir, S. and Osburn, B. I. (1992b). The complete sequence of genome segment 10 from a virulent strain of USA bluetongue virus serotype 11. In: *Bluetongue, African horsesickness and related orbiviruses*. Walton, T.E. and Osburn, B. I. (Eds.). CRC Press, Boca Raton, p 471-476.

De Sá, R. O., Zellner, M., and Grubman, M. J. (1994). Phylogenetic analysis of segment 10 of African horsesickness virus and cognate genes from other orbiviruses. *Virus Research* **33**(2), 157-165.

Deléage, G. and Roux, B. (1987). An algorithm for protein secondary structure prediction based on class prediction. *Protein Engineering* **1**, 289-294.

Deléage, G., Clerc, F. F, and Roux B. (1989). ANTHEPROT: IBM PC and Apple Macintosh versions. *CABIOS* **5**(2), 159-160.

Deléage, G., Clerc, F. F, Roux B. and Gautheron, D. C. (1988). ANTHEPROT: a package for protein sequence analysis using a microcomputer. *CABIOS* **4**(3), 351-356.

Doms, R. W., Lamb, R. A., Rose, J. K. and Helenius, A. (1993). Folding and assembly of viral membrane proteins. *Virology* **193**, 545-562.

Du Toit, R. M. (1944). The transmission of blue-tongue and horse sickness by *Culicoides*. *Onderstepoort Journal of Veterinary Science and Animal Industry* **19**, 7-16.

Eaton, B. T. and Crameri, G. S. (1989). The site of bluetongue virus attachment to glycophorins from a number of animal erythrocytes. *Journal of General Virology* **70**, 3347-3353.

Eaton, B. T., Gould, A. R., Hyatt, A. D., Cauper, B. E. H., Martyn, J. C. and White, J. R. (1991). A bluetongue serogroup-reactive epitope in the amino-terminal half of the major core protein VP7 is accessible on the surface of the bluetongue virus particle. *Virology* **180,** 696-687.

Els, H. J. (1973). Electron microscopy of bluetongue virus RNA. *Onderstepoort Journal of Veterinary Research* **40**(2), 73-76.

Els, H. J. and Verwoerd, D. W. (1969). Morphology of bluetongue virus. *Virology* **38**, 213-219.

Emanuelsson, O., Nielson, H., Brunak, S. and Von Heijne, G. (2000). Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. *Journal of Molecular Biology* **300**, 1005-1016.

Erasmus, B. J. (1985). The history of bluetongue. In: *Bluetongue and Related Orbiviruses.* Proceedings of an International Symposium. Barber, T. L., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 7-12.

Erasmus, B. J., Adelaar, T. F., Smit, J. D., Lecatsas, G. and Toms, T. (1970). The isolation and characterization of Equine Encephalosis virus. *Bulletin de l'Office International des Epizooties* **74**, 781-789.

Estes, M. K. and Cohen, J. (1989). Rotavirus gene structure and function. *Microbiological Reviews* **53**(4), 410-449.

Farrell, R. E. (1993). RNA Methodologies. *A Laboratory Guide for Isolation and Characterization*. Academic Press, Inc., San Diego, California, USA.

French, T. J., Inumaru, S. and Roy, P. (1989). Expression of two related nonstructural proteins of bluetongue virus (BTV) type 10 in insect cells by a recombinant baculovirus: production of polyclonal ascitic fluid and characterization of the gene product in BTV-infected BHK cells. *Journal of Virology* **63**(8), 3270-3278.

Geourjon, C. and Deléage, G. (1995). ANTHEPROT 2.0: a three-dimensional module fully coupled with protein sequence analysis methods. *Journal of Molecular Graphics* **13**, 209-212.

Geourjon, C., Deléage, G., and Roux B. (1991). ANTHEPROT: an interactive graphics software for analyzing protein structures from sequences. *Journal of Molecular Graphics* **9**, 188-190.

Gerdes, G. H. and Pieterse, L. M. (1993). The isolation and identification of Potchefstroom virus: a new member of the equine encephalosis group of orbiviruses. *Journal of the South African Veterinary Association* **64**(3), 131-132.

GIBCO BRL Life Technologies. (1995). *Instruction Manual: BAC-TO-BAC™ Baculovirus Expression Systems.*

Gonzalez, H. A. and Knudson, D. L. (1987). Genetic relatedness of Corriparta serogroup viruses. *Journal of General Virology* **68**, 661-672.

Gonzalez, H. A. and Knudson, D. L. (1988). Intra- and inter-serogroup genetic relatedness of orbiviruses. I. Blot hybridization of viruses of Australian serogroups. *Journal of General Virology* **69**(1), 125-134.

Gorman, B. M. (1979). Variation in orbiviruses. *Journal of General Virology* **44**, 1-15.

Gorman, B. M. (1983). On the evolution of orbiviruses. *Intervirology* **20**, 169-180.

Gorman, B. M. and Taylor, J. (1985). Orbiviruses. In: *Virology*. Fields, B. N. *et al.* (Eds.). Raven Press, New York, p. 907-925.

Gorman, B. M., Taylor, J., Walker, P. J., Davidson, W. L. and Brown, F. (1981). Comparison of bluetongue type 20 with certain viruses of the bluetongue and Eubenangee serological groups of orbiviruses. *Journal of General Virology* **57**(2), 251-261.

Gould, A. R. (1988). Nucleotide sequence of the Australian bluetongue virus serotype 1 RNA segment 10. *Journal of General Virology* **69**, 945-949.

Gould, A. R., McColl, K. A. and Pritchard, L. I. (1992). Phylogenetic relationships between bluetongue viruses and other orbiviruses. In: *Bluetongue, African horsesickness and related orbiviruses*. Proceedings of the second international symposium. Walton, T. E. and Osburn, B. I. (Eds.), CRC Press, London, p 452-460.

Greiner, E. C., Barber, T. L., Pearson, J. E., Kramer, W. L. and Gibbs, P. J. (1985). Orbiviruses from *Culicoides* in Florida. In: *Bluetongue and Related Orbiviruses*. Proceedings of an International Symposium. Barber, T. L., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 195-200.

Grimes, J., Basak, A. K., Roy, P. and Stuart, D. I. (1995). The crystal structure of bluetongue virus VP7. *Nature* **373,** 167-170.

Grubman, M. J. and Lewis, S. A. (1992). Identification and characterization of the structural and nonstructural proteins of African horsesickness virus and determination of the genome coding assignments. *Virology* **186**(2), 444-451.

Guirakhoo, F., Catalan, J. A. and Monath, T. P. (1995). Adaptation of bluetongue virus in mosquito cells results in overexpression of NS3 proteins and release of virus particles. *Archives of Virology* **140**, 967-974.

Hewat, E. A., Booth, T. F., Loudon, P. T. and Roy P. (1992). Three-dimensional reconstruction of baculovirus expressed bluetongue virus core-like particles by cryo-electron microscopy. *Virology* **189,** 10-20.

Higgins, D. G. and Sharp, P. M. (1988). CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene* **73**, 237-244.

Higgins, D. G. and Sharp, P. M. (1989). Fast and sensitive multiple sequence alignments on a microcomputer. *CABIOS* **5**, 151-153.

Higgins, D. G., Thompson, J. D. and Gibson, T. J. (1996). Using CLUSTAL for multiple sequence alignments. *Methods in Enzymology* **266**, 383-402.

Hofmann, K. and Stoffel, M. (1993). TMbase – a database of membrane spanning protein segments. *Biological Chemistry Hoppe-Seyler* **347**, 166.

Hofmann, K., Bucher, P., Falquet, L. and Bairoch, A. (1999). The PROSITE database, its status in 1999. *Nucleic Acids Research* **27**(1), 215-219.

Holmes, I. H., Boccardo, G., Estes, M. K., Furuichi, M. K., Hoshino, Y., Joklik, W. K., McCrae, M., Mertens, P. P. C., Milne, R. G., Samal, K. S. K., Shikata, E., Winton, J. R., Uyeda, I. and Nuss, D. L. (1995). Family Reoviridae. In: *Classification and Nomenclature of Viruses. Sixth Report of the International Committee on Taxonomy of Viruses. Archives of Virology Supplement* **10**. Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A., Summers, M. D. (Eds.). Springer Verlag, Wien New York, p 208-237.

Hopp, T. P. and Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proceedings of the National Academy of Sciences USA* **78**, 3824-3828.

Howell, P. G., Groenewald, D., Visage, C. W., Bosman, A-M., Coetzer, J. A. W. and Guthrie, A. J. (2002). The classification of seven serotypes of equine encephalosis virus and the prevalence of homologous antibody in horses in South Africa. *Onderstepoort Journal of Veterinary Research* **69**(1), 79-93.

Huang, I. J., Hayama, E., Jeong, Y. J., Li, J. K. (1993). Conservation of the segment 4 gene sequence and of a leucine zipper motif in VP4 among five US bluetongue viruses. *Virology* **195**(2), 772-779

Huismans, H. and Cloete, M. (1987). A comparison of different cloned bluetongue virus genome segments as probes for the detection of virus-specified RNA. *Virology* **158**(2), 373-380.

Huismans, H. and Howell, P. G. (1973). Molecular hybridization studies of the relationships between different serotypes of bluetongue virus and on the difference between the virulent and attenuated strains of the same serotype. *Onderstepoort Journal of Veterinary Research* **40**(3), 93-194.

Huismans, H. and Van Dijk, A. A. (1990). Bluetongue virus structural components. *Current Topics in Microbiology and Immunology* **162**, 21-41.

Huismans, H., Bremer, C. W. and Barber, T. L. (1979). The nucleic acid and proteins of epizootic haemorrhagic disease virus. *Onderstepoort Journal of Veterinary Research* **46**(2), 95-104.

Huismans, H., Cloete, M. and Le Roux, A. (1987a). The genetic relatedness of a number of individual cognate genes of viruses in the bluetongue and closely related serogroups. *Virology* **161**(2), 421-428.

Huismans, H., Van Dijk, A. A. and Bauskin, A. R. (1987b). *In vitro* phosphorylation and purification of a nonstructural protein of bluetongue virus with affinity for single-stranded RNA. *Journal of Virology* **61**(11), 3589-3595.

Huismans, H., Van Staden, V. and Nel, L. H. (1992). Molecular comparison of the cognate genes of bluetongue, African horse sickness, and epizootic hemorrhagic disease viruses which encode nonstructural proteins NS2 and NS3. In: *Bluetongue, African horsesickness and related orbiviruses.* Proceedings of the second international symposium. Walton, T.E. and Osburn, B. I. (Eds.), CRC Press, Boca Raton, p 348-357.

Hwang, G. -Y., Yang, Y. -Y., Chiou, J. -F. and Li, J. K. -K. (1992). Sequence conservation among the cognate nonstructural NS3/3A protein genes of six bluetongue viruses. *Virus Research* **23**(1-2), 151-161.

Hyatt, A. D., Brookes, S. M., Gould, A. R. and Eaton, B. T. (1992). Morphogenesis of bluetongue viruses: development of a model for the site of virus synthesis, translocation, and release from infected tissue culture cells. In: *Bluetongue, African horsesickness and related orbiviruses.* Proceedings of the second international symposium. Walton, T.E. and Osburn, B. I. (Eds.), CRC Press, Boca Raton, p 358-365.

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

Hyatt, A. D., Gould, A. R., Coupar, B. and Eaton, B. T. (1991). Localization of the non-structural protein NS3 in bluetongue virus-infected cells. *Journal of General Virology* **72**(9), 2263-2267.

Hyatt, A. D., Zhao, Y. and Roy, P. (1993). Release of bluetongue virus-like particles from insect cells is mediated by BTV nonstructural protein NS3/NS3A. *Virology* **193**(2), 592-603.

Iwata, H., Yamagawa, M. and Roy, P. (1992). Evolutionary relationships among the gnat-transmitted orbiviruses that cause African horse sickness, bluetongue, and epizootic hemorrhagic disease as evidenced by their capsid protein sequences. *Virology* **191**(1), 251-261.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. and Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends in Biochemical Sciences* **23**, 403-405.

Jensen, M. J. and Wilson, W. C. (1995). A model for the membrane topology of the NS3 protein as predicted from the segment 10 of epizootic haemorrhagic virus serotype 1. *Archives of Virology* **140**(4), 799-805.

Jensen, M. J., Cheney, I. W., Thompson, L. H., Mecham, J. O., Wilson, W. C., Yamakawa, M., Roy, P. and Gorman, B. M. (1994). The smallest gene of the orbivirus, epizootic hemorrhagic disease, is expressed in virus-infected cells as two proteins and the expression differs from that of the cognate gene of bluetongue virus. *Virus Research* **32**(3), 353-364.

Jessup, D. A. (1985). Epidemiology of two *Orbiviruses* in California's native wild ruminants: preliminary report. In: *Bluetongue and Related Orbiviruses*. Proceedings of an International Symposium. Barber, T. L., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 43-45.

King, L. A. and Possee, R. D. (1992). *The Baculovirus Expression System: A Laboratory Guide*, Chapman and Hall, London.

Kozak, M. (1981). Possible role of flanking nucleotides in recognition of the AUG initiator codon by eukaryotic ribosomes. *Nucleic Acids Research* **9**(20), 5233-5252.

Kozak, M. (1984). Compilation and analysis of sequences upstream from the translational site in eukaryotic mRNAs. *Nucleic Acids Research* **12**(2), 857-872.

Kozak, M. (1989). The scanning model for translation: an update. *The Journal of Cell Biology* **108**, 229-241.

Kozak, M. (1991). Structural features in eukaryotic mRNAs that modulate the initiation of translation. *Journal of Biological Chemistry* **266**, 19867-19870.

Kumar, S., Tamura, K., Jakobsen, I. B. and Nei, M. (2001). MEGA2: molecular evolutionary genetics analysis software. *Bioinformatics* **17**(12), 1244-1245.

Kyte, J. and Doolittle, R. F. (1982). A simple method for displaying the hydrophobic character of a protein. *Journal of Molecular Biology* **157**, 105-132.

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

Laviada, M. D., Arias, M. and Sánchez-Vizcaíno, J. M. (1993). Characterization of African horsesickness virus serotype 4-induced polypeptides in Vero cells and their reactivity in Western immunoblotting. *Journal of General Virology 74*(1), 81-87.

Laviada, M. D., Casal, I., Vela, C. and Sánchez-Vizcaíno, J. M. (1994). Use of non-structural protein NS3 of African horse sickness virus for differentiation between vaccinated and infected animals. In: *OIE Scientific Conference on the control of foot and mouth disease, African horsesickness and contagious bovine pleuropneumonia*. Gaborone, Botswana, 20-23 April 1994.

Laviada, M. D., Roy, P., Sánchez-Vizcaino, J. M. and Casal, I. J. (1995). The use of African horse sickness virus NS3 protein, expressed in bacteria, as a marker to differentiate infected from vaccinated horses. *Virus Research* **38**(2-3), 205-218.

Lecatsas, G., Erasmus, B. J. and Els, H. J. (1973). Electron microscopic studies on equine encephalosis virus. *Onderstepoort Journal of Veterinary Research* **40**, 53-58.

Lee, J. W. and Roy, P. (1986). Nucleotide sequence of a cDNA clone of RNA segment 10 of bluetongue virus (serotype 10). *Journal of General Virology* **67**(12), 2833-2837.

Levy, J. A., Fraenkel-Conrat, H. and Owens, R. A. (1994). *Virology*. Third Edition. Prentice Hall Inc., New Jersey, USA.

Lewis, S. H. and Grubman, M. J. (1990). Bluetongue virus surface exposure of VP7. *Virus Research* **16**, 7-26.

Lipman, D. J. and Pearson, W. R. (1985). Rapid and sensitive protein similarity searches. *Science* **227**, 1435-1441.

Lodish, H, Baltimore, D., Berk, A., Zipursky, S. L., Matsudaira, P. and Darnell, J. (1995). *Molecular Cell Biology*. Third Edition. Scientific American Books, USA.

Lord, C. C., Venter, G. J., Mellor, P. S., Paweska, J. T. and Woolhouse, M. E. (2002). Transmission patterns of African horse sickness and equine encephalosis viruses in South African donkeys. *Epidemiology and Infection* **128**(2), 265-275.

Loudon, P. T. and Roy, P. (1992). Interaction of nucleic acids with core-like and subcore-like particles. *Virology* **191**, 231-236.

Lubroth, J. (1992). The complete epidemiological cycle of African horsesickness: our incomplete knowledge. In: *Bluetongue, African horse sickness and related orbiviruses*. Proceedings of the second international symposium. Walton, T. E. and Osburn, B. I. (Eds.). CRC Press, Boca Raton, p 982-989.

Luckow, V. A. and Summers, M. D. (1988). Trends in the development of baculovirus expression vectors. *Biotechnology* **6**, 47-55.

Luckow, V. A., Lee, S. C., Barry, G. F. and Olins, P. O. (1993). Efficient generation of infectious recombinant baculoviruses by site-specific transposon mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *Journal of Virology* **67**, 4566-4579.

Lupas, A. (1996). Prediction and analysis of coiled-coil structures. *Methods in Enzymology* **266**, 513-525.

Lupas, A., Van Dyke, M. and Stock, J. (1991). Predicting coiled coils from protein sequences. *Science* **252**, 1162-1164.

Martin, L.-A., Meyer, A. J., O'Hara, R. S., Fu, H., Mellor, P. S., Knowles, N. J. and Mertens, P. P. C. (1998). Phylogenetic analysis of African horse sickness virus segment 10: sequence variation, virulence characteristics and cell exit. *Archives of Virology Supplement* **14**, 281-293.

McCrae, M. A. and Joklik, W. K. (1978). The nature of the polypeptide encoded by each of the ten double-stranded RNA segments of reovirus type 3. *Virology* **89**, 578-593.

McLaughlin, S. and Aderem, A. (1995). The myristoyl-electrostatic switch: a modulator of reversible protein-membrane interaction. *Trends in Biochemical Sciences* **20**, 272-275.

Mertens, P. (2001). Orbiviruses and bluetongue virus. *Encyclopedia of Life Sciences*. Nature Publishing Group, England.

Mertens, P. P. C. and Sangar, D. V. (1985). Analysis of the terminal sequence of the genome segments of four orbiviruses. *Virology* **140**, 55-67.

Mertens, P. P. C., Arella, M., Attoui, H., Belloncik, S., Bergoin, M., Boccardo, G., Booth, T. F., Chiu, W., Diprose, J. M., Duncan, R., Estes, M. K., Gorziglia, M., Gouet, P., Gould, A. R., Grimes, J. M., Hewat, E., Hill, C. I. H., Hoshino, Y., Joklik, W. K., Knowles, N., López Ferber, M. L., Malby, R., Marzachi, C., McCrae M. A., Milne, R. G., Nibert, M., Nunn, M., Omura, T., Prasad, B. V. V., Pritchard, I., Samal, S. K., Schoehn, G., Shikata, E., Stoltz, D. B., Stuart, D. I., Suzuki, N., Upadhyaya, N., Ueyada, I., Waterhouse, P., Williams, C. F., Winton, J. R. and Zhou, H. Z. (2000). Reoviridae. In: *Virus Taxonomy: Classification and Nomenclature of Viruses. Seventh Report of the International Committee on Taxonomy of Viruses.* Van Regenmortel, M. H. V., Fauquet, C. M., Bishop, D. H. L., Carsten, E. B., Estes, M. K., Lemon, S. M., Manilof, J., Mayo, M. A., McGeoch, D. J., Pringle, C. R. and Wickner, R. B. (Eds.). Academic Press, San Diego, p 395-480.

Mertens, P. P. C., Brown, F. and Sangar, D. V. (1984). Assignment of the genome segments of bluetongue virus type 1 to the proteins which they encode. *Virology* **135**(1), 207-217.

Meyer, J. C., Bergmann, C. C. and Bellamy, A. R. (1989). Interaction of rotavirus cores with the nonstructural glycoprotein NS28. *Virology* **171**, 98-107.

Moss, S. R., Jones, L. D. and Nutall, P. A. (1992). Comparison of the nonstructural protein, NS3, of tick-borne and insect-borne orbiviruses. *Virology* **187**(2), 841-844.

Murphy, F. A., Borden, E. C., Shope, R. E. and Harrison, A. (1971). Physicochemical and morphological relationships of some arthropod-borne viruses to bluetongue virus – a new taxonomic group. Electron microscopic studies. *Journal of General Virology* **13**, 273-288.

Nel, L. H. and Huismans, H. (1991). Synthesis of virus-specified tubules of epizootic haemorrhagic disease virus using a baculovirus system. *Virus Research* **19**, 139-152.

Nel, L. H., Picard, L. A. and Huismans, H. (1990). A characterization of the nonstructural protein from which the virus-specified tubules in epizootic haemorrhagic disease are composed. *Virus Research* **18**, 219-230.

Nevill, E. M., Erasmus, B. J. and Venter, G. J. (1992). A six-year survey of viruses associated with *Culicoides* biting midges throughout South Africa (Diptera: Ceratopogonidae). In: *Bluetongue, African horsesickness and related orbiviruses*. Proceedings of the second international symposium. Walton, T.E. and Osburn, B. I. (Eds.). CRC Press, Boca Raton, p 314-319.

Newton, K., Meyer, J. C., Bellamy, A. R. and Taylor, J. A. (1997). Rotavirus nonstructural glycoprotein NSP4 alters plasma membrane permeability in mammalian cells. *Journal of Virology* **71**(12), 9458-9456.

Nielsen, H., Engelbrecht, J., Brunak, S. and Von Heijne, G. (1997). Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Engineering* **10**(1), 1-6.

O'Hara, R. S., Meyer, A. J., Burroughs, J. N., Pullen, L., Martin, L.-A. and Mertens, P. P. C. (1998). Development of a mouse model system, coding assignments and identification of the genome segments controlling virulence of African horse sickness virus serotypes 3 and 4. *Archives of Virology Supplement* **14**, 259-279.

Oldfield, S., Adachi, A., Urakawa, T., Hirasawa, T. and Roy, P. (1990). Purification and characterization of the major group-specific core antigen VP7 of bluetongue virus synthesized by a recombinant baculovirus. *Journal of General Virology* **71**, 2649-2656.

Oldfield, S., Hirasawa, T. and Roy, P. (1991). Sequence conservation of the outer capsid protein, VP5, of bluetongue virus, a contrasting feature to the outer capsid protein VP2. *Journal of General Virology* **72**, 449-451.

Ozawa, Y. (1985). Bluetongue and related orbiviruses: overview of the world situation. In: *Bluetongue and Related Orbiviruses*. Proceedings of an International Symposium. Barber, T. L., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 13-20.

Page, R. D. M. and Holmes, E. C. (1998). *Molecular Evolution A phylogenetic approach*. Blackwell Science Ltd, Oxford.

Parker, J. M. R., Guo, D. and Hodges, R. S. (1986). New hydrophilicity scale derived from high-performance liquid chromatography peptide retention data: correlation of predicted surface residues with antigenicity and X-ray-derived accessibility sites. *Biochemistry* **25**, 5425-5432.

Parsonson, I. M. and Snowdon, W. A. (1985). Bluetongue, epizootic haemorrhagic disease of deer and related viruses: current situation in Australia. In: *Bluetongue and Related Orbiviruses.* Proceedings of an International Symposium. Barber, T. L., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 27-35.

Paweska, J. (1999). A fatal case of equine encephalosis in the Port Elizabeth district, South Africa. *AnimalNet*, 29 April 1999.

Pedley, S., Mohamed, M. E. H. and Mertens, P. P. C. (1988). Analysis of genome segments from six different isolates of bluetongue virus using RNA-RNA hybridisation: a generalised coding assignment for bluetongue viruses. *Virus Research* **10**(4), 381-390.

Pierce, C. M., Balasuriya, U. B. R. and MacLachlan, N. J. (1998). Phylogenetic analysis of the S10 gene of field and laboratory strains of bluetongue virus from the United States. *Virus Research* **55**(1), 15-27.

Poruchynsky, M. S. and Atkinson, P. H. (1991). Rotavirus protein rearrangements in purified membrane-enveloped intermediate particles. *Journal of Virology* **65**, 4720-4727.

Potgieter, A. C., Steele, A. D. and Van Dijk, A. A. (2002). Cloning of complete genome sets of six dsRNA viruses using an improved cloning method for large dsRNA genes. *Journal of General Virology* **83**, 2215-2223.

Rao, C. D. and Roy, P. (1983). Genetic variation of bluetongue virus serotype-11 isolated from host (sheep) and vector (*Culicoides variipennis*) at the same site. *American Journal of Veterinary Research* **44**(5), 911-914.

Rao, C. D., Kiuchi, A. and Roy, P. (1983). Homologous terminal sequences of the genome double-stranded RNAs of bluetongue virus. *Journal of Virology* **46**(2), 378-383.

Riegler, L., Venter, G. J., O'Hara, R., Mellor, P. S. and Mertens, P. P. (2000). Variation in NS3 of African horsesickness can influence transmission by culicoides vector species. *Seventh International Symposium on Double-stranded viruses*, Palm Beach, Aruba, December, 2000.

Rost, B. (1996a). PHD: predicting one-dimensional protein structure by profile based neural networks. *Methods in Enzymology* **266**, 525-539.

Rost, B. (1996b). PredictProtein. *Methods in Enzymology* **266**, 525-539.

Rost, B. and Sander, C. (1993). Prediction of protein secondary structure at better than 70% accuracy. *Journal of Molecular Biology* **232**, 584-599.

Rost, B. and Sander, C. (1994). Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* **19**, 55-72.

Rost, B., Fariselli, P. and Casadio, R. (1996). Topology prediction for helical transmembrane proteins at 86% accuracy. *Protein Science* **7**, 1704-1718.

Rost, B., Sander, C. and Schneider, R. (1994). Redefining the goals of protein secondary structure prediction. *Journal of Molecular Biology* **235**, 13-26.

Roy, P. (1989). Bluetongue virus genetics and genome structure. *Virus Research* **13**(3), 179-206.

Roy, P. (1992). Bluetongue virus proteins. *Journal of General Virology* **73**(12), 3051-3064.

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

Roy, P., Fukusho, A., Ritter, G. D. and Lyon, D. (1988). Evidence for genetic relationship between RNA and DNA viruses from the sequence homology of a putative polymerase gene of bluetongue virus with that of vaccinia virus: conservation of RNA polymerase genes from diverse species. *Nucleic Acids Research* **16**(24), 11759-11767.

Roy, P., Marshall, J. A. A. and French, T. J. (1990). Structure of the bluetongue virus genome and its encoded proteins. *Current topics in Microbiology and Immunology* **162**, 43-87.

Roy, P., Mertens, P. P. C. and Casal, I. (1994). African horse sickness virus structure. *Comparative Immunology Microbiology and Infectious Diseases* **17**(3-4), 243-273.

Sailleau, C., Moulay, S. and Zientara, S. (1997). Nucleotide sequence comparison of the segments S10 of the nine African horsesickness virus serotypes. *Archives of Virology* **142**(5), 965-978.

Sailleau, C., Seignot, J., Davoust, B., Cardinale, E., Fall, B., Hamblin, C. and Zientara, S. (2000). African horse sickness in Senegal: serotype identification and nucleotide sequence determination of segment S10 by RT-PCR. *The Veterinary Record* **146**(4), 107-108.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989). *Molecular cloning. A laboratory manual.* Second Edition. Cold Spring Harbor Laboratory Press.

Sander, C. and Schneider, R. (1991). Database of homology-derived structures and the structural meaning of sequence alignment. *Proteins* **9**, 56-68.

Sanger, F., Nicklen, S. and Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences USA* **74**, 5463-5467.

Sebastiani, F., Meiswinkel, R., Gomulski, L. M., Guglielmino, C. R., Mellor, P. S., Malacrida, A. R. and Gasperi, G. (2001). Molecular differentiation of the Old World *Culicoides imicola* species complex (Diptera, Ceratopogonidae) inferred using random amplified polymorphic DNA markers. *Molecular Ecology* **10**, 1773-1786.

Silverman, L. and Resh, M. D. (1992). Lysine residues form an integral component of a novel NH2-terminal membrane targeting motif for myristylated $pp60^{v\text{-}src}$. *Journal of Cellular Biology* **119**, 415-425.

Sonnhammer, E. L. L. and Kahn, D. (1994). Modular arrangement of proteins as inferred from analysis of homology. *Protein Science* **3**, 482-492.

Sonnhammer, E., Von Heijne, G. and Krogh, A. (1998). A hidden Markov model for predicting transmembrane helices in protein sequences. *Proceedings of the Sixth International Conference on Intelligent Systems for Molecular Biology* (ISMB98), 175-182.

St George, T. D. (1985). The search for bluetongue viruses in Australia. In: *Bluetongue and Related Orbiviruses*. Proceedings of an International Symposium. Barber, L. T., Jochim, M. M. and Osburn, B. I. (Eds.), Alan R. Liss, Inc., New York, p 295-305.

Stoltz, M. A., Van der Merwe, C. F., Coetzee, J. and Huismans, H. (1996). Subcellular localization of the nonstructural protein NS3 of African horsesickness virus. *Onderstepoort Journal of Veterinary Research* **63**(1), 57-61.

Sugiyama, K., Bishop, D. H. L. and Roy, P. (1981). Analyses of the genomes of bluetongue viruses recovered in the United States. I. Oligonucleotide fingerprint studies that indicate the existence of naturally occurring reassortant BTV isolates. *Virology* **114**, 210-217.

Swanepoel, R., Erasmus, B. J., Williams, R. and Taylor, M. B. (1992). Encephalitis and chorioretinitis associated with neurotropic African horsesickness virus infection in laboratory workers. Part III. Virological and serological investigations. *South African Medical Journal* **81**(9), 458-461.

Taylor, J. A., Meyer, J. C., Legge, M. A., O'Brien, J. A., Street, J. E., Lord, V. J., Bergmann, C. C. and Bellamy, A. R. (1992). Transient expression and mutational analysis of the rotavirus intracellular receptor: the C-terminal methionine residue is essential for ligand binding. *Journal of Virology* **66**(6), 3566-3572.

Taylor, J. A., O'Brien, J. A. and Yeager, M. (1996). The cytoplasmic tail of NSP4, the endoplasmic reticulum-localized non-structural glycoprotein of rotavirus, contains distinct virus binding and coiled-coil domains. *EMBO*, **15**(17), 4469-4476.

Theodoridis, A., Nevill, E. M., Els, H. J. and Boshoff, S. T. (1979). Viruses isolated from *Culicoides* midges in South Africa during unsuccessful attempts to isolate bovine ephemeral fever virus. *Onderstepoort Journal of Veterinary Research* **46**, 191-198.

Theron, J., Uitenweerde, J. M., Huismans, H. and Nel, L. H. (1994). Comparison of the expression and phosphorylation of the non-structural protein NS2 of three different orbiviruses: evidence for the involvement of a ubiquitous cellular kinase. *Journal of General Virology* **75**, 3401-3411.

Thomas, C.P, Booth, T.F. and Roy, P. (1990). Synthesis of bluetongue viral-coded phosphoprotein and formation of inclusion bodies by recombinant baculoviruses in insect cells: it binds the single-stranded RNA species. *Journal of General Virology* **71,** 2073-2083.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997). The Clustal_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* **25**(24), 4876-4882.

Tian, P., Hu, Y., Schilling, W. P., Lindsay, D. A., Eiden, J. and Estes, M. K. (1994). Nonstructural glycoprotein NSP4 of rotavirus affects intracellular calcium levels. *Journal of Virology* **68**, 251-257.

Tomori, O., Baba, S., Adu, F. and Adeniji, J. (1992). An overview and perspective on orbivirus disease prevalence and occurrence of vectors in Africa. In: *Bluetongue, African horsesickness and related orbiviruses.* Proceedings of the second international symposium. Walton, T. E. and Osburn, B. I. (Eds.), CRC Press, Boca Raton, p 23-33.

Turnbull, P. J., Cormack, S. B. and Huismans, H. (1996). Characterization of the gene encoding core protein VP6 of two African horsesickness virus serotypes. *Journal of General Virology* **77**, 1421-1423.

Uitenweerde, J. M., Theron, J., Stoltz, M. A. and Huismans, H. (1995). The multimeric non-structural NS2 protein of bluetongue virus, African horsesickness virus and epizootic haemorrhagic disease virus differ in their single stranded RNA-binding ability. *Virology*, **209**, 624-632.

Van Dijk, A. A. and Huismans, H. (1988). *In vitro* transcription and translation of bluetongue virus mRNA. *Journal of General Virology* 69(3), 573-581.

Van Niekerk, M., Freeman, M., Paweska, J. T., Howell, P. G., Guthrie, A. J., Potgieter, A. C., Van Staden, V., and Huismans, H. (2003). Variation in the NS3 gene and protein in South African isolates of bluetongue and equine encephalosis viruses. *Journal of General Virology* (in press)

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

Van Niekerk, M., Van Staden, V., Van Dijk, A. A. and Huismans, H. (2001b). Variation of African horsesickness virus nonstructural protein NS3 in southern Africa. *Journal of General Virology* **82**, 149-158.

Van Staden, V. and Huismans, H. (1991). A comparison of the genes which encode non-structural protein NS3 of different orbiviruses. *Journal of General Virology* **72**(5), 1073-1079.

Van Staden, V., Smit, C. C., Stoltz, M. A., Maree, F. F. and Huismans, H. (1998). Characterization of two African horse sickness virus nonstructural proteins, NS1 and NS3. *Archives of Virology Supplement* **14**, 251-258.

Van Staden, V., Stoltz, M. A. and Huismans, H. (1995). Expression of nonstructural protein NS3 of African horsesickness virus (AHSV): evidence for a cytotoxic effect of NS3 in insect cells, and characterization of the gene products in AHSV infected Vero cells. *Archives of Virology* **140**(2), 289-306.

Van Staden, V., Theron, J., Greyling, B. J., Huismans, H. and Nel, L. H. (1991). A comparison of the nucleotide sequences of cognate NS2 genes of three different orbiviruses. *Virology* **185**(1), 500-504.

Venter, G. J., Groenewald, D. M., Paweska, J. T., Venter, E. H. and Howell, P. G. (1999). Vector competence of selected South African *Culicoides* species for the Bryanston serotype of equine encephalosis virus. *Medical and Veterinary Entomology* **13**, 393-400.

Verwoerd, D. W. (1969). Purification and characterization of bluetongue virus. *Virology* **38**, 203-212.

Verwoerd, D. W. and Huismans, H. (1969). On the relationship between bluetongue, African horsesickness and reoviruses: hybridization studies. *Onderstepoort Journal of Veterinary Research* **36**(2), 175-180.

Verwoerd, D. W., Louw, H and Oellermann, R. A. (1970). Characterization of bluetongue virus ribonucleic acid. *Journal of Virology* **5**(1), 1-7.

Viljoen, G. J. and Huismans, H. (1989). The characterization of equine encephalosis virus and the development of genomic probes. *Journal of General Virology* **70**, 2007-2015.

Von Heijne, G. (1992). Membrane protein structure prediction: hydrophobicity analysis and the positive inside rule. *Journal of Molecular Biology* **225**, 497-494.

Vreede, F. T. and Huismans, H. (1994). Cloning, characterization, and expression of the gene that encodes the major neutralization-specific antigen of African horsesickness virus serotype 3. *Journal of General Virology* **75**, 3629-3633.

Vreede, F. T. and Huismans, H. (1998). Sequence analysis of RNA polymerase gene of African horse sickness virus. *Archives of Virology* **143**, 413-419.

Wade-Evans, A. M. (1990). Complete sequence of genome segment 10, encoding the NS3 protein, of bluetongue virus, serotype 1 from South Africa. *Nucleic Acids Research* **18**(16), 4920.

Williams, R., Du Plessis, D. H. and Van Wyngaardt, W. (1993). Group-reactive ELISAs for detecting antibodies to African horsesickness and equine encephalosis viruses in horse, donkey, and zebra sera. *Journal of Veterinary Diagnostic Investigation* **5**(1), 3-7.

Worley, K. C., Wiese, B. A. and Smith, R. F. (1995). BEAUTY: an enhanced BLAST-based search tool that integrates multiple biological information resources into sequence similarity search results. *Genome Research* **5**(2), 173-184.

Wu, C. H., Zhao, S., Chen, H. L., Lo, C. J. and McLarty, J. (1996). Motif identification neural design for rapid and sensitive protein family search. *CABIOS* **12**(2), 109-118.

Wu, C., Shivakumar, S., Shivakumar, C. V. and Chen, S. (1998). GeneFIND web server for protein family identification and information retrieval. *Bioinformatics* **14**(2), 223-224.

Wu, X., Chen, S. -Y., Iwata, H., Compans, R. W. and Roy, P. (1992). Multiple glycoproteins synthesized by the smallest RNA segment (S10) of bluetongue virus. *Journal of Virology* **66**(12), 7104-7112.

Yamakawa, M., Kubo, M. and Furuuchi, S. (1999). Molecular analysis of the genome of Chuzan virus, a member of the Palyam serogroup viruses, and its phylogenetic relationships to other orbiviruses. *Journal of General Virology* **80**, 937-941.

Zhang, M., Zeng, C. Q., Dong, Y., Ball, J. M., Saif, L. J., Morris, A. P. and Estes, M. K. (1998). Mutations in the rotavirus nonstructural glycoprotein NSP4 are associated with altered virus virulence. *Journal of Virology* **72**(5), 3666-3672.

Zhao, H., Lindqvist, B., Garoff, H., Von Bonsdorff, C. H. and Liljestrom, P. (1994). A tyrosine-based motif in the cytoplasmic domain of the alphavirus envelope protein is essential for budding. *EMBO*, **13**(18), 4204-4211.

Zientara, S., Sailleau, C., Moulay, S., Plateau, E. and Cruciere, C. (1993). Diagnosis and molecular epidemiology of the African horsesickness virus by the polymerase chain reaction and restriction patterns. *Annales de Recherches Veterinaires* **24**, 385-395.

Zientara, S., Sailleau, C., Plateau, E., Moulay, S., Mertens, P. P. C. and Crucière, C. (1998). Molecular epidemiology of African horse sickness virus based on analyses and comparisons of genome segments 7 and 10. *Archives of Virology Supplement* **14**, 221-234.