Assembly of Full-Length cDNA, and Heterologous Expression, of *Nudaurelia β Virus* RNA

Gary Joseph Luke

Assembly of Full-Length cDNA, and Heterologous Expression, of *Nudaurelia β Virus* RNA

Submitted in Fulfilment of the Requirements for the Degree of

MASTER OF SCIENCE
of

RHODES UNIVERSITY

by

GARY JOSEPH LUKE

January 2001

TABLE OF CONTENTS

ACKNOWLEDGEMENTS

I would like to acknowledge God Almighty, through Jesus Christ, without whom I would not be where I am today.

I am very grateful for the invaluable help of my supervisor, Professor Don Hendry. His continuing advice and suggestions guided my research to reach its objective.

Dr. Karl Gordon, for motivating via the internet to persist, and helping with unexplainable problems. I would also like to thank him for supplying most of the plasmids and primers used in the project.

Professor Greg Blatch and his research group, for their continued input, encouragement and constructive criticism which was invaluable.

Dr. Rosemary Dorrington and her research group, for their guidance and technical assistance throughout my research.

I would also like to thank my wife Tina, for her help, understanding and encouragement throughout, and spending many hours proof-reading my rough drafts.

Last but not least, I must thank my little girl Kristine for not keeping us awake all night.

ABSTRACT

Nudaurelia beta virus (NβV) is a monopartite genome virus belonging to the family Tetraviridae. Its host range has been found to be limited to a single insect order, the Lepidoptera (moths and butterflies). The single-stranded positive-sense RNA genome consists of 6625 nucleotides containing two open reading frames (ORFs). The 5' proximal ORF of 5778 nucleotides encodes a protein of 215 kDa containing three functional domains characteristic of RNA-dependent RNA polymerase. The 3' proximal ORF, of 1836 nucleotides, encodes the 66 kDa capsid precursor protein and overlaps the replicase gene by more than 99% and is in the +1 reading frame relative to the replicase reading frame.

The full-length cDNA construct of the NβV genome was assembled using a homologous overlapping PCR linking method. The starting material consisted of seven overlapping pieces that were constructed for sequencing. Due to the degradation of the full-length RNA obtained from virus extracted from field-collected *Nudaurelia cytherea capensis* larvae other alternative methods needed to be applied. Sub-cloning using restriction enzyme sites also required an alternative method being used, due to the abundance of restriction sites of the same type in the NβV genome. This led to the use of a method similar to "*DNA Shuffling"* where overlapping pieces were connected using a modified PCR protocol. After the construction of the NβV genome, the full-length PCR product was cloned and checked for large insertion and deletions that could have resulted from the PCR amplification.

The heterologous expression of the NβV capsid protein linked to a fusion protein (Glutathione S-transferase) in *E.coli,* confirmed the authenticity of the prescribed capsid gene ORF. The expression showed that the virus protein was subjected to protease digestion in DH5α *E.coli,* suggesting that the protein was insoluble in the cell cytoplasm. The capsid gene expression in a modified *E.coli* strain, Epicurian Coli BL21-CodonPlus (DE3)-RIL, resulted in high levels of the correct molecular weight protein with minimal

degradation. The modified strain was designed for over-expression of eukaryotic protein with lowered protease activity.

The above results have opened the way for further research that would yield valuable insight into the molecular biology and replication strategy of the NβV in cell cultures.

LIST OF FIGURES

CHAPTER 1:

CHAPTER 3:

CHAPTER 4:

CHAPTER 5:

LIST OF TABLES

CHAPTER 1:

CHAPTER 4:

CHAPTER 5:

APPENDICES:

ABBREVIATIONS

%	-	percent
µ (prefix)	-	micro (10^{-6})
µg	-	microgram
µl	-	microlitre
+	-	positive-sense
A_{260}	-	absorbance at 260 nanometre
A_{280}	-	absorbance at 280 nanometre
Ab	-	antibody
bp	-	base pairs
cDNA	-	complementary DNA
CP	-	coat protein
DNA	-	deoxyribonucleic acid
dNTP	-	2' -deoxynucleoside- 5' -triphosphate
E	-	extinction coefficient
EDTA	-	ethylene diamine tetra-acetic acid
et al.	-	et alia (and others)
Fig.	-	figure
g	-	gram
h	-	hour
HaSV	-	Helicoverpa armigera stunt virus
kb	-	kilobase pairs
kDa	-	kilodalton
l	-	litre
M	-	molar
m/v	-	mass per volume
mg	-	milligram
min	-	minutes

ml	-	millilitre
mm	-	millimetres
mM	-	millimolar
mRNA	-	messenger ribonucleic acid
MW	-	molecular weight
NBV	-	Nudaurelia beta virus
ng	-	nanogram
nt	-	nucleotide
^{O}C	-	degrees Celsius
PAGE	-	polyacrylamide gel electrophoresis
PBS	-	phosphate buffered saline
PCR	-	polymerase chain reaction
PEG	-	polyethylene glycol
pH	-	negative logarithm of the hydrogen ion concentration
RNA	-	ribonucleic acid
rpm	-	revolutions per minute
RT	-	room temperature
s	-	second
SDS	-	sodium dodecyl sulphate
ss	-	single-stranded
Temed	-	N, N, N', N' -tetramethylethylenediamine
Tris	-	tris (hydroxymethyl) amino methane
tRNA	-	transfer ribonucleic acid
UV	-	ultraviolet
v/v	-	volume per volume
VLP	-	virus-like particles
w/v	-	weight per volume

CHAPTER 1

REVIEW OF LITERATURE

1.1. TETRAVIRIDAE

(i) INTRODUCTION

This family of viruses, formerly referred to as the *Nudaurelia β virus* group, consists of ten members and eight unassigned possible members (Hanzlik & Gordon, 1997). The tetraviruses are confined to a single insect order, the *Lepidoptera* (moths and butterflies), making them the only recognized virus family whose host range is restricted to insects (Moore, 1991; Reinganum, 1991). Little is known about the ecology of these viruses although large-scale epizootics that have sharply reduced host populations have been observed (Moore, 1991). Agricultural interest in the possible utility of the tetraviruses stems from the ability of these viruses to infect the midgut cells, rapidly retarding the growth of the host and eventually liquefying the body contents. The insect hosts of the tetraviruses are among the most highly damaging worldwide insect pests, for which new control strategies are urgently required. More precise knowledge of the structures that allow these tetraviruses to survive a degradative milieu after ingestion and enter the cells lining its host's midgut may provide leads for the design of novel, proteinaceous, control factors (Christian *et al*., 1993). Such factors, after expression in transgenic crop plants, would be able to mimic these abilities of the viruses and deliver deleterious contents into the midgut cells of harmful host, thus serving to protect the plant.

Structural studies of these viruses have centred on the unique T = 4 symmetry (tetra) found in the icosahedral structure of their capsids, the distinguishing feature of the *Tetraviridae* family (Francki *et al.,* 1991). The single-stranded positive-sense RNA genome of the tetraviruses is encapsidated in a icosahedral structure. It was thought until recently that the capsid contained only a monopartite, single-stranded RNA genome,

however recently, a subgroup of tetraviruses with bipartite genomes was characterized and showed to have bipartite single-stranded RNA genomes with capped nonpolyadenylated strands of 5.5 kb (RNA 1) and 2.5kb (RNA 2) and no 3' blockage (Agrawal & Johnson, 1992; Hanzlik *et al*., 1993). Reports of complete genome sequences of the tetraviruses from both groups and other data that is emerging about the pathology and replication strategies appear to have broad implications for areas as diverse as cell and structural biology, virus evolution, and agricultural and medical biotechnology (Hanzlik & Gordon, 1997).

(ii) HISTORY

In the 1940s Tooke and Hubbard (1941) noted the annual occurrence of a disease of the larvae of the pine tree emperor moth, *Nudaurelia cytherea capensis,* described then as *polyhedral wilt* which was characterized by flaccidity and internal liquefaction of the larvae. The infection of the larval population was so severe and widespread that dead larvae literally formed a carpet beneath the trees.

In 1965, Grace and Mercer from Australia discovered larvae from *Antheraea eucalypti* reared in the laboratory that had died. The cause of death could not be revealed by routine pathological investigations. However electron microscopy studies of the blood and tissues of the newly dead larvae revealed large numbers of spherical, virus-like particles. The particles were packed in close arrays in the midgut cells of the larvae.

In 1968, Hendry *et al* of South Africa noted with the aid of electron microscopy that larvae of the *Nudaurelia cytherea capensis,* a common pest of *Pinus radiata,* had died due to infection in the midgut cells of the larvae by a non-inclusion virus. Juckes (1970) found that diseased *N. cytherea* were infected by five small RNA viruses which were termed Nudaurelia γ,δ,α,ε, and β viruses. Longworth and Harrap in 1968, discovered similar *enonoccluded* spherical virus that infected four types of saturniid species, *Philosamia cynthia x ricini, Hyalophora cecropia, Antheraea pernyi* and *Actias selene.* Each tetravirus has been named according to its insect host. The extensive biophysical characterization of *Nudaurelia β virus* (NβV) lead to the type name of the viruses

becoming the *Nudaurelia β-like virus* group. In 1982, the name *Nudaurelia β-like virus* was formally recognized (Matthews, 1982).

Finch *et al.*, in 1974 supplied the first three-dimensional reconstruction of NβV, showing a capsid with a icosahedral surface symmetry T = 4, and a diameter of 35-40 nm. Studies by Struthers and Hendry in 1974 showed that NβV contained 11% single-stranded RNA with a nucleotide composition of 28% guanylic acid, 23% uridylic acid, 25% cytidylic acid, and 24% adenylic acid. Further studies by Struthers and Hendry (1974) showed that the small 1.8 MDa single-stranded +RNA genome is loosely packed inside the non-enveloped, icosahedral capsid. In 1985, Hendry *et al.* isolated another virus from Nudaurelia carcasses, the *Nudaurelia ω virus* (NωV), which had biophysical properties similar to those of NβV but differed by having a second ss RNA genomic strand. These findings lead to a renaming of the group. The lattice symmetry of the T=4 capsid structure, which is a feature of all the viruses, created the idea of using the Greek name *tettares,* four, to form the basis for the new group name *Tetraviridae* (Francki *et al.,* 1991). NωV subsequently becomes the type virus for a second genus within the *Tetraviridae*, the ω-like genus (Murphy *et al.,* 1995), which joined the β-like genus.

(iii) CLASSIFICATION

In the above section on the history of *Tetraviridae* it was mentioned that the genus naming of these viruses changed based on the number of the genomic strands. In the past most of the classification was based on the morphological and biophysical characteristics of each virus. Serological relationship to the NβV, the type member, was the ultimate method of classifying a newly found virus, as being a tetravirus. Members of the tetraviruses were and still are named after the insect host species of the original isolate. In 1997, Hanzlik & Gordon compiled a listing of the tetraviruses (Table 1-1) on the grounds of the specifications laid down by the International Committee on the Taxonomy of Viruses (ICTV). Viruses were grouped on the basis of (i) a major protein component of >60 kDa produced from (ii) a separate genomic RNA combined with (iii) the biological property of having an insect host (Murphy *et al*., 1995).

Table 1-1: Classification of the Family *Tetraviridae* (based on Hanzlik & Gordon, 1997)

Virus name	Abbreviation	Host family	Origin
Genus: Omegatetravirus			
Nudaurelia capensis ω virus	NωV	*Saturniidae*	South Africa
Helicoverpa armigera Stunt virus	HaSV	*Noctuidae*	Australia
Genus: Betatetravirus			
Nudaurelia capensis β virus	NβV	S*aturniidae*	South Africa
Antheraea eucalypti β virus	AeV	*Saturniidae*	Australia
Darna trima virus	DtV	*Limacodidae*	Malaysia
Dasychira pudidunda virus	DpV	*Lymantriidae*	UK
Pilosamia cynthia x *ricini* virus	PxV	*Saturniidae*	UK
Pseudoplusia includens virus	PiV	*Noctuidae*	USA
Thosea asigna virus	TaV	*Limacodidae*	Malaysia
Trichoplusia ni virus	TnV	*Noctuidae*	USA
Unassigned possible members			
Acherontia atropas virus	AaV	*Sphingidae*	Canary Islands
Agraulis vanillae virus	AvV	*Nymphalidae*	Argentina
Callimorpha quadripuntata virus	CqV	*Arctiidae*	UK
Eucocytis meeki virus	EmV	*Noctuidae*	Papua New Guinea
Euploea corea virus	EcV	*Danaidae*	Australia/ Germany
Hypocritae jacobeae virus	HjV	*Arctiidae*	UK
Lymantria ninayi virus	LnV	*Lymantriidae*	Papua New Guinea
Nudaurelia ε virus	NεV	*Saturniidae*	South Africa

Since Hanzlik and Gordon (1997) compiled this classification the two genus names have been changed from ω-like group to Omegatetravirus and the β-like group has been changed to Betatetravirus (Mayo & Pringle, 1998). However with molecular work on tetraviruses yielding new discoveries, the tetraviruses might get more members or lose their name to a new taxon. For example *Nudaurelia* ε *virus* (Juckes, 1979), which is

morphologically similar to the tetraviruses, could be found to be organized along lines that differ from the two known genera.

(iv) SEROLOGY

The application of serological methods to the study of the small RNA viruses of insects has followed three broad approaches (Moore & Tinsley, 1982).

a) The use of antisera raised against known isolates of a virus to screen for other members of the same serotype.

b) Survey of insects for the geographical and/or temporal distribution of a known virus type.

c) Screening mammalian sera for the presence of antibodies that react with known insect viruses.

These serological applications apply significantly for the *Tetraviridae.*

a) Most known or possible tetraviruses have been assigned to the genus of betatetraviruses due to their serological relationship to at least one of a group of well-characterized β-like viruses, each member having a positive serological reaction to at least one other member of the group (reviewed from Moore, 1991; Reinganum, 1991; Hendry & Agrawal, 1993).

b) The serological interaction between the various viruses from different countries around the world was first reported by Juckes, *et al.*, (1973). He showed that there were serological relationships between different viruses with a given geographical area. In 1994, Hendry & Agrawal illustrated the serological relationships between the tetraviruses, which had been examined at that time (Table 1-2).

Table 1-2: The serological relationship among the Tetraviridae (Hendry & Agrawal, 1994).

	ANTISERUM							
	NβV	DpV	DtV	TnV	TaV	PxV	AeV	PiV
VIRUSES								
NβV	++	++	+	-	+	+	+	*
DpV	++	++	+	(+)	*	*	*	*
DtV	+	+	++	-	+	+	*	*
TnV	(+)	-	-	++	*	*	*	-
TaV	+	*	+	*	+	+	*	*
PxV	+	*	+	*	+	+	*	*
NωV	-	-	*	*	*	*	-	*
AeV	+	*	*	*	*	*	+	-
PiV	*	*	*	-	*	*	-	+

Key: ++, strong positive reaction; +, positive reaction; (+), weak reaction;

-, negative reaction; *, not tested.

For full virus names and geographical origin, see Table 1, pg.4.

c) MacCallum *et al.*. (1979) reported the presence of mammalian antibodies to a tetravirus, Darna trima virus (DtV). The antibodies to DtV were detected in sera from several species of domestic animals (including cows, sheep and pigs). The antibodies to DtV were also isolated in a small percentage of human sera obtained from both Malaysia and the U.K. These antibodies were present in both sexes in whom no associated illness was identified.

(v) VIRUS-HOST INTERACTION

In 1965, Grace and Mercer noted a high mortality amoung laboratory reared *Antheraea eucalypti* and pathological examination revealed large numbers of spherical, viruslike particles. These examinations showed that the infected larval midgut was the only organ obviously affected (Figure 1-1). This was confirmed by Greenwood and Moore (1984), who used enzyme-linked immunosorbent assay (ELISA) combined with electron microscopy (EM) to examine the tissues of the infected larvae. The EM detected virus particles in the foregut in addition to the midgut.

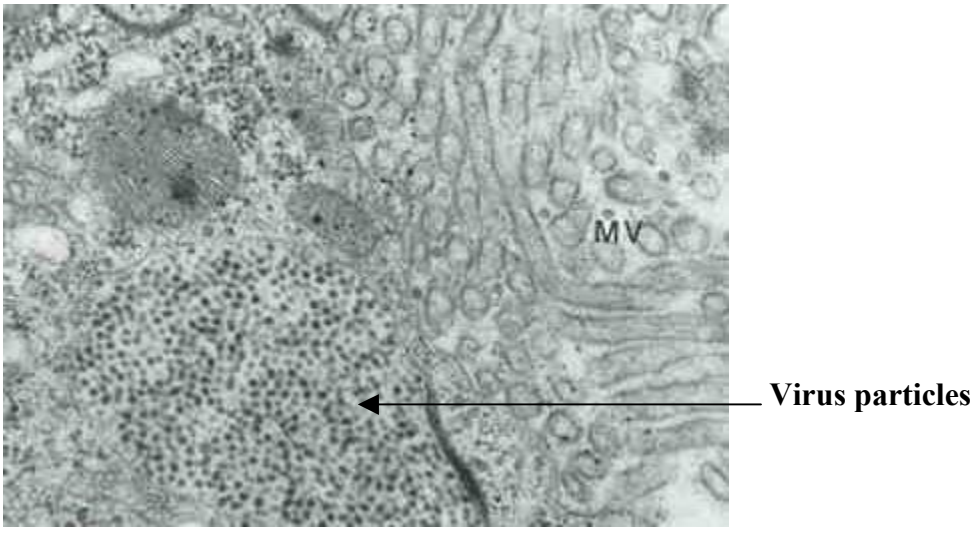


Figure 1-1: Section of midgut cell of *Trichoplusia ni* showing membrane-bound arrays of TnV in the cytoplasm. MV = micovilli (Magnification x 50000)(Reinganum, 1991).

Transmission of the virus appears to be both horizontal via ingestion and vertical from parent to progeny, although evidence strongly suggest the former. How the latter could occur in the wild is not clear, emerging larvae possibly becoming infected as they bite their way out of eggs superficially contaminated with virus (transovum transmission). Virus could also be present in insect frass, but whether this could play a role in transmission is unclear. Another possibility is transovarial transmission from infected adults (Hendry & Agrawal, 1994).

The pathological response of larvae to virus infection depends upon the particlar virus-host combination, larval age, and virus dose. Neonate larvae of *Trichoplusia ni* fed with doses of TnV supported a chronic infection where the only gross symptom of disease was a marked reduction in larval and pupal size (Vail *et al.*, 1983). Since the virus multiplies primarily in the cells of the midgut (Figure 1-1), the size reduction is most probably due to starvation. On the other hand, virus administered to late instar larvae has little or no effect and the virus can persist as an inapparent infection in succeeding generations (Reinganum, 1991).

Most other tetraviruses cause severe symptoms, usually resulting in death, in host larvae of all instars. Insects infected with AeV, NβV, NωV, DpV, PxV and DtV become moribund, usually accompanied by discolouration and flaccidity. Vomiting can occur accompanied by the voiding of loose or discolored faeces. The infected larvae often hang from branches by the last pair of prolegs or fall to the ground. The final stages of the disease are often characterized by complete liquefaction of the internal organs. Infected larvae which manage to pupate often die before reaching the adult form, or emerge as underdeveloped or malformed adults (Hendry & Agrawal, 1994).

(vi) VIRION PROPERTIES

The basic biophysical properties that characterize the *Tetraviridae* are: a particle diameter that varies between 35-40 nm; an icosahedral shape containing 240 copies of a single polypeptide protein with a mol. wt. of about 65 000 daltons and arranged with T=4 symmetry; a single-stranded RNA genome with an average mol. wt. of about $2.0 \times 10^{6.}$ The majority of tetraviruses have a monopartite genome and to date two members have a bipartite genome. The more precise physical and chemical properties of a few members of the *Tetraviridae* (Table 1-3*)* have been compiled by Hanzlik & Gordon, (1997). The properties of the other members of the *Tetraviridae* have to date not been determined.

Table 1-3: Physical & Chemical Properties of Tetraviruses (Hanzlik & Gordon, 1997)

Virus	Diameter (mn)	Density (g/ml)	Sedimentation coefficient (S)	Capsid protein (M_r)	RNA ($M_r \times 10^3$)
Genus: Betatetravirus					
Nudaurelia capensis β	39.7	1.295	210	61000 (Major = 60572 Da) (Minor = 7975 Da)	1800 (6.5 kb)
Antherea eucalypti	32	n.d.	215	n.d.	+
Trichoplusia ni	35-38	1.3	200	67000-68000	1900
Darna trima	35-38	1.289	199	62000-66000	+
Thosea asigna	35	1.275	194	60800	+
Philosamia cynthia x ricini	35	1.275	206	62400	+
Dasychira pudibunda	38	1.31	n.d.	66000	1800
Pseudoplusia includens	38	1.33	190	55000	1900
Nudaurelia capensis ε	40.1	1.285	217	61000	n.d.
Genus: Omegatetravirus					
Nudaurelia capensis ω	41.0	1.285	n.d.	Major = 62019 Da Minor = 7817 Da	RNA1=5.3kb RNA2=2.4kb
Helicoverpa armigera stunt	38	1.296	n.d.	Major = 63378 Da Minor = 7309 Da	RNA1=5.3kb RNA2=2.5kb

Key: +, contains RNA; n.d., not determined

(vii) STRUCTURE

The first example of a non-enveloped virus capsid with a T=4 icosahedral surface symmetry was reported by Finch *et al.* (1974), using electron micrographs of NβV. The micrographs showed that the icosahedral surface of the virus was constructed from 240 protein structural subunits clustered in trimers with four Y-shaped trimers per icosahedral face. Of the other members of the *Tetraviridae,* only NωV (Agrawal & Johnson, 1992) has been shown to have the same icosahedral T=4 quasi symmetry.

Olson & Baker (1990) constructed three-dimensional structures of frozen-hydrated NβV and NωV using electron microscopy and image analysis. The two virus particles (Figure 1-2) at 3.2-nm resolution showed the hallmark of T=4 symmetry, namely, 20 triangular faces, each having four trimeric aggregates. Each face has distinct structures at points of symmetry protruding approximately 4 nm above the contiguous surface. However the NβV face is distinct by being more planar and having three prominent pits.

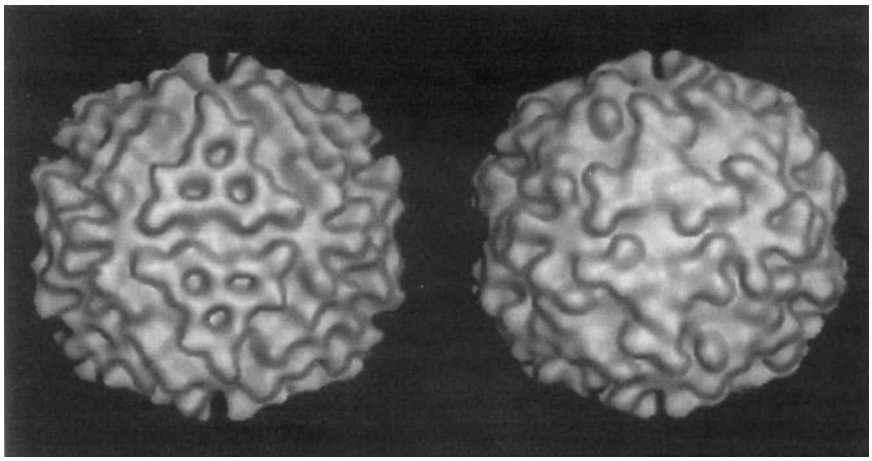


Figure1-2: Image reconstruction of frozen-hydrated *Nudaurelia* viruses analysis by E.M. The virus particle on the left is NβV and on the right NωV. (Olson & Baker, 1990)

The *Tetraviridae* are apparently the only viruses whose capsid structure is solely organized with T=4 icosahedral symmetry. The subunit packing of the T=4 contiguous capsid shell is similar to the T=3 shell (Figure 1-3). The capsid consists of multiple domains created by protein subunits. Crystallographic studies of RNA viruses that infect eukaryotic cells have shown that all capsid proteins contain an eight-stranded antiparallel

β-barrel motif that forms the contiguous shell of the particle. Although the image reconstruction gives no detail of the subunit organization in the contiguous shell portion of the NβV capsid, it has been proposed that it may be formed by β-barrels arranged according to architectural principles similar to those found in the shells of T=3 viruses (Olson *et al.*, 1990).

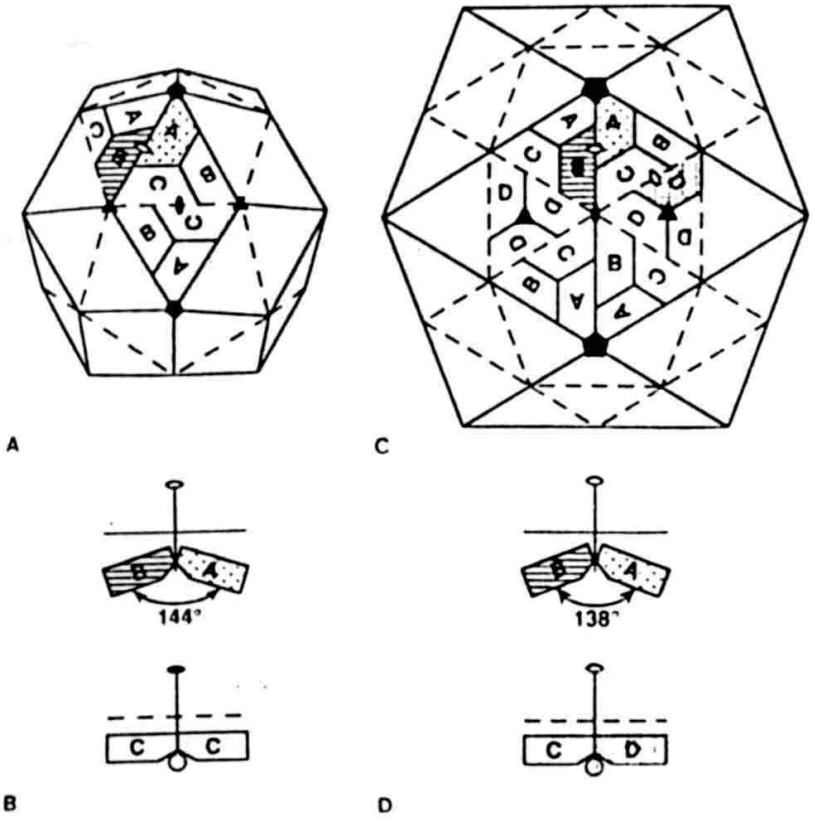


Figure 1-3: Diagrammatic representations of the subunit associations accounting for contiguous shells in T=3 (A & B) and T=4 icosahedrons (C & D) (Olson *et al.*, 1990)

The geometry of the T=3 quasi-equivalent structure is explained by examining the difference between an icosahedral twofold axis (solid ellipse), which relates two C subunits, and a quasi twofold axis (open ellipse), which relates A and B subunits. The C subunits lie in the same plane, but the A and B subunits are in planes with a dihedral angle of 144^{O} . The T=4 structure contains two types of quasi twofold axes. Like the T=3 structure, one type relates subunits (C and D) in the same plane, but the other related subunits (A and B) in different planes with a dihedral angle of 138^{O}. The contiguous shell of a hypothetical T=4 structure may be constructed with the A, B, and C subunits from

the quasi-equivalent trimer of the T=3 asymmetric unit and with the D subunit that is related to the C subunit by quasi twofold symmetry. Thus, the model asymmetric unit consists of subunits A, B, C, and D, the four required by the T=4 lattice symmetry of the tetraviral icosahedral capsid (Olson *et al.*, 1990).

(viii) MOLECULAR COMPARISON OF THE TWO TETRAVIRAL GENERA

Sequence analysis of members of the two genera of the *Tetraviridae* show they are clearly related but have significant differences in their genomes (Hanzlik & Gordon, 1997). The method of gene expression differs between the mono and bipartite genomes. The betatetravirus genus has a large 5' proximal open reading frame (ORF) that encodes a protein containing three functional domains characteristic of RNA-dependent RNA polymerase. The 3' proximal ORF, which encodes capsid precursor proteins that are translated from a sub-genomic RNA, overlaps the replicase gene by more than 99% (Figure 1-4) (Gordon *et al.*, 1999). After virus assembly, each of the 240 capsid protein precursors is cleaved into a major protein of 61 kDa and a minor protein of 8 kDa.

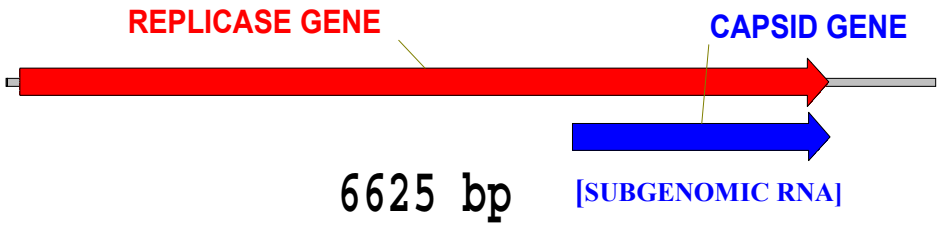


Figure 1-4: Genome organization of NβV ORFs.

On the other hand, the bipartite genomes of the omegatetraviruses translate the RNA-dependent RNA polymerase from the single ORF in RNA 1. The RNA 2 encodes two proteins with partially overlapping open reading frames. Its first AUG codon is located about 280 nucleotides from the 5' end and initiates a 157-residue protein of unknown function, whereas the next ORF initiates at about nucleotide 366 and encodes the virus capsid protein precursor (Figure 1-5). Each precursor is cleaved after assembly, like the betatetravirus, into a major protein of 62 kDa and a minor protein of 8 kDa.

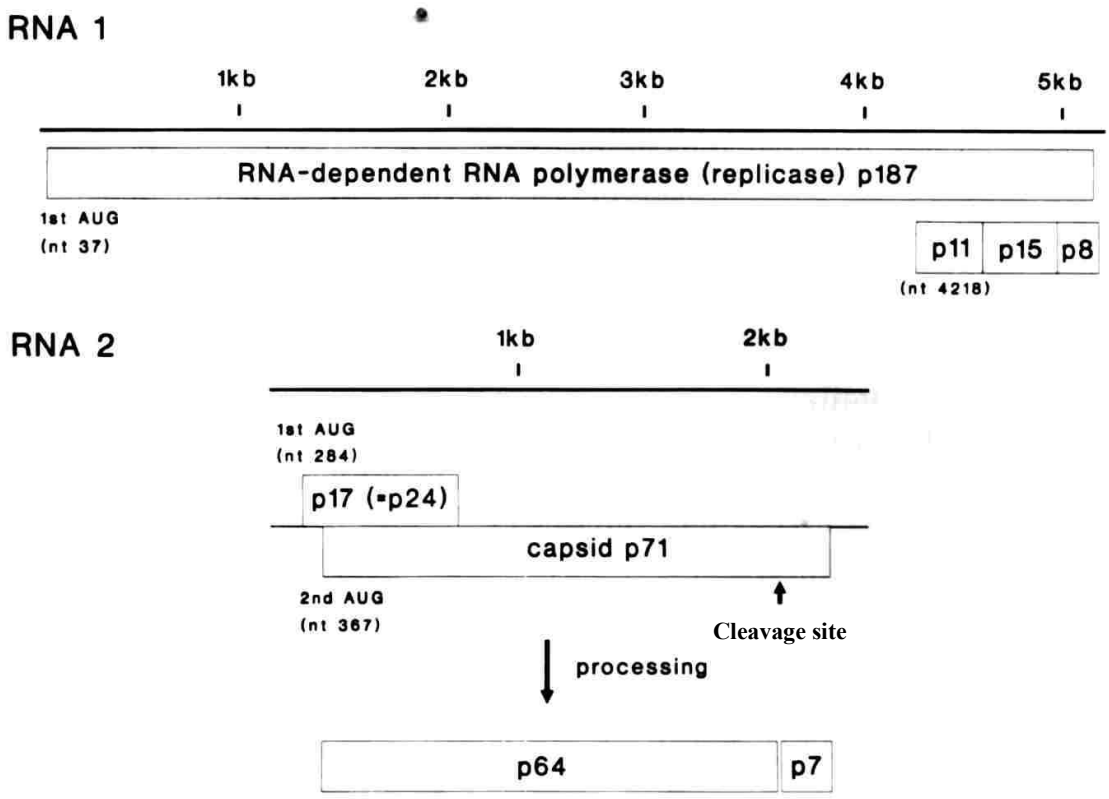


Figure 1-5: Genome organization of HaSV RNAs 1 and 2 (Hanzlik & Gordon, 1997)

Comparison of the sequences of NβV and another member of the *Tetraviridae*, *Helicoverpa armigera* stunt virus (HaSV), shows identities of 31.6% for the replicase and 24.5% for the capsid protein (Gordon *et al.*, 1999). The results clearly indicate that these two viruses from different genera are related.

The genomic RNAs of *Tetraviridae* are all likely to carry a cap structure at the 5' terminus. The 3' terminus of the genome are nonpolyadenylated and can be folded into a tRNA-like secondary structure that has a valine anticodon; the tRNA-like structure lacks a pseudoknot in the aminoacyl stem (Hanzlik and Gordon, 1997), a feature common to both genera of tetraviruses.

1.2. THE BIOLOGY OF THE PINE TREE EMPEROR MOTH, *NUDAURELIA CYTHEREA* (LEPIDOPTERA: SATURNIIDAE)

(i) DESCRIPTION OF *NUDAURELIA CYTHEREA*

The pine emperor moth is one of the more important defoliating insects attacking *Pinus radiata,* an exotic conifer grown for timber production, in the Western, Southern and Eastern Cape forest region of Southern Africa (Figure 1-6). In some localities widespread defoliation occurs annually but in others this insect is scarce and never assumes pest proportions. *Pinus radiata* was introduced on a commercial scale in the Western Cape in about 1883. By 1885 this indigenous insect had adapted itself so well to this new host that a serious outbreak occurred that year (Geertsema, 1975).

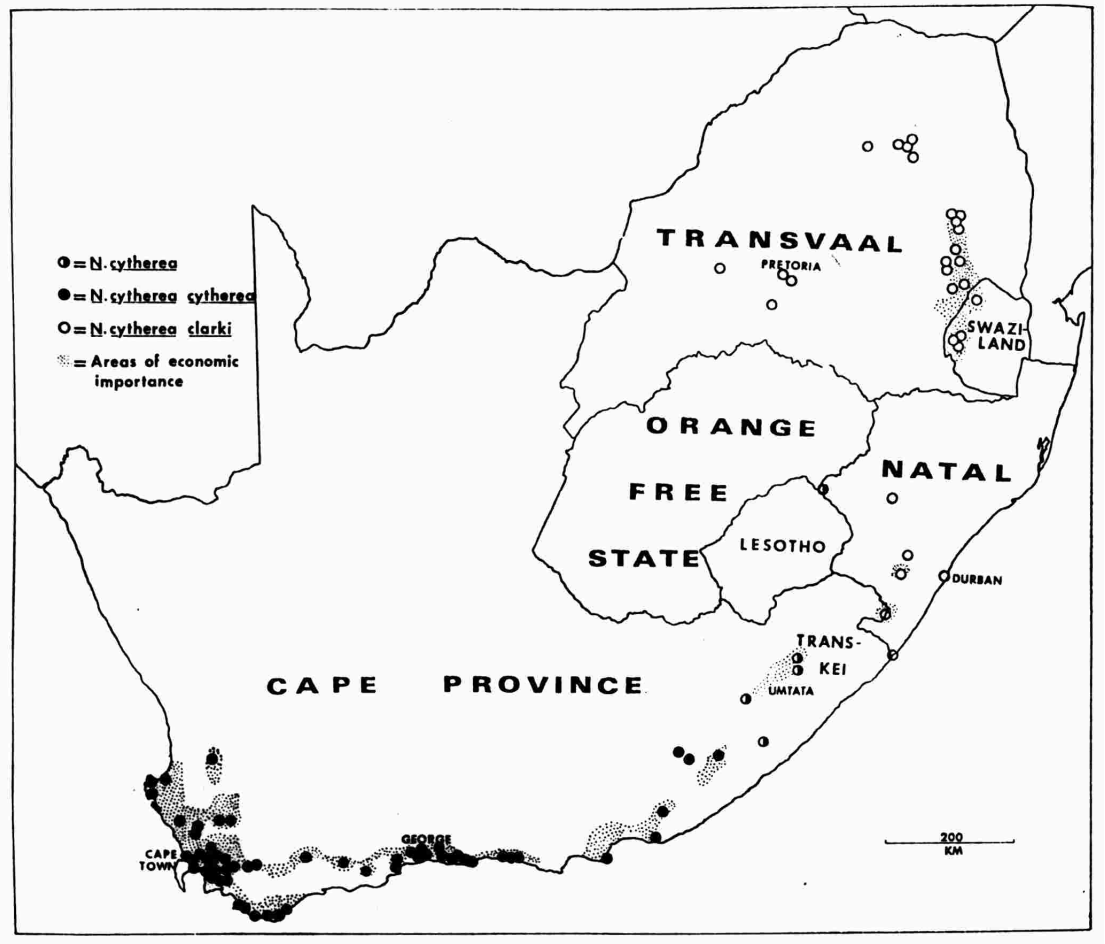


Figure 1-6: Geographical distribution in Southern Africa of *Nudaurelia cyntherea* (Fabr.) (Geertsema & Giliomee, 1972).

Later studies showed that two sub-species occurred in southern Africa, namely *Nudaurelia cytherea cytherea* (Fabr.) and *Nudaurelia cytherea clarki*. The former was previously known as *Nudaurelia cytherea capensis* (Stoll) and the latter as *Nudaurelia cytherea cytherea* (Fabr.). Tooke & Hubbard (1941) reviewed the occurrence of infestations in the Western Cape up to 1941. *Nudaurelia* was first reported near Franschhoek in 1929. In 1931 a light infestation was present on the lower mountain slopes but the infestation spread gradually until it covered 600 acres by 1937. Defoliation of the trees usually occurs from June to early December and ranges from slight to total loss of foliage. In the Western Cape this period of insect attack also covers the active growing period of the main host, *P.radiata* (Figure 1-7). Although trees seldom die as a result of defoliation, these trees nevertheless receive a setback and are more vulnerable to attack by other destructive agencies. Attacks by this insect are not confined to a single season and infestations usually remain in the same locality for some years (Geertsema, 1975).

Figure1-7: Diseased and healthy larvae of *Nudaurelia cytherea cytherea* (Fabr.) on *Pinus radiata (*Geertsema, 1970).

(ii) LIFE CYCLES

(a) The egg stage :

The eggs of *Nudaurelia cytherea cytherea* are oval to round in shape. The average length of the egg is about 2.92mm and width 2.52mm. Eggs are creamy-white in colour and covered to a varying extent with a purplish brown secretion, which cements the eggs to one another and to the surface on which the eggs are deposited (Figure 1-8). The eggs are usually laid in clusters, up to 45 eggs, on the foliage of food plants (Geertsema, 1975).

Figure 1-8: Eggs from the *Nudaurelia cytherea cytherea* (Fabr.) (Geertsema, 1970).

(b) The larval stage:

The larvae of the *N.cytherea cytherea* fall into various instars.

Instar I

Ground colour of body orange-brown when newly hatched, turning reddish brown later. Mild-segmental areas of body segments black and forming a dorsal saddle-shaped region on every segment. Head shiny-black with black-tipped white setae. Setae on prothoracic

shield mostly pointing anteriorly. Thoracic legs black to dark brown The larval length varies from about five to twelve millimetres (Figure1-9) (Geertsema, 1975).

Figure1-9: *Nudaurelia cytherea* larvae at instar I

Instar II

General ground colour reddish-brown with a ventral extension of the dorsal saddle-shaped areas incorporating the spiracles. Yellow pearl spots of various sizes and shapes present and showing up on the darker ground colour. Spiracles with some yellow pearl spots on the dorsal peritreme. Prothoracic and anal shields black to brownish-black. The size of the second instar varies from 11 to 17 mm (Geertsema, 1975).

Instar III

Ground to reddish-brown in colour. The presence of more and larger sized pearl spots on the body surface. Blueish spots present in the anterior and posterior lateral edges of segments; where yellow and blue spots intermingle, a light greenish effect is obtained. On the dorsal surface only a few pearl spots are present and the black ground colour of this area is retained. Peritreme of spiracle black with the central area orangey-yellow. The larval length varies from 17 to 24 mm (Geertsema, 1975).

Instar IV

The general reddish-brown shade of the body mostly replaced or obscured by the presence of a concentration of yellow, blue and green pearl spots on the various body segments. The darker brown or blackish shade of the mid-segmental lateral area has almost disappeared. Pearl spots are now clearly visible on the ventral surface of abdominal segments 8 and 9. General length of larva ranges from 23 to 62 mm (Geertsema, 1975).

Instar V

Ground to reddish-brown colour with intersegmental areas of light brown to yellowish-brown. Pearl spots, large and covering almost entire lateral body segment except for the segmental edges and ventral surface. General length of larva ranges from 61 to 143 mm (Figure 1-10) (Geertsema, 1975).

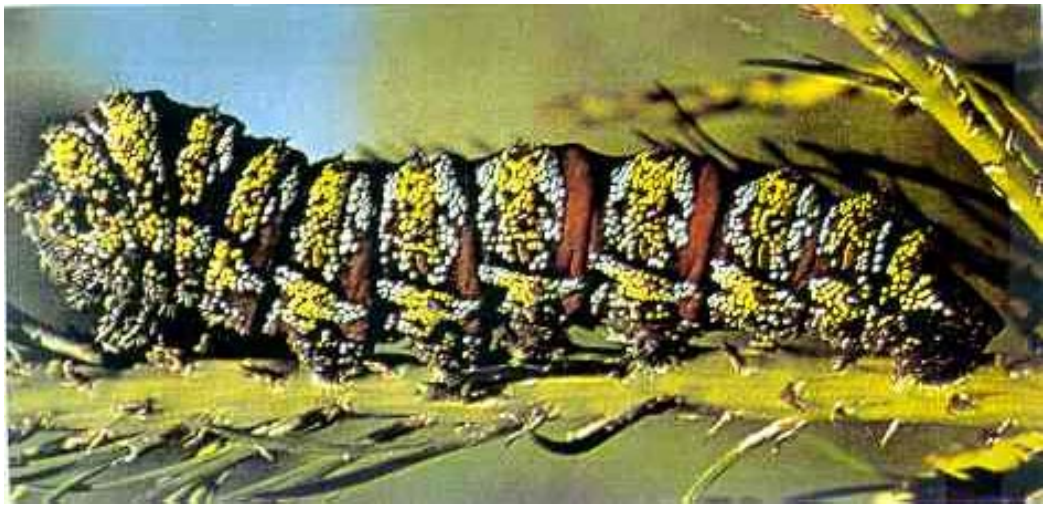


Figure 1-10: The final larval instar (V) of the *Nudaurelia cytherea*

(c) The pupal stage:

The pupa is obtect and enclosed in a pupal cell in the ground. When first formed the pupal case is soft and light yellow in colour, but the pupa turns blackish in colour (Figure 1-11) on hardening after a period ranging from three to six days. The pupal shape is cylindrical with the cephalic end rounded. Male pupae have the genital opening ventromedially on the ninth abdominal segment and this orifice is flanked by a pair of oval swellings. The genital aperture of the female pupa is present ventromedially on the anterior margins of the eighth and ninth abdominal segments. The average length of the male pupae ranges from 5.61 to 3.33 centimeters (cm) and width from 2.05 to 1.25 cm. The female average length ranges from 5.92 to 3.51 and width from 2.17 to 1.33 cm (Geertsema, 1975).

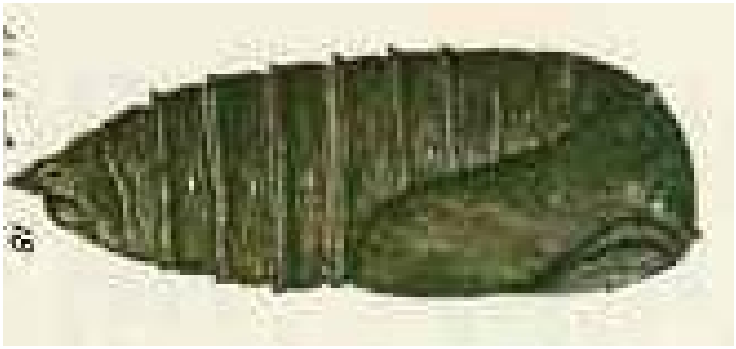


Figure 1-11: *Nudaurelia cytherea* pupa (Geertsema, 1970)

(d) The adult stage:

Adults are large and robust. The shape of the antennae easily distinguishes the sexes, those of the male being broadly pectinate, while those of the female are longer and almost lineal. Thorax of both sexes is pubescent and maroon-red to brown on underside. Abdomen is of lighter shade than general colour; the underside is darker. Tibiae and tarsi are slate black but femora maroon. Ground colour of wings is variable, ranging from slate black to brown to yellow. Heavy green or reddish suffusion sometimes present in especially the medial wing area. Fore wings of males have a slightly falcate apex; female wing apices are more rounded. Wings are divided into three distinct regions by the ante-

and postmedial lines. Ante- and postmedial lines enclose round eyespots on both wings. Eyespot consists of central hyaline area surrounded from the inside outwards by concentric yellowish, black and greyish-white rings. The average wing span for male is between 13.59 ± 1.25 cm and females 14.32 ± 0.94 cm (Geertsema, 1975).

MALE **FEMALE**

Figure 1-12. The *Nudaurelia cytherea cytherea* adult form (Geertsema, 1970)

(iii) ECOLOGICAL IMPORTANCE

The larvae of *Nudaurelia cytherea* are polyphagous, their host plants being a mixture of indigenous and exotic plants. The indigenous plants are preferred but, if found intergrowing in *Pinus radiata* (exotic) plantations, ovipositing adults prefer the exotic plants. Food shortages forces the larvae to attempt feeding on unsuitable plants. A true host plant is one on which the insect is bred to maturity. *N. cytherea* sometimes causes a great deal of damage to the forest industry and regular spraying campaigns have to be undertaken to prevent extensive defoliation (Figure 1-13).

Figure 1-13: Spraying *P.radiata* for *N. cytherea* larvae

Damage to windbreaks and smaller plantings of *P. radiata* as well as to some exotic and indigenous trees and shrubs used for the reclamation of coastal sand dunes is caused annually by the larvae. Varying degrees of defoliation occur and this leads to corresponding damage to trees, i.e. scorching of the bark, deformation of the crown and young trees, and probably loss of wood increment. In the case of continuous defoliation the trees may suffer so severely as to cause them to lose their vigour and die. Unhealthy and struggling trees are also more attractive to secondary insects. Young trees, 1-2 years of age, are rarely defoliated although eggs may be laid on them. Trees from 3 to 15 years old are most prone to attack and suffer the most injury. In the older trees defoliation still takes place but foliage loss is not so obvious due to it being too hard for the freshly emerged larvae to feed on (Geertsema & Giliomee, 1972).

1.2. NUDAURELIA β VIRUS

(i) PHYSICAL PROPERTIES

The data in table 1-3 on page 9 indicates that the capsids of the ω-like and β-like viruses are composed of two protein components. However, closer examination of the NβV with more modern techniques shows its capsid to also contain a second protein. Hanzlik & Gordon (1997) reported that the data, based on amino acid sequencing of the coat proteins of both the known ω-like viruses and nucleotide sequencing of the ORFs, showed the ω-like capsids to be composed of a 63-kDa protein with a still uncharacterised amino-terminal blockage and an unmodified, highly basic 7- to 8-kDa peptide. These two proteins were shown to be produced by cleavage of the precursor protein at a specific site between adjacent asparagine and phenylalanine residues. It was also noted that expression of the coat protein gene in insect cells using baculovirus expression showed that the cleavage occurs upon assembly of the capsid.

A similar situation existed for NβV, suggested by Agrawal and Johnson (1992) and confirmed with NβV peptide and genomic sequencing (Gordon *et al.*, 1999). The NβV capsid is composed of proteins of 60,572 Da and 7995 Da, resulting from cleavage between adjacent asparagine and glycine residues.

(ii) MOLECULAR BIOLOGY AND SEQUENCING OF NβV

The sequencing of the NβV was done using fragmented NβV RNA that was reverse transcribed into cDNA using random primers and then cloned into p*Bluescript* vectors. Sequence analysis (Gordon *et al.*, 1999) confirmed the NβVs 6625-nucleotide RNA genome (Appendix 1) to be dicistronic, encoding both the replicase at the 5' end and the capsid protein at the 3' end. The initiation codon of the ORF for the replicase gene is located at nucleotide 93-95. This ORF encodes a 1925-amino-acid protein with a calculated molecular mass of 215 kDa. The replicase gene covers 5778 nucleotides or

87% of the NβV genomic RNA (Figure 1-4), terminating at stop codon residues 5868-5870.

The start of the capsid gene, which extents from nucleotides 4039 to 5874, is located within the replicase gene. The two reading frames overlap by 1827 nucleotides, which represents virtually the whole capsid gene. The capsid gene is in the 1+ reading frame (Appendix 1) relative to that of the replicase gene.

The ORF of the capsid gene commences 2587 nucleotides from the 3' end of the NβV RNA, primer extension analysis of this section of the genome showed nucleotide 3970 to be the 5' end of the subgenomic RNA. This makes the length of the subgenomic RNA 2656 nucleotides, which is in close agreement with that determined by Northern blotting (Hanzlik & Gordon, 1997). Therefore the capsid AUG codon is located 70 nucleotides downstream of the start of the subgenomic RNA identified by 5'-RACE. Future experiments testing the *in vitro* translation of *in vitro* transcripts of cDNA clones corresponding to this portion of the virus genome will demonstrate whether the subgenomic RNA found is an efficient mRNA for the capsid protein. There are no other significant ORFs on the NβV sequence, including the 636 nucleotide stretch between the end of the capsid gene and the beginning of the tRNA-like structure at nucleotide 6514 (Gordon *et al.*, 1999).

The 3' terminal structure (Figure 1-14) of 120 nucleotides has a valine anticodon , similar to the tRNA-like secondary structures identified on the HaSV RNA. Therefore the lack of a poly (A) tail on the 3' terminus suggest this might be a prominent structural feature, with one purpose being to protect the RNA from exonucleolytic degradation (Hanzlik & Gordon, 1997).

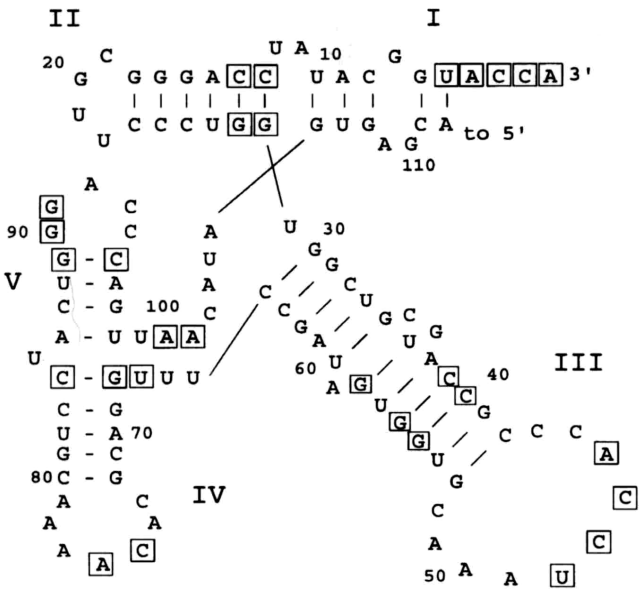


Figure 1-14: The tRNA-like secondary structure present at the 3' end of the NβV RNA. The structure is presented in an L form, as for tRNAs. Nucleotides common to the sequences and structures of both HaSV RNAs (Gordon *et al.*, 1995) and NβV are shown boxed. The numbering of each sequence is from the 3' end; the terminal residue is shown as an A.

(iii) REPLICATION OF NβV

In 1991, du Plessis *et al.* set out to identify the proteins synthesized *in vitro* using NβV genome as a message, and to determine whether any possible subgenomic mRNAs, or their double-stranded replicative forms, could be detected in infected tissue. This was the first step towards determining the mechanism by which the NβV expresses it capsid and replicase gene. Their findings suggested that a gene for a 140-kDa protein was located at the 5'end of the RNA and that the capsid protein was towards the 3' end of the genome. The virus genomic RNA would therefore be the mRNA for the nonstructural protein, which was concluded to be the virus-encoded replicase. A subgenomic RNA corresponding to the 3' region of the genomic RNA was invoked as one possible way in which the capsid gene would be translated. An alternative suggestion was that the

downstream capsid gene could be translated by internal initiation of translation on genomic RNA.

Hanzlik & Gordon (1997) showed, using Northern blotting of total virus RNA, that a 3' probe corresponding to the 3' region of the RNA genome bound to both the 6.5 kb genome and to a 2.5-kb RNA which they labelled the NβV subgenome.

Therefore, one can conclude that NβV uses its genomic RNA as a mRNA for its replicase protein, which is translated from an ORF beginning at the first initiation codon. The gene located toward the 3' end of the genome, and almost completely overlapping the replicase gene, encodes the capsid precursor by translation from a subgenomic RNA.

1.3. OBJECTIVES AND RESEARCH GOALS

(i) FULL-LENGTH cDNA CONSTRUCT OF THE *NUDAURELIA β VIRUS* GENOME

Constructing a full-length cDNA clone of the NβV genome into an expression vector would enable examination of the expression of the replicase and capsid ORFs. The sequencing of the NβV genome (Gordon *et al.*, 1999) has laid the foundation for the construction of a full-length copy of the virus genome. Cloned cDNA fragments representing the entire genome are available with specific primers for PCR and cDNA synthesis. The full-length clone could be used to study the genomic and sub-genomic transcriptional factors (mRNA template start sites and transcriptional promoters, etc.)

A major hindrance to the study of NβV in the past has been the lack of a cell culture system supporting NβV replication, which has caused NβV to be the subject of more speculation than experimentation. The RNA used in the past has been degraded which has resulted in problems when trying to synthesize long pieces of cDNA. The cDNA that has been made from the extracted RNA has averaged between 500 to 1500 bp. The overlapping pieces of cDNA have been cloned into vectors and used to sequence the

NβV genome (Gordon, *et al.*,1999), a joint project between CSIRO, Australia and Rhodes University, South Africa.

The objective of this project is to assemble the different cDNA clones (Figure 1-15) to construct a full-length template of 6625 bp. This would lead the research into cloning the full-length template into different expression vectors and examining the methods of replication. The ultimate goal would be the synthesis of infectious transcripts from the cloned full-length template.

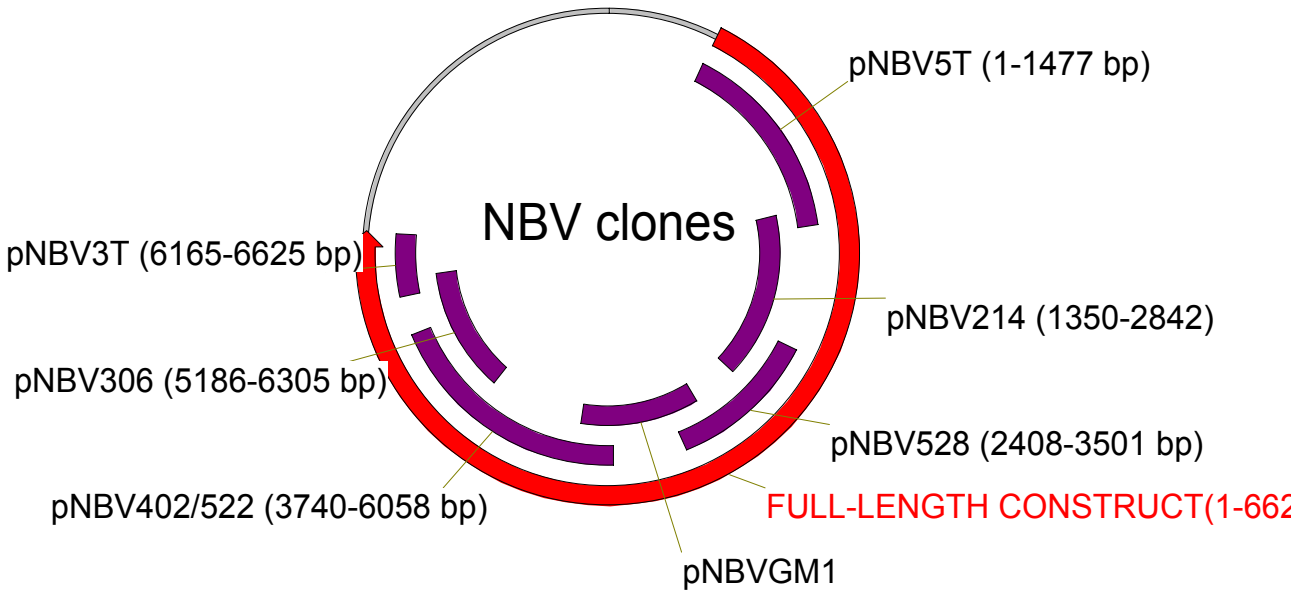


Figure 1-15: Diagram showing the different NβV clones (purple) and a (full-length construct) of the different clone (red).

(ii) EXPRESSION OF NβV CAPSID PROTEIN IN PROKARYOTES

In 1995, Agrawal & Johnson used *E.coli* to verify the expression of NωV coat protein from RNA2 before cloning it into baculovirus expression vectors.

Therefore, while constructing a full-length cDNA clone, an attempt will be made to isolate, subclone and express the capsid gene. The capsid ORF has been identified from the published sequence (Gordon *et al.*, 1999), which will assist in the isolation and cloning of the capsid gene. Successful expression of NβV coat protein will also confirm the fidelity of the assembled coat protein ORF.

The overproduction of heterologous proteins in *Esherichia coli* is often accompanied by misfolding that leads to insoluble aggregates within inclusion bodies of the cytoplasm. This leads to degradation of the protein and often-complete breakdown. The plan to overcome the insolubility of the protein and therefore increase the protein expression would be to clone the capsid gene with a fusion protein. The fusion partner could greatly improve the solubility of the passenger protein (capsid protein) that would otherwise accumulate within inclusion bodies of the cell cytoplasm. Once the capsid ORF expresses NβV precursor protein with a fusion protein and thereby confirms the authenticity of the prescribed capsid ORF, the plan would be to turn to eukaryotic systems and attempt to assemble a capsid structure with 240 copies of the cleaved post-translated precursor protein (Figure 1-16).

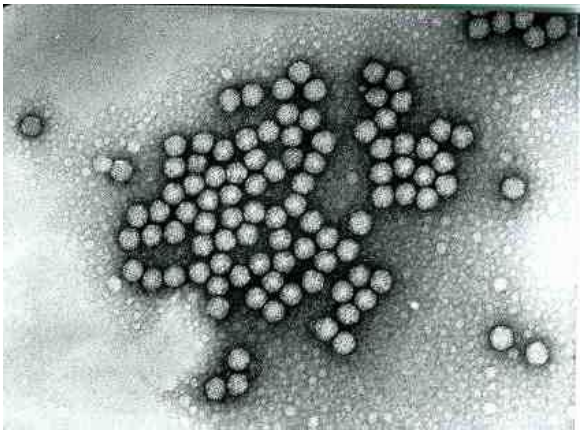


Figure 1-16: Icosahedral *Nudaurelia* β capsid structures

CHAPTER 2
Materials & Methods

2.1 MATERIALS:

The plasmids as well as most of the primers, were kindly provided by Drs K. Gordon and T. Hanzlik, CSIRO, Canberra. Additional primers were ordered from the Synthetic DNA laboratory, UCT. Plasmid diagrams of the different clones used for the research with various restriction enzymes sites are available in the appendices (constructed using InfoMax, Inc. Vector NTI Deluxe v4.0.1) as are sequences and binding sites of the different oligonucleotide primers engineered for PCR. See appendices for materials, buffers, stock solutions and media.

2.2 METHODS:

2.2.1 VIRUS EXTRACTION (Morris, 1979)

Frozen virus infected *Nudaurelia cytherea capensis* larvae collected by Prof. D.A. Hendry in November 1996, were thawed and weighed. Double the larvae weight of 0.5 M Tris, 0.01 M EDTA at pH 7.5 was added together with half the weight of CCL_4 / 0.2% Mercaptoethanol. The larvae were then homogenised on ice using the Du Pont Mini-Sorval Tissue Homogeniser and centrifuged for 15 minutes at 8000g (7500 rpm) using the Beckman centrifuge with the JA-20 rotor. This removed larger cellular debri and denatured proteins from the lysate. The supernatant that contained the virus was removed and made 8% with respect to polyethylene glycol (PEG) 6000 and 0.1M NaCl. Once the chemicals were dissolved, the solution was placed in the refrigerator (+/- 45 min.) until a precipitate formed, which was collected by centrifugation at 10000 rpm for 10 min in a Beckman JA-20 rotor. The pellet was resuspended in 0.1M TE buffer (+/- 20 ml). The solution was then spun at 10000 rpm for 10 min and the supernatant (containing the virus) placed into a fresh tube and the pellet discarded. The virus was then pelleted

through 1 ml of 30% sucrose in a Beckman L8-80M ultracentrifuge using a SW41 swinging bucket rotor at 41 000 rpm for 96 minutes. (The time required to pellet virus particles through a density gradient for swinging bucket rotors was calculated using k' factor in Appendix 2). The supernatant was discarded and the pellet resuspended in 500 µl of 0.05 M TE buffer. The resuspended mix was then placed in the fridge for +/- 2 hrs (preferably O/N) to allow the virus to dissolve. The resuspended material was then transferred to 1.5 ml Eppendorf tubes and centrifuged for 10 min at maximum speed.

Purification of virus was achieved by layering the supernatant on a discontinuous CsCl gradient (equal volumes of 60% and 30% CsCl made up with 0.05 M TE, pH 7.5, see Appendix 2) and centrifuging at 30000 rpm overnight in a swinging bucket rotor. 0.4 ml Fractions were removed with the aid of an ISCO density gradient fractionator (Flow rate 1ml/min). Absorbance readings of each fraction were done at 260 nm using the Shimadzu UV-160A spectrophotometer to determine which fraction contained the virus as well as the virus concentration (see Appendix 2).

2.2.2 RNA EXTRACTION

After CsCl centrifugation, all the fractions containing virus were pooled, diluted in TE buffer and centrifuged at 40000 rpm for 60 min. (determined by using k' factor in Appendix 2) using the Beckman L8-80M ultracentrifuge with the SW65 swinging bucket rotor. The supernatant was poured off and discarded. The pellets were resuspended in 400 µl of 0.05 M TE buffer and 50 µl removed for further analysis by protein electrophoresis. To the remaining 350 µl of resuspended virus, 35 µl of 10% SDS was added to make a final concentration of 1% SDS. 400 µl of buffer-saturated phenol was then added, vortexed for 60 seconds and then centrifuged for 2 min at maximum speed in a microfuge. The clear aqueous layer above the phenol was removed and placed in a clean 1.5 ml Eppendorf tube to which another 400 µl of buffer-saturated phenol was added followed by vortexing and centrifugation as before. The aqueous layer was then removed and placed into another 1.5 ml Eppendorf tube and extracted twice with 400 µl of chloroform: isoamyl (24:1). The final aqueous layer was removed and 0.1 volume of

3M sodium acetate was added with 2.5 volumes of 95% ethanol. The mixture was stored at -20^{O}C overnight to precipitate the RNA. The 1.5 ml Eppendorf tube was centrifuged in a microfuge for 20 minutes at 4^{O}C and the supernatant removed. The pellet was washed with 70% cold ethanol and air dried (+/- 15 min.). The pellet was then resuspended in 20 µl DEPC-treated water and stored at -20^{O}C for further RNA analysis.

2.2.3. DENATURING RNA GEL ELECTROPHORESIS

A 1% agarose/formaldehyde gel was cast (see Appendix 3). 1 µl of RNA sample was prepared by adding 2 µl of 10X MOPS, 3.5 µl of 37% formaldehyde, 10 µl of formamide and made up to 20 µl with DEPC-treated water. The mixture was heated to 65^{O}C for 15 minutes and then cooled on ice. To the 20 µl mixture, 4 µl of loading buffer (see Appendix 3) with 1 µl of 10% ethidium bromide was added. Once the gel had set, the comb was removed, and the gel was placed in 1X MOPS electrophoresis running buffer. The samples were loaded with a marker (Roche RNA Molecular Weight Marker II, see appendix 3 for size distribution) into appropriate wells and electrophoresed in a Hybaid Electro-4 gel system at approximately 100V for +/- 30 minutes. After electrophoresis was completed the gel was examined on a UV transilluminator for visualisation of RNA bands.

2.2.4 POLYACRYLAMIDE GEL ELECTROPHORESIS OF PROTEINS

Electrophoresis was carried out on denatured proteins in polyacrylamide gels to allow separation of proteins based on their molecular sizes. SDS-PAGE gels were made according to standard protocols (see appendix 2) with a 4% stacking gel and a 12% resolving gel. Samples were mixed at a ratio of 1:4 with SDS loading buffer and heated to 95^{O}C for 5 minutes to denature the proteins. Electrophoresis was carried out using the Omeg Scientific Vertical Electrophoresis Apparatus with 1X bath buffer (Appendix 2) at approximately 100V, until the desired separation had been achieved. To visualise the protein bands, the gel was immersed in Coomassie blue staining solution for 2 hours at room temperature, the staining solution was removed and the gel rinsed in destain.

2.2.5. MAKING cDNA FROM RNA

During the procedure gloves were worn at all times to prevent RNase activity. 2 µl of RNA (~ 1 µg) was added to 1 µl of reverse primer and made up to 10 µl with DEPC-treated water. The mix was incubated at 70^{O}C for 10 min, put on ice for 1 min and spun briefly. To the mixture, 4 µl of 5X 1^{st} Strand buffer was added, 2 µl of 0.1M DTT and 1µl of 10 mM dNTP's, this was mixed together by pipetting up and down and spun briefly. The mixture was then incubated at 42^{O}C for 2 min and 1µl of Gibco BRL® Superscript™ II RNase H^{-} Reverse Transcriptase enzyme was added and incubated at 42^{O}C for 60 min. The cDNA template could then be amplified by taking 2µl of the mixture and adding a forward primer upstream from the reverse primer and performing PCR amplification (see 2.2.8 below).

2.2.6. DNA AGAROSE GEL ELECTROPHORESIS

Agarose gels were made with 1X TBE buffer. 3 µl of ethidium bromide solution was added per 70 ml volume of molten agarose before pouring into the gel tray and inserting the comb. Once set, the comb was removed, the gel was submerged in 1X TBE, and DNA samples with loading buffer were placed into the wells. The samples were electrophoresed in a Hybaid Electro-4 gel system at approximately 100V for 1 hour. The DNA was viewed on a UV transilluminator and photographed and analysed using a Kodak DNA/RNA Gel Analysis System.

2.2.7. DNA EXTRACTION FROM LOW-MELTING POINT AGAROSE GELS

The restriction enzyme digest or PCR product was electrophoresed in a low-melting point (LMP) gel. The gel was placed on the UV transilluminator to visualise the specific DNA band. The fragment was excised from the LMP gel using a clean scalpel (the gel was kept on the UV box for as short a time as possible to reduce UV damage to the DNA). The fragment was purified from the LMP gel using a Macherey-Nagel Nucleospin Extract kit. The extracted fragment was quantified using a Pharmacia Genequant spectrophotometer and visualised on a DNA agarose gel.

2.2.8 POLYMERASE CHAIN REACTION

PCR was performed using a Roche Molecular Biochemicals Expand High Fidelity PCR System, which is composed of an enzyme mix containing thermostable *Taq* DNA polymerase and a *Pwo* proofreading polymerase. This system generates products with a 3' single A overhang, which is useful for PCR cloning. Two master mixes of 30 µl were prepared separately (see appendix 4) and mixed together on ice in a thin walled PCR tube and overlaid with mineral oil. The sample was placed in a Hybaid Omnigene Thermocycler and PCR performed. The specific cycle profiles for the optimum melting temperature of the primers and the time for elongation of the DNA templates are all summarised in Appendix 4. Cycles were kept to a maximum of twenty cycles to reduce the chances of mutations occuring.

2.2.9. PCR LINK OF OVERLAPPING DNA FRAGMENTS

PCR LINK is a process of linking two different fragments of DNA that have homologous overlapping termini by a primer extension process. The two fragments are added together in a PCR mix with a forward primer of the one fragment and a reverse primer of the other fragment. The first few PCR cycles denature the fragments into single strands and as the temperature drops to the annealing temperature the overlapping homologous regions anneal. When the temperature reaches elongation temperature the central overlapping region that has annealed acts as a primer and the polymerase copies ("fills in") the single-stranded sections in both directions to form a double-stranded full-length fragment, the length of the two fragments. Once the fragments are linked the forward and reverse primers added to the mix will amplify the linked fragment (see Fig. 4.11).

2.2.10. PCR PRODUCT CLONING

The Promega pGEM®-T Easy Vector System has been designed to clone PCR products with a 3' single A overhang. The pGEM®-T Easy Vector (appendix 5) has been linearised using *EcoRV* and has a single 3' T overhang inserted at the ends which greatly improves the efficiency of PCR product ligation into the plasmid by creating a semi-cohesive site. The protocol of ligating PCR product to pGEM®-T Easy Vector is given in

appendix 5. The recombinant clones inactivate the α-peptide product of the lacZ gene and can therefore be identified using blue/white colour screening.

2.2.11. BLUNT-ENDING OF FRAGMENTS WITH AN OVERHANG

The fragment was cut from the plasmid vector by using *BamHI* (Roche Molecular Biochemicals) restriction enzyme digestion (see Appendix 5), which generated cohesive ends. The cut fragment was gel purified to remove the vector and the cohesive ends were then filled in with dNTP's using the Roche Molecular Biochemicals T4 DNA Polymerase method (see Appendix 5). The blunt-ended fragment was gel purified again to remove any dNTP's and T4 DNA polymerase. The fragment was then ready to be cloned into any blunt-cutter restriction site.

2.2.12. BLUNT-END CLONING INTO PLASMID VECTORS

The vector was linearised using a *SmaI* (Promega) that is a blunt-end cutter. The ends were then dephosphorylated using Shrimp Alkaline Phosphatase (Amersham Life Science) enzyme (see protocol in Appendix 5) to prevent re-ligation of the vector without the insert during the ligation process. The linearised vector was gel purified to remove any phosphatase enzyme that might inhibit the ligation process. The insert and the vector were quantified using a Pharmacia Genequant and visualised on an agarose gel. The ratio of insert to vector as well as the ligation mix protocol used for blunt-ended cloning is given in Appendix 5. After the ligation incubation period, the cloned plasmid was transformed into competent *E.coli* cells. (To monitor the ligation efficiency, linearised vector DNA both phosphorylated and dephosphorylated was subjected to ligation and transformed.)

2.2.13. ELECTRO-TRANSFORMATION OF ELECTROCOMPETENT *E.COLI*

The ligation buffer salts needed to be diluted before electroporation; therefore 1 µl of ligation reaction was added to 5 µl of distilled water. The mixture was incubated at $68^{O}C$ for 1 min, vortexed briefly and cooled to room temperature; this inactivated the ligase and improved the transformation efficiency. The Stratagene Cloning System Electroporation 1000 was set at 1800V. Sufficient 40 µl aliquots of electrocompetent

cells (see Appendix 5 for preparation method) stored at -80^{O}C were thawed, and 80 µl sterile distilled water was added. 40 µl of the diluted cells were used per electroporation (each frozen aliquot was sufficient for three transformations).

At the electroporator, 1.5 µl of each ligation mix was placed into separate 1.5 ml Eppendorf tubes with 40 µl of the diluted electrocompetent *E.coli* cells and mixed by pipetting up and down. The entire mix was transferred to a cold Stratagene electroporation cuvette, tapped to ensure that the liquid had reached the bottom, and wiped with a tissue to remove any condensation (the moisture can cause arcing at the cuvette). The cuvette was then placed in a cold cuvette holder, and inserted into the electroporator. The "Pulse" button was pressed twice and released when the buzzer sounded.

The cuvette was quickly removed and 200 µl of SOC medium (Appendix 5) added to the cuvette and mixed well. The transformed cells in the SOC medium were returned to their original 1.5 ml Eppendorf tubes and incubated at 37^{O}C for 30 minutes to allow for the cells to recover. The cells in the SOC medium were then plated out onto Luria Agar (LA) + antibiotic plates and incubated overnight at 37^{O}C. (To monitor the transformation efficiency, competent *E.coli* cells were transformed with 1 pg of supercoiled plasmid and from the number of colonies that grew the transformation efficiency could be calculated, see Appendix 5 for calculation).

2.2.14. PLASMID ISOLATION: ALKALINE-LYSIS (Sambrook, 1989)

Separated colonies from the transformation plate were picked off with sterile toothpicks and inoculated in different 5 ml Luria broth (LB) + antibiotic broth tubes and incubated overnight with shaking at 37^{O}C. 1.5 ml of each culture was transferred to 1.5ml Eppendorf tubes and centrifuged for one minute. The medium was removed and a further 1.5 ml of culture added and centrifuged for 1 min. The medium was then removed again leaving the pellets as dry as possible.

The pellets were resuspended in 100 µl of ice-cold lysozyme solution (Appendix 5) by vortexing and incubated at room temperature for five minutes. 200 µl of NaOH/SDS (Appendix 5) was added and mixed by inverting the tubes two to three times (without vortexing). The mixture was then stored on ice for five minutes. 150 µl of acetate neutralizing solution (Appendix 5) was added and vortexed gently in an inverted position for 10 seconds. The mixture was stored on ice for five minutes. The tubes were centrifuged at 15000rpm for five minutes at 4^{O}C.

The supernatant was removed into clean tubes and mixed by vortexing with equal volumes of phenol/IAC. This was centrifuged and the upper phase removed into clean 1.5 ml Eppendorf tubes. The upper phase was then mixed with equal volumes of IAC only and centrifuged to separate the phases. The upper phase was then removed again into clean 1.5 ml Eppendorf tubes.

Precipitation of the DNA was done by mixing twice the volume of 90% ethanol to the upper phase and vortexing for 10 seconds. The tubes were then stored at -20^{O}C for two hours. The precipitate was collected by centrifugation at 13000 g for thirty minutes. The supernatant was discarded and the DNA pellets washed with 70% ethanol. The ethanol was removed completely and the pellet allowed to air dry for about fifteen minutes. The dry DNA pellets were dissolved in 40 µl TE/RNase (Appendix 5) and stored at -20^{O}C.

2.2.15. PLASMID ISOLATION: "SMART PREPS" (Berghammer, 1993)

Separated colonies from the transformation plate were picked off with sterile toothpicks and inoculated into different 5 ml LB + antibiotic broth tubes and incubated overnight at 37^{O}C. 1.5 ml of each culture was transferred to 1.5 ml Eppendorf tubes and centrifuged for one minute. The medium was removed and a further 1.5 ml of culture added and centrifuged for one minute. The medium was then removed again leaving the pellet as dry as possible.

The pellets were resuspended in 50µl Smart Buffer (see Appendix 5) and incubated at room temperature for 30 minutes. The tubes were then incubated at 100^{O}C for 90 seconds

and transferred to ice for five minutes. The cellular debris, protein and chromosomal DNA were pelleted by centrifugation at 13000 g for ten minutes. The supernatant containing the plasmid DNA of the different cultures was removed into clean 1.5 ml Eppendorf tubes and stored at -20^{O}C.

2.2.16. PROTEIN EXPRESSION BY INDUCTION

A 25 ml overnight culture of transformed *E.coli* was transferred into a fresh 225ml LB + antibiotic broth flask. It was replaced on the shaker at 37^{O}C until the OD_{600} reached between 0.6 - 1.0. At this stage a 1ml sample was taken and labelled as the uninduced sample. The culture was then induced with 250µl of 1M IPTG. After each hour the OD_{600} and a 1ml sample was taken and labelled. The cells were then spun down and resuspended in 50µl phosphate buffered saline for every 0.5 absorbance unit; the calculation to determine final PBS resuspension volume went as follows:

$$\frac{A_{600}}{0.5} \times 50\mu l = \text{Final resuspension volume}$$

Once the samples were resuspended in PBS, 20 µl was taken for protein gel electrophoresis and the rest stored at -20^{O}C.

2.2.17. PURIFICATION OF EXPRESSED PROTEIN

A 25ml overnight culture was transferred into a fresh 225 ml LB + antibiotic broth flask. It was replaced on the shaker at 37^{O}C until the OD_{600} reached between 0.6 - 1.0.
The culture was induced with 250 µl of IPTG and placed on the shaker until optimal expression was reached (the time was calculated from the results in 2.2.16). The culture was then spun down (JA14 rotor) at 5000 rpm for 10 min. The supernatant was removed and the pellet resuspended in 2.5 ml PBS with 2.5 µl 1M PMSF (protease inhibitor). The resuspended cells were aliquoted into three 1.5 ml Eppendorf tubes. The tubes were placed on ice and sonicated three times for 30 second intervals with 1min break between intervals. The tubes were then spun for 20 min at 13000 g. The supernatant was removed and pooled together in a 15 ml tube.

Then 250 µl of 50% slurry Glutathione Beads (Roche Molecular Biochemicals) was added to the pooled sample and placed on a rocker at $4^{O}C$ for 45 min. The beads were then spun down at the lowest possible speed and the supernatant removed. The beads were then resuspended in 15 ml PBS and mixed by hand to wash the beads (this process was repeated twice). In the final wash, once the beads had been spun down, the supernatant was removed and the beads resuspended in 250 µl elution buffer. The elution buffer removed the expressed protein linked to the GST fusion protein that had bonded to the Glutathione beads. The supernatant was aliquoted into 50 µl samples in 1.5 ml Eppendorf tubes.

2.2.18. WESTERN BLOT USING CHEMILUMINESCENCE

The proteins were resolved on a SDS-PAGE gel as described in section 2.2.4 of Methods. The gel was removed from the glass plates and rinsed in transfer buffer. The gel was then stacked in a Biorad Western Transfer Chamber Cassette as described in the transfer set up in Appendix 7. During the transfer the buffer was kept as cold as possible to improve the transfer of protein to the nitrocellulose membrane. After the transfer was complete, the nitrocellulose membrane was stained with Ponceau S stain for 2 min and destained with distilled water to visualise the protein. It was then washed twice with TBS.

The membrane was then transferred to 5% block (Appendix 7) for 1hour with shaking at room temperature or overnight at $4^{O}C$. It was then transferred for 1 hour with shaking to the primary antibody supplied by Prof. D.A. Hendry (anti-NβV serum made in rabbit) that was diluted in 5% block (1 in 500 dilution). After incubation the membrane was given two 20 min washes in TBST and then transferred to 5% block for 20 min with shaking (this was repeated). Then it was transferred for 30 min with shaking to the secondary antibody, supplied in the Roche Molecular Biochemicals Chemiluminescence Western Blotting Kit (anti-rabbit/mouse), which was diluted in 5% block (1 µl in 12.5 µl 5% block). The membrane was then given four 15 min washes with TBST. During the washes, the chemiluminescence detection reagent supplied in the Roche Molecular Biochemicals Chemiluminescence Western Blotting Kit was prepared (50µl solution B

added to 5 ml solution A at room temperature) and left to stand for 30 min. The detection reagent was then added to the membrane blot and sealed in a clear plastic bag. The blot was then exposed to x-ray film in a Kodak X-ray Film Cassette for 10 to 60 seconds. The exposure could be repeated for a longer time with another piece of film. The film was then developed in the dark room.

CHAPTER 3

Purification of NβV from *Nudaurelia Cytherea Capensis* Larvae

3.1 INTRODUCTION

The isolation and purification of NβV from infected larvae is achieved by a series of centrifugation steps in which the virus is sedimented through 30% sucrose and finally purified using a caesium chloride (CsCl) density gradient. The use of isopycnic or equilibrium density gradient centrifugation allows the separation of particles based on their differences in densities. This method relies strictly on the different buoyant densities of virus, proteins, nucleic acids, and cellular material. Each type of molecule sediments down the density gradient to the level at which its density equals that of the gradient. With centrifugation, the NβV was localised as a sharp band in the CsCl gradient at a density of 1.295 g/ml (Hendry *et al*., 1985). At this point fractions of the caesium chloride gradient were taken and quantified for virus particles using UV spectrophotometry.

The collected fractions from the CsCl density gradient were analysed by SDS polyacrylamide gel electrophoresis (PAGE) to confirm that the sample taken contained NβV protein molecular weight, 61 kDa.

3.2 RESULTS

(i) Caesium Chloride Density Centrifugation

After 16 hours of CsCl density gradient centrifugation bands could be visualised within the gradient by placing the gradient above a light in a dark room. The virus material had resolved into two major bands and three minor bands (see figure 3-1). The two major bands were expected to be NωV at the top and NβV just below it, as the omega virus which has a density of 1.285 g/ml, is slightly less dense than the beta virus, which has a

density of 1.295 g/ml (Hendry *et al.*, 1985). Three minor bands, were observed at higher density in the gradient than the two main viral bands, which have been theorised to be omega and beta virus particles that have been altered structurally by the pH and CsCl salt concentrations. It has been proposed that the Cs^+ ions penetrate a proportion of the NωV and NβV particles thus making them more dense. (Hendry, *et al*., 1985). Alternatively, they could represent particles with higher RNA content.

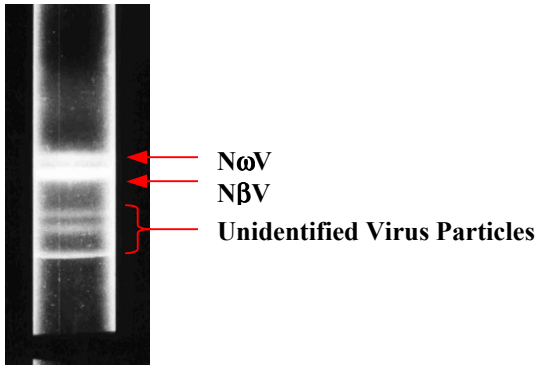


Figure 3-1: Purified virus bands in a CsCl density gradient

(ii) Analysis by Ultra Violet Absorption

From the CsCl density gradient tube, the fractionator removed 0.4 ml samples into 29 Eppendorf tubes. The optical density of each tube was read at 260 nm in order to determine the virus concentration. The UV spectrophotometer readings were plotted on a graph versus their fraction numbers to determine which tubes contained the highest concentration of viral particles. (Figure 3-2). Tubes 13 to 16 were found to have high concentrations of virus particles and were therefore pooled as sample. This sample which contained the two major bands at the lower density and hence was a mixture of omega and beta virus. Tubes 17 to 19 indicated a slightly lower concentration of virus particles and were pooled as sample 2 which contained the three minor bands.

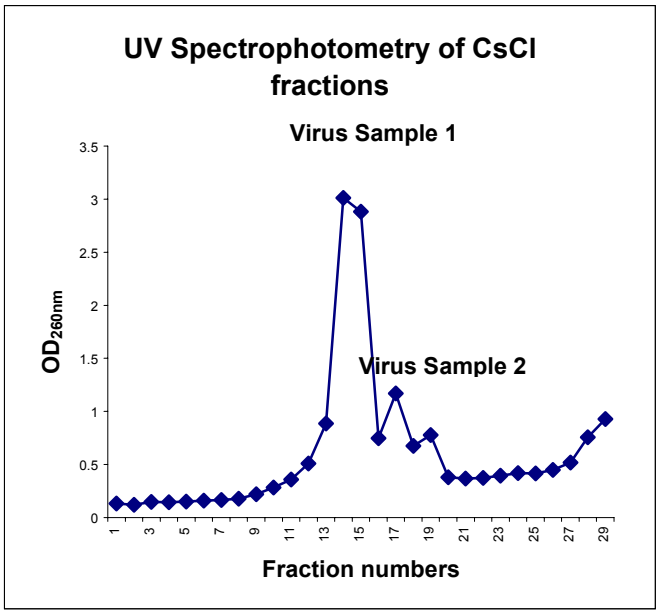


Figure 3-2: Virus fractions from a CsCl gradient plotted against the OD_{260nm} readings

Sample 1 was used for RNA extraction. Further protein gel analysis of sample 1 confirmed that the two major fractions in the density gradient that were pooled together were NβV and NωV, respectively.

(iii) Protein Gel Analysis

After the acrylamide gel had been destained, two clear bands of the correct molecular weight could be seen in lanes 2 and 3 (figure 3-3). The top band was thought to be the capsid protein of NωV, as it appeared to have a molecular weight of 62 kDa (Agrawal & Johnson, 1992) and the stronger lower band the NβV, as it appeared to have a molecular weight of 61 kDa (Struthers & Hendry, 1974). Although the molecular weight marker was not very clear, the bands were visible enough to determine the molecular weights of the two bands in lanes 2 and 3. Lane 2 was loaded with 4 µl of sample 1 and lane 3 with 8 µl loaded. From the PAGE analysis, it could also be concluded that there was a higher concentration of beta virus present in the sample than omega virus as the β-virus band was much stronger than the ω-virus band. This agrees with the relative band intensities seen in Fig. 3.1.

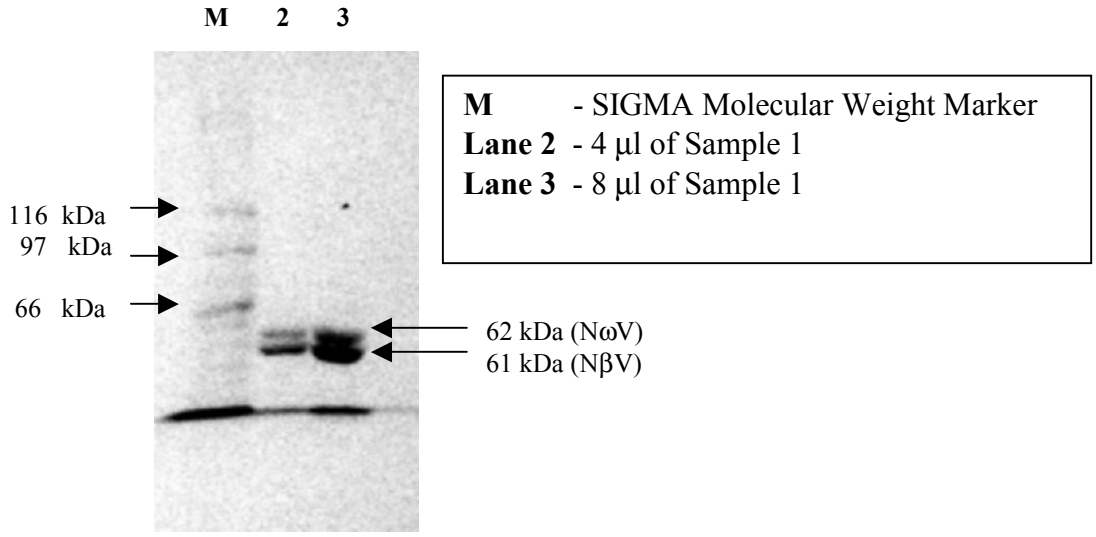


Figure 3-3: 8% SDS PAGE gel of Sample 1

3.3 DISCUSSION

From the results one could conclude that the purification of virus particles from *Nudaurelia cytherea capensis* larvae resulted in the extraction of at least two different viruses. The CsCl density gradient ultracentrifugation resulted in two major bands of NωV and NβV, respectively. The minor bands at higher density could only be explained as a result of the penetration of Cs^+ ions into a proportion of the viral particles, which must have occurred during the CsCl purification procedure causing some of the viral particles to become more dense than others (Hendry, *et al*., 1985). [see 3.2 (i)] Alternatively the denser bands could represent particles with increased RNA content.

These results could be confirmed by taking smaller fractions from the density gradient tube, thereby isolating each band from the CsCl gradient into separate Eppendorf tubes. Acrylamide gel electrophoresis followed by Western blotting using both beta and omega antibodies would indicate the type of virus in each band. The presence of beta or omega virus would indicate that there had been some alteration that caused an increase in the density of the virus particle. Alternatively, nucleic acid probes specific for either beta or omega RNA could be used to identify the RNA in the heavier fractions.

The UV absorbance results of the fractions demonstrated that the bands in the density gradient had a high concentration of virus particles. PAGE analysis of the fractions showed that the two major bands were two different species of virus, NωV and NβV, which were co-infecting the *Nudaurelia cytherea capensis* larvae, as demonstrated for the first time by Hendry and others in 1985. The other possibility could be that different species of viruses were infecting different insects and once the insects were pooled together the result demonstrated two major bands of different viruses.

The aim from here was to extract RNA from the virus particles and use primers to construct a full-length cDNA copy of the NβV genome for cloning into a plasmid vector. The fact that NωV RNA would be present after the extraction created no problem because the primers used were designed for NβV specifically and the presence of small amounts of omega RNA would not have made a difference in the construction of NβV cDNA.

CHAPTER 4

Constructing a Full-Length cDNA Clone of the NβV Genome

SECTION 4.1: Extraction of RNA for cDNA Synthesis

4.1.1 INTRODUCTION

Little is known about the replication of the NβV owing, in large, to the lack of a suitable permissive cell culture systems. As a result, in order to study the genes that encode the different viral proteins, the isolation of a full-length copy of the virus genome is required. As the degradation of RNA in the virus capsids over time is fairly rapid in frozen material, it is essential that freshly collected infected larvae be used in order to increase the chance of extracting an intact full-length single strand of RNA. The larvae that were used for the isolation and purification of virus particles were collected in November 1996 by Prof. D.A. Hendry. For this reason, the chances of obtaining a full-length single strand of RNA were diminished.

The virus extraction from infected *Nudaurelia cytherea capensis* larvae in the previous chapter yielded a high concentration of purified virus particles. As a result, attempts were made to extract intact full-length RNA from these purified virus particles. The first step was to extract RNA from the isolated virus particles (sample 1). This resulted in a mixture of virus RNA genomes (i.e. NβV and NωV). The RNA was then visualised in a denaturing agarose gel to determine its concentration and quality. Attempts were then made to synthesize cDNA copies of the full-length NβV RNA using a reverse transcriptase enzyme with a NβV 3' terminus primer (NBV3TRZ, see Appendix 4) which would not recognise the NωV sequence.

The cDNA template constructed using the reverse transcriptase procedure would be used to make double stranded DNA with the aid of a second forward primer specific to the 5'-

end through polymerase chain reaction (PCR). The PCR procedure would amplify the targeted sequence of the double stranded DNA with the aid of a set of forward and reverse oligonucleotide primers and DNA polymerase. PCR amplification allows for the visualization of the targeted sequence in an agarose gel.

4.1.2 RESULTS

(i) RNA Extraction

The RNA that was extracted from the virus capsids was of low quality (Fig. 4-1). The chance of extracting a full-length template of the NβV genome from this was poor. The band (lane 1) produced in the 1% agarose/formaldehyde gel was very faint and difficult to size against the RNA marker, which (as is apparent on the gel) had become somewhat degraded. The size distribution of the RNA marker bands should have gone from 7.4 to 1.6 kb (see Appendix 3), therefore the top of the RNA smudge in the marker lane could be said to be about 7.4 kb. Therefore the faint band in lane 1 was estimated to be about 6.6 kb in size, the size of the NβV genome.

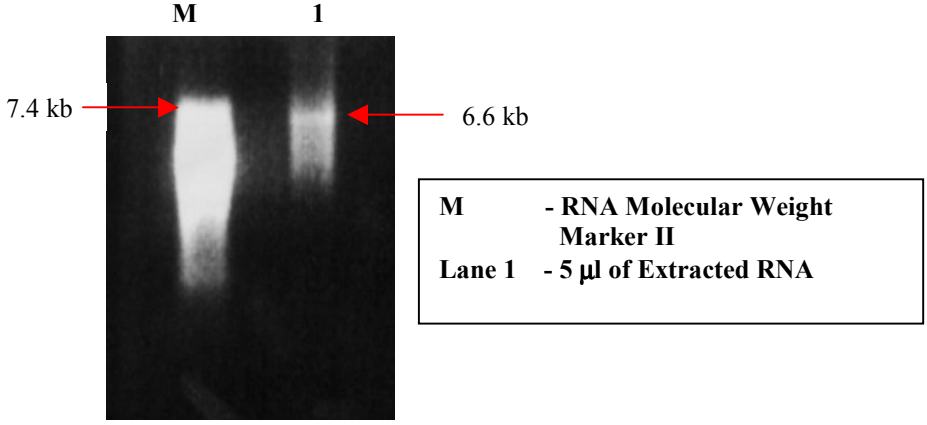
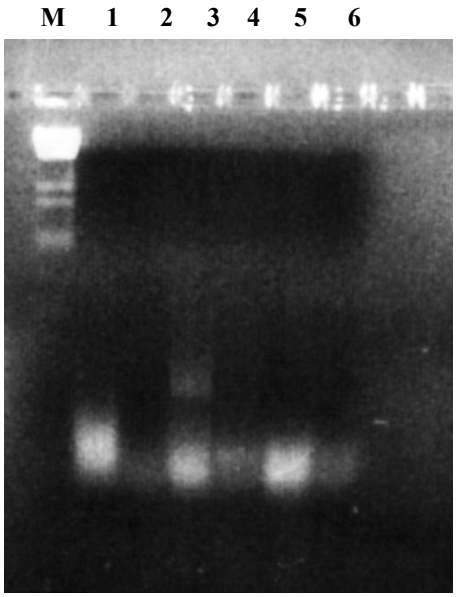


Figure 4-1: Extracted viral RNA run on a denaturing agarose gel.

The appearance of a faint band at about 6.6 kb suggested that a low concentration of full-length RNA single-strands were present and therefore attempts were made to synthesize cDNA from the extracted RNA.

(ii) cDNA Synthesis and Amplification by PCR

Attempts to construct a full-length cDNA copy of the NβV from the extracted RNA were not successful. PCR primers from different areas of the genome were also used which resulted in no amplification of any product (Fig. 4-2). This confirmed that there was no synthesis of cDNA from the extracted RNA product.



M - λ/BstE II marker
Lanes 1 & 2 - PCR reaction with NBV5T & NBV3TRZ primers (Full-length)
Lanes 3 & 4 - PCR reaction with NBV1408F & NBV4050R primers (2648 bp)
Lanes 5 & 6 - PCR reaction with NBV5T & NBV4050R primers (4048 bp)

Note: λ/BstE II marker fragment sizes are available in Appendix 4, p.106

Figure 4-2: 1% Agarose gel of attempted PCR amplification from the cDNA template synthesised from extracted RNA

4.1.3 DISCUSSION

It is clear from the above results that the RNA extracted from the virus particles was degraded. The faint band of RNA at 6.6 kb failed to synthesize a full-length construct of cDNA using reverse transcription.

Virologists have encountered an ongoing problem with RNA quality when working with virus particles extracted from *Nudaurelia cytherea capensis* (D. Hendry pers. comm.). The RNA that is extracted is very degraded and difficult to work with. Researchers have however been able to synthesize cDNA from short sections of the extracted RNA and clone the cDNA into plasmid vectors. The whole NβV genome is represented by seven

overlapping sections (Fig. 1-14) that have been cloned into vectors (Appendix 6). These sections were used by Gordon *et al.* (1999) to sequence the NβV genome. This left the option of trying to assemble the NβV genome by linking the overlapping clones. The method that was attempted to link the genome fragments consisted of progressively connecting two sections together by sub-cloning with restriction enzyme sites in the overlapping region.

SECTION 4.2. Linking the NβV Genome by Sub-Cloning in Restriction Enzyme Sites

4.2.1 INTRODUCTION

In 1997, D. Hendry started the project of linking the genome together by sub-cloning. Sub-cloning at restriction enzyme sites in the overlapping regions is a method used to link different sections of DNA together. The drawback of using restriction enzymes to link pieces together is that the restriction site of the enzyme used must only be present in the overlapping region. This was a major problem as the NβV genome has been found to have many common restriction sites, making sub-cloning difficult (Fig. 4-3).

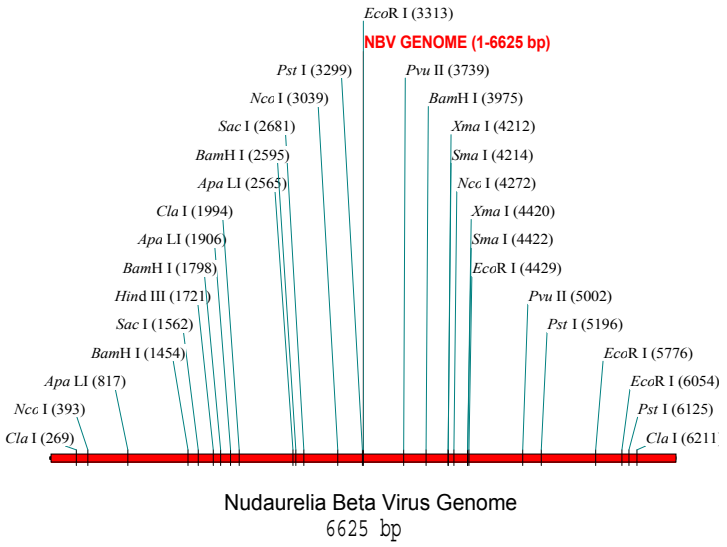


Figure 4-3: Diagram demonstrating a few recurring restriction enzyme sites in the NβV genome.

Once a restriction site had been used to link two sections together, the problem of not being able to use that enzyme in another section limited the number of enzymes that could be used.

At this stage the problem with linking the different sections together was that the available clones did not cover the entire genome. In the middle region of the genome, from position 3505 to 3739, the plasmids used for sequencing could not be traced (Fig. 4-4). This was a vital area of the genome, without which it would be impossible to link the different sections of the genome together to create a full-length construct of the NβV genome.

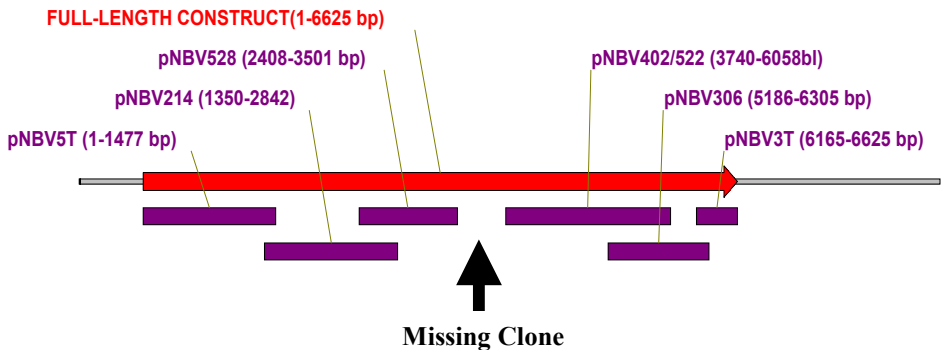


Figure 4-4: The NβV clones available.

Reverse transcription needed to be done on the extracted RNA in order to synthesize cDNA of this missing section. This work was done in conjunction with a student completing her B.Sc Honours degree, Michelle Bowker. The aim was to utilise the PCR reaction to synthesize the missing section. Once the product was retrieved it would be cloned into a PCR cloning vector, pGEM-T Easy Vector, and stored at -80^{O}C.

In order to link the three upstream clones (see Fig. 4-4), plasmid maps were examined in order to determine how to remove the NBV5T insert from pFASTBAC1 (Appendix 5) and clone it into pBluescript KS (pBSKS) (Appendix 5). This would make it possible to use the multiple cloning sites for subsequent subcloning of the other inserts that had also been cloned into pBSKS. Once this had been achieved, NBV214 (Appendix 6) would be excised from its plasmid with restriction enzymes and sub-cloned into the new

NBV5T/pBSKS vector. This would link NBV5T (Appendix 6) to NBV214 creating a cloned insert of 2.5 kb. The plan was then to sub-clone NBV528 (Appendix 6) into the new 2.5 kb insert and create a construct containing the 5' 3500 bases of the NβV genome within pBSKS (Fig. 4-5).

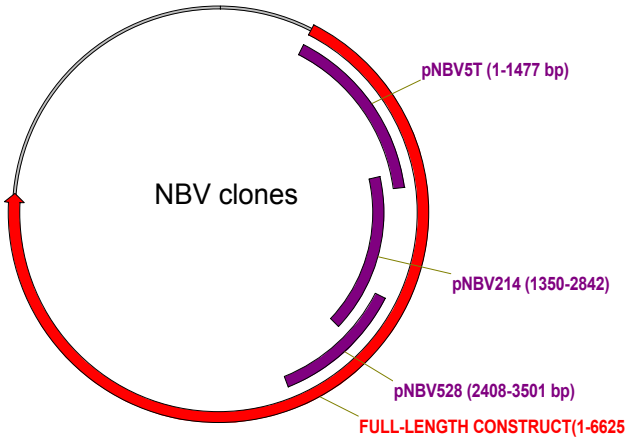


Figure 4-5: Diagram representing the sub-cloning plan to assemble the 5' half of the NβV genome.

4.2.2. RESULTS

(i) Synthesis of Plasmid GM1

Michelle Bowker synthesized a cDNA template using reverse primer NBVCAP1 (see Appendix 4), situated 4317 bp into the genome. The template was then added to a PCR mix containing a forward primer, NBVREP2 (see Appendix 4). After a number of cycles, the PCR mix was examined in a 1% agarose gel (Fig. 4-6) for the amplification of a band of the correct size (1069 bp).

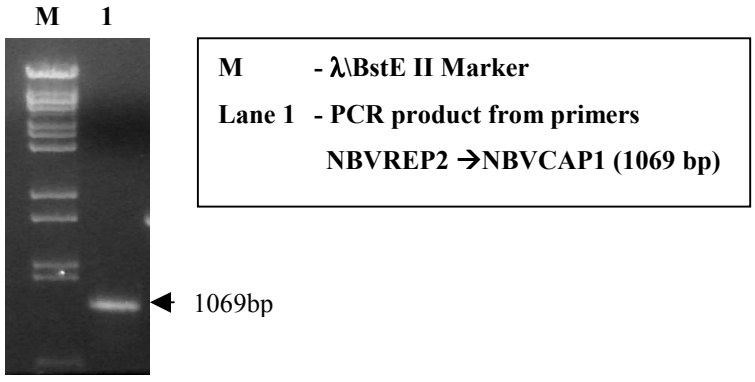


Figure 4-6: PCR product amplified from cDNA synthesized from NβV RNA

The PCR product was excised from the LMP agarose gel, purified, and then cloned into pGEM-T Easy Vector and transformed into DH5α. The plasmid isolation preps were then checked by digestion with *Pst I* (Roche Molecular Biochemicals) restriction enzyme. PCR fragments that had been cloned in the correct orientation would yield two bands of 3041 & 1047 bp, whereas a fragment in the incorrect orientation would yield two bands of 4013 & 75 bp (Fig. 4-7). After confirming which cloned plasmids were in the correct orientation, stock cultures were set up and stored at -80^{O}C and the plasmid labelled pGEM-GM1. Futher restriction digests were done on pGEM-GM1 in order to check that the cloned insert had not developed any PCR mutations during the amplification procedure and also that the insert matched the published sequence taken from RNA extracted from an earlier *Nudaurelia cytherea capensis* larvae isolate. All the restriction digests used gave the expected results.

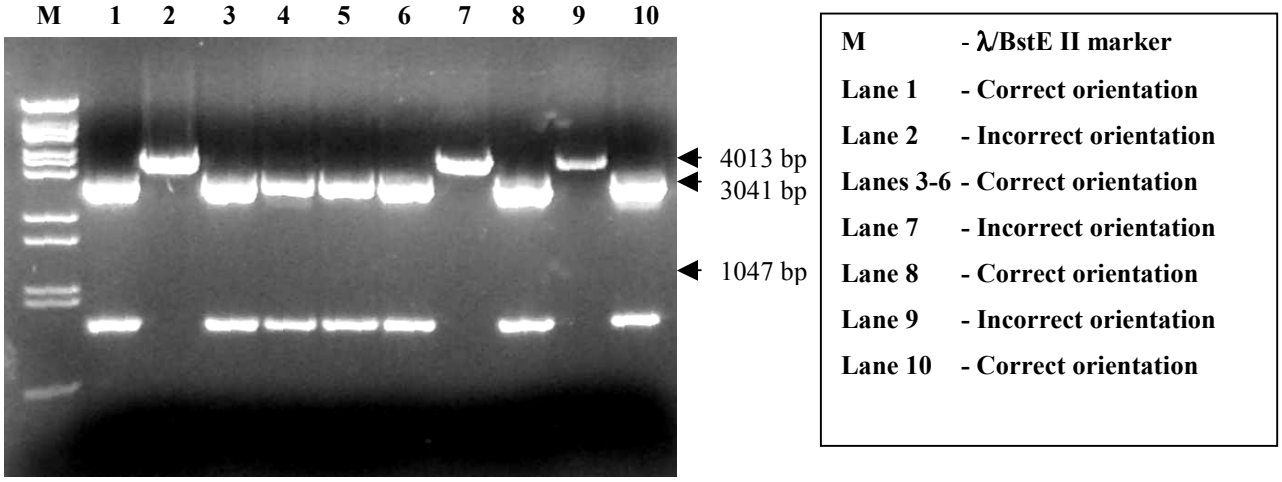


Figure 4-7: *Pst I* restriction digest check of pGEM-GM1 for correct orientation cloned isolates.

(ii) Extracting NBV5T from pFASTBAC1 and Sub-Cloning into pBluescript KS^{+}

The NBV5T insert was removed from pFASTBAC1 using *BamHI* and *EcoRI*. This created different cohesive ends, thereby increasing the sub-cloning efficiency into pBSKS. The excised insert of 1477 bp was gel purified (Fig. 4-8).

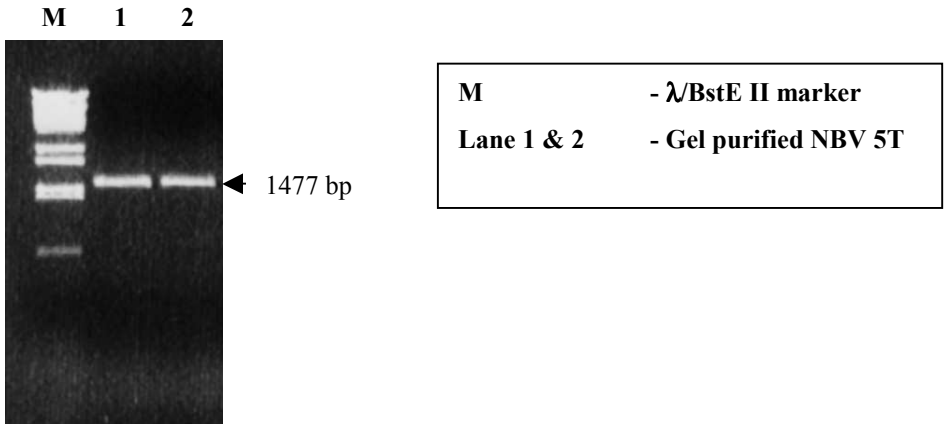


Figure 4-8: 1% Agarose gel of purified NBV5T excised from pFASTBAC1 using *BamHI* and *EcoRI*.

The pBSKS was also linearised using *BamHI* and *EcoRI* and purified from a LMP gel. The insert and plasmid were added to a ligation mix (see Appendix 5) and transformed into DH5α. Seven colonies were picked and "smart prepped" to obtain plasmid DNA. The plasmids were then digested with *BamHI* and *EcoRI,* which created an insert band of 1477 bp and a vector band of 2961 bp (Fig. 4-9).

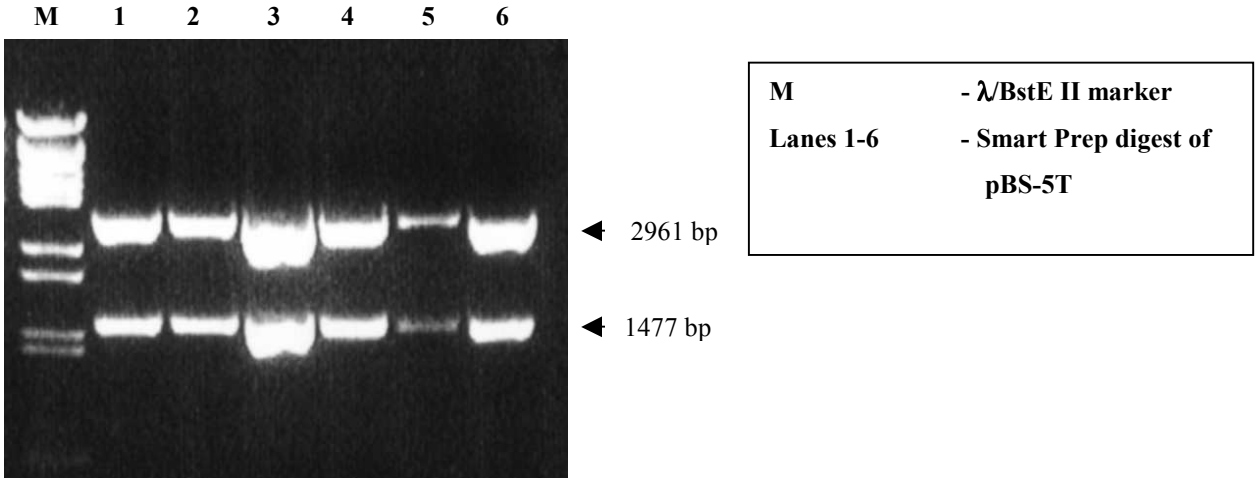


Figure 4-9: 1% Agarose gel of pBS-5T plasmid digested with *EcoRI* and *BamHI.*

(iii) Excising NBV214 from pBluescript KS using *BamHI* and Sub-Cloning into pBS-5T

Plasmid NBV214 was cut with a *BamHI* restriction enzyme and was expected to yield three bands of 3179 bp, 1141 bp and 111 bp. But after the digest, the gel had four bands all of the incorrect size (Fig. 4-10)

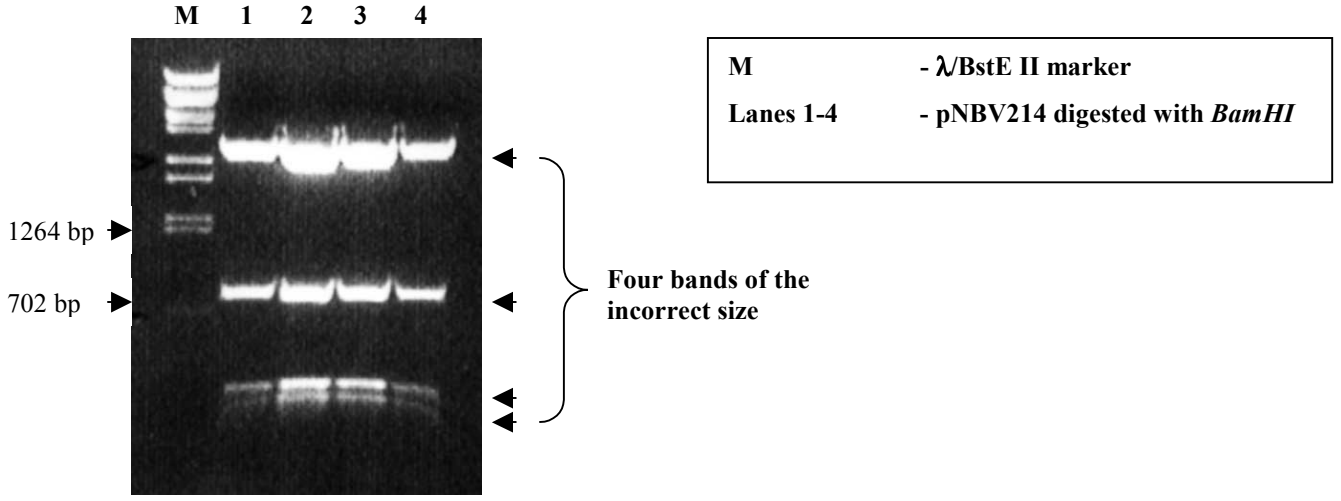


Figure 4-10: *Bam HI* restriction digest of pNBV214 run on a 1% agarose gel.

From the above results, it was evident that the *BamHI* restriction enzyme was cutting NBV214 insert at another restriction site, thereby reducing the 1141 bp band to smaller bands (as visible in the gel). It was at this stage that Dr. Karl Gordon at CSIRO in Australia was contacted and asked to check the sequencing results for a possible polymorphism. His response confirmed that a sequence polymorphism at position 1799 nucleotide, had caused an "A" to be replaced by a "T" thereby creating a new *BamHI* site. The restriction digestion of pNBV214 therefore resulted in the formation of four bands of 3058bp, 797bp, 344 bp and 252 bp, respectively. The predicted restriction digestion results confirmed the above results, thereby affirming the sequencing polymorphism error.

This meant that the *BamHI* site could no longer be used for sub-cloning the NBV214 insert into pBS-5T, and that other restriction sites would have to be looked at for sub-cloning purposes.

4.2.3. DISCUSSION

After examining the NβV clones carefully, it was clear that other methods would have to be applied to link the inserts of the different clones together. The number of restriction sites present precluded any possible attempts at linking the genome together by sub-cloning.

The construction and cloning of the missing section, pGEM-GM1 heralded the fact that the whole genome was now available in seven sections. The reality was to find a method to assemble the pieces together to construct a full-length copy of the NβV genome.

A *Nature* paper by Stemmer, (1994) demonstrated a technique used by the author called "*DNA shuffling*". This technique linked pieces of fragmented overlapping DNA together by PCR. It was decided to attempt, using this technique, to link the overlapping pieces of the cloned inserts together. The technique was slightly altered and only required twenty PCR cycles to link two overlapping pieces of DNA together (refer to chapter 2 for DNA linking method). The reason for reducing the number of cycles was to reduce the chances of a mutation occuring, which could cause the genome sequence to be altered.

SECTION 4.3: Linking the NβV genome Segments using a PCR technique.

4.3.1. INTRODUCTION

The principle of linking overlapping pieces of DNA together is based on the standard PCR technique, which requires that two homologous overlapping regions anneal and thermostable DNA polymerase elongates the annealed section in a 3' direction by primer extension. Once the annealed section has elongated, the standard 5'-end and 3'-end oligonucleotide primers anneal to the newly synthesised ends and amplify the linked templates (Fig. 4-11).

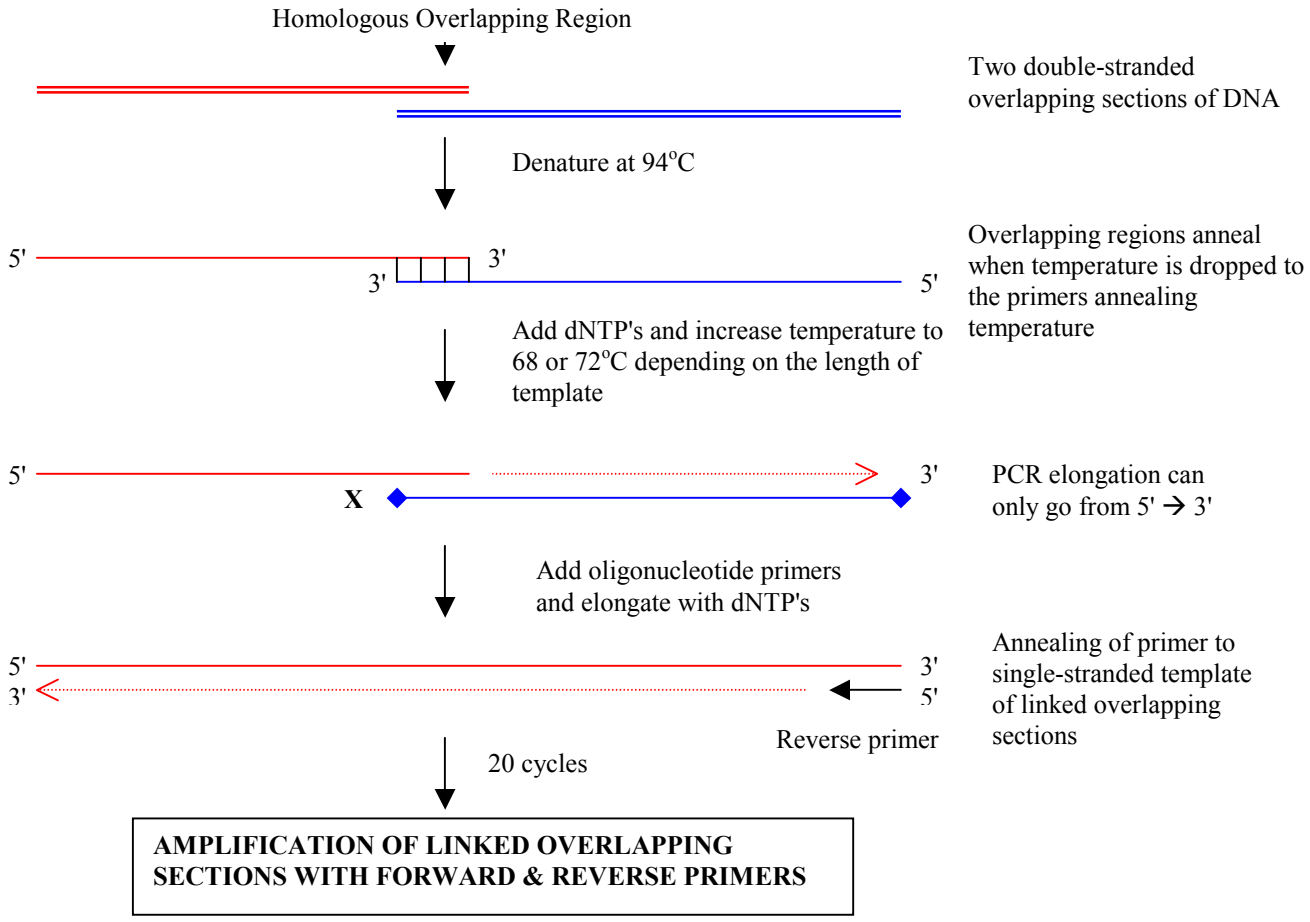


Figure 4-11: Flow-diagram demonstrating the linking technique of overlapping sections.

Mutations that occur during PCR cycles can only be minimized by reducing the number of cycles and by using a 3'-5' exonuclease containing proofreading polymerase in combination with a thermostable *Taq* DNA polymerase enzyme (see section 2.2.8). The proofreading polymerase enzyme results in a 3-fold increased in the fidelity of the synthesized DNA. This reduces the error rate to 8.5 x 10^{-6} nucleotides compared to *Taq* DNA polymerase that has an error rate of 2.6 x 10^{-5} nucleotide. But mutations would still occur during the PCR assembling of the full-length construct of the NβV genome. This would result in amino acid alteration during translation of the genome and would therefore affect the protein structures of the virus. Therefore the fully constructed beta genome would have to be sequenced. This would identify any mutation or mismatches in the NβV sequence. Once the mutations have been identified site-directed mutagenesis could be done to correct the errors.

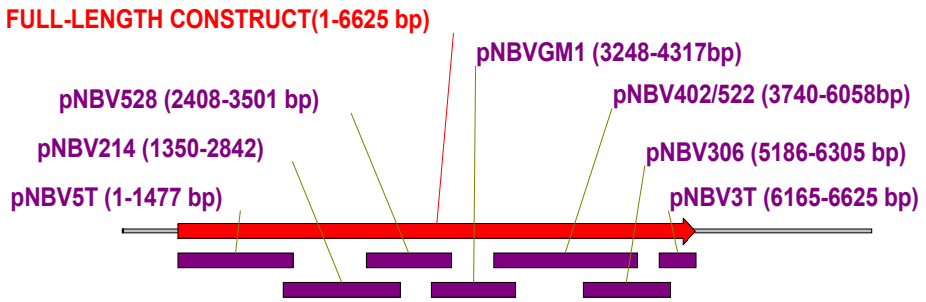


Figure 4-12: Diagram demonstrates the seven overlapping clones.

The plan from the above diagram was to link the 5'-end overlapping clones first, therefore linking NBV5T with NBV214 and NBV528. This would link half the genome together. After which the 3'-end could be linked using pNBVGM1, pNBV402/522, pNBV306 and pNBV3T thereby assembling the other half of the genome. When the genome is assembled into two sections of about 3300 bp each, the plan would be to link the two pieces together and construct the full-length piece. The full-length construct would be cloned into a PCR cloning vector and checked with several restriction enzymes. PCR amplification would also be used to check the binding of the primers in areas that cover the length of the genome.

4.3.2. RESULTS

(i) Assembling the 5'-Half of the NβV Genome using a PCR Linking Technique

Gel purifying the excised inserts from the vector reduced the possibility of mismatches or insertion occuring during the subsequent PCR reaction. The inserts were extracted either by restriction enzyme digest or PCR amplification. The reason PCR was used instead of cutting with restriction enzymes for most of the inserts was because of the difficulty with multiple restriction sites being present in the inserts (as discussed in Section 4.2).

The first approach was to excise inserts NBV214 and NBV528 using restriction enzymes and then link the two fragments. Plasmid NBV214 was cut using Promega *SacII*, which should have created fragments of 2929 bp and 1502 bp. But the digest resulted in bands of the incorrect size. After analysing the plasmid maps closely it was discovered that the insert NBV214 was cloned in the incorrect orientation. This meant that the plasmid NBV214 had to be double digested using *SacII* and *XhoI* (Roche Molecular Biochemicals), thus creating bands of 2885 bp, 1509 bp and 37 bp. The 1509 bp fragment containing the NBV214 insert was gel purified from a LMP gel (Fig. 4-13). Plasmid NBV528 was cut using *BamHI* and *ClaI,* thus creating fragment bands of 2924 bp, 936 bp and 195 bp. The 936 bp band, containing the cloned insert NBV528 was gel purified (Fig. 4-13).

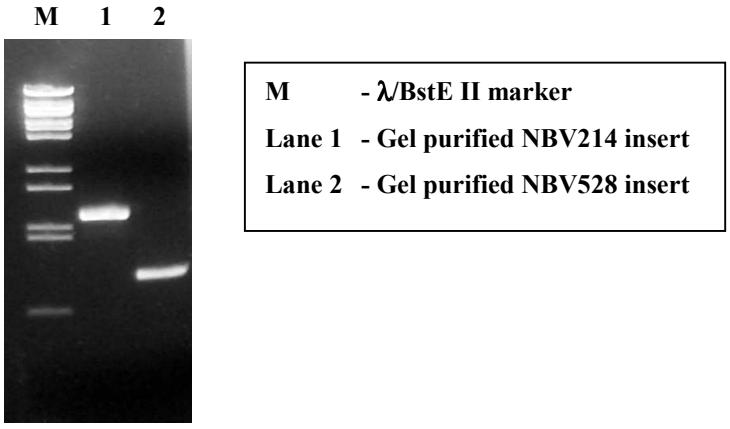


Figure 4-13: 1% Agarose gel of purified NBV214 and NBV528 inserts.

The plan was then to link the purified fragments using the PCR linking technique. End primers NBV1408F and NBV3309R (Appendix 4) were used to amplify the newly linked fragment (Fig. 4-14).

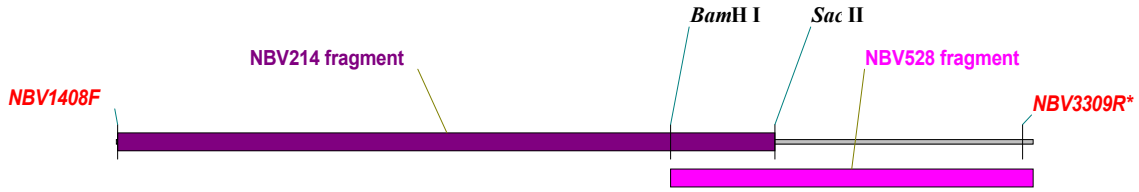


Figure 4-14: Diagram demonstrating the PCR linking technique for NBV214 and NBV528 fragments.

The two fragments were placed into a PCR mix with the forward and reverse primers. Twenty cycles were run on the thermocycler at the specified settings (Appendix 4) and the mix was then viewed in a 1% agarose gel for amplification of the 1977 bp link product (Fig. 4-15).

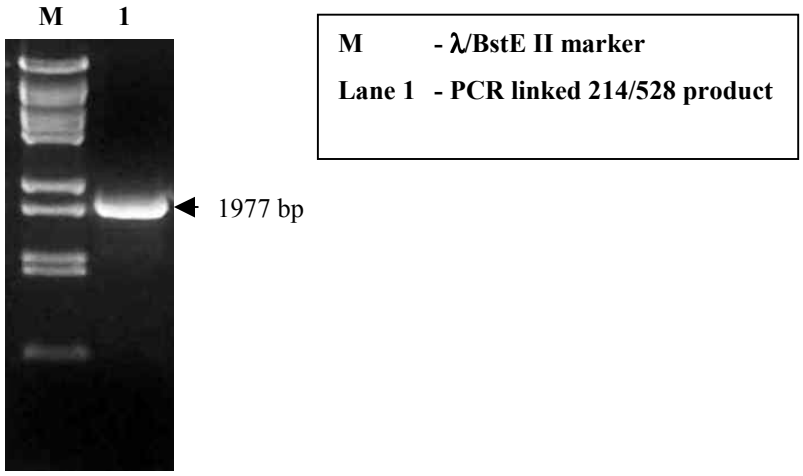


Figure 4-15: NBV 214/528 PCR link in a 1% Agarose gel

The linked product was then purified by gel extraction. The next step was to link NBV5T insert to the 214/528 PCR product. The extraction of the NBV5T insert from the FASTBAC1 plasmid (pFB5T) was done using *BamHI* and *EcoRI*. The digestion

produced three fragments of 4768 bp, 1463 bp and 22 bp. The 1463 bp fragment, containing the NBV5T insert was gel purified (Fig. 4-16).

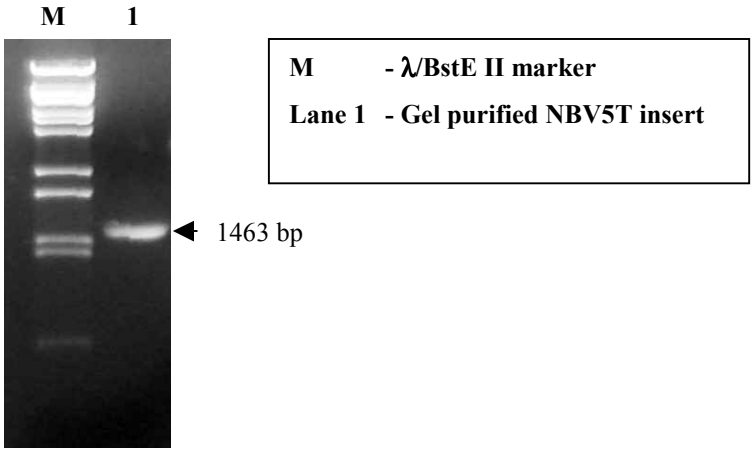


Figure 4-16: 1% agarose gel of purified NBV5T insert.

The NBV5T insert was added to the 214/528 product in a PCR mix with NBV5TPCR and NBV3309R primers (Appendix 4). The tube was placed in a thermocycler with the specified settings (Appendix 4) and programmed for a twenty-cycle run. After the run was complete the PCR mix was examined in a 1% agarose gel for a band 3377 bp in size (Fig. 4-17).

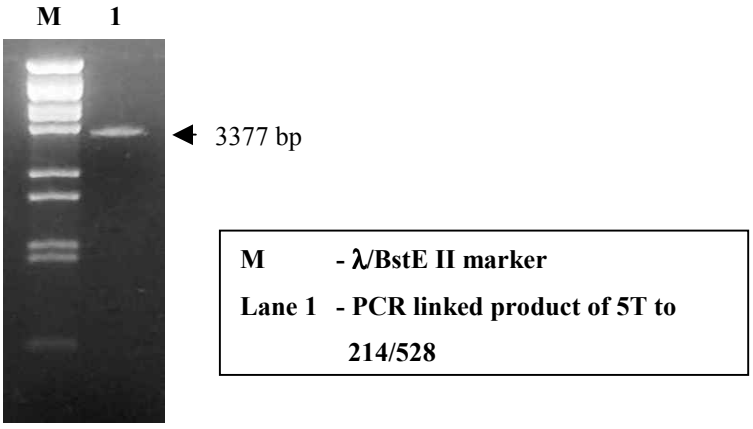


Figure 4-17: PCR link of 5T insert to 214/528 linked product in a 1% agarose gel.

Because this product was produced using the reverse primer NBV3309R, it was decided to label the new 3377 bp product, NBV3309. The product was then cloned into pGEM-T

Easy Vector and checked with a number of different restriction enzyme to examine the authenticity of the PCR product, NBV3309. The digest gave all the expected results, which confirmed that this PCR technique was working. The next approach was to link the 3'-end of the genome to form two pieces of equal length.

(ii) Assembling the 3'-Half of the Genome using the PCR Linking Technique

This approach was slightly different because there were four pieces and all had to be extracted using PCR amplification. The plan was to link the GM1 to 402/522 and then link 306 to 3T (see Fig. 4-12). Once these two fragments were constructed; the plan was to link them together to create the other half of the genome.

The first step was to extract NBV-GM1 from its vector using PCR amplification (NBVREP2 and NBVCAP1 primers were used) and then to purify the PCR product in a LMP gel. This approach was also used for the NBV402/522 insert (using NBV4039F and NBVBCAPR primers). The gel-purified products were examined in a 1% agarose gel to confirm the presence of the GM1 band of 1069 bp and the 402/522 band of 1838 bp.

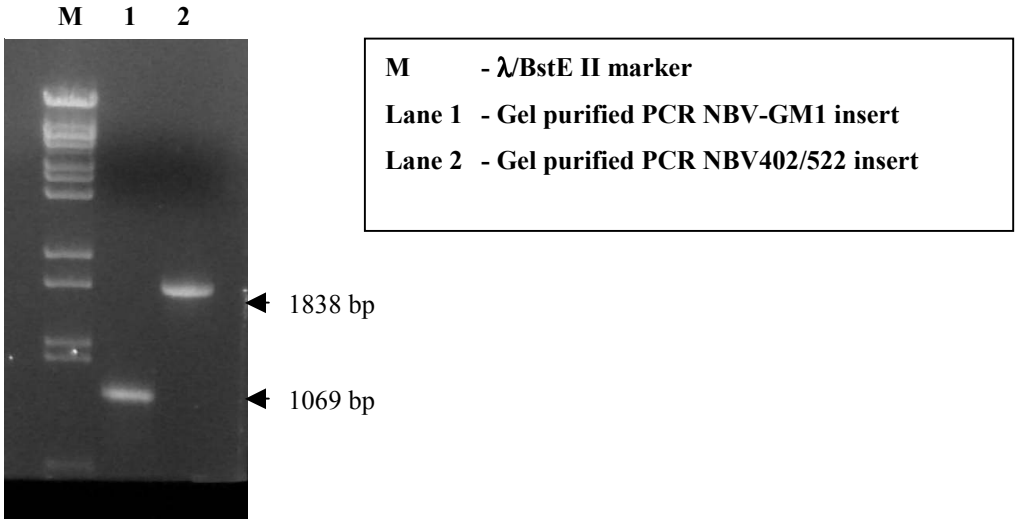


Figure 4-18: NBV-GM1 and NBV402/522 gel purified product in 1% agarose.

The two purified products were then added into a PCR mix with primers NBVREP2 and NBVBCAPR. The tube was placed in the thermocycler at the specified settings (Appendix 4) for twenty cycles. The PCR mix was then examined in a 1% agarose gel for an amplification band size of 2629 bp (Fig. 4-19).

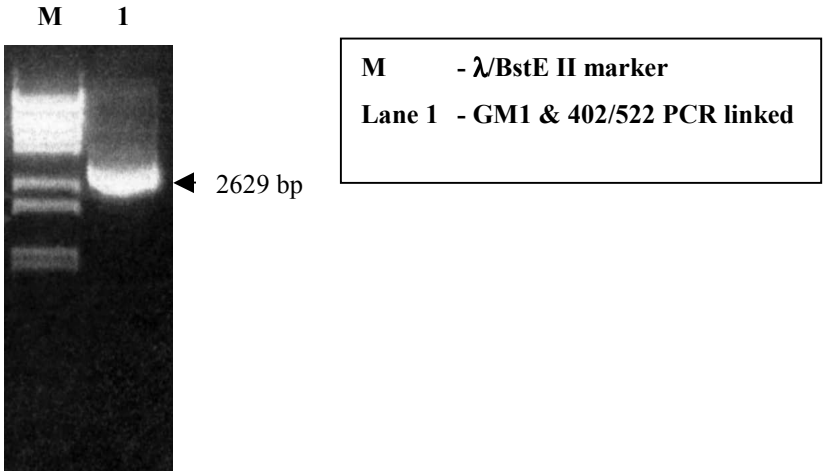


Figure 4-19: 1% agarose gel demonstrating the PCR link of NBV-GM1 to NBV402/522

The next step was to link NBV306 to NBV3T. The inserts were amplified from the plasmids using PCR (primers NBV5202F and NBV6281R were used for insert 306 and primers NBV6165F and NBV3TRZ were used for insert 3T, see Appendix 4). The PCR products were gel purified from a LMP gel and viewed in a 1% agarose gel (Fig. 4-20). PCR 306 created a fragment of 1095 bp and PCR 3T a fragment of 406 bp, respectively.

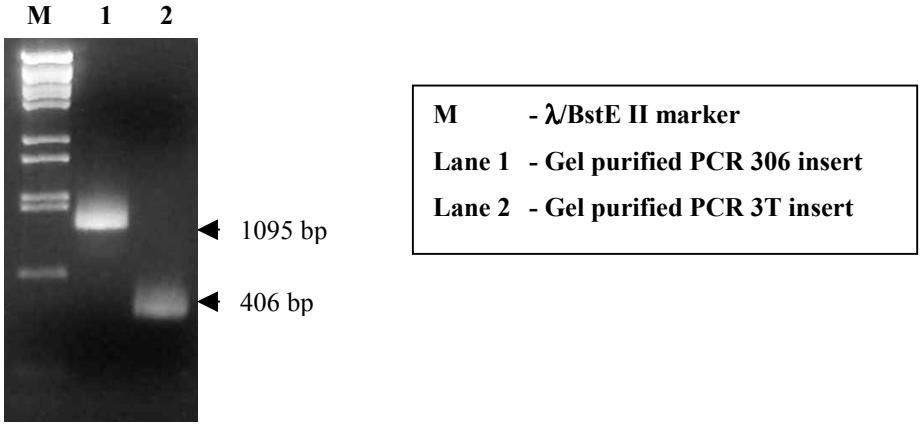


Figure 4-20: Gel purified PCR fragments of NBV306 & NBV3T inserts in a 1% agarose gel.

The two purified fragments were added together in a PCR mix with end primers NBV5202F and NBV3TRZ. The tube was placed in a thermocycler with the specified settings (Appendix 4). Twenty cycles were performed and then examined on a 1% agarose gel, the expected band size was 1423 bp (Fig. 4-21).

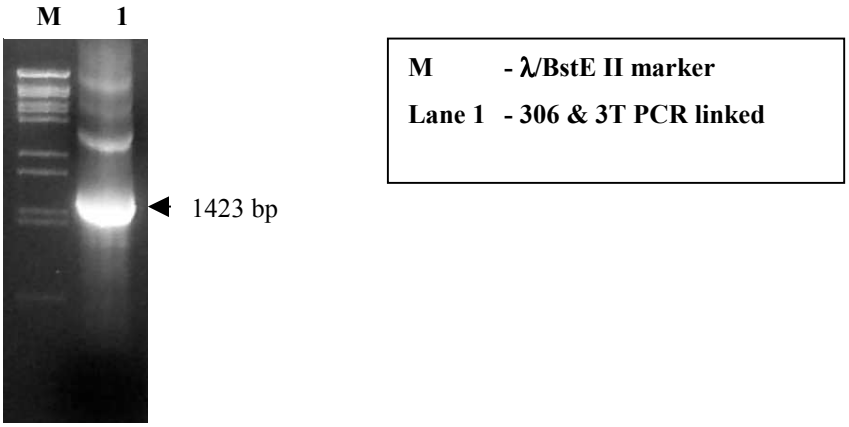


Figure 4-21: The PCR link of 306 to 3T demonstrated in a 1% agarose gel

GM1 & 402/522 PCR linked product was named NBV2629 and the 306 & 3T linked fragment was named NBV306/3T. Both these new fragments were gel purified and added together in a PCR mix with NBVREP2 and NBV3TRZ primers. The mix was placed in a thermocycler at specified settings and run for twenty cycles. After the run was complete, the PCR mixture was run on a 1% agarose to confirm the presence of a 3377 bp size fragment, which was designated NBV3377 (Fig. 4-22).

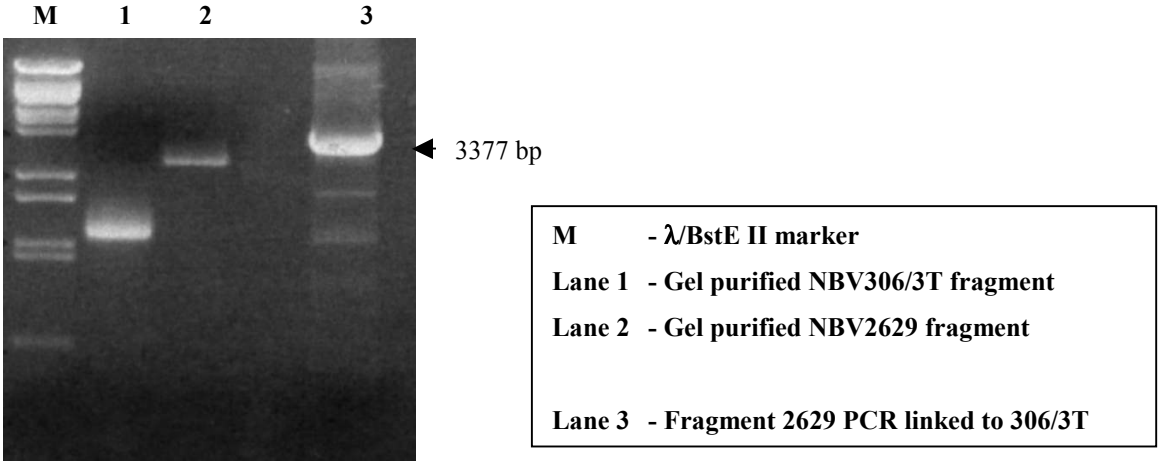


Figure 4-22: 1% agarose gel demonstrating the PCR link of NBV2629 to NBV306/3T.

(iii) PCR Link of the 5'-Half to the 3'-Half to construct a Full-Length PCR Fragment

The final approach was to link NBV3309 to NBV3377. This link was problematic as the overlapping region between the two large fragments was only 129 bp (Fig. 4-23).

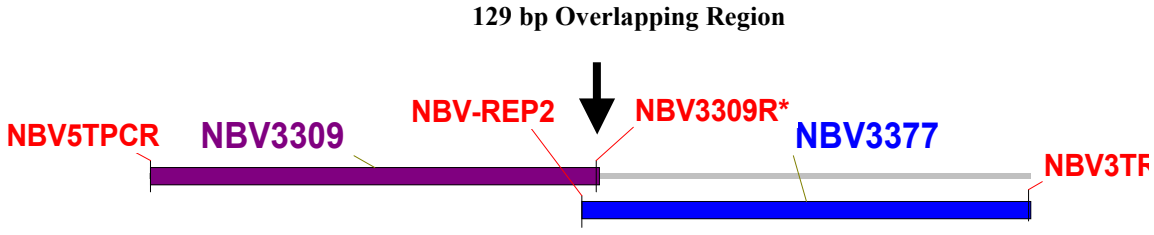


Figure 4-23: Diagram illustrating the PCR link of NBV3309 and NBV3377

After a few attempts at PCR amplification, using different elongation temperatures and by adjusting the annealing temperature, two bands appeared in the 1%agarose gel. The top band appeared to be around 6.6 kb and the other about 1.8 kb (Fig. 4-24).

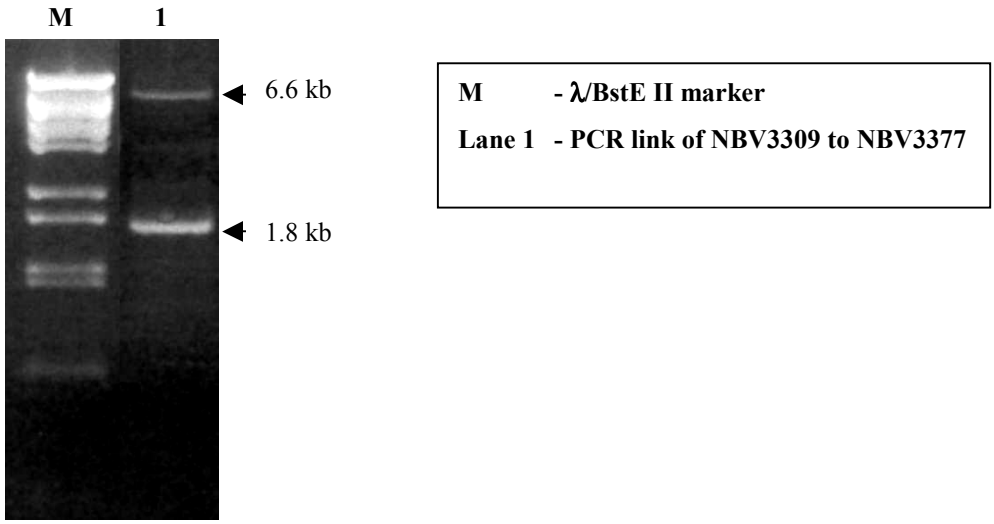


Figure 4-24: NBV3309 linked to NBV3377 in a 1% agarose gel.

This PCR link was repeated yielding the same results, indicating that some form of unspecific binding was occuring. The plan was then to gel purify the top band and check it with restriction enzymes. However, the concentration of the extracted product was too low to clone into a vector.

The next plan was to modify the end primer NBV3TRZ; this primer had a ribozyme binding sequence, which was removed and re-designed to anneal eighteen nucleotides into the 3' end (Appendix 4). A PCR link was attempted with the new primer (NBV3TR) (Appendix 4) but still no amplification could be achieved. The next step was to lengthen the overlapping region. The 5'-half was easier to lengthen by attaching NBV-GM1 insert to NBV3309 using PCR linking. The new NBV3309ext fragment stretched from NBV5T PCR primer to NBVCAP1 reverse primer. The PCR reaction was examined on a 1% agarose gel for a band of 4317 bp in length (Fig. 4-25).

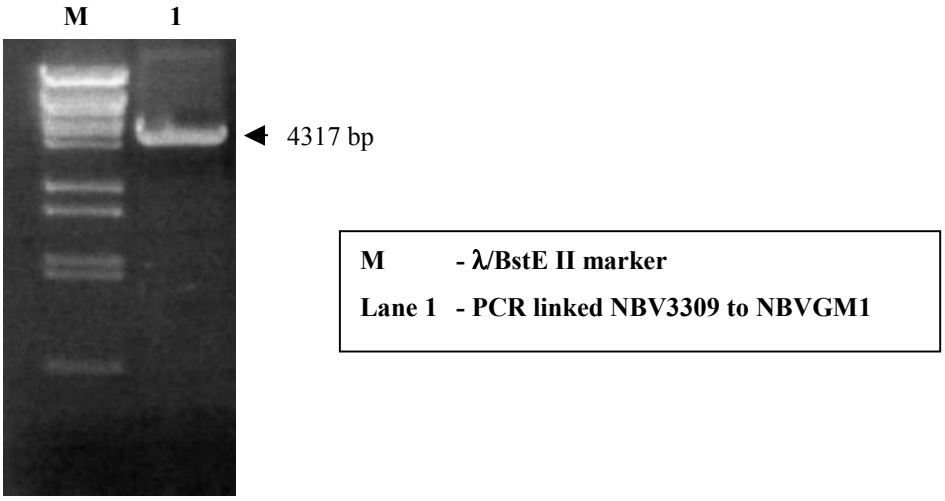


Figure 4-25: 1% agarose gel of the PCR link of NBV3309 to NBVGM1

The new fragment created a 1069 bp overlap between NBV3309ext and NBV3377, which was expected to improve the specificity of annealing in the overlapping regions. The fragment was gel purified and added to a PCR mix with NBV3377. The end primers NBV5TPCR and NBV3TR were added, and the tube was placed in the thermocycler for twenty cycles. One run was done at an elongation temperature of 68^{O}C and another at 72^{O}C. The two mixes were visualised in a 1% agarose gel for amplification of the linked fragment, which created a band of 6625 bp (Fig. 4-26).

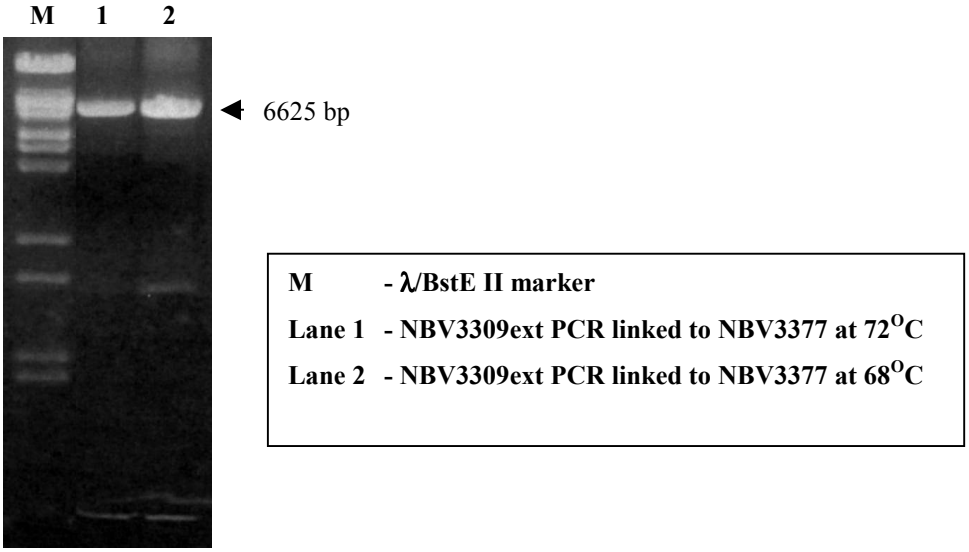


Figure 4-26: The full-length cDNA construct of the NβV genome in a 1% agarose gel.

It is clear from the above gel that using an elongation temperature of 68^{O}C, for fragments over 3 kb, increases the concentration of the DNA amplification. The new full-length PCR fragment was named PCR6625. The PCR product would now have to be cloned into a PCR cloning vector and checked by cutting with restriction enzymes. The genome could also be checked with a range of PCR primers for amplification. This would check for any possible mutations that have caused loss of restriction sites or loss of annealing of primers, thus preventing the amplification of a specific area of the genome.

(iv) Cloning PCR6625 into pGEM-T Easy Vector and performing a Restriction Enzyme Digest and PCR Primer checks

The PCR6625 fragment was gel purified and incubated overnight in a ligation mix with pGEM-T Easy vector. The plasmids were transformed into *E.coli* DH5α. Two colonies grew which were mini-prepped for plasmid isolation. 1 µl of the mini-prep was digested with *Sac I* (Roche Molecular Biochemicals) (Fig. 4-27).

Figure 4-27: Plasmid map demonstrating the *SacI* sites for PCR6625 cloned into pGEM-T Easy Vector in the correct orientation.

If PCR6625 was cloned in the correct orientation, the *SacI* fragments would be 4530 bp, 3994 bp and 1119 bp. If the fragment was cloned in the incorrect orientation the fragments would be 6918 bp, 1606 bp and 1119 bp. The mini-preps were digested and checked on a 1% agarose gel (Fig. 4-28).

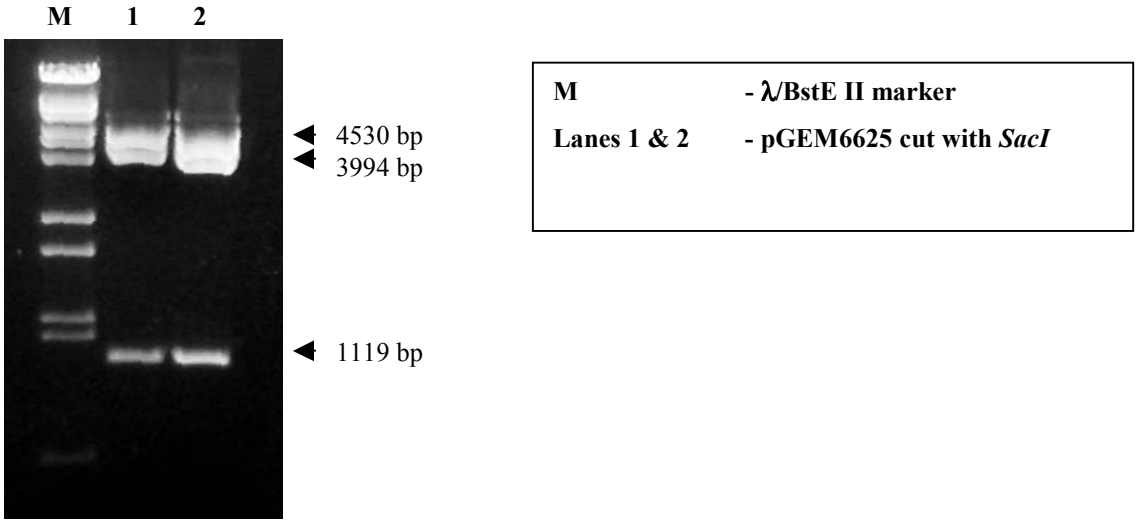


Figure 4-28: 1% Agarose gel of a *SacI* digest of pGEM6625 to check that PCR6625 was cloned in the correct orientation.

Now that it had been confirmed that both mini-preps contained PCR6625 cloned in the correct orientation, it was decided to check pGEM6625 with a number of restriction enzymes. The table below indicates the different restriction enzymes used with the expected fragment sizes. (Table 4-1).

Table 4-1: Fragment sizes for different restriction enzyme digest of pGEM6625

RESTRICTION ENZYME	**FRAGMENT SIZE (bp) of pGEM6625**
ClaI	4217/ 3701/ 1725
EcoRI	3320/ 3000/ 1347/ 1116/ 582/ 278
PstI	6288/ 1897/ 929/ 529
XbaI	9643
XhoI	4188/ 2527/ 1814/ 1114
SacI	4530/ 3994/ 1119
BanI	2211/ 1286/ 1184/ 1087/ 1076/ 809/ 514/ 497/ 359/ 337/ 207/ 66
AcsI	2339/ 1696/ 1397/ 1040/ 738/ 661/ 609/ 582/ 278/ 227/ 76

1 µl of the mini-prep was used for each digestion and visualised in a 0.7 % agarose gel run overnight at 30 volts (Fig. 4-30). The restriction enzymes covered most of the cloned insert (Fig. 4-29) in an attempt to determine if mutations had occurred during the PCR assembling process.

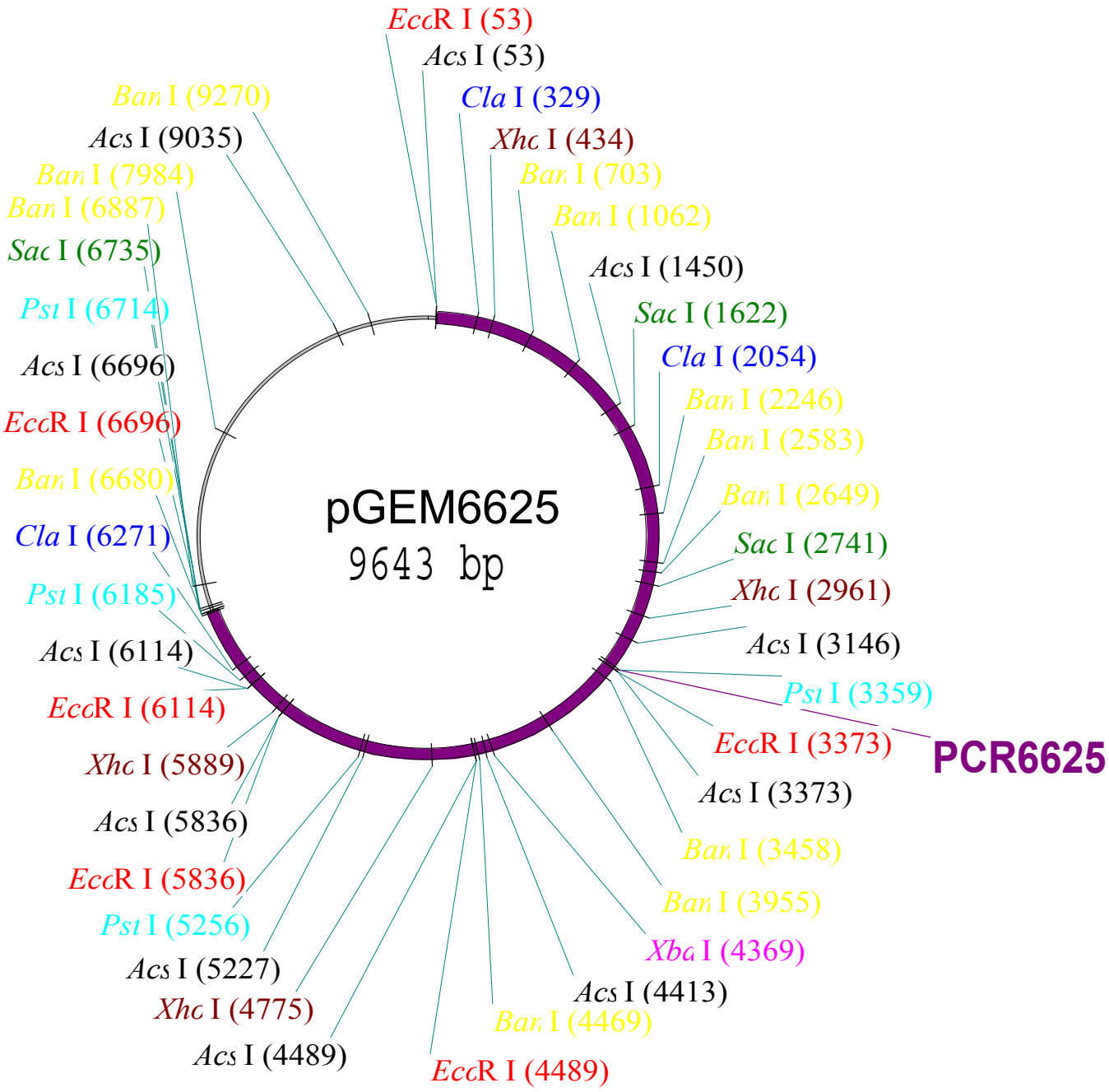


Figure 4-29: Plasmid map showing the different restriction enzymes sites used to check pGEM6625.

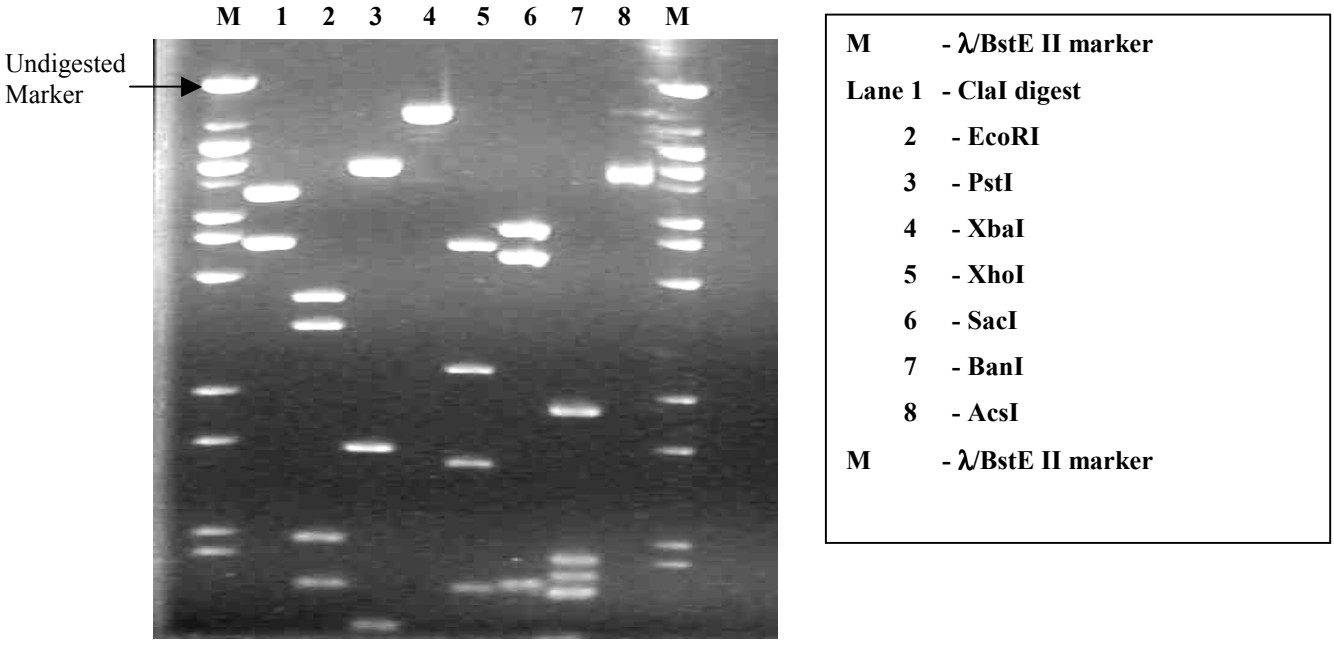


Figure 4-30: 0.7% agarose gel of pGEM6625 digested with eight different restriction enzymes.

Unfortunately, the *AcsI* restriction enzyme had expired and therefore gave no result except supercoiled DNA; the enzyme was checked in another digest (Fig. 4-31). One factor that was evident from the long gel differentiation was that the marker had only been partially digested and had a band of undigested lambda DNA. The gel results showed that, except for *AcsI* enzyme that had expired, *ClaI* was the only enzyme that gave the incorrect restriction band sizes. Only 2 bands were seen on the gel, rather than the 3 bands, which were expected. The plasmid map of pGEM6625 was analysed to determine which *ClaI* restriction site contained the problem. The *ClaI* restriction site at nucleotide 269 was found to be faulty. The site was checked on the original plasmid, pFBNBV5T, by digestion with *ClaI*. *A*nother double digest with *ClaI* and *HindIII* was also done which would cut with the *HindIII* even if the *ClaI* site was faulty, thereby linearizing the plasmid (Fig. 4-31).

Table 4-2: Restriction enzyme check of pFBNBV5T for faulty sites

RESTRICTION ENZYME	FRAGMENT SIZES (bp) of pFBNBV5T
ClaI	6253
ClaI + *HindIII*	4971/ 1282
AcsI	3628/ 1399/ 496

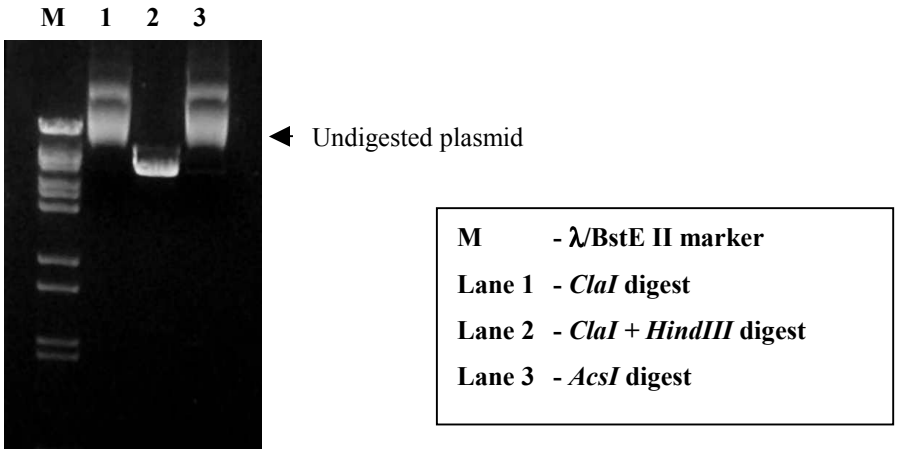


Figure 4-31: Restriction digest of pFBNBV5T to check for a faulty *ClaI* site and ascertain whether *AcsI* had expired. The digest was run on a 1% agarose gel for 1 hour.

The *ClaI* digest of pFBNBV5T (Appendix 6) showed that the site at nucleotide 269 was faulty. The *AcsI* digest also showed that the enzyme had expired. Closer analysis showed that a "G" base preceded the *ClaI* site in the reverse sequence causing it to be methylated, thus blocking the endonuclease activity of the enzyme and preventing it from cutting the plasmid. The result confirmed that no mutation had occurred in the sequence.

Another set of digests was done on pGEM6625 to check if any other possible mutations had occurred. Another seven enzymes were used which covered a broad area of the cloned PCR insert (Fig. 4-32).

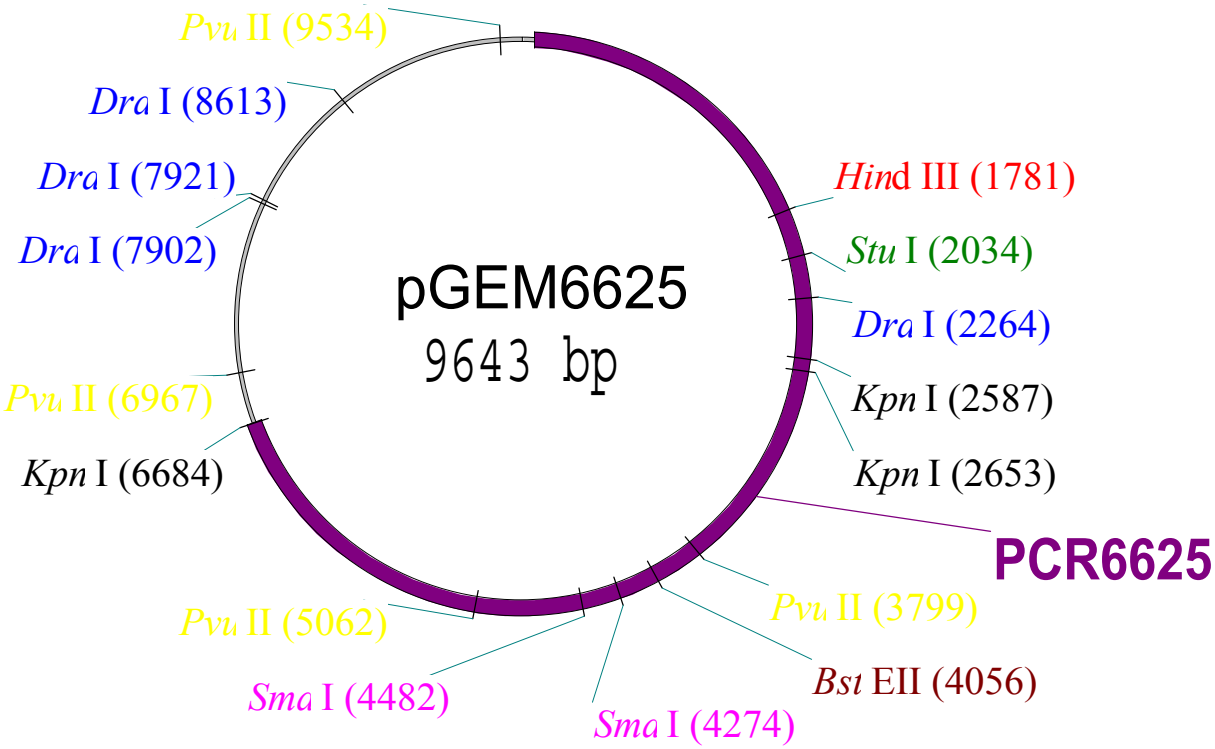


Figure 4-32: Plasmid map showing the seven different restriction sites on pGEM6625.

The table below indicates the possible fragment sizes of the different enzymes used to check pGEM6625 (Table 4-3).

Table 4-3: Fragment sizes for different restriction enzyme digests of pGEM6625

RESTRICTION ENZYME	FRAGMENT SIZE (bp) of pGEM6625
SmaI	9435/ 208
HindIII	9643
DraI	5643/ 3294/ 692/ 19
BstE II	9643
KpnI	5546/ 4031/ 66
PvuII	3908/ 2567/ 1905/ 1263
StuI	9643

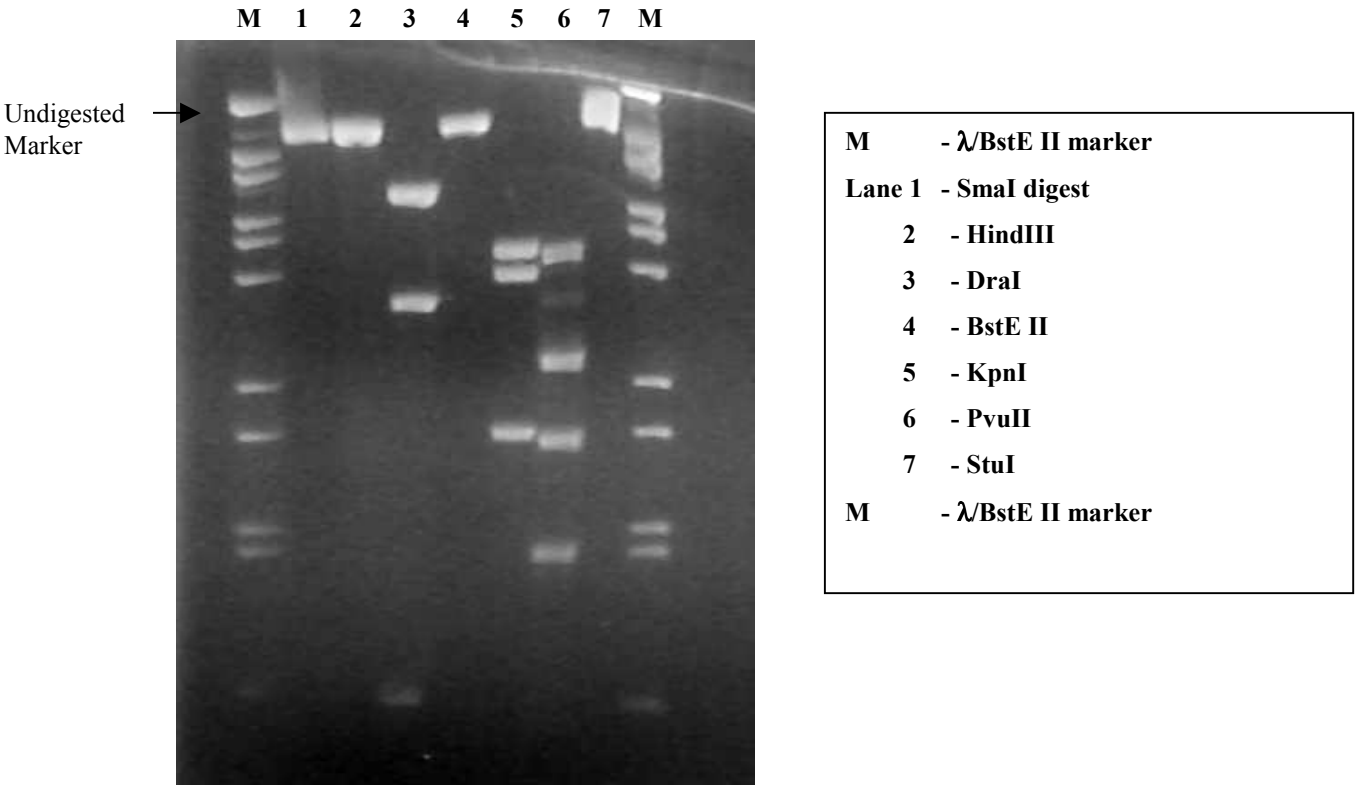


Figure 4-33: 0.7% agarose gel run overnight with seven different restriction digests to check the cloned insert PCR6625.

KpnI was the only enzyme to give unexpected results. All the other restriction enzymes gave the expected results. The *KpnI* result suggested that an extra site had been created close to the 5'-end. pFBNBV5T was digested with *KpnI* to determine if the extra site was present in the original plasmid. The digest was run on a 1% agarose gel and should have resulted in a single band of 6253 bp in size. It was calculated that if an extra *KpnI* site was present in the original plasmid, two bands of about 5200 bp and 1000 bp would be visualised in the gel (Fig. 4-34).

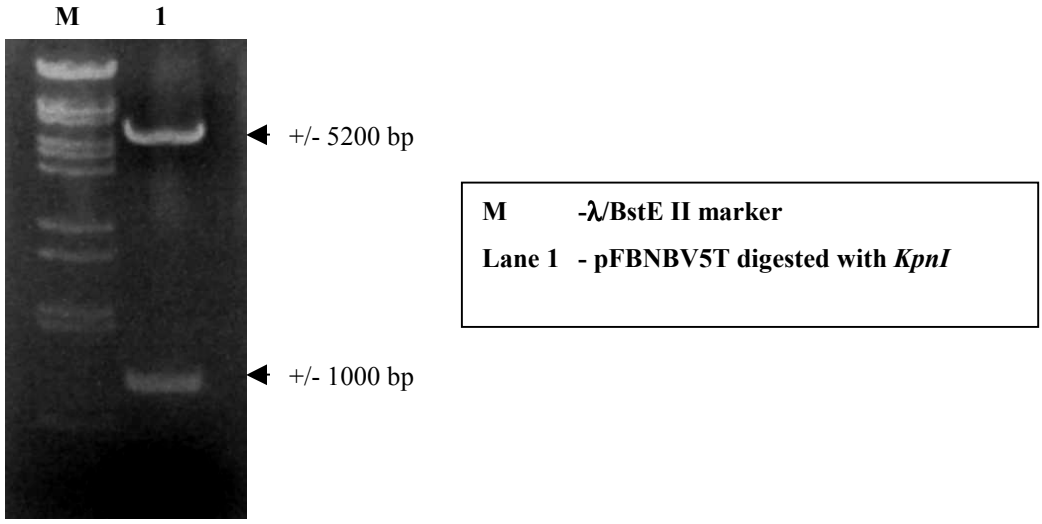


Figure 4-34: 1% agarose gel demonstrating a *KpnI* restriction digestion of pFBNBV5T.

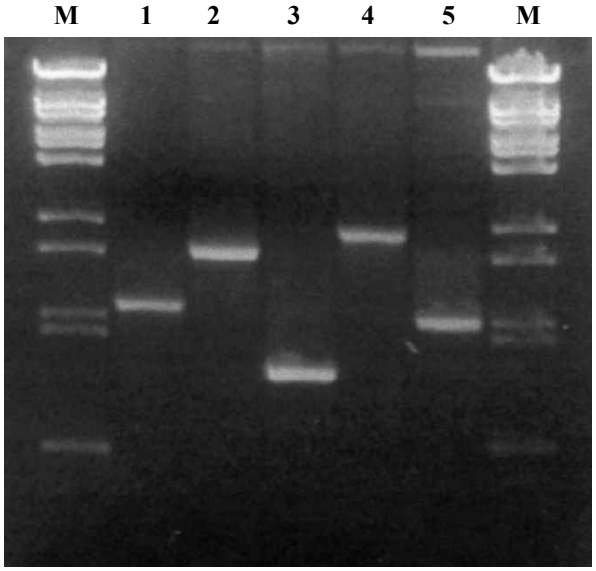
From the above result it was clear that there was an extra *KpnI* site and that it had originated from the original plasmid. It was calculated to be at about nucleotide 550, where a possible sequence alteration could have created the extra site. Dr. Karl Gordon was once again contacted to check his sequence diagrams. He concluded that there could have been a possible sequencing error but that the plasmid he had used to sequence that 5'-end was different to the plasmid used to PCR link. The pFBNBV5T which was used to link the genome was created by Dr. Terry Hanzlik using PCR, which could have altered the base at nucleotide 550 and caused the mutation which led to the formation of the *KpnI* site. Further investigations would have to be done in order to determine the cause of this extra restriction enzyme site.

The final check of the full-length clone was with different PCR primers. Five sets of forward and reverse primers were used that created PCR products that covered the entire cloned insert. The primers used and the expected fragment sizes are tabulated in Table 4-4.

Table 4-4: The PCR reactions used to check pGEM6625

PCR REACTIONS	PRIMERS (see appendix 4)	EXPECTED FRAGMENT SIZE (bp)
Reaction 1	F → NBV5TPCR R→ NBVREP1	1430 bp
Reaction 2	F → NBV1408F R → NBV3309R	1977 bp
Reaction 3	F → NBVREP2 R → NBVCAP1	1069 bp
Reaction 4	F → NBV4039F R → NBV6281R	2258 bp
Reaction 5	F → NBV5202F R → NBV3TR	1423bp

The primers were added to the standard PCR mix and the tube was placed into the thermocycler for 20 cycles. The reaction mixes were run in a 1% agarose gel (Fig. 4-35).



M - λ/BstE II marker
Lane 1 - PCR reaction 1 (1480 bp)
Lane 2 - PCR reaction 2 (1977 bp)
Lane 3 - PCR reaction 3 (1069 bp)
Lane 4 - PCR reaction 4 (2258 bp)
Lane 5 - PCR reaction 5 (1423 bp)
M - λ/BstE II marker

Figure 4-35: Different PCR reactions checking the authenticity of pGEM6625.

It was clear from the PCR reaction that the primers were binding in the expected areas and amplifying fragments of the correct size. All the different primer reactions gave the

correct results, confirming that all the sections of the virus genome were present in the new full-length construct of the NβV genome, and that no duplications or large deletions had occurred during the PCR reactions.

4.3.3. DISCUSSION

The PCR method used to link the genome segments has the drawback of possibly generating mutations, but as seen in the previous sections in this chapter, this was the only way remaining to link the segments. Now that the full-length genome has been constructed, there are methods that can be used to remove problem areas that may have been created by PCR linking. One such problem area is the creation of a *KpnI* site at nucleotide 550. The problem could have originated when the plasmid was created using PCR or be due to a sequencing polymorphism. The other areas of the full-length clone that have been checked have all indicated that no mutations have occurred. The final check would be to sequence the clone in both directions. This would indicate the presence of mutations that might have occurred during the assembly process. Further analysis of any mismatches would have to be done in order to determine if they were sequencing errors or if they originated during the linking process.

Having a full-length cloned NβV genome that matches the published sequence will lay the pathway for future research into the NβV genome. For many years, the problem of not having a full-length NβV genome to work with has hampered virologists. Once all the errors have been identified and rectified that problem will no longer exist. The future work on the NβV genome would be to research the genes that code for the synthesis of infectious viral particles.

The cloned insert pNBV402/522 contained the capsid open reading frame. This plasmid permitted the expression and isolation of the capsid protein, and laid the groundwork for future investigations to express the genes that control the synthesis of infectious virus particles.

CHAPTER 5

Expressing and Purifying NβV Capsid Protein in a Heterologous System

5.1. INTRODUCTION

Numerous studies have been done on amino acid sequence analysis of the NβV capsid protein (see Gordon *et al.,*1999), but minimal work has been done on actually cloning the capsid gene into an expression vector and isolating beta capsid protein. Analyses done thus far indicate that the virus capsid gene is translated from a separately transcribed 2656-nucleotide subgenomic RNA. It has also been shown that the capsid gene AUG start codon lies only 70 nucleotides downstream of the start of the subgenomic RNA. Amino acid sequence analyses calculate the capsid gene to encode a protein of 612 amino acids and to have a calculated molecular mass of 66,405 Da (Gordon *et.al*, 1999). The translation of the capsid gene is in the +1 reading frame relative to that of the replicase gene, which is in the +3 reading frame.

The aim of this study was to clone the capsid gene into an *E.coli* heterologous system and then to express and purify the capsid precursor protein. pNBV4039 (Appendix 6) was chosen as the source of the capsid gene because this plasmid was synthesised by K. Gordon from the pNBV402/522 (Appendix 6) with corresponding *BamHI* sites to assist cloning. The capsid gene was excised from pNBV4039 and cloned into a prokaryotic expression vector, pGEX. Due to the difficulty of expressing foreign polypeptides in bacterial systems, it was decided to use an expression vector that binds a fusion protein to the N-terminus of the capsid protein. The fusion protein thereby stabilizes the protein by making it more soluble within the bacteria, thus preventing any aggregation and denaturation.

The pGEX vector is a fusion-protein expression vector that expresses proteins as a product fused to Glutathione S-Transferase (GST). The *lac* repressor binds to the p_{tac}

promoter, repressing the expression of the GST fusion protein. Upon induction with IPTG, derepression occurs and the GST fusion protein is expressed. The capsid gene can be inserted into the polylinker located at the end of the GST gene (Smith and Corcoran, 1994). This polylinker has unique restriction endonuclease sites for *BamHI, SmaI* and *EcoRI* followed by termination codons in all three frames (Appendix 5). The polylinker expresses protease cleavage sites so that the cloned protein can be released from the GST moiety. Three vectors were available that had the restriction endonuclease sites in different frames, therefore the vector in which the capsid protein expressed in-frame with the GST protein had to be chosen (Appendix 5).

Only two of the three possible vectors were accessible for cloning, pGEX-2T and pGEX-3X. Analyses of the different possible cloning restriction sites of the different vectors indicated that pGEX-3X (Appendix 5) remained in-frame with the capsid ORF using the *SmaI* site. The capsid gene would have to be removed from pNBV4039 using *BamHI* and the protruding overhangs filled-in to create a blunt-end that could be cloned into the *SmaI* site. Once the capsid gene had been cloned into the pGEX-3X vector, the newly constructed plasmid vector would then have to be transformed into a bacterial cell line that would produce optimal expression of the capsid protein.

A specific method would need to be used to identify that the over expressed protein was in fact the NβV precursor synthesised as a fusion product. Chemiluminescence could be used to flag the expressed protein on the western blot membranes after incubation with NβV antibodies. Another method that could be used would be to bind the over expressed protein to glutathione slurry-beads and thereby purify the capsid/fusion protein from the *E.coli* cells. The purified expressed protein would then be labelled with NβV antibodies on a western blot membrane, which could then be developed for visualisation using chemiluminescence; this would confirm that no unspecific bacterial cell binding had occurred. Another experiment to be attempted would be to over express using a genetically modified bacterial strain (Epicurian Coli BL21-CodonPlus (DE3)-RIL) that has been designed to express eukaryotic protein. These bacterial cells permit the efficient, high-level expression of heterologous proteins in *Escherichia coli*.

5.2 RESULTS

5.2.1. Cloning the Capsid Gene into a Fusion Protein Expression Vector

The first approach was to choose a vector that would allow for the cloning of the capsid gene in-frame with the GST protein. pGEX-3X (Appendix 5) was chosen because the capsid gene cloned into the *SmaI* site would remain in-frame with the GST fusion protein (Fig. 5-1).

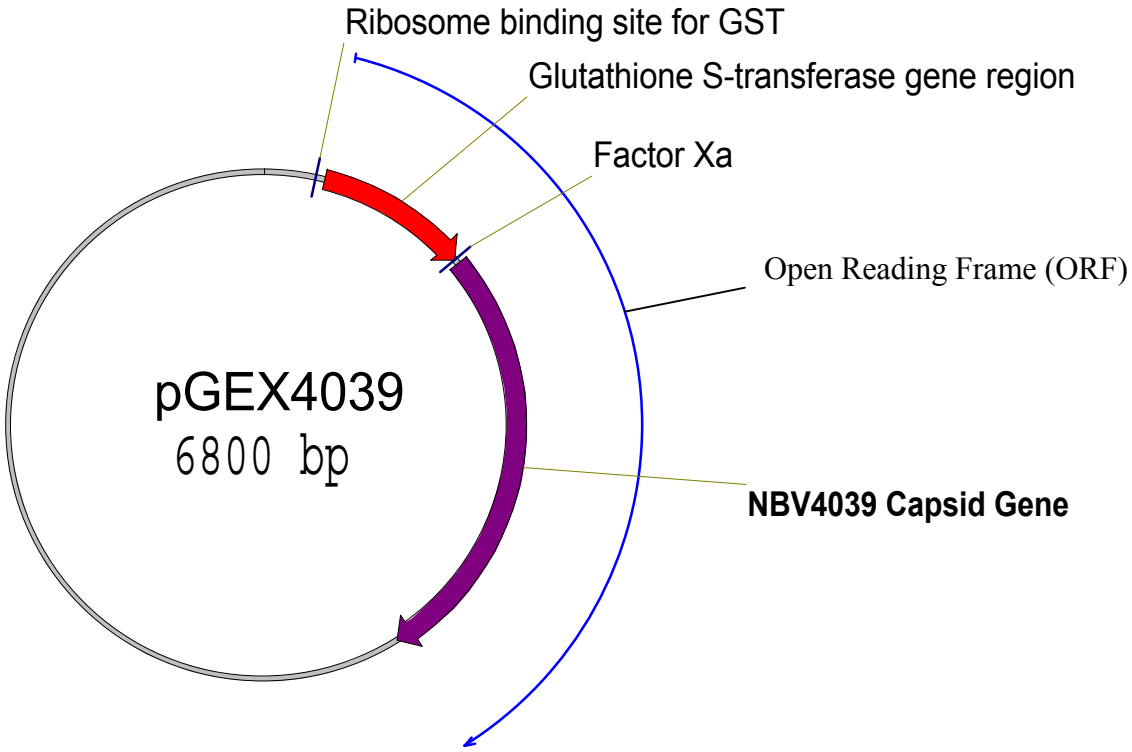


Figure 5-1: Diagram representing the plasmid map showing that NBV4039 (capsid gene) cloned into pGEX-3X created an ORF that read from the AUG of the GST fusion protein to the stop codon in the capsid gene.

The capsid gene was excised from its current plasmid, pBluescript KS, by using the primer engineered *BamHI* sites at the ends of the capsid gene. This created 5'-end overhangs at either side of the excised fragment. The fragment was then placed in a mix with dNTP's and DNA polymerase and the overhangs filled-in to create blunt-ends. The

cloning vector was then linearised at the *SmaI* site. The blunt-ended fragment and the linearised vector were gel purified and run on a 1% agarose gel to determine the relative DNA concentrations (Fig. 5-2).

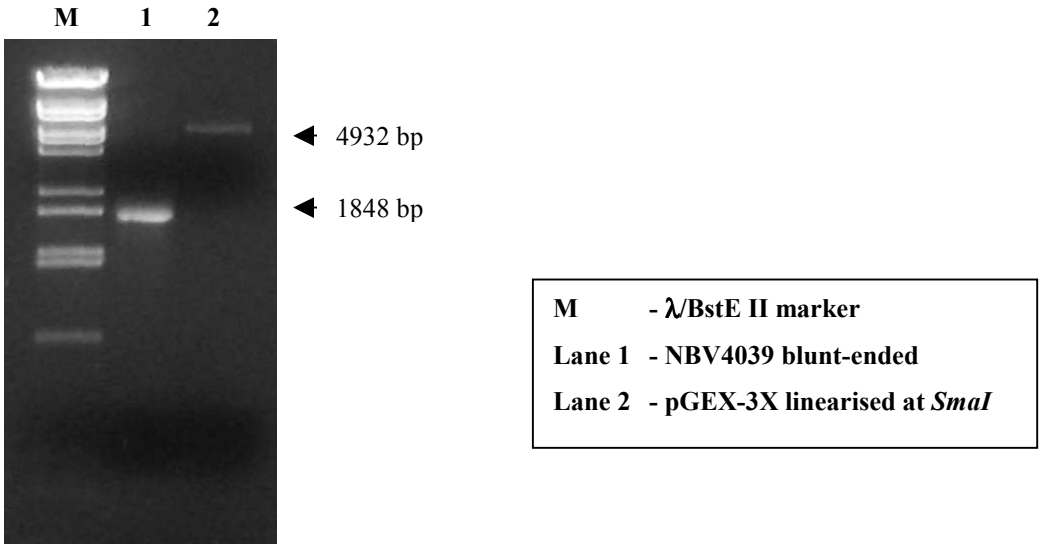
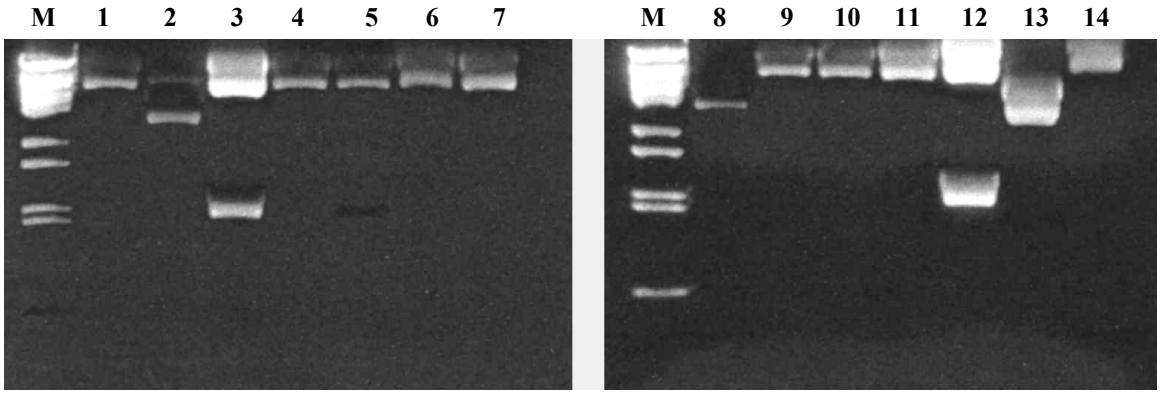


Figure 5-2: 1% agarose gel demonstrating the concentration of the capsid gene fragment and the fusion protein vector prior to the ligation reaction.

The two fragments (2 µl of insert: 1 µl of vector) were placed in a ligation mix overnight at room temperature. The following day the ligation mix was transformed into DH5α cells, plated out and incubated overnight at 37^{O}C with control plates. The total plate count was 32 colonies. Fourteen colonies were picked off and grown overnight at 37^{O}C. "Smart-preps" were done on the overnight broth for plasmid isolation. The plasmids were checked for cloned inserts using *EcoRI* (Table 5-1) and run on a 1% agarose gel for visualisation of bands (Fig. 5-3).

Table 5-1: Fragment sizes of *EcoRI* restriction digest of pGEX4039.

Restriction Enzyme digest of pGEX4039	Fragment Sizes for Correct Orientation (bp)	Fragment Sizes for Reverse Orientation (bp)	Fragment size for no insert (bp)
EcoRI	5348/ 1347/ 105	5047/ 1347/ 406	4952



M	**- λ/BstE II marker**
Lane 1-14	**- Smart-preps digested with *EcoRI***

Figure 5-3: *EcoRI* restriction digest to identify the plasmid with the capsid gene cloned in the correct orientation.

The above gel was not very clear but from the results it looked like lanes 3, 5 (lane 5 had a band of the correct size but did not appear in the digital photograph) and 12 could contain the insert cloned in the correct orientation. The plasmid extract in lane 14 was also checked to confirm that there was plasmid only. The plasmid extract of lanes 3, 5, 12 and 14 were digested again in *EcoRI* and *PstI,* but using less DNA to confirm that the clones contained inserts with the correct orientation (Table 5-2). The digest were run in a 1% agarose gel (Fig. 5-4).

Table 5-2: Fragment sizes of *EcoRI & PstI* restriction digest of plasmid extracts 3, 5, 12 & 14 to confirm correct orientation.

Restriction Enzyme of plasmid extracts 3,5, 12 & 14	**Fragment Sizes for Correct Orientation (bp)**	**Fragment Sizes for Incorrect Orientation (bp)**	**Fragment Sizes for no insert**
EcoRI	5348/ 1347/ 105	5047/ 1347/ 406	4952
PstI	5154/ 1646	4674/ 2126	4952

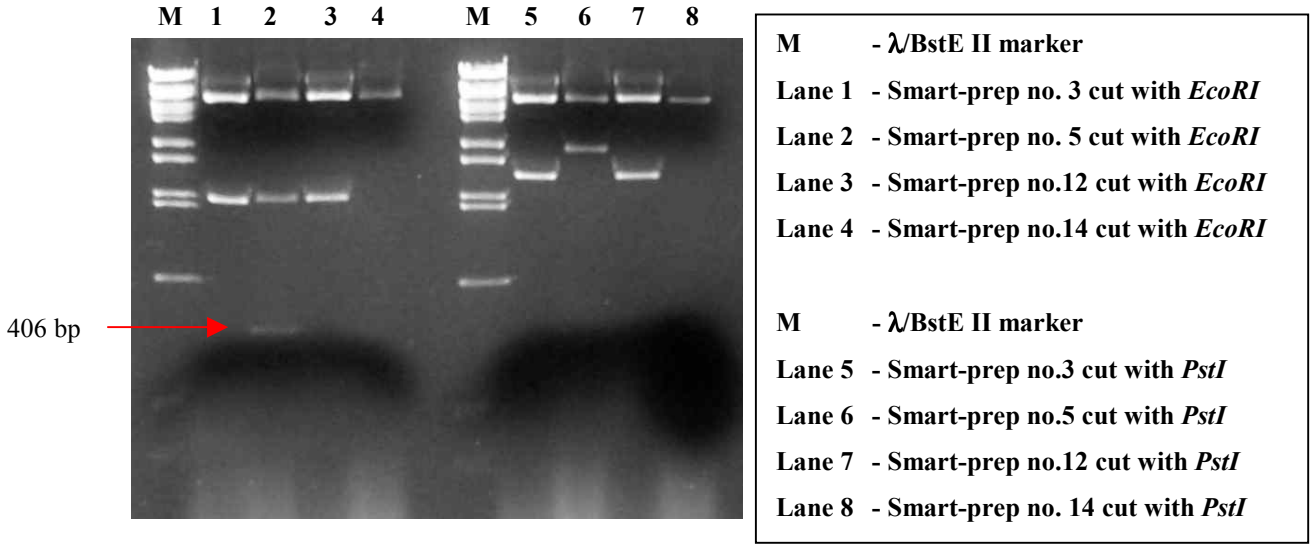


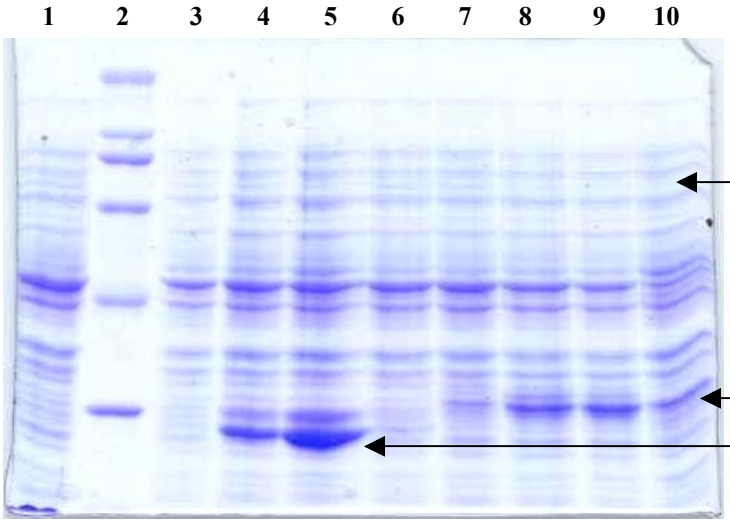
Figure 5-4: *EcoRI & PstI* restriction digest confirming that plasmid extracts 3,5, 12 & 14 had the cloned insert in the correct orientation.

The results from the above gel indicate that the capsid gene had been successfully cloned into the pGEX vector in the correct orientation in extracts 3 and 12. Extracts 4 had the insert but in the incorrect orientation. It was also confirmed that prep 14 contained the pGEX-3X vector but no cloned insert. It was decided to use this transformed vector as the control strain, as it had no cloned capsid gene. The DH5α bacterial strain with plasmids pGEX4039 and pGEX-3X were grown up overnight in a broth and stock cultures were made that were stored at -80^{O}C.

5.2.2. Expression of Capsid Protein with GST Fusion Protein in DH5α Cells

Two overnight cultures of pGEX4039 and pGEX-3X in DH5α were transferred to fresh broth flasks. Once the correct OD at 600nm was reached, IPTG was added to the cultures, which derepressed the p_{tac} promoter and started expression of the GST fusion protein. Samples were taken every hour for the first six hours and after 16 hours. These were all stored at -20^{O}C. After the expression experiment had been completed, the samples were run in two 12% acrylamide gels. One gel was stained with Coomassie stain and the other

was transferred by Western Blotting to a nitrocellulose membrane. The total molecular weight of the expressed capsid precursor protein linked to the fusion protein was 93 900 Da (precursor capsid protein - 66 400 Da + GST protein 27 500 Da). The Coomassie stained gel was then destained and placed on a light box to visualise the protein bands (Fig. 5-5).



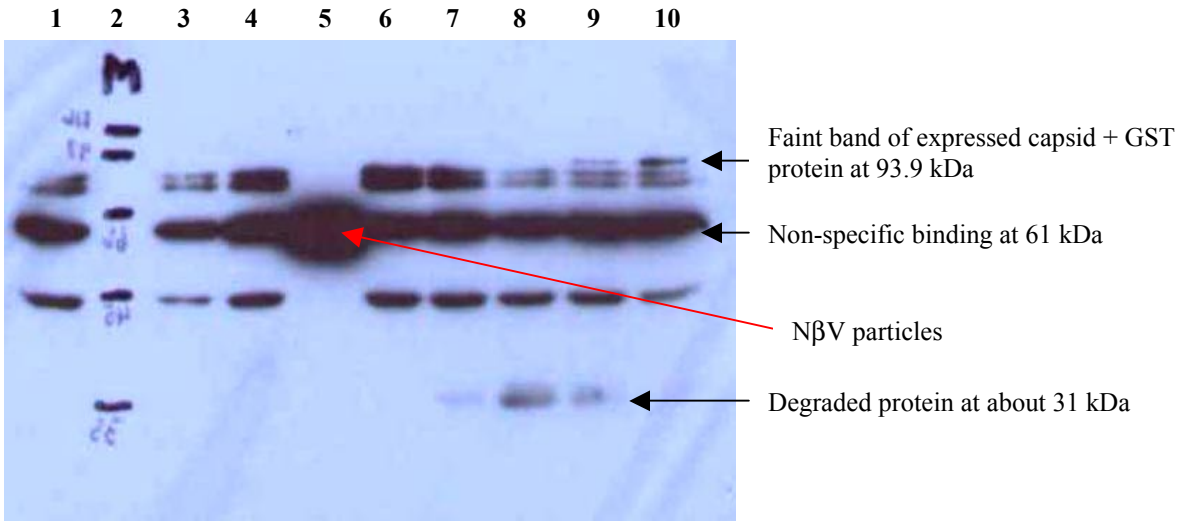
Lane 1	**- DH5α only**
2	**- BIORAD Molecular Weight Marker**
3	**- pGEX-3X (Uninduced)**
4	**- pGEX-3X (Induced for 1 hr)**
5	**- pGEX-3X (Induced for 6 hrs)**
6	**- pGEX4039 (Uninduced)**
7	**- pGEX4039 (Induced for 1 hr)**
8	**- pGEX4039 (Induced for 2 hrs)**
9	**- pGEX4039 (Induced for 6 hrs)**
10	**- pGEX4039 (Induced overnight)**

Note: BIORAD molecular weight marker fragments sizes are available in Appendix 2, p.102

Figure 5-5: Extracted protein from the pGEX4039 expression experiment run in a 12% acrylamide gel stained with Coomassie stain.

The over expression of the GST protein by the plasmid pGEX-3X indicated that the conditions in the cell were favourable for protein production. The above Coomassie gel also suggested that the protein expressed by the DH5α bacterial strain was degrading to a

31-kDa protein. This could only be confirmed using Western Blot, which would flag the degraded protein with NβV antibodies. The other 12% acrylamide gel was transferred using western blot to a nitrocellulose membrane. The membrane was incubated in a 1:500 dilution of NβV antibody and then incubated with an anti-rabbit POD-labelled secondary antibody. The POD (horseradish peroxidase) catalyses the oxidation of luminol present in the detection reagent. This causes an activated intermediate reaction product to form, which decays to the ground state by emitting light. The emitted light from the nitrocellulose membrane is then detected on an X-ray film, the film being exposed to the membrane for 10 minutes (Fig. 5-6).



Lane 1	**- DH5α only**
2	**- BIORAD Molecular Weight Marker**
3	**- pGEX-3X (Uninduced)**
4	**- pGEX-3X (Induced for 6 hrs)**
5	**- Positive Control (NβV)**
6	**- pGEX4039 (Uninduced)**
7	**- pGEX4039 (Induced for 1 hr)**
8	**- pGEX4039 (Induced for 4 hrs)**
9	**- pGEX4039 (Induced for 6 hrs)**
10	**- pGEX4039 (Induced overnight)**

Figure 5-6: X-ray film of a Western blot developed using chemiluminescence demonstrating the expression of Capsid + GST protein in DH5α cells.

Besides all the unspecific binding of the NβV antibody that only appeared after long exposures, a faint band appeared on the Western blot membrane in the 94-kDa region, which increased over time, suggesting that expression occurred in DH5α. The increasing band at 31 kDa on the Coomassie gel, suggested that degradation of the capsid/fusion protein was occuring. Therefore, the probable reason the capsid/fusion protein band was so faint was because of degradation by the DH5α bacterial strain. As the production of eukaryotic protein in *E.coli* produces polypeptide structures, which are foreign to the prokaryotic cell, protease activity increases causing degradation of the expressed protein. The non-specific binding in lanes 1, 3, 4, 6-10 at 61 kDa detected on the Western blot could likely be due to polyclonal antibodies produced by the rabbit, which are binding to *E.coli* protein (G. Blatch, pers. comm.).

It was therefore suggested that the plasmid pGEX4039 and pGEX-3X be transformed into BL21-CodonPlus- (DE3)-RIL bacterial strain, thereby increasing the expression of the capsid/fusion protein.

5.2.3. Transformation and Expression of pGEX4039 and pGEX-3X in Epicurian Coli BL21-CodonPlus- (DE3)-RIL Cells

Efficient production of heterologous proteins in *E.coli* is frequently limited by the rarity of certain tRNA species that are more abundant in the organism from which the heterologous proteins are derived. High-level expression of a heterologous protein in *E.coli* can deplete the pool of rare tRNA species and stall translation. BL21-CodonPlus (DE3) -RIL cells contain extra copies of the *argU, ileY* and *leuW* tRNA genes, which encode tRNA species that recognise the arginine codons AGA and AGG, the isoleucine codon AUA, and the leucine codon CUA. The availability of these tRNA species allows for the high-level expression of many eukaryotic recombinant genes that are poorly expressed in conventional prokaryotic strains. Another feature of this bacterial strain is that they naturally lack the Lon protease, which degrades recombinant protein. In addition, this strain has been engineered to be deficient in a second protease, the OmpT protein (Stratagene catalogue, 1999).

The purpose of this investigation was to over express the capsid protein in order to confirm that the sequence in pNBV4039 was in fact the authentic capsid ORF of NβV genome. From the expression experiment in DH5α cells, there was some indication that the capsid protein was being expressed. To enhance this expression, the pGEX4039 was transformed into BL21-CodonPlus (DE3)-RIL cells. The pGEX4039 and pGEX-3X were transformed into new cells and grown up in broth overnight. Alkaline-lysis plasmid isolation preps were done on the broths and 1µl of the prep digested with *PstI* restriction enzyme. BL21-CodonPlus (DE3)-RIL strain contains the *ColE1*-compatible pACYC-based plasmid. This plasmid contains extra copies of the *argU, ileY,* and *leuW* tRNA genes. It is 3582 bp in size and could be observed when analysing the mini-prep by gel electrophoresis. The pGEX4039 digested with *PstI* was expected to produce two bands of 5154 bp and 1646 bp, while the pGEX-3X was expected to produce only one band of 4952 bp. The digests were run in a 1% agarose gel for visualisation of the restriction fragments (figure 5-7).

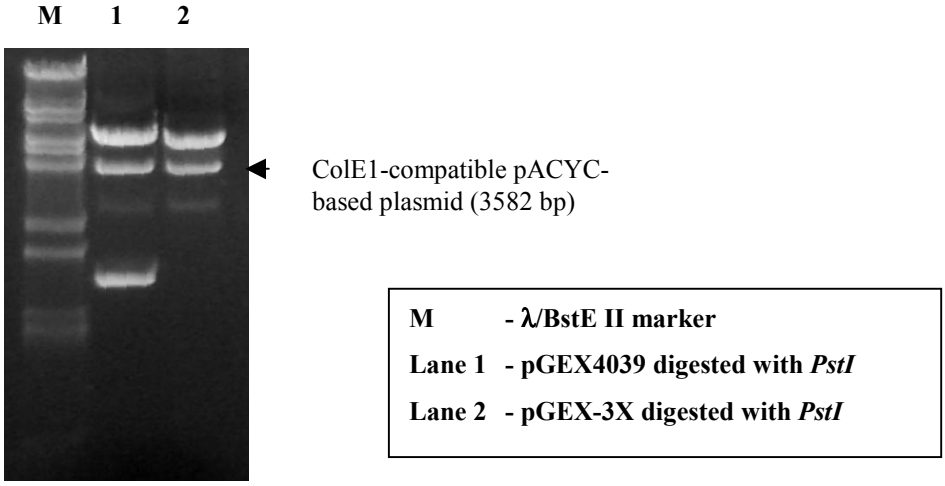
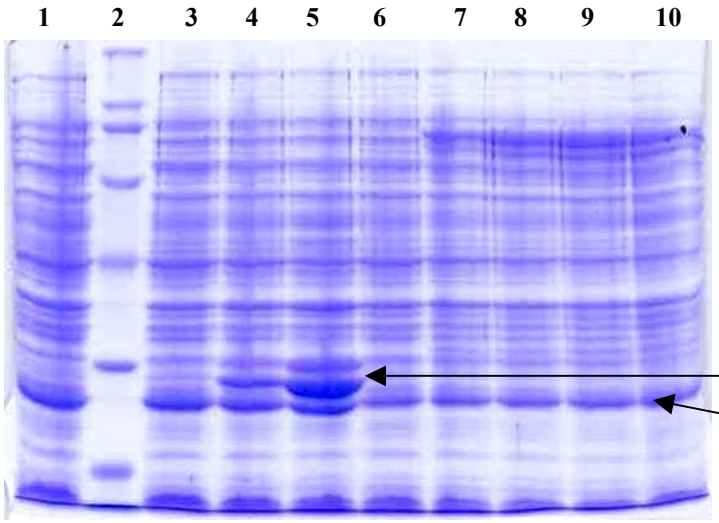


Figure 5-7: 1% agarose gel of *PstI* digests of pGEX4039 & pGEX-3X. The plasmids were isolated from BL21-CodonPlus (DE3)-RIL cells.

Before inducing the expression of the capsid/fusion protein, an overnight culture of BL21-CodonPlus (DE3)-RIL cells, containing pGEX4039, was set up in the presence of the selective antibiotic for pGEX, namely ampicillin. Chloramphenicol (34 µg/ml) was added to the culture to ensure the presence of the pACYC-based plasmids. The overnight culture was then added to a fresh culture flask and allowed to grow at 37^{O}C until the

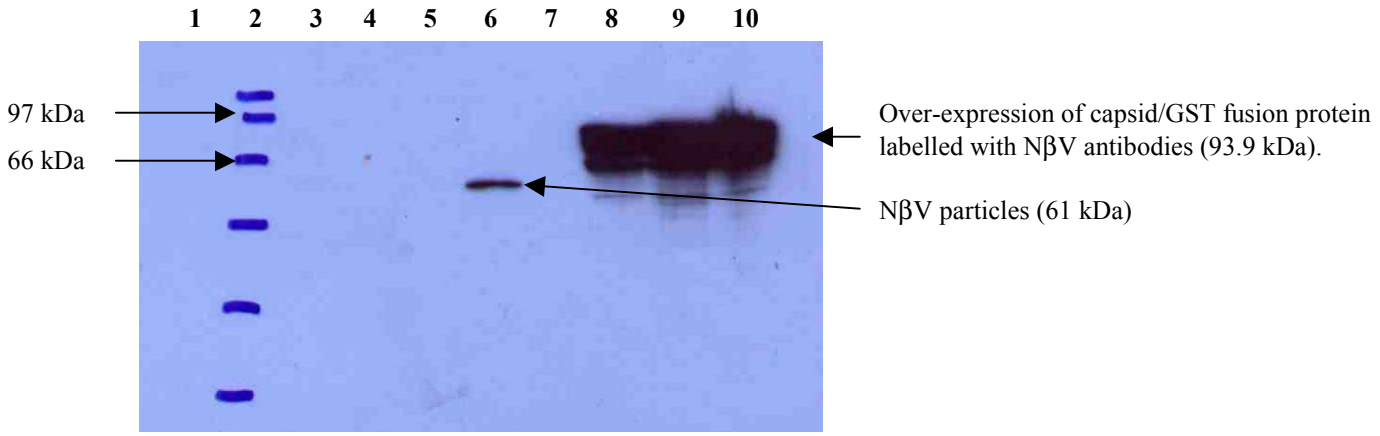
required OD at 600nm had been reached. IPTG was then added to induce expression and hourly samples were taken for the first six hours, after which one sample was taken overnight. The samples were then prepared in SDS-PAGE loading buffer and boiled for 5minutes. After loading the samples were run in two 12% resolving acrylamide gels. The one gel was stained in a Coomassie stain for two hours and then destained to visualise the protein bands (figure 5-8).



Lane 1	**- BL21-CodonPlus (DE3)-RIL cells only**
2	**- BIORAD Molecular Weight Marker**
3	**- pGEX-3X (Uninduced)**
4	**- pGEX-3X (Induced for 1 hr)**
5	**- pGEX-3X (Induced for 6 hrs)**
6	**- pGEX4039 (Uninduced)**
7	**- pGEX4039 (Induced for 1 hr)**
8	**- pGEX4039 (Induced for 4 hrs)**
9	**- pGEX4039 (Induced for 6 hrs)**
10	**- pGEX4039 (Induced overnight)**

Figure 5-8: 12% acrylamide gel stained with Coomassie stain demonstrating the protein bands from the pGEX4039 capsid/ GST expression experiment in BL21-CodonPlus (DE3)-RIL cells.

A band could be seen on the Coomassie stained gel at 93,9 kDa suggesting the over-expression of capsid/GST protein. This result could only be confirmed by Western blot analysis. The increased expression of the GST protein indicated that conditions were improved for protein expression. Another interesting result was that the 31 kDa band present in the DH5α expression experiment was not present. The second 12% resolving gel protein bands were transferred to a nitrocellulose membrane by Western blot. The membrane was incubated with primary and secondary antibodies and then visualised using chemiluminescence and developed onto an X-ray film. The primary antibody concentration was 1:500 and the X-ray film was exposed in a cassette for 5 seconds (figure 5-9).



Lane 1	**- BL21-CodonPlus (DE3)-RIL cells only**
2	**- BIORAD Molecular Weight Marker**
3	**- pGEX-3X (Uninduced)**
4	**- pGEX-3X (Induced for 1 hr)**
5	**- pGEX-3X (Induced for 6 hrs)**
6	**- Positive Control (NβV particles)**
7	**- pGEX4039 (Uninduced)**
8	**- pGEX4039 (Induced for 4 hrs)**
9	**- pGEX4039 (Induced for 6 hrs)**
10	**- pGEX4039 (Induced overnight)**

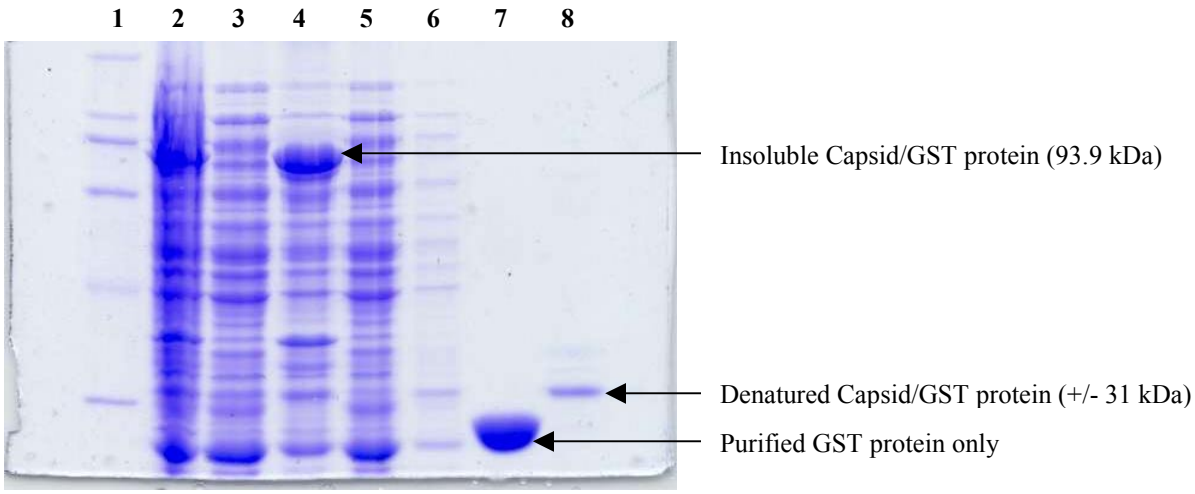
Figure 5-9: X-ray film of a Western blot developed using chemiluminescence demonstrating the expression of Capsid + GST protein in BL21-CodonPlus (DE3)-RIL cells.

It was clear from the Western blot analysis that the prominent protein band on the Coomassie gel was due to the over-expression of the capsid protein fused to GST. The X-ray film could not be exposed to the membrane for longer than 5 seconds due to the strong light emission from the over-expressed band. The Western blot also confirmed that the antibody was binding specifically to the capsid protein as the other lanes with over-expression of the GST protein only, had no light emission on the X-ray film.

This experiment was taken one-step further to determine if the capsid protein, fused to the GST protein, was soluble in the bacterial cells. GST agarose slurry-beads were used to bind the soluble capsid/GST protein from solution after the cells had been sonicated. This method would purify the expressed protein from the bacterial cells and confirm the results obtained from this section's expression experiment. (Section 5.2.3.)

5.2.4. Purification of the Capsid/GST Fusion Protein from BL21-CodonPlus (DE3)-RIL cells

In the expression experiment, using BL21-CodonPlus (DE3)-RIL cells, it was decided that after induction with IPTG the optimal expression of capsid/GST protein occurred after 6 hours. Purification of the GST fusion protein alone was done as a control. The purification of the protein required that a new 250 ml LB broth flask was inoculated and allowed to grow until the required OD had been reached. The bacteria were then induced with IPTG to activate expression and allowed to incubate for 6 hours. After 6 hours the 250 ml culture was spun down and the capsid/GST protein was purified by the procedure described in chapter 2. During the course of the purification procedure, samples were taken at different stages to determine if the protein was not removed from the cells during sonication, thereby indicating possible insolubility. At the end of the procedure the GST beads were mixed with an elution buffer to elute the capsid/GST protein from the beads. The beads were then spun down. The supernatant and the samples taken during the course of the purification were mixed with SDS-PAGE loading buffer and electrophoresed in two 12% resolving gels. The one gel was stained in Coomassie stain and the other prepared for Western blot analysis. After destaining the Coomassie gel was placed on a light box in order to visualise protein bands (Fig. 5-10).



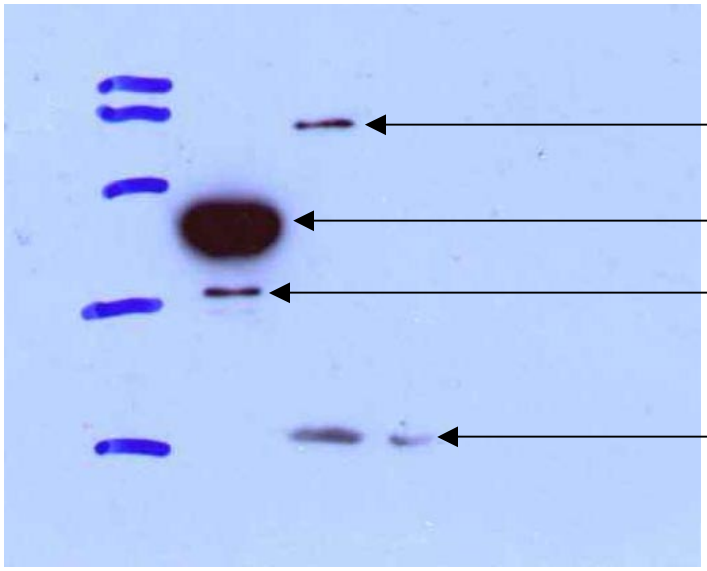
Lane 1	**- BIORAD Molecular Weight Marker**
2	**- Sample taken from total resuspend cells mass in PBS**
3	**- Sample taken from supernatant after cell were spun down after sonication**
4	**- Sample taken from the sonicated total cell mass resuspended in PBS**
5	**- Sample taken from supernatant of GST-slurry bead wash no.1**
6	**- Sample taken from supernatant of GST-slurry bead wash no.2**
7	**- Purified GST fusion protein eluded from slurry beads**
8	**- Purified Capsid/GST protein eluded from slurry beads**

Figure 5-10: 12% acrylamide gel stained with Coomassie stain demonstrating the proteins present at the various purification stages of capsid/GST protein.

The above gel showed some interesting results. In lane 4 it was clear that large amounts of expressed protein remained bound in the *E.coli* cells. Either the protein was not soluble or the protein was not released during sonication. Another theory could be that the protein was binding to the cell membrane and not released. The membrane-bound theory was supported by the high levels of denatured protein (lane 4) and the low levels of capsid/GST protein that bound to the GST-slurry beads and which eluted out in the elution buffer (lane 8). In the previous section, whole cells were denatured using the SDS-PAGE loading buffer before electrophoresis, and showed little denaturation (Fig. 5-8), indicating that the expressed protein was not being denatured in the cells. Also the control experiment using GST fusion protein demonstrated a large degree of purification,

suggesting that the sonication of the cells was enough to release the protein and not cause denaturation.

To determine the level of capsid/GST protein that had been purified, the other 12% resolving gel was transferred to a nitrocellulose membrane using the Western blot technique, using the same protocol used in the previous experiments. The primary antibody concentration remained at 1:500. The X-ray film was exposed for 1 minute and developed for visualisation of the protein bands (Fig. 5-11).



Lane 1	**- BIORAD Molecular Weight Marker**
2	**- Positive Control (NβV)**
3	**- Purified Capsid/GST protein after elution from beads (20 µl of sample was loaded)**
4	**- Purified Capsid/GST protein after elution from beads (10 µl of sample loaded)**
5	**- Purified GST protein only (20 µl loaded)**
6	**- Purified GST protein only (10 µl loaded)**

Figure 5-11: X-ray film demonstrating the Western blot of a 12% resolving gel electrophoresed with purified capsid/GST protein and GST protein only.

The Western blot confirmed that low levels of capsid/GST-expressed protein were being purified. The film also showed even amounts of degraded capsid/GST proteins, which suggests that denaturation had occurred. The purification also confirmed that the expressed protein was NβV capsid precursor protein, as the NβV antibodies bound specifically to the expressed polypeptide protein translated from pGEX4039, thereby confirming the authenticity of the capsid gene open reading frame.

5.3. DISCUSSION

Removing the NβV coat protein ORF from pNBV4039 and cloning it into a prokaryotic expression vector that expressed the cloned insert as a fusion protein, led to some interesting results. From the expression experiment in DH5α cells, the level of expression was low due to possible proteolytic degradation in the cells. It was clear on the Western blot that the level of degradation lessened over time and the level of capsid/GST protein increased. This could have been due to lowered cell metabolism, which was caused by nutrient depletion in the broth and thereby lowering of the proteolytic activity. The level of degradation was clear in the Coomassie stained gel, which had a band of degraded protein at about 31000 Da. Interestingly enough, in all the experiments where denaturation occurred, the capsid/GST protein was degraded to about 31 000 Da while still retaining its reactivity with antibodies to NβV.

A possible theory could be that the GST fusion protein of 27 500 Da, which is soluble and stable in prokaryotic cells, was not affected by protease activity. This fusion protein could have retained a protease-resistant part of the capsid protein after protease digestion. The action of the proteases is targeted to specific amino acid bonds from the carboxy-terminus, which therefore breaks the polypeptide chain at specific amino acids from the C to the N-terminus. It is possible that a few of the amino acids attached to the GST protein were not affected by these proteases and were recognised by the antibodies.

When the plasmid vector pGEX4039 containing the capsid gene was placed into an *E.coli* strain with lowered proteolytic activity (BL21-CodonPlus (DE3)-RIL cells), the level of

complete 93,9 kDa polypeptide structures increased remarkably. This suggested that the protease activity in the DH5α had a significant influence in the degradation of the capsid/GST protein. Another factor that could also have had a significant affect on the level of expression from the different expression experiments was the raised level of rare tRNAs in BL21-CodonPlus (DE3)-RIL cells. The lack of these rare tRNAs in prokaryotic cells affects protein accumulation levels, mRNA and plasmid stability and, in extreme cases, inhibits protein synthesis and cell growth due to premature termination during translation (Baneyx, 1999).

From the results of the purification experiment, another possible theory could be that the capsid proteins are membrane bound to the endoplasmic reticulum (ER) in the insect making the polypeptide chain more stable. Before some proteins are folded into their mature structures, they are very unstable in the cytoplasm of the *E.coli* (Baneyx, 1999). Protein purification from the cells requires sonication causing the release of the capsid/GST protein. Due to the instability of the capsid polypeptide chains, denaturation of the NβV capsid protein occurred up to a region close to the stable GST protein thereby creating a "stable" denatured capsid/GST protein of about 31 000 Da. Another interesting experiment that could be done at a later stage would be to determine at which stage during the purification procedure the degraded protein started to become visible. This would also indicate if the high frequency sound of sonication had any influence on the degradation of the expressed protein.

Even though there were levels of degradation of the capsid/GST protein, the results confirmed that the predicted open reading frame for the expression of the NβV capsid protein is correct. The NβV antibodies that bound to the over-expressed protein produced in a heterologous system confirmed that the precursor protein had NβV immunoglobulin binding sites.

CHAPTER 6

Final Discussion & Future Prospects

6.1. Full-Length cDNA of the NβV Genome

Attempting to extract a single-stranded RNA template of 6.6 kb from larvae that have been stored for over two years at -20^{O}C has been an extremely difficult task and has led to the use of an alternative method of constructing a full-length copy of the genome. The only alternative was to use the available sequencing clones that overlapped, and sub-clone pieces together or to use the PCR linking technique. After a number of sub-cloning attempts, the abundance of restriction enzyme sites scattered across the NβV genome resulted in this technique being aborted. The only other method was the PCR linking technique. The obvious negative feature of using PCR to assemble pieces together was the possible mutations that could occur. This procedure had to be altered by reducing the number of cycles and using a proofreading polymerase to reduce the number of possible mutations to the NβV genome. With the alterations, the linking procedures become very effective. The overlapping inserts linked with remarkable ease and when checked with restriction enzymes, the different digests produced the expected fragments. The only changes that were noted were present in the original plasmids prior to the linking procedure. Sequencing the fragment in both directions to isolate the possible mutations could not be carried out due to lack of available time prior to writing this thesis. This would clearly be one of the future prospects for the full-length PCR clone.

Once pGEM6625 has been sequenced and checked against the original sequence published by Gordon *et al.* (1999), the mutations would then be corrected using a site-directed mutagenesis technique. The plan would then be to use the corrected full-length copy of the NβV genome to express the replicase and capsid proteins. Cloning the NβV genes into a eukaryotic expression vector would be an interesting experiment. Previous research done on other members of the *Tetraviridae* has resulted in the maturation of

virus-like particles in eukaryotic cells following the expression of the virus capsid gene. Either using a baculovirus expression vector in an insect cell line or cloning into a yeast expression vector and inducing expression could yield valuable insights into the molecular biology of NβV. This should induce the production of full-length genomic RNA templates encapsidated in virus-like particles. Virus extraction and RNA studies would follow to determine that the cloned copy of the NBV genome is re-creating single-stranded mRNAs that are imitating the original virus RNA strand (producing genomic and sub-genomic RNA). It would also be interesting to determine if the extracted virus particles containing the genomic RNA strand would be infectious if fed to the *Nudaurelia cytherea capensis* larvae.

Firstly, one would need to express the replicase protein; this would confirm the start and stop co-ordinates of the replicase gene. This could be done in the same way as was used for identifying the ORF of the capsid gene. By cloning the replicase gene into a prokaryotic expression vector, one could induce the gene to express a replicase protein of 215 kDa. This would also confirm the fidelity of the full-length clone that had been corrected according to the published NβV sequence.

The second approach would be to analyse the mRNA transcribed from the full-length cloned copy. This would require that mRNA be extracted from transfected eukaryotic cells once sub-genomic transcription had been induced. This would produce fresh RNA that could be extracted and analysed on RNA denaturing gels using Northern blotting. The Northern blots would identify the genome and sub-genomic RNA templates. cDNA copies of the sub-genomic RNA could be sequenced to identify the exact sub-genomic transcriptional start site. This would also help to isolate the transcription promoters for subgenomic and genomic RNA production. The replication strategy of the omegatetraviruses is understood in some detail largely due to the HaSV work. The little that is known about the betatetraviruses is what has been deduced from the NβV sequence. Having a full-length clone of NβV will hopefully enable us to answer questions (as mentioned above) about the NβV molecular biology and replication strategy.

6.2. NβV Capsid Protein Expression

The expression of the NβV precursor protein in a heterologous system indicated that the cloned open reading frame for the capsid protein of the NβV genome was functional. The expression experiment was characterized by the degradation of the expressed proteins in *E.coli* DH5α. This indicated that the overproduction of precursor protein was unstable, probably due to misfolding and resultant segregation into insoluble aggregates that were degraded by proteases.

A follow-up experiment that could be done would be to slow the metabolism of the cell down by lowering the growth culture temperature. The earlier experiment in DH5α cells showed that depletion of nutrients over time in a Luria broth culture resulted in a lowered metabolism, which increased the level of expression of the 93.9 kDa protein. By lowering the temperature to 25^{O}C, one could possibly improve the quality of the precursor protein and reduce its degradation in the cytoplasm. Another interesting result would be to clone only the ORF of the capsid gene without a fusion protein into a prokaryotic expression vector and transform into BL21-CodonPlus (DE3) -RIL cells. One would then induce the cells at a lower culture temperature, taking samples every hour. Analysis of the expressed proteins could be done on a PAGE gel using Western blot and comparing to authentic NβV particles. This would indicate that the precursor protein of 66 kDa level undergoes maturation and possible post-translational cleavage into mature capsid proteins of 61 and 5 kDa. Agrawal and Johnson published in 1995 that the changes in the size of the protein during maturation are caused by post-translational cleavage into the minor and major protein.

The next step would be to take the capsid gene and clone it into a eukaryotic expression system. Work done by Agrawal and Johnson in 1995, showed that encapsidation of the mRNA occurred during the assembly of NωV viral capsids in a baculovirus expression system. Johnson and colleagues (Johnson *et al.,* 1989) listed the functions of viral structural proteins: (i) the specific recruitment of viral nucleic acid, (ii) assembly around this nucleic acid into a stable form, (iii) the specific recognition of and binding to a susceptible cell, and (iv) release of the viral genome into a cell environment that will

support replication. The above points are relevant when studying the capsid structures of the NβV. Therefore it would be beneficial to study the NβV capsid structures using a baculovirus expression system, *in vivo,* which provides for the compartmentalization and post-translational machinery found in cell types in which they are normally synthesized. Another interesting aspect that could be determined with the capsid structure research would be to compare the differences of cloning the full-length copy and cloning the capsid gene only and the different capsid structures that are produced. This would determine if there are possibly untranslated regions (UTR) that could have an affect in the assembly of the 240 identical subunits into an icosahedral capsid structure.

6.3 Future Prospects

The past 3 years of research on the NβV has shown some significant advances for future work that would help in understanding the molecular biology and replication of the virus. The confirmation of the authenticity of the coat protein ORF would direct research into further analysis on the NβV icosahedral capsid structure. It would be interesting to determine if the NβV has capsid protein surface domains similar to those of the HaSV, that are implicated in cell binding, as belonging to the Ig protein superfamily (Hanzlik & Gordon, 1997). This would have practical applications in medical biotechnology research.

Further heterologous expression of the replicase protein would confirm the fidelity of the NβV sequence and help understand the replication and assembly of monopartite genome tetraviruses. The replicase protein could also be used to manufacture antiserum from rabbits for future *in vivo* work on NβV replication. HaSV is at present the only tetravirus that can readily be obtained from laboratory-reared insects (Hanzlik & Gordon, 1997). The prospects of cloning the full-length genome of NβV into a baculovirus expression system has much significance for the future.

APENDICES AND REFERENCES

APPENDIX 1: MATERIALS

Table 1: List of materials and their suppliers

PRODUCTS	SUPPLIER
λ DNA	Roche Molecular Biochemicals
Acrylamide	Promega Corporation
Agar-Agar Powder	Biolab Diagnostics (PTY) Ltd
Agarose	Promega Agarose LE, Analytical Grade
Ampicillin Powder	Roche Molecular Biochemicals
APS	Promega Corporation
Bis-Acrylamide	Roche Molecular Biochemicals
Boric Acid	Saarchem (PTY) Ltd
Bromophenol Blue	Merck Laboratory Suppliers (PTY) Ltd
Carbon Tetrachloride (CCl_4)	Saarchem-Holpro Analytic (PTY) Ltd
Casium Chloride	Roche Molecular Biochemicals
Chloramphenicol Powder	Roche Molecular Biochemicals
Coomassie Brilliant Blue G	Sigma Chemical Co
DEPC	Fluka
dNTP's	Roche Molecular Biochemicals
EDTA	Merck Laboratory Suppliers (PTY) Ltd
Ethanol	Saarchem (PTY) Ltd
Formaldehyde	Merck N.T. Laboratory Suppliers (PTY) Ltd
Glacial Acetic acid	Saarchem (PTY) Ltd
Glucose-D	Saarchem (PTY) Ltd
Glycerol	Saarchem (PTY) Ltd
Glycine	Sigma Chemicals Co
HCl	Saarchem (PTY) Ltd
IPTG	Roche Molecular Biochemicals
Low Melting Point Agarose	Whitehead Scientific Agarose Low Melt LMZ
Lysozym	Roche Molecular Biochemicals
β-Mercaptoethanol	Merck Laboratory Suppliers (PTY) Ltd
Methanol	Saarchem (PTY) Ltd
Mops	Roche Molecular Biochemicls Mops
Nacl	Saarchem (PTY) Ltd
NaOH	Saarchem (PTY) Ltd
Nitrocellulose Paper	Amersham Pharmacia Biotech Hybond-C Extra Nitrocellulose Paper
Phenol	Saarchem (PTY) Ltd
SDS	Roche Molecular Biochemicals
Sodium Acetate	Merck Laboratory Suppliers (PTY) Ltd
Sucrose	Merck Laboratory Suppliers (PTY) Ltd
TEMED	Promega Corporation
Tris Base	Roche Molecular Biochemicals Tris
Tryptone	Biolab Diagnostics (PTY) Ltd
TWEEN 20	Merck Laboratory Suppliers (PTY) Ltd
X-Gal	Roche Molecular Biochemicals
Xylene Cyanol	Electran BDH Chemicals (PTY) Ltd
Yeast Extract	Biolab Diagnostics (PTY) Ltd

APPENDIX 2: VIRUS PURIFICATION AND ANALYSIS

1. CENTRIFUGATION

❖ The length of time to spin was calculated as follows:

$$t(\text{hrs}) = \frac{k}{S}$$

Where:

t = sedimentation time (hours)

S = sedimentation coefficient (Svedberg unit), which is the velocity, v (cm/sec), of the particle in a gravitational field of unit intensity (1cm/sec^2).

k = factor for calculating the time required to sediment a particle of known S value in a particle rotor at its maximum speed. k value for fixed angle rotors, refer to table below.

k factor for 65 fixed-angle rotor = 45

S value for NβV = 210

Thus:

$$t = \frac{45}{210} = 0.21 \text{ hours or } 13 \text{minutes to pellet N}\beta\text{V particles at } 65{,}000 \text{ RPM}$$

Table 2: k Factors for Fixed-Angled Rotors

ROTOR	MAX. RPM	r_{max} (cm)	r_{min} (cm)	k AT MAX. SPEED
75 Ti	75,000	8.0	3.7	35
65	65,000	7.8	3.68	45
60 Ti	60,000	9.0	3.7	63
50 Ti	50,000	8.1	3.8	77
50	50,000	7.1	3.7	66
42.1	42,000	9.9	3.9	134
40.3	40,000	8.0	4.8	81
40.2	40,000	8.0	3.5	131
40	40,000	8.1	3.8	120
35	35,000	10.4	3.5	225
30.2	30,000	9.4	6.3	113
30	30,000	10.5	5.0	209
21	21,000	12.0	6.0	398
19	19,000	13.3	4.4	776
15	15,000	14.2	4.07	1405
Batch	16,000	9.7	4.4	783

The **k'** is a guide to the time required to move a band of particles to the bottom of the tube through a density gradient. k' factors are used in swinging bucket rotor and the particle density as well the S (sedimentation coefficient) must be known.
For k' values for swinging bucket rotors, refer to table below.

k' factor for SW41 rotor for a NβV particle with a density of 1.3g/ml = 335
S value of NβV = 210

- The length of time to spin NβV particles through a density gradient is calculated as follows:

$$T = \frac{k'}{S} = \frac{335}{210} = 1.6 \text{ hours or } 96 \text{ minutes}$$

∴ It would take 96 minutes to pellet NβV particles through a density gradient at 41,000 rpm.

Table 3: k and k' Factors for Swinging-Bucket Rotors

ROTOR	k' FACTORS									k FACTOR
	PARTICLE DENSITY									
	1.1	1.2	1.3	1.4	1.5	1.6	1.7	1.8	1.9	
SW 65 Ti	261	143	126	120	116	113	112	110	109	46
SW 56 Ti	313	170	150	142	137	134	132	131	129	55
SW 50.1	342	164	162	153	148	145	143	141	140	59
SW 41 Ti	689	379	335	317	307	300	295	292	289	132
SW 40 Ti	754	416	368	349	338	330	325	321	318	137
SW 39	640	348	307	290	281	275	270	267	264	112
SW 36	894	491	435	411	398	390	383	379	375	157
SW 27	1492	816	722	683	661	646	636	629	622	265
SW 27.1	1709	945	838	793	768	752	740	731	724	310
SW 25.2	1862	1024	907	858	830	813	800	790	783	335
SW 25.1	1881	1035	917	867	840	822	809	799	791	338

2. BUFFERS

❖ 0.5 M Tris/0.01 M EDTA, pH7.5 (10X TE stock buffer)

60.55 g Tris
3.72 g EDTA
Adjust pH to 7.5 with conc. HCl
Make up to 1 litre with distilled water.

❖ Purification buffer

To the 66 g of larvae add :-
130 ml 0.5 M Tris/0.01 M EDTA, pH 7.5
33 ml Carbon Tetrachloride (CCl_4) and 0.2% Mercaptoethanol (0.2 ml)

3. CAESIUM CHLORIDE GRADIENT

Caesium chloride (CsCl) solutions were made by dissolving 3.6 g (30%) and 7.2 g (60%) CsCl in 12 ml 0.05 M TE, pH 7.5 respectively. 6.0 ml of each solution was added to a 12.5 ml cellulose acetate centrifuge tube using a gradient mixing device.

4. DETERMINATION OF VIRUS CONCENTRATION

Virus concentration was determined using the equation:

$$\frac{1 \text{ mg/ml}}{\mathbf{E}_{260\text{ nm}}} = \mathbf{3.1}$$

Where **E** represents the extinction coefficient for NβV at 260 nm and at a concentration of 1ug/ml.

5. SDS-POLYACRYLAMIDE GEL ELECTROPHORESIS

❖ Resolving Gel Buffer Stock Solution (1.5 M Tris-HCl, pH 8.8):

18.15 g Tris in 100 ml distilled water, adjusted to pH 8.8 with conc. HCl

❖ Stacking Gel Buffer Stock Solution (0.5 M Tris-HCl, pH 6.8):

6.0 g Tris in 100 ml distilled water, adjusted to pH 6.8 with conc. HCl

❖ Bath Buffer (10X):

15.15 g Tris
72.05 g Glycine
5.00 g SDS
500 ml Distilled water
Dilute 10X before use.

- Acrylamide Stock (30%):
  30 g Acrylamide
  0.8 g Bis-Acrylamide
  100 ml Distilled water

- SDS Loading Buffer:
  3.2 g 10% SDS
  0.8 ml β-mercaptoethanol
  1.6 ml Glycerol (100%)
  0.8 ml 0.2% Bromophenol Blue
  2.0 ml 0.5M Tris-HCl, pH 6.8
  7.6 ml Distilled water
  16.0 ml

  Store at room temperature. **Dilute sample 1:4 and heat to 95^{O}C for 5 minutes**.

- 10% SDS (Sodium Dodecyl Sulphate):
  5 g SDS
  50 ml Distilled water

- 10% APS (Ammonium persulphate):
  0.1 g APS
  1 ml Distilled water
  Prepare fresh each time.

- Resolving or Separating Gel:

Table 4: Resolving Gel Concentrations

	7.5%	**12%**	**15%**
Deionised water	9.7 ml	6.7 ml	4.7 ml
1 M Tris-HCl, pH 8.8	5 ml	5 ml	5 ml
10% SDS	200 ul	200 ul	100 ul
Acrylamide	5 ml	8 ml	10 ml
10% APS	100 ul	100 ul	100 ul
TEMED	20 ul	20 ul	20 ul
Total	20 ml	20 ml	20 ml

Butanol saturated water is layered over the resolving gel to form an immiscible layer.

❖ Stacking Gel (4%):

Table 5: 4% Stacking Gel

Deionised water	**6.1 ml**
0.5 M Tris-HCl, pH 6.8	2.5 ml
10% SDS	100 ul
Acrylamide	1.33 ml
10% APS	100 ul
TEMED	20 ul
Total	10 ml

❖ Acrylamide Plug:
4 ml Acrylamide stock (30%)
100 ul 10% APS
40 ul TEMED

❖ Agarose Plug:
0.04 g Agarose (1%)
1.6 ml 1.5 M Tris-HCl, pH 8.8
4.0 ml

Melt the agarose in microwave.

❖ Staining Solution (Coomassie):
2.0 g Coomassie Brilliant Blue
450 ml Methanol
100 ml Glacial Acetic Acid
450 ml Distilled water

Stirred overnight and then filtered.

❖ Destaining Solution:
450 ml Methanol
100 ml Glacial Acetic Acid
450 ml Distilled water

6. MOLECULAR WEIGHT MARKERS:

Table 6: SIGMA High Molecular Weight Marker for SDS-PAGE Gels (daltons)

Proteins	Approx. Molecular Weights
Myosin	205 000
β-galactosidase	116 000
Phosphorylase	97 400
Bovine Albumin	66 000
Egg Albumin	45 000
Carbonic Anhydrase	29 000

Table 7: BIORAD Molecular Weight Marker for SDS-PAGE Gels (daltons)

Proteins	Approx. Molecular Weights
Myosin	200 000
β-galactosidase	166 250
Phosphorylase	97 400
Serum Albumin	66 200
Ovalbumin	45 000
Carbonic Anhydrase	31 000
Trypsin Inhibitor	21 500
Lysozyme	14 400
Aprotinin	6 500

APPENDIX 3: RNA EXTRACTION

1. DEPC-TREATED WATER (RNASE FREE):

0.5 ml DEPC (Sigma Diethyl Pyrocarbonate, $C_6H_{10}O_5$) was added to 1l distilled water and left to stand overnight. Deactivate DEPC by autoclaving for 30 minutes.

2. TRIS-HCL BUFFER-SATURATED PHENOL:

Melt 200 g phenol at 68^{O}C for +/- 20 minutes. Add 0.1% hydroxyquinoline (0.2 g). An equal volume of 0.5 M Tris-HCl, pH8.0 added and the mixture was stirred on a magnetic stirrer for 15 minutes. The mixture was allowed to stand till the two phases separated**.** The upper phase was removed using a pasteur pipette. An equal volume of 0.1 M Tris-HCl, pH 8.0 was added to the phenol, stirred for 15 minutes, and the upper layer removed as before. The extractions were repeated until the pH of the phenolic phase was > 7.8. Finally, 0.1 volume of 0.1 M Tris-HCl, pH 8.0 containing 0.2% mercaptoethanol was layered on top of the phenol. The Tris-HCl buffer-saturated phenol was stored at 4^{O}C in a dark bottle.

3. DETERMINATION OF RNA CONCENTRATION:

The fact that RNA absorbs at 260 nm enables quantification by the following calculation:

[RNA] ug/ml = OD_{260} x 40

Where: OD_{260} = absorbance at 260 nm
40 = extinction coefficient of RNA

[RNA] ug/ml = 0.089 x 40 x 300 (dilution factor)
= 1068 ug/ml or 1.07 ug/ul

4. RNA GEL ELECTROPHORESIS:

❖ 10X MOPS buffer (pH 7.0):

41.85 g MOPS (0.2M)
2.05 g Sodium Acetate (50mM)
1.86 g EDTA (10mM)
made up to 500 ml with distilled water, adjust pH to 7.0 with NaOH and autoclaved.

❖ 1% Agarose/Formaldehyde gel:

Dissolve 1 g agarose in 85 ml distilled water by microwaving for ~ 2 min. Allow to cool to approximately 65^{O}C. Add 10 ml pre-warmed 10X MOPS buffer and 5 ml pre-warmed Formaldehyde (pH 3.4). Cast gel

- RNA Loading Buffer (10X):
  50% glycerol
  1 mM EDTA, pH 8.0
  0.4% Bromophenol Blue
  0.4% Xylene Cyanol

- Ethidium Bromide:
  A 10 ug/ml stock solution was stored at $4^{O}C$.

- RNA Marker:
  Roche Molecular Weight Marker II
  Size Distribution: 7.4/5.3/2.8/1.9/1.6 kb
  Working Concentration: Mix 1 mg/ml total RNA in H_2O and load 5 µl per lane.

APPENDIX 4: DNA ANALYSIS

1. DNA GEL BUFFERS:

- 10X Tris-Borate-EDTA (TBE) Stock:

  54.0 g Tris Base
  27.5 g Boric Acid
  20 ml 0.5M EDTA (pH 8.0)

  Make up to 1l with distilled water and adjust pH to 8.0 with conc. HCl.

- 10X Tris-Acetic Acid-EDTA (TAE) Stock:

  48.4 g Tris Base
  11.4 ml Glacial Acetic Acid
  20 ml 0.5 M EDTA (pH 8.0)

  Make up to 1l with distilled water and adjust pH to 8.0 with conc. HCl.

- TE (10 mM Tris/HCl, pH 8.0·1 mM EDTA):

  1.21 g Tris Base
  2 ml 0.5 M EDTA (pH 8.0)

  Make up to 1l with distilled water and adjust pH to 8.0 with conc. HCl. Sterilise by autoclaving. (pH the Tris to pH 8.0 before adding the EDTA and top up to 1l).

2. DNA AGAROSE GELS:

- 1% Agarose Gel:

  0.7 g Agarose Powder
  7 ml 10X TBE
  63 ml Distilled water

  Dissolve in microwave for +/- 2 minutes. The gel is allowed to cool before adding 2 ul of 10 ug/ml ethidium bromide solution. Cast gel into a gel tray and allow to set.

- 1% Low-Melting Point Agarose Gel:

  0.7 g LMP Agarose Powder
  7 ml 10X TAE
  63 ml Distilled water

  Dissolve in microwave for +/- 2 minutes. The gel is allowed to cool before adding 2 ul of 10 ug/ml ethidium bromide solution. Cast gel into a gel tray and allow to set.

❖ DNA Loading Buffer (6X):
0.25% Bromophenol blue
0.25% Xylene Cyanol
30% Glycerol

❖ DNA Molecular Weight Marker (Lambda DNA-*Bst*E II):
200 ul λ DNA (0.25 ug/ul)
24 ul 10X Buffer
10 ul BstE II
234 ul (50ug of DNA)

- Incubate at 60^{O}C for 5 hrs.
- Clean with Phenol/IAC (Isoamyl-Chloroform)
- Clean again with IAC

Stock Solution:-
200 ul λ/*Bst*E II (50ug of DNA)
550 ul TE**,** pH 8.0
150 ul Loading Buffer
900 ul

Aliquot into three 1.5 ml Eppendorf tubes (300ul each), one kept at 4^{O}C and the other two stored at -20^{O}C.

Table 8: λ/*Bst*E II Fragment Sizes (bp)

FRAGMENTS	BASE PAIRS
1	8,454
2	7,242
3	6,369
4	5,686
5	4,822
6	4,324
7	3,675
8	2,323
9	1,929
10	1,371
11	1,264
12	702
13	224
14	117

3. PCR (POLYMERASE CHAIN REACTION):

- 10 mM dNTP's:
  Take 30 ul of each nucleotide solution (dATP, dTTP, dCTP and dGTP) into the same Eppendorf tube. Add 180 ul of DEPC-treated water bringing the volume to 300 ul. Aliquot into five new Eppendorf tubes (60 ul into each), store at -20^{O}C.

- General Mix Protocol:
  Two master mixes are prepared as follows.

Master mix I:

1.5 µl	10mM dNTP's
x µl	Template DNA (0.1-0.75ug)
1.0 µl	BSA
2.4 µl	25mM $MgCl_2$
2.0 µl	10 µM Forward Primer
2.0 µl	10 µM Reverse Primer
x µl	DEPC-treated water
30.0 µl	

Master mix II:

6.0 µl	10X Buffer (without $MgCl_2$)
0.9 µl	Polymerase Enzyme Mix
x µl	DEPC-treated water
30.0 µl	

Mix the two mixes together and overlay with mineral oil before placing into thermocycler.

- Thermocycler Settings:

Table 9: PCR Thermocycler Settings

1x	Denature template 1min at 94^{O}C
15x	Denaturation at 94^{O}C for 15s Annealing for 30s at specific primer melting temp. Elongation at 72^{O}C for time duration depending on the length of the template.
1x	Prolong elongation of 7min at 72^{O}C

Elongation temperature depends on the length of amplification product: 72^{O}C are used for amplification up to 3 kb; 68^{O}C are used for amplification >3kb.

Table 10: PCR Elongation Times

Elongation time	**45s**	**1 min**	**2 min**	**4 min**	**8min**
PCR fragment length (kb)	up to 0.75	1.5	3	6	10

❖ Oligonucleotide Primers:

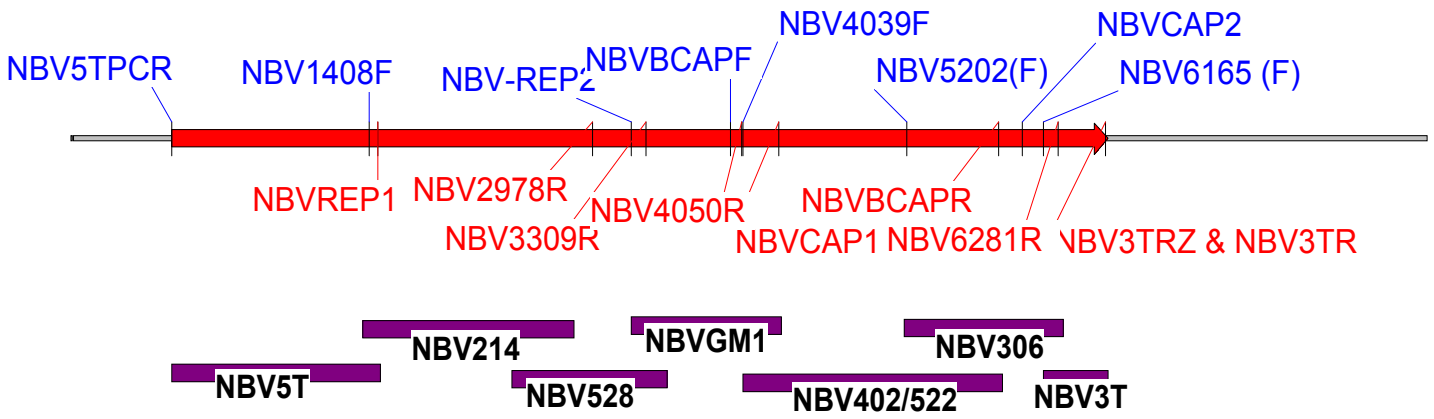


Figure 1: The different PCR primers (forward primers in blue and reverse primers in red) and a map illustrating their binding sites in the NβV genome.

NBV5TPCR F (31-mer) [1-31bp] Tm Value: 68.1 deg C

5' -GATTCTCCCACTTGATTGTACGGAGAACCCG- 3'

NBV1408F (19-mer) [1400-1418bp] Tm Value: 55.0 deg C

5' -CGACGAGGAGGTCGTACGC- 3'

NBVREP1 R (18-mer) [1480-1477bp] Tm Value: 49.2 deg C

5' -TTCACCACTTTATCGGCG- 3'

NBV2978R (19-mer) [2981-2978bp] Tm Value: 53.9 deg C

5' -CTGTCGACTTGCTCGGGC- 3'

NBVREP2 F (18-mer) [3248-3265bp] Tm Value: 50.7 deg C

5' -GAATGGGCAGTCAGACCG- 3'

NBV3309R (22-mer) [3377-3356bp] Tm Value: 47.4 deg C

5' -GACCCAGTCATAAGAACAGAAG- 3'

NBVBCAPF (52-mer) [3957-3995bp] Tm Value: 87.8 deg C

5' -GGGGGGATCCACAATGAGCGACCTACACTTGGACCCCGGTTTTGTCATGCGG-3'

Bam HI

NBV4039F (32-mer) [4039-4056bp] Tm Value: 79.3 deg C

5' -GGGGGGATCCGGAGATGGATGCCAACGTGCAG- 3'

Bam HI ▲ 4039 (Capsid gene start codon) →

NBV4039SD (32-mer) [4039-4056bp] Tm Value: 80.0 deg C

5' -AGGAGGATCCGGAGATGGATGCCAACGTGCAG- 3'

Shine-Dalgarno ▲— 4039 (Capsid gene start codon) →

NBV4050R (18-mer) [4048-4031bp] Tm Value: 52.1 deg C

5' -TGGCATCCATCTTGTCCG- 3'

NBVCAP1 R (18-mer) [4317-4300bp] Tm Value: 42.0 deg C

5' -CGGGTCTAGATAGTCGTG- 3'

NBV5018R (18-mer) [5019-5002bp] Tm Value: 49.2 deg C

5' -CGCAGTCTGGAACACCAG- 3'

NBV5202F (18-mer) [5202-5219bp] Tm Value: 50.0 deg C

5' -CAACTCAATCAAGATTCC- 3'

NBVBCAPR (30-mer) [5857-5877bp) Tm Value: 65.6 deg C

5' -GGGGGGATCCTACAAGCTAAGGTTCCCTAT- 3'

Bam HI ▲— 5877 (Capsid Gene Stop Codon) →

NBVCAP2 F (18-mer) [6024-6041bp] Tm Value: 47.6 deg C

5' -ATCAACTCGACCCAATCG- 3'

NBV6165R (18-mer) [6165-6182bp] Tm Value: 52.0 deg C

5' -CCGCGGTGTTAAATCTTA- 3'

NBV6281R (38-mer) [6297-6275bp] Tm Value: 77.6 deg C

5' -GGGGGGATCCGCACTCTGGGCATGTGTGAAGG- 3'

Bam HI ▲— 6297 →

NBV3TRZ (47-mer) [6625-6608bp] Tm Value: 88.7 deg C

5' -GGCCGGCGCCCGGGCGACATATGCCGGACTGGTACCGTATAGGTCCC- 3'

Partial Ribozyme Sequence ▲— 6625 →

NBV3TR (18-mer) [6625-6608bp] Tm Value: 56.0 deg C

5' -TGGTACCGTATAGGTCCC- 3'

APPENDIX 5: RECOMBINANT DNA TECHNIQUES

1. BACTERIAL GROWTH MEDIA

- Luria-Bertani Broth (LB):

10 g Tryptone
5 g Yeast Extract
10 g NaCl

The above mixture was dissolved in 1l distilled water and autoclaved for 20 minutes.

- LB + Agar (LA) Plates:

5 g Tryptone
2.5 g Yeast extract
5 g NaCl
7.5 g Agar-Agar Powder

The above mixture was dissolved in 500 ml distilled water (+/- 30 plates) and autoclaved for 20 minutes.

- LB + Antibiotic:

10 g Tryptone
5 g Yeast Extract
10 g NaCl

The above mixture was dissolved in 1l distilled water and autoclaved for 20 minutes. Allow media to cool and add 1ml (1µl antibiotic stock per ml media) antibiotic.

- LA + Antibiotic:

5 g Tryptone
2.5 g Yeast extract
5 g NaCl
7.5 g Agar-Agar Powder

The above mixture was dissolved in 500 ml distilled water (+/- 30 plates) and autoclaved for 20 minutes. Allow agar to cool and add 0.5 ml (1µl antibiotic stock per ml media) antibiotic. Antibiotic in media only has 1 month fridge (4^{O}C) life.

❖ SOC medium:

2.0 g	Tryptone
0.5 g	Yeast extract
0.05 g	NaCl
10 ml	250 mM KCl

Dissolve the above in 85 ml distilled water, pH to 7.0 and adjust the volume to 98 ml. Autoclave for 20 minutes. Allow to cool and add 1 ml 2 M $MgCl_2$ and 1ml 2 M glucose stock that has been filter sterilized, bringing each to a final concentration of 20 mM. Filter the complete medium through a 0.2 mm filter unit.

2. STOCK SOLUTIONS:

❖ Ampicillin Stock (100mg/ml):

1 g	Ampicillin Powder

The above was dissolved in 10 ml sterile distilled water. Aliquot into 10 Eppendorf tubes and freeze at -20^{O}C. Working concentration = 100 ug/ml.

❖ Chloramphenicol Stock (34mg/ml):

0.34 g	Chloramphenicol Powder

The above was dissolved in 10 ml of absolute ethanol. Aliquot into 10 Eppendorf tubes and freeze at -20^{O}C. Working concentration = 34 ug/ml.

❖ 1M IPTG (Isopropylthiogalactoside):

1.19 g	IPTG powder

Dissolve in 5 ml distilled water. Filter sterilise and aliquot into 5 Eppendorf tubes, store at -20^{O}C. Dilute 1000X into broth for final concentration of 1 mM. For blue/white expression on LA + antibiotic plates spread 10 µl and allow to absorb for 30 min at 37^{O}C before use.

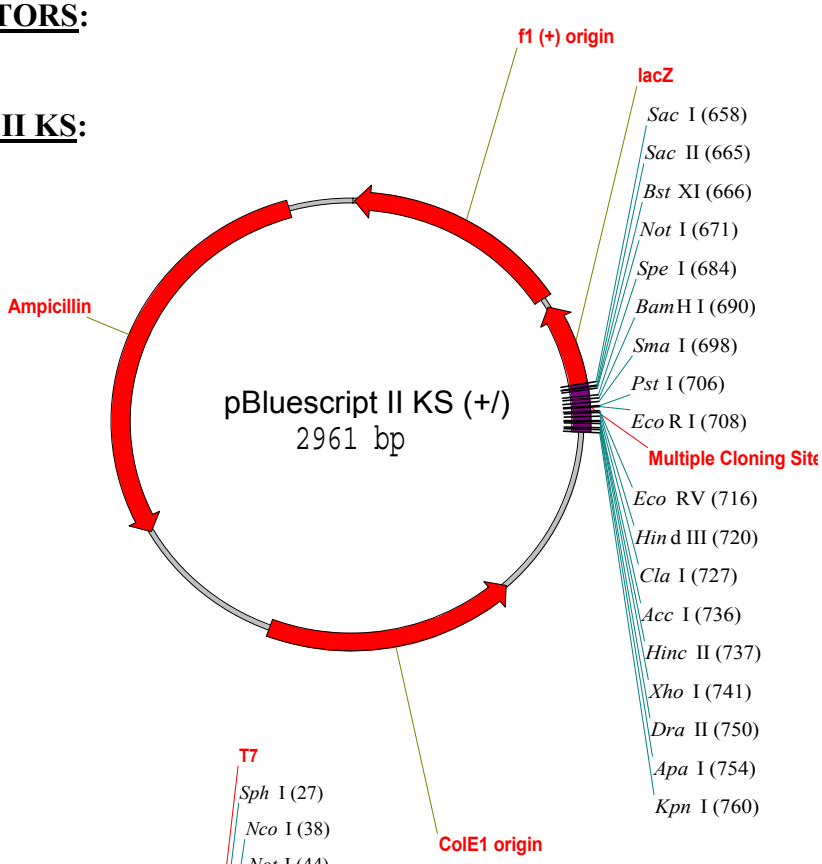
❖ X-Gal:

100 mg	5-bromo-4-chloro-3-indolyl-β-D-galactoside

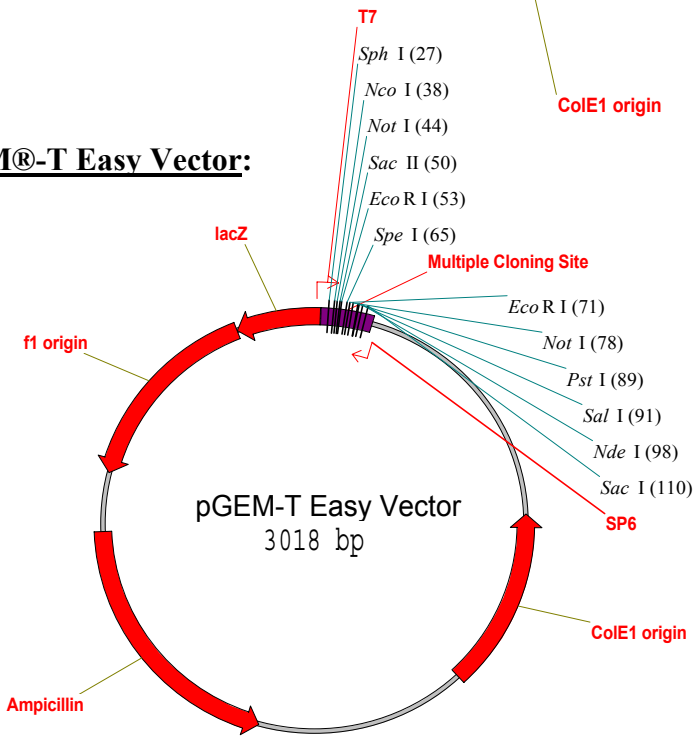
Dissolve in 2 ml N,N'-dimethyl-for-mamide. Cover with aluminium foil and store at -20^{O}C. For blue/white expression on LA + antibiotic plates spread 20 µl and allow to absorb for 30 min at 37^{O}C before use.

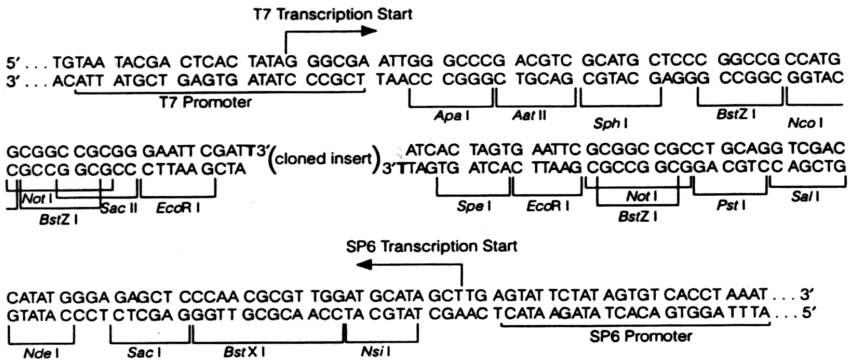
3. CLONING VECTORS:

❖ pBluescript® II KS:

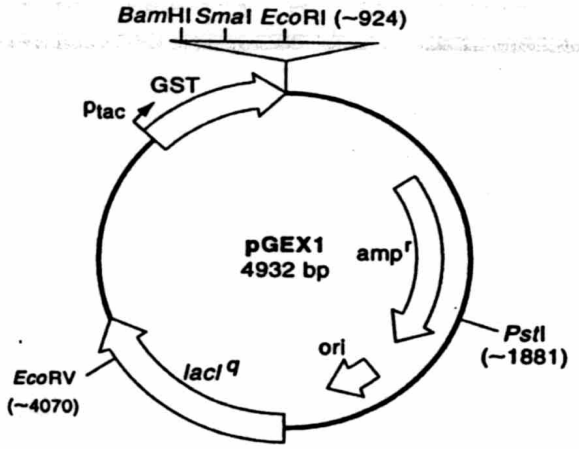


❖ pGEM®-T Easy Vector:





❖ **pGEX**



	Pro	Lys	Ser	Asp	Pro	Arg	Glu	Phe	Ile	Val	Thr	Asp	*STOP*		
pGEX1	CCA	AAA	TCG	GAT	CCC	CGG	GAA	TTC	ATC	GTG	ACT	GAC	TGA	CGA	TCT G
				BamHI		*SmaI*		*EcoRI*							

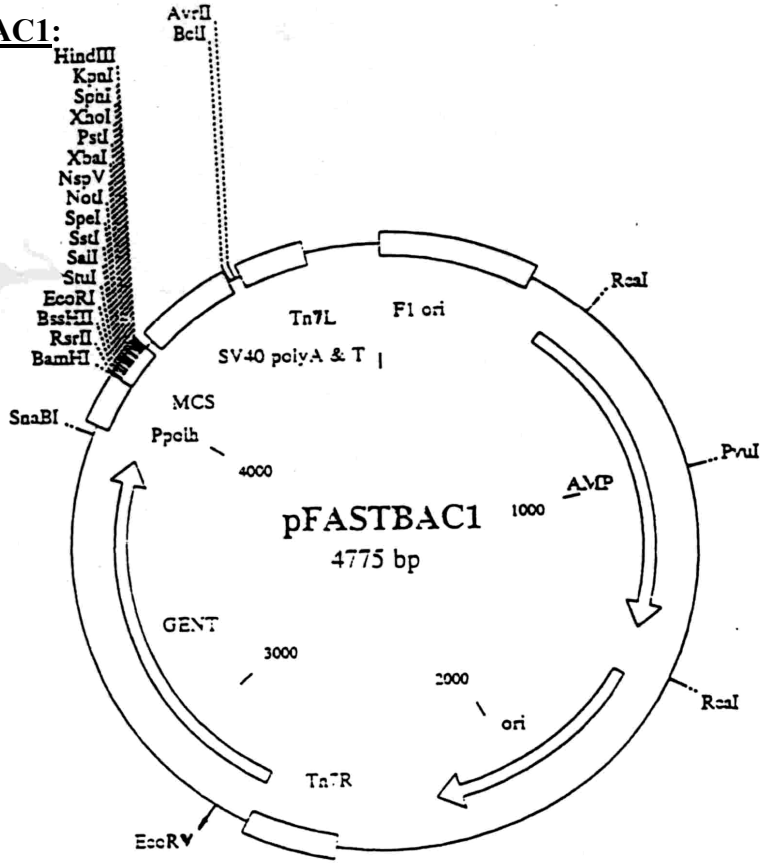
thrombin

	Pro	Lys	Ser	Asp	Leu	Val	Pro	Arg	Gly	Ser	Pro	Gly	Ile	His	Arg	Asp	*STOP*
pGEX2T	CCA	AAA	TCG	GAT	CTG	GTT	CCG	CGT	GGA	TCC	CCG	GGA	ATT	CAT	CGT	GAC	TGA	CTG
									BamHI		*SmaI*		*EcoRI*				

factor Xa

	Pro	Lys	Ser	Asp	Leu	Ile	Glu	Gly	Arg	Gly	Ile	Pro	Gly	Asn	Ser	Ser	*STOP*
pGEX3X	CCA	AAA	TCG	GAT	CTG	ATC	GAA	GGT	CGT	GGG	ATC	CCC	GGG	AAT	TCA	TCG	TGA	CTG
										BamHI		*SmaI*		*EcoRI*			

❖ **pFASTBAC1:**



4. LIGATION OF PLASMID VECTOR AND INSERT DNA:

❖ Vector : Insert Ratio:

$$\frac{\text{ng of vector} \times \text{kb size of insert}}{\text{kb size of vector}} \times \text{molar ratio of } \frac{\text{insert}}{\text{vector}}$$

= ng of insert

1:1 or 1:3 molar ratio of vector : insert usually work well.

❖ Quantification of DNA:

The concentration of double stranded DNA was determined by reading the absorbance at 260 nm and 280 nm. 5 µl of the DNA is mixed with 995 µl TE, pH 8.0 (1:200 dilution), blank with TE first. The purity of the DNA was determined by dividing the 260 reading by the 280, which should equal 1.8. The concentration was determined by reading the absorbance at 260 nm and using the following equation:

$1\ OD_{260nm} = 50\mu g/ml$

❖ DNA Precipitation:

To x volume of DNA:-
Add 1/10 the volume of 3M sodium acetate (NaOAC) and 2.5 the volume of 95% ice cold ethanol. Incubate at -20^{O}C for 1 hour or overnight. Microfuge the sample for 15 minutes. Wash the pellet twice in ice cold 70% ethanol. Air dry the pellet at room temperature for 15-20 minutes and resuspend the pellet in 20-50 µl TE, pH 8.0.

❖ T4 DNA Polymerase Protocol:

0.2-5 µg	DNA
1.0 µl	10mM dNTP's
2.0 µl	10X T4 DNA Polymerase Buffer
5 U/µg DNA	T4 DNA Polymerase
x µl	DEPC-treated water
20 µl	

Incubate for 5 minutes at 37^{O}C.
Inactivate the enzyme by heating at 75^{O}C for 10 minutes.

❖ Dephosphorylation Protocol:

General formula for calculating the picomoles of ends of linear double-stranded DNA is:-

(µg DNA/kb size of DNA) x 3.04 = pmol of ends

x pmol of ends	Linearised DNA
10 µl	10X Shrimp Alkaline Phosphatase Buffer
1-2 µl	Shrimp Alkaline Phosphatase (1U/pmol of ends)
x µl	DEPC-treated water
100 µl	

- Incubate for 15 minutes at 37^{O}C.
- Inactivate SAP enzyme for 15 minutes at 65^{O}C
- Add another 1-2 µl SAP enzyme
- Incubate for 15 minutes at 37^{O}C
- Inactivate SAP enzyme for 15 minutes at 65^{O}C.

❖ Standard Ligation Protocol:

150 ng	Insert DNA
50 ng	Vector DNA
1 U	T4 DNA Ligase (for blunt ends use 2 U)
1 µl	10X Ligase Buffer
x µl	DEPC-treated water
10 µl	

The follow is a guideline for incubation temperature and time:
Blunt ends: 22^{O}C for 4 to 16 hours (overnight)
Protruding ends: 22^{O}C for 3 hours or 4^{O}C for 16 hours (overnight)

❖ Ligation Protocol for pGEM-T Easy Vectors:

Table 11: Ligation mix for PCR Cloning into pGEM-T Easy Vectors

	Standard Reaction	**Positive Control**	**Background Control**
2X Rapid Ligation Buffer	5 µl	5 µl	5 µl
pGEM-T Easy Vector (50 ng)	1 µl	1 µl	1 µl
PCR Product (insert ratio 3:1)	x µl	-	-
Control Insert DNA	-	2 µl	-
T4 DNA Ligase (3 Weiss units/µl)	1 µl	1 µl	1 µl
deionized water to a final volume of	10 µl	10 µl	10 µl

Mix the reaction by pipetting. Incubate the reactions for 1 hour at room temperature or alternatively for maximum number of transformants, incubate the reactions overnight at 4^{O}C.

5. TRANSFORMATION OF DNA INTO BACTERIA:

❖ Preparation of Electrocompetent E.coli

1. Set up a 25 ml culture of E.coli cells (DH5α or BL21-DE3) from a glycerol stock or fresh LA plate and incubate at 37^{O}C overnight with shaking.
2. Inoculate a 250 ml culture with the 25 ml O/N culture. Grow for +/- 60-90 min. with shaking until the OD_{600nm} = 0.65
3. Place on ice for 30 min. (*it is imporatnt that the cells remain on ice throughout this whole procedure)*.
4. Pellet the cells by centrifugation at 6750 rpm (JA14) for 15 min at 4^{O}C.
5. Resuspend the cells in 250 ml of ice-cold sterile water.
6. Pellet the by centrifugation at 6750 rpm for 15 min at 4^{O}C.
7. Resuspend the cells in 125 ml of ice-cold sterile water.
8. Pellet the by centrifugation at 6750 rpm for 15 min at 4^{O}C.
9. Resuspend the cells in 5 ml of 10% glycerol which has been diluted using sterile water.
10. Pellet the by centrifugation at 6750 rpm (JA20) for 15 min at 4^{O}C.
11. Resuspend the cells in 750 µl of 10% glycerol which has been diluted using sterile water.
12. Aliquot 40 µl quantities into 1.5 ml Eppendorf tubes. Freeze immediately using liquid nitrogen.
13. Store the aliquots at -80^{O}C

❖ Calculation of Transformation Efficiency:

Transformation efficiency (cfu/ug) =

$$\frac{\text{cfu on control plate}}{\text{ng of supercoiled vector plated}} \times \frac{10^3 \text{ng}}{1\mu\text{g}} \times \text{final dilution plated}$$

6. PLASMID ISOLATION:

❖ Lysozyme Solution:

4.5 ml	20% Glucose
2.0 ml	0.5 M EDTA
2.5 ml	1 M Tris/HCl, pH 8.0

Make up to 100 ml with distilled water and sterilize by Millipore filtration (0.22µ). Store at 4^{O}C and immediately before use, add 4 mg/ml of lysozyme powder.

❖ NaOH / SDS Lysis Solution:
Immediately before use, prepare the following solution (for 12 mini-preps):

3.6 ml	Distilled water
0.2 ml	20% SDS
0.4 ml	2 M NaOH

❖ Acetate Neutralizing Solution:

29.4 g	Potassium Acetate
11.5 ml	Glacial Acetic Acid

Make up to 100 ml with distilled water. Sterilize by Millipore filtration (0.22µ) and store at 4^{O}C.

❖ TE / RNase:
Immediately before use prepare the following dilution:

1 ml	TE, pH 8.0
2 µl	10 mg/ml RNase A

❖ Smart Buffer:

100 µl	1 M Tris/HCl, pH 8.0
100 µl	100 mM EDTA
3.75 ml	40% Sucrose
200 µl	5 mg/ml BSA
500 µl	4 mg/ml RNase A
20 mg	Lysozyme

Make up to 10 ml with distilled water and dispense into 0.5 ml aliquotes. Freeze at -20^{O}C. Do not thaw and re-freeze.

7. RESTRICTION ENZYME DIGESTS:

Restriction enzymes require different reaction buffers and different incubation temperatures. Some enzymes also require the addition of 1%BSA (bovine serum albumin).

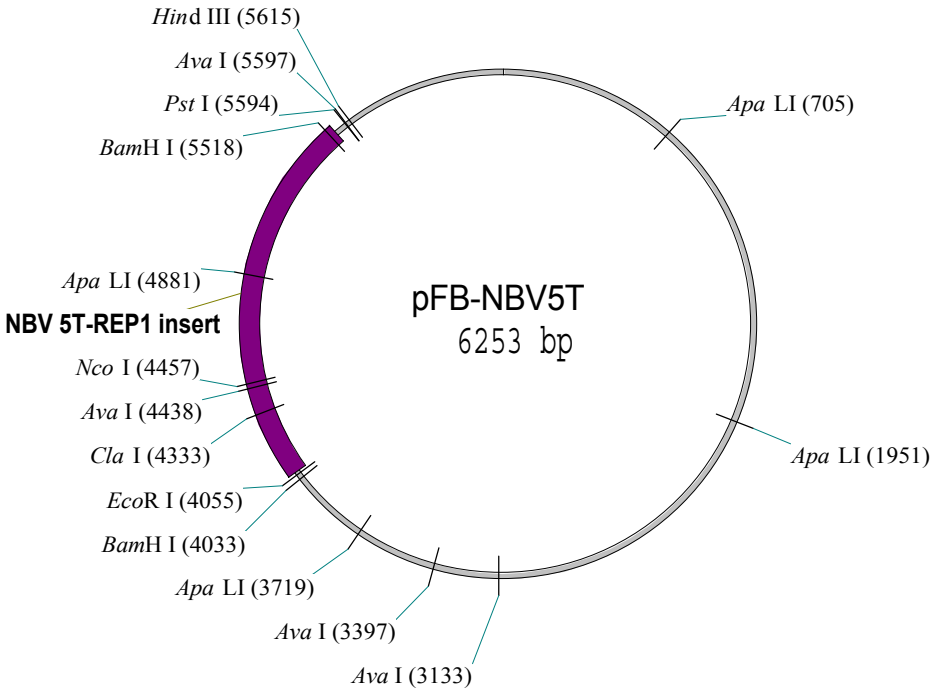
Master mix:

x µl	DNA
2.0 µl	10X Restriction Enzyme Buffer
1 U/µg DNA	Restriction Enzyme
x µl	DEPC-treated water
20.0 µl	

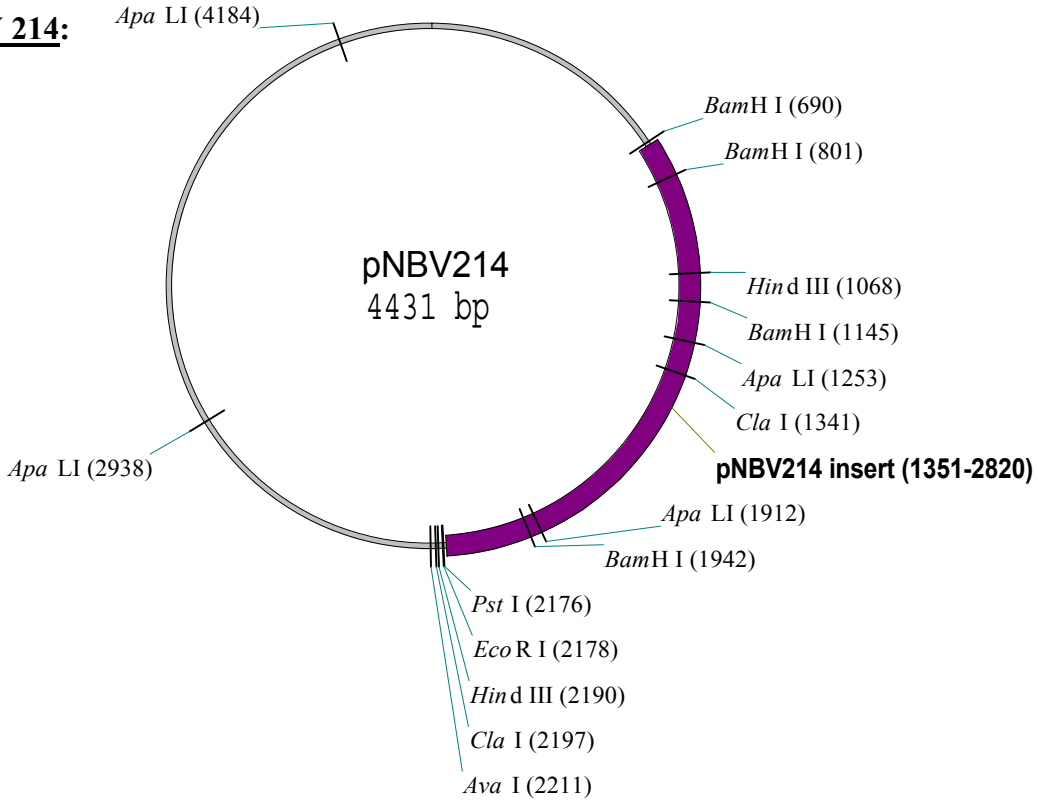
Incubate at the required temperature for +/- 2-3 hours.

APPENDIX 6: NβV CLONES

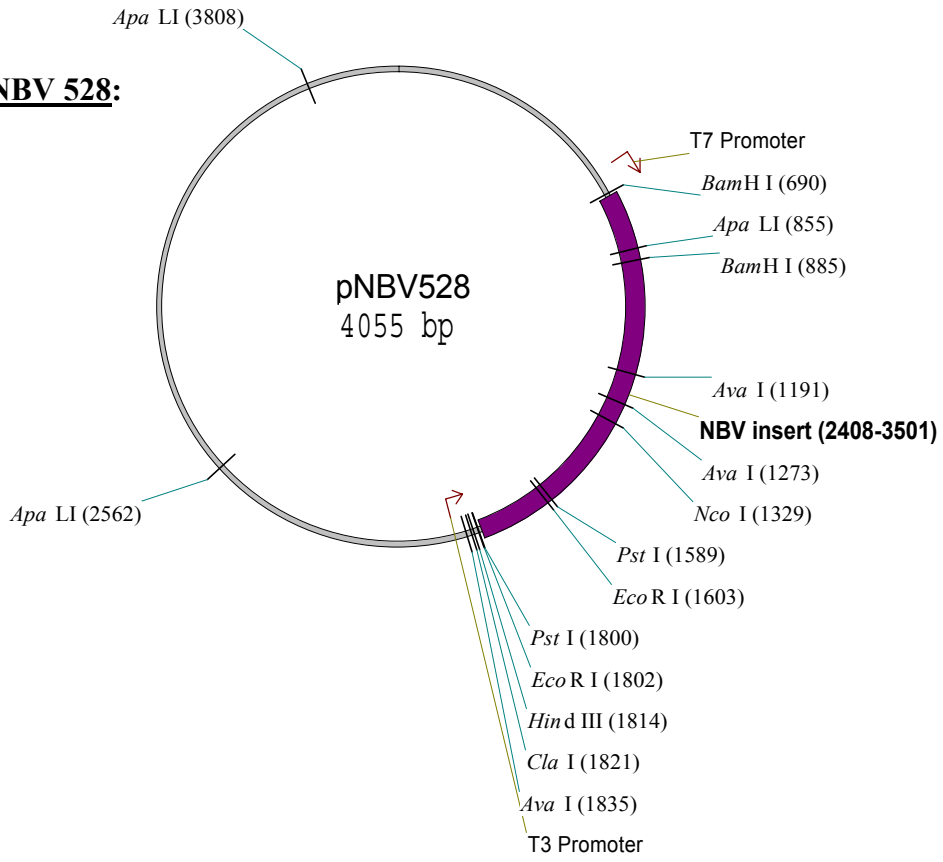
1. pNBV 5T:



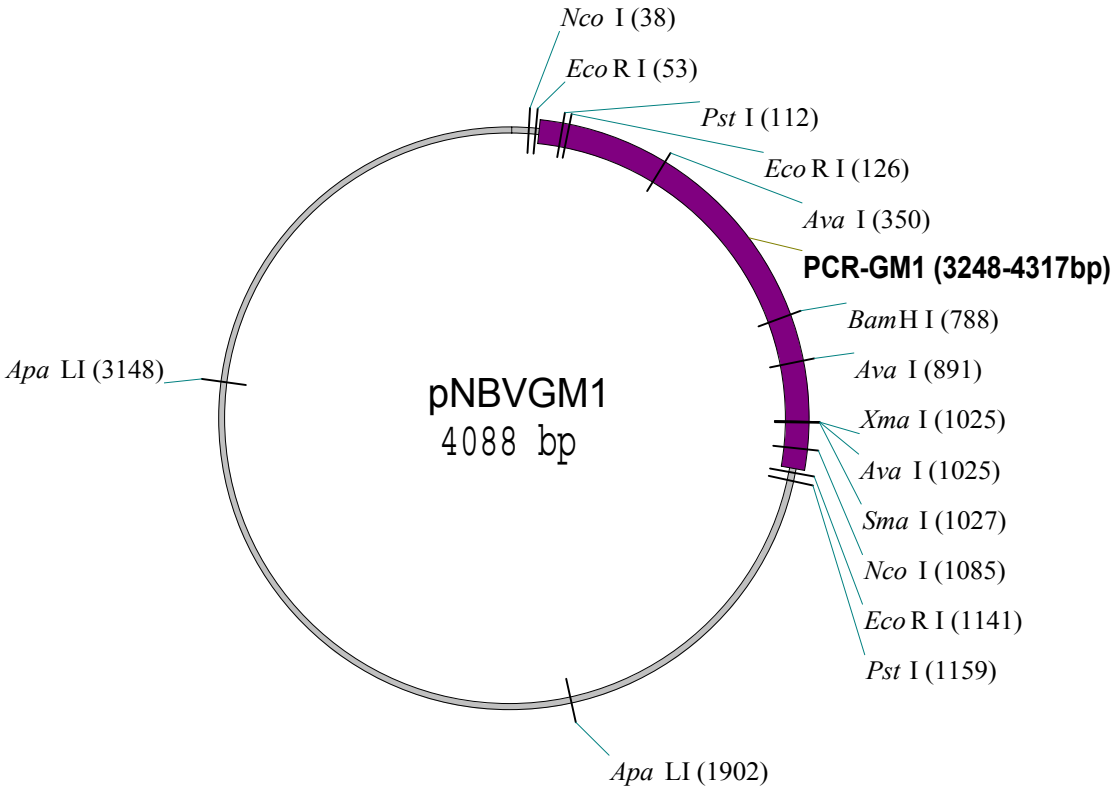
2. pNBV 214:



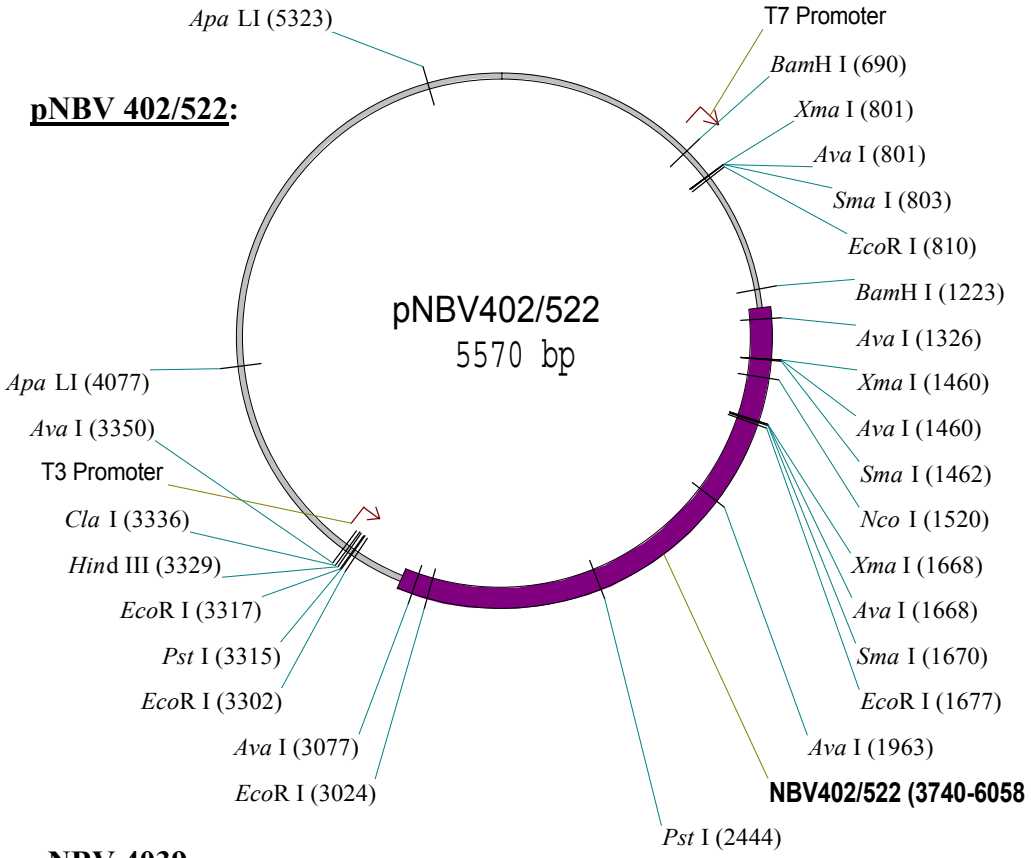
8. pNBV 528:



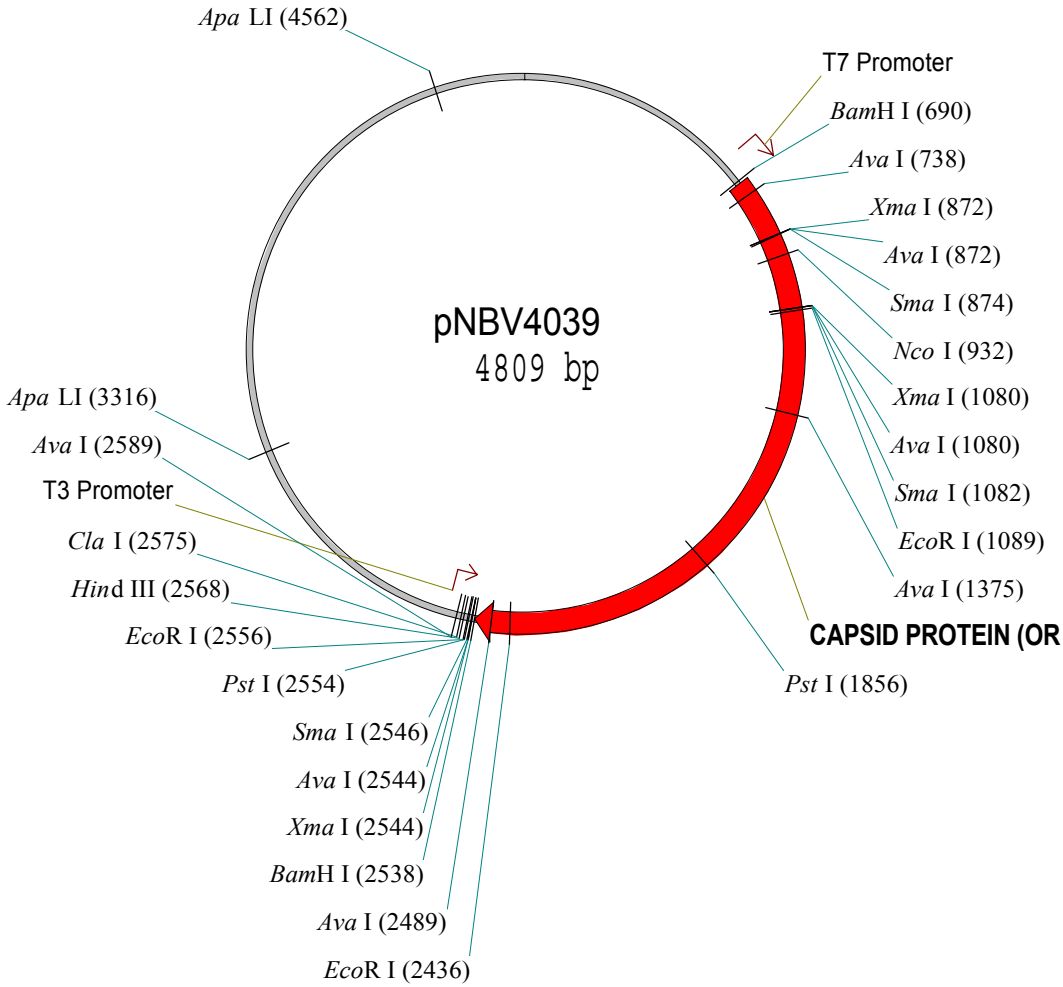
5. pNBV GM1:



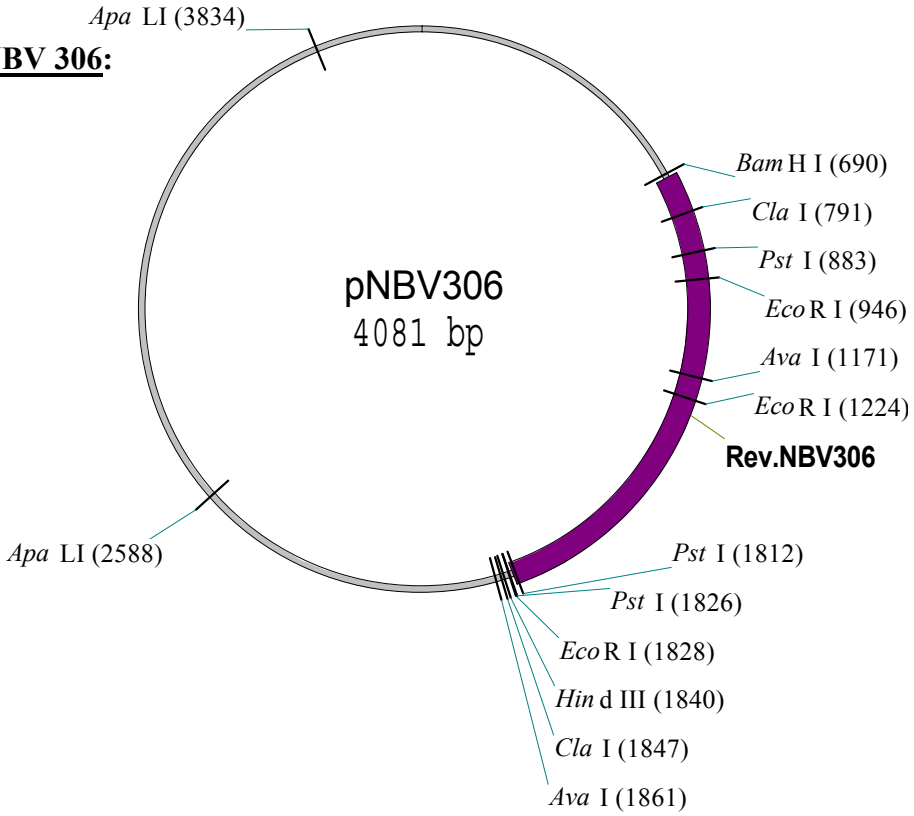
6. **pNBV 402/522:**



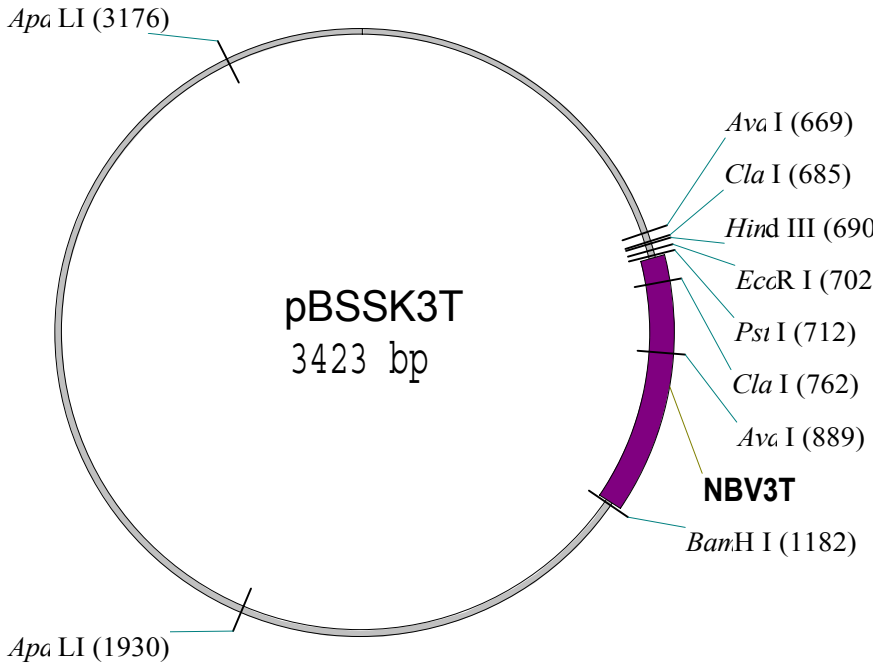
7. **pNBV 4039:**



8. pNBV 306:



9. pNBV 3T:



APPENDIX 7: WESTERN BLOT ANALYSIS

1. BUFFER AND STOCK SOLUTIONS:

❖ Transfer Buffer:

3.03 g	Tris Base (25 mM)
14.4 g	Glycine (192 mM)
200 ml	Methanol (20%)

Make up to 1l with distilled water. Prepare fresh for every run and chill.

❖ Tris Buffered Saline (TBS):

6.05 g	Tris Base (50 mM)
8.76 g	NaCl (150 mM)

Made up to 1l with distilled water and pH adjusted to 7.5 with conc. HCl.

❖ TBS-Tween 20 (TBST):

Dissolve 1 ml Tween 20 in 1 litre TBS

❖ 5% Block:

Dissolve 5 g of non fat dry milk powder in 100 ml TBS

❖ Ponceau S Stain:

0.5%	Ponceau S
1%	Glacial Acetic Acid

Make in 100 ml distilled water. Stain the membrane for 2 minutes, then destain in distilled water.

2. TRANSFER SET UP

Clear panel (+ electrode)
Fibre pad (saturated in transfer buffer)
2X Whatman no.1 Fiter paper (saturated in transfer buffer)
Nitrocellulose membrane (soaked in buffer for 30 min)
Gel (equilibrated in buffer)
2X Whatman no.1 Filter paper (saturated in transfer buffer)
Fibre pad (saturated in transfer buffer)
Black panel (- electrode)

Run at 100V for 1hour in small apparatus and 3 hours in large apparatus.

APPENDIX 8: *NUDAURELIA β VIRUS* GENOME WITH TRANSLATED ORFS

```
  1  GATTCTCCCA CTTGATTGTA CGGAGAACCC GAAACACGGT GGAGGGGGCC
     CTAAGAGGGT GAACTAACAT GCCTCTTGGG CTTTGTGCCA CCTCCCCCGG
                          Start of Replicase gene ORF
  +3                                          Val MetGlu
                                              ]----------
 51  ACCGATCCCA AGCCTTGATA TCGCTTGTGT TAGGCACAAG TGATGGAAGA
     TGGCTAGGGT TCGGAACTAT AGCGAACACA ATCCGTGTTC ACTACCTTCT

  +3 pAlaSerLys GlnLeuArgVal LeuAspAla GlnGluArg AlaLysAla
     ----------------------------------------------------
101  CGCAAGCAAG CAGCTCCGCG TCCTGGATGC CCAGGAGCGC GCGAAGGCCG
     GCGTTCGTTC GTCGAGGCGC AGGACCTACG GGTCCTCGCG CGCTTCCGGC

  +3 laPheGlnLeu AspPheIle AlaSerValGlu ThrLeuGlu AspAlaGln
     ----------------------------------------------------
151  CCTTCCAACT CGACTTCATA GCCTCTGTCG AGACTTTGGA AGACGCTCAG
     GGAAGGTTGA GCTGAAGTAT CGGAGACAGC TCTGAAACCT TCTGCGAGTC

  +3 GluLysTyrGlu GlyMetMet PheArgSer GlyThrLysLeu ProSer
     ----------------------------------------------------
201  GAGAAGTACG AGGGCATGAT GTTTCGCAGT GGCACGAAAC TGCCATCAAC
     CTCTTCATGC TCCCGTACTA CAAAGCGTCA CCGTGCTTTG ACGGTAGTTG

  +3 rHisIleLys LeuAlaIleAsp LeuArgVal AlaGluLys AspLeuArg
     ----------------------------------------------------
251  CCATATTAAG TTGGCAATCG ATCTGAGAGT TGCGGAGAAA GATCTACGCC
     GGTATAATTC AACCGTTAGC TAGACTCTCA ACGCCTCTTT CTAGATGCGG

  +3 rgHisValLys AsnValPro ThrValLeuGlu IleGlyPro SerValGlu
     ----------------------------------------------------
301  GGCACGTTAA GAATGTACCG ACAGTGCTGG AAATTGGACC CAGTGTTGAG
     CCGTGCAATT CTTACATGGC TGTCACGACC TTTAACCTGG GTCACAACTC

  +3 SerValArgTyr AlaValGln ThrArgAsp LysGluArgVal HisGly
     ----------------------------------------------------
351  AGCGTGCGTT ACGCTGTGCA GACTCGAGAC AAGGAGAGAG TCCATGGCTG
     TCGCACGCAA TGCGACACGT CTGAGCTCTG TTCCTCTCTC AGGTACCGAC

  +3 sThrPheSer AspAlaArgAsp AsnLeuArg HisAsnLys IleGlyTyr
     ----------------------------------------------------
401  CACCTTCTCC GACGCGCGTG ATAACCTCCG CCACAATAAG ATCGGTTATG
     GTGGAAGAGG CTGCGCGCAC TATTGGAGGC GGTGTTATTC TAGCCAATAC

  +3 luAlaHisTyr AspArgLys IleGlyProAsp AlaAlaLeu LeuAlaAla
     ----------------------------------------------------
451  AAGCCCATTA CGACAGAAAG ATTGGACCTG ACGCCGCCCT TCTGGCCGCT
     TTCGGGTAAT GCTGTCTTTC TAACCTGGAC TGCGGCGGGA AGACCGGCGA
```

```
  +3 GlyIleProThr AspThrPhe CysValAsp GlyPheSerAsn CysGlu
     ---------- ---------- ---------- ---------- ----------
 501 GGTATCCCAA CTGACACCTT CTGTGTCGAC GGCTTCTCCA ATTGCGAGTA
     CCATAGGGTT GACTGTGGAA GACACAGCTG CCGAAGAGGT TAACGCTCAT

  +3 rGlnSerPro LeuAlaIleAla CysHisSer LeuTyrPro AspGlyGlu
     ---------- ---------- ---------- ---------- ----------
 551 CCAATCCCCC CTCGCCATTG CCTGCCACTC ACTTTACCCC GATGGGGAAA
     GGTTAGGGGG GAGCGGTAAC GGACGGTGAG TGAAATGGGG CTACCCCTTT

  +3 erAsnSerIle MetAspVal AlaLysGlyMet AlaLeuHis GlyThrHis
     ---------- ---------- ---------- ---------- ----------
 601 GTAATAGTAT TATGGACGTG GCTAAAGGCA TGGCTCTCCA CGGCACCCAC
     CATTATCATA ATACCTGCAC CGATTTCCGT ACCGAGAGGT GCCGTGGGTG

  +3 ValIleTyrAla TrpMetHis LeuProVal GluLeuLeuThr LeuThr
     ---------- ---------- ---------- ---------- ----------
 651 GTGATATATG CGTGGATGCA TCTGCCCGTG GAACTGCTAA CGCTCACCGA
     CACTATATAC GCACCTACGT AGACGGGCAC CTTGACGATT GCGAGTGGCT

  +3 pAlaAspAsn IlePheGluGly TyrSerIle ArgPheGlu GluThrGly
     ---------- ---------- ---------- ---------- ----------
 701 TGCAGACAAT ATTTTTGAAG GGTATAGCAT TCGGTTTGAG GAGACAGGGG
     ACGTCTGTTA TAAAAACTTC CCATATCGTA AGCCAAACTC CTCTGTCCCC

  +3 laLeuProCys ThrLysArg ArgLysAlaIle PheSerGly TyrAsnAsp
     ---------- ---------- ---------- ---------- ----------
 751 CACTACCCTG CACCAAGAGG AGGAAAGCCA TATTCTCCGG TTATAACGAT
     GTGATGGGAC GTGGTTCTCC TCCTTTCGGT ATAAGAGGCC AATATTGCTA

  +3 PheGlySerAla TyrValHis AspAlaHis HisTrpAlaGly TrpLeu
     ---------- ---------- ---------- ---------- ----------
 801 TTTGGTTCGG CCTATGTGCA CGATGCCCAC CATTGGGCTG GTTGGCTTAA
     AAACCAAGCC GGATACACGT GCTACGGGTG GTAACCCGAC CAACCGAATT

  +3 sHisArgGly ValAspThrPro TyrGlyPhe SerIleLeu IleAspIle
     ---------- ---------- ---------- ---------- ----------
 851 GCATCGGGGA GTAGACACCC CGTATGGCTT CTCCATATTG ATCGACATAC
     CGTAGCCCCT CATCTGTGGG GCATACCGAA GAGGTATAAC TAGCTGTATG

  +3 lnGlnArgPhe GlyMetHis ThrLysLeuLys IleThrArg GlyHisSer
     ---------- ---------- ---------- ---------- ----------
 901 AACAGAGGTT CGGTATGCAC ACGAAATTAA AGATCACCCG TGGGCACAGC
     TTGTCTCCAA GCCATACGTG TGCTTTAATT TCTAGTGGGC ACCCGTGTCG

  +3 SerGlySerIle ThrThrVal PheProLeu SerLysLeuGly LeuIle
     ---------- ---------- ---------- ---------- ----------
 951 AGTGGCAGTA TCACCACCGT GTTCCCGTTG TCGAAATTGG GCTTGATCTG
     TCACCGTCAT AGTGGTGGCA CAAGGGCAAC AGCTTTAACC CGAACTAGAC

  +3 pValProAsn IleValLysIle MetTyrPro LysAlaLys HisGluPro
     ---------- ---------- ---------- ---------- ----------
1001 GGTGCCGAAC ATAGTCAAAA TAATGTACCC TAAAGCCAAA CACGAGCCGG
     CCACGGCTTG TATCAGTTTT ATTACATGGG ATTTCGGTTT GTGCTCGGCC
```

```
   +3 luTyrIleVal ThrAspLys LysLysTyrGlu GlyValCys ValTyrVal
      ---------- ---------- ---------- ---------- ----------
1051  AGTACATCGT CACGGATAAG AAGAAGTATG AAGGCGTTTG CGTGTACGTC
      TCATGTAGCA GTGCCTATTC TTCTTCATAC TTCCGCAAAC GCACATGCAG


   +3 GlyThrArgVal GlnSerSer GlyLysSer IleThrLeuAla GluIle
      ---------- ---------- ---------- ---------- ----------
1101  GGAACGAGGG TGCAAAGTTC CGGCAAGTCT ATTACGCTCG CTGAGATTGT
      CCTTGCTCCC ACGTTTCAAG GCCGTTCAGA TAATGCGAGC GACTCTAACA

   +3 lGlnTyrIle ArgThrArgLeu ThrArgIle IleLeuAsn GlyThrVal
      ---------- ---------- ---------- ---------- ----------
1151  TCAATACATC CGAACAAGAT TAACACGCAT CATTCTGAAT GGCACTGTCC
      AGTTATGTAG GCTTGTTCTA ATTGTGCGTA GTAAGACTTA CCGTGACAGG

   +3 isGluLysThr TrpThrIle AlaGluGlnAsp IleGluArg LeuAlaVal
      ---------- ---------- ---------- ---------- ----------
1201  ACGAGAAAAC GTGGACCATA GCAGAGCAAG ACATTGAGCG ACTTGCCGTT
      TGCTCTTTTG CACCTGGTAT CGTCTCGTTC TGTAACTCGC TGAACGGCAA

   +3 SerIleMetPhe ArgLysAsn ValGluArg AlaValSerGlu LysAla
      ---------- ---------- ---------- ---------- ----------
1251  AGCATTATGT TCCGCAAGAA TGTGGAACGC GCTGTGTCTG AAAAGGCACT
      TCGTAATACA AGGCGTTCTT ACACCTTGCG CGACACAGAC TTTTCCGTGA

   +3 uMetArgAla GlnLysLysCys LysSerAla GluLysGln AlaLeuLeu
      ---------- ---------- ---------- ---------- ----------
1301  GATGAGAGCG CAGAAGAAGT GCAAGAGCGC TGAAAAACAA GCGCTGCTGC
      CTACTCTCGC GTCTTCTTCA CGTTCTCGCG ACTTTTTGTT CGCGACGACG

   +3 roValTrpMet ArgArgIle AlaAsnTrpPhe GlnAspLys PheGlnIle
      ---------- ---------- ---------- ---------- ----------
1351  CAGTTTGGAT GCGGAGGATC GCCAATTGGT TCCAAGACAA ATTTCAAATC
      GTCAAACCTA CGCCTCCTAG CGGTTAACCA AGGTTCTGTT TAAAGTTTAG

   +3 AspGluGluVal ValArgLys ArgTyrLeu GluCysLeuLys AlaGln
      ---------- ---------- ---------- ---------- ----------
1401  GACGAGGAGG TCGTACGCAA GCGCTACCTT GAGTGTCTCA AGGCGCAACC
      CTGCTCCTCC AGCATGCGTT CGCGATGGAA CTCACAGAGT TCCGCGTTGG

   +3 oTrpIleHis AlaAspLysVal ValAsnCys GluThrLys ArgTyrAsn
      ---------- ---------- ---------- ---------- ----------
1451  CTGGATCCAC GCCGATAAAG TGGTGAACTG CGAGACCAAG CGCTATAACC
      GACCTAGGTG CGGCTATTTC ACCACTTGAC GCTCTGGTTC GCGATATTGG

   +3 roThrValAla GluValGly ProLysAsnHis LeuLeuAla ThrThrGly
      ---------- ---------- ---------- ---------- ----------
1501  CTACTGTCGC CGAGGTGGGT CCTAAGAATC ATTTGCTCGC CACTACCGGA
      GATGACAGCG GCTCCACCCA GGATTCTTAG TAAACGAGCG GTGATGGCCT

   +3 LeuArgGluLeu GlnArgGlu IleProSer AlaAsnGluPro GlnAsp
      ---------- ---------- ---------- ---------- ----------
1551  TTGCGCGAGC TCCAAAGGGA AATACCCAGT GCTAACGAAC CGCAAGATAG
      AACGCGCTCG AGGTTTCCCT TTATGGGTCA CGATTGCTTG GCGTTCTATC
```

```
   +3 gGlyAlaLys AlaTrpHisSer AlaHisAla AspLeuAsp IleTyrAla
      ---------- ---------- ---------- ---------- ----------
1601  AGGAGCCAAG GCATGGCACT CCGCTCACGC CGATCTCGAC ATTTACGCCG
      TCCTCGGTTC CGTACCGTGA GGCGAGTGCG GCTAGAGCTG TAAATGCGGC

   +3 luGlyLeuArg LeuAspSer AlaLysGluAla AlaAlaGly LysGlnSer
      ---------- ---------- ---------- ---------- ----------
1651  AGGGACTCCG ACTCGACTCC GCTAAAGAGG CAGCAGCGGG TAAGCAGTCC
      TCCCTGAGGC TGAGCTGAGG CGATTTCTCC GTCGTCGCCC ATTCGTCAGG

   +3 LeuAlaIleThr LeuGlnGln AlaPheGln ValLeuGlyLys ThrLys
      ---------- ---------- ---------- ---------- ----------
1701  CTGGCGATCA CATTGCAGCA AGCTTTCCAA GTCCTAGGGA AGACCAAGTG
      GACCGCTAGT GTAACGTCGT TCGAAAGGTT CAGGATCCCT TCTGGTTCAC

   +3 sGluGlyCys AsnAsnIleGlu IleGluTyr TrpThrGly ProProGly
      ---------- ---------- ---------- ---------- ----------
1751  CGAGGGGTGC AACAACATTG AAATCGAGTA CTGGACCGGA CCCCCCGGAT
      GCTCCCCACG TTGTTGTAAC TTTAGCTCAT GACCTGGCCT GGGGGGCCTA

   +3 erGlyLysSer ArgAlaAla LysProArgPhe AlaAspLeu GlnGlyGly
      ---------- ---------- ---------- ---------- ----------
1801  CCGGGAAATC CAGGGCCGCT AAGCCGAGAT TTGCAGATTT GCAGGGGGGC
      GGCCCTTTAG GTCCCGGCGA TTCGGCTCTA AACGTCTAAA CGTCCCCCCG

   +3 ValLeuTyrCys AlaProThr ArgThrLeu ArgAspAlaLeu AspGlu
      ---------- ---------- ---------- ---------- ----------
1851  GTGTTGTACT GCGCCCCTAC GCGCACGCTG CGCGACGCCC TCGACGAAAG
      CACAACATGA CGCGGGGATG CGCGTGCGAC GCGCTGCGGG AGCTGCTTTC

   +3 rValValHis ProSerArgVal CysThrTyr HisAsnAla LeuHisVal
      ---------- ---------- ---------- ---------- ----------
1901  CGTCGTGCAC CCTTCCCGTG TTTGCACTTA CCACAACGCA CTGCATGTCG
      GCAGCACGTG GGAAGGGCAC AAACGTGAAT GGTGTTGCGT GACGTACAGC

   +3 laAlaLysGlu SerGlyAsn ArgProPheAsp ValIleVal IleAspGlu
      ---------- ---------- ---------- ---------- ----------
1951  CTGCCAAAGA GTCTGGCAAT AGGCCTTTTG ACGTTATCGT CATCGATGAA
      GACGGTTTCT CAGACCGTTA TCCGGAAAAC TGCAATAGCA GTAGCTACTT

   +3 AlaGluThrThr ProAlaCys TyrValGly ThrMetHisHis AlaSer
      ---------- ---------- ---------- ---------- ----------
2001  GCGGAGACGA CGCCGGCTTG CTACGTAGGT ACGATGCATC ATGCATCGCC
      CGCCTCTGCT GCGGCCGAAC GATGCATCCA TGCTACGTAG TACGTAGCGG

   +3 oSerSerArg IleValCysLeu GlyAspPro HisGlnIle GlyTyrIle
      ---------- ---------- ---------- ---------- ----------
2051  TAGTAGTAGG ATCGTCTGTC TGGGCGATCC GCACCAGATC GGTTACATCG
      ATCATCATCC TAGCAGACAG ACCCGCTAGG CGTGGTCTAG CCAATGTAGC

   +3 spPheSerAsp ArgLysAsp AspLeuLysPro PheSerIle IleAlaAla
      ---------- ---------- ---------- ---------- ----------
2101  ACTTTTCGGA TCGAAAAGAC GATTTGAAAC CTTTCAGTAT CATAGCAGCC
      TGAAAAGCCT AGCTTTTCTG CTAAACTTTG GAAAGTCATA GTATCGTCGG
```

```
   +3 GluCysArgThr ArgArgPhe AsnThrThr TyrArgCysPro GlnAsp
      ----------------------------------------------------
2151  GAATGTCGCA CTCGTAGGTT TAACACCACT TATAGGTGCC CACAAGACGT
      CTTACAGCGT GAGCATCCAA ATTGTGGTGA ATATCCACGG GTGTTCTGCA

   +3 lLeuAsnLeu ProIlePheLys ThrLeuTyr ProAspAla IleSerPhe
      ----------------------------------------------------
2201  TTTAAACTTG CCCATATTCA AAACTCTATA CCCGGACGCG ATATCGTTCA
      AAATTTGAAC GGGTATAAGT TTTGAGATAT GGGCCTGCGC TATAGCAAGT

   +3 erLysGlnLeu ThrSerIle ArgTyrLeuThr ArgAlaArg SerValThr
      ----------------------------------------------------
2251  GCAAACAATT GACTAGCATC CGTTACCTCA CACGGGCAAG ATCAGTTACC
      CGTTTGTTAA CTGATCGTAG GCAATGGAGT GTGCCCGTTC TAGTCAATGG

   +3 ArgThrArgHis AlaGlnThr LeuThrGln AspGlnLysPro HisSer
      ----------------------------------------------------
2301  CGAACACGCC ACGCTCAGAC CCTGACGCAG GACCAAAAGC CACATTCGGA
      GCTTGTGCGG TGCGAGTCTG GGACTGCGTC CTGGTTTTCG GTGTAAGCCT

   +3 uProProVal ThrAlaHisGlu ProGlnAla ArgArgThr AspValIle
      ----------------------------------------------------
2351  ACCGCCAGTG ACCGCGCATG AGCCGCAGGC ACGACGTACG GACGTTATAG
      TGGCGGTCAC TGGCGCGTAC TCGGCGTCCG TGCTGCATGC CTGCAATATC

   +3 alHisTyrAla GlyThrLeu ProGluArgAla LeuLeuGlu LysValArg
      ----------------------------------------------------
2401  TGCATTACGC CGGCACTTTA CCCGAAAGGG CACTGTTAGA GAAGGTGCGG
      ACGTAATGCG GCCGTGAAAT GGGCTTTCCC GTGACAATCT CTTCCACGCC

   +3 HisIleAsnVal AlaLeuThr ArgHisThr AsnAlaLeuTyr IleArg
      ----------------------------------------------------
2451  CATATAAACG TCGCGTTGAC TCGGCACACA AACGCCCTAT ATATCAGGGA
      GTATATTTGC AGCGCAACTG AGCCGTGTGT TTGCGGGATA TATAGTCCCT

   +3 pGluSerGlu LysGlyGluLeu ValProSer LeuMetThr ProProSer
      ----------------------------------------------------
2501  CGAAAGTGAA AAAGGAGAGT TGGTACCTTC ATTAATGACA CCGCCAAGCT
      GCTTTCACTT TTTCCTCTCA ACCATGGAAG TAATTACTGT GGCGGTTCGA

   +3 rpSerThrTyr ArgCysThr ProValAspLys GlnMetVal ProAspPro
      ----------------------------------------------------
2551  GGAGCACTTA TCGGTGCACC CCCGTTGACA AGCAAATGGT ACCGGATCCG
      CCTCGTGAAT AGCCACGTGG GGGCAACTGT TCGTTTACCA TGGCCTAGGC

   +3 ValAlaValGlu ArgGluAsn GlySerSer GlyProCysAsp SerHis
      ----------------------------------------------------
2601  GTTGCAGTGG AGCGAGAGAA CGGATCGTCT GGTCCGTGTG ACTCCCACCA
      CAACGTCACC TCGCTCTCTT GCCTAGCAGA CCAGGCACAC TGAGGGTGGT

   +3 sIleGlyAla IleThrIleLeu GlnGluLeu GlyLysLeu ThrAspThr
      ----------------------------------------------------
2651  TATCGGCGCG ATTACTATAT TGCAAGAGCT CGGCAAATTA ACGGATACGA
      ATAGCCGCGC TAATGATATA ACGTTCTCGA GCCGTTTAAT TGCCTATGCT
```

```
   +3 ysGlyValArg ValPheGlu SerGluAlaVal ProThrAla HisArgArg
      ----------------------------------------------------
2701  AAGGCGTACG AGTATTTGAA TCCGAAGCCG TCCCAACCGC TCACCGGCGC
      TTCCGCATGC TCATAAACTT AGGCTTCGGC AGGGTTGGCG AGTGGCCGCG

   +3 ValValLeuAsp GlyAsnLeu AspSerGly ProAspArgTyr ProMet
      ----------------------------------------------------
2751  GTAGTGCTTG ACGGCAACCT CGATTCAGGG CCCGATCGTT ACCCGATGTA
      CATCACGAAC TGCCGTTGGA GCTAAGTCCC GGGCTAGCAA TGGGCTACAT

   +3 rGlnPheThr AsnLeuArgGly ThrLysTyr ThrAsnIle LysAspAsn
      ----------------------------------------------------
2801  TCAGTTCACT AACCTCCGCG GGACCAAATA CACGAATATC AAGGACAACC
      AGTCAAGTGA TTGGAGGCGC CCTGGTTTAT GTGCTTATAG TTCCTGTTGG

   +3 lnGlnAlaLeu HisThrLeu ValGlyArgTyr AlaArgLys IleAsnSer
      ----------------------------------------------------
2851  AACAAGCGTT GCATACGCTC GTCGGCCGGT ATGCACGCAA GATAAACAGC
      TTGTTCGCAA CGTATGCGAG CAGCCGGCCA TACGTGCGTT CTATTTGTCG

   +3 SerSerArgGlu AspAlaGlu PheAspVal LysArgIleThr AlaArg
      ----------------------------------------------------
2901  TCGAGCCGAG AGGACGCCGA GTTTGACGTT AAGAGAATCA CAGCCAGGCT
      AGCTCGGCTC TCCTGCGGCT CAAACTGCAA TTCTCTTAGT GTCGGTCCGA

   +3 uLysGluTrp IleProPheArg ThrAlaGlu ProGluGln ValAspSer
      ----------------------------------------------------
2951  CAAGGAATGG ATTCCTTTTA GAACGGCAGA GCCCGAGCAA GTCGACAGTT
      GTTCCTTACC TAAGGAAAAT CTTGCCGTCT CGGGCTCGTT CAGCTGTCAA

   +3 ysPheAlaAsp AlaMetGln LysIleAlaGlu ArgGlyHis GlyValAsp
      ----------------------------------------------------
3001  GCTTTGCCGA CGCCATGCAA AAGATCGCCG AACGCGGCCA TGGCGTCGAT
      CGAAACGGCT GCGGTACGTT TTCTAGCGGC TTGCGCCGGT ACCGCAGCTA

   +3 AspIleGluAsp PheTrpSer AsnGluGly GlnArgIleSer TyrHis
      ----------------------------------------------------
3051  GACATCGAGG ACTTCTGGTC GAACGAAGGC CAAAGAATTT CTTACCACCT
      CTGTAGCTCC TGAAGACCAG CTTGCTTCCG GTTTCTTAAA GAATGGTGGA

   +3 uLysGlyGln GlnLysValMet AspProThr LysLeuLys LeuGlyGln
      ----------------------------------------------------
3101  TAAGGGCCAG CAAAAAGTCA TGGACCCCAC CAAACTGAAA CTTGGACAAG
      ATTCCCGGTC GTTTTTCAGT ACCTGGGGTG GTTTGACTTT GAACCTGTTC

   +3 lyIleSerAla HisGluLys CysAlaAsnIle AlaLeuSer AlaTrpVal
      ----------------------------------------------------
3151  GTATCTCCGC GCATGAAAAA TGCGCTAACA TTGCCCTCAG CGCGTGGGTG
      CATAGAGGCG CGTACTTTTT ACGCGATTGT AACGGGAGTC GCGCACCCAC

   +3 ArgIleIleGln AspGlnMet SerThrSer GluLysPheIle PheAla
      ----------------------------------------------------
3201  AGGATTATCC AAGATCAGAT GAGCACGTCA GAGAAGTTCA TCTTCGCGAA
      TCCTAATAGG TTCTAGTCTA CTCGTGCAGT CTCTTCAAGT AGAAGCGCTT
```

```
   +3 nGlyGlnSer AspArgAspThr MetSerIle IleGluAla ArgLeuGln
      --------------------------------------------------
3251  TGGGCAGTCA GACCGCGATA CCATGTCTAT CATTGAGGCA CGCCTGCAGG
      ACCCGTCAGT CTGGCGCTAT GGTACAGATA GTAACTCCGT GCGGACGTCC

   +3 luLysAlaArg GluPheLys SerIleAspIle LysGluPhe AspThrVal
      --------------------------------------------------
3301  AGAAGGCGCG GGAATTCAAA TCTATAGATA TCAAGGAGTT CGATACGGTA
      TCTTCCGCGC CCTTAAGTTT AGATATCTAT AGTTCCTCAA GCTATGCCAT

   +3 HisAsnTrpVal SerIleLeu ValPheSer TrpArgCysAsp ArgGly
      --------------------------------------------------
3351  CATAACTGGG TCAGTATTCT TGTCTTCTCG TGGCGTTGCG ACCGTGGGTG
      GTATTGACCC AGTCATAAGA ACAGAAGAGC ACCGCAACGC TGGCACCCAC

   +3 sProGluHis LeuIleGluTyr PheGluLys ArgSerLys SerArgThr
      --------------------------------------------------
3401  CCCAGAGCAC CTTATCGAGT ATTTCGAGAA ACGCTCGAAA AGCCGGACGC
      GGGTCTCGTG GAATAGCTCA TAAAGCTCTT TGCGAGCTTT TCGGCCTGCG

   +3 euSerSerArg IleGlySer ValAspValSer PheMetLeu AspSerGly
      --------------------------------------------------
3451  TCTCAAGCCG CATAGGAAGC GTCGACGTTA GCTTCATGCT CGATTCTGGC
      AGAGTTCGGC GTATCCTTCG CAGCTGCAAT CGAAGTACGA GCTAAGACCG

   +3 AlaValTrpThr IleAlaArg AsnThrLeu PheAlaSerGly LeuMet
      --------------------------------------------------
3501  GCTGTCTGGA CCATTGCCAG AAACACCTTA TTTGCCTCGG GTCTTATGCT
      CGACAGACCT GGTAACGGTC TTTGTGGAAT AAACGGAGCC CAGAATACGA

   +3 uAlaLeuPhe ValGlyValAsp PheIleAla AlaLysGly AspAspVal
      --------------------------------------------------
3551  CGCCCTTTTC GTCGGCGTCG ATTTCATCGC GGCGAAAGGC GATGATGTCT
      GCGGGAAAAG CAGCCGCAGC TAAAGTAGCG CCGCTTTCCG CTACTACAGA

   +3 heLeuAlaGly AsnAsnLeu TyrLeuAspAla GluArgLeu ArgMetGly
      --------------------------------------------------
3601  TCCTCGCAGG GAATAATTTG TACTTGGACG CAGAACGGCT TCGCATGGGA
      AGGAGCGTCC CTTATTAAAC ATGAACCTGC GTCTTGCCGA AGCGTACCCT

   +3 SerTyrLeuAla AlaAsnAsn LeuLysIle GluLysThrAla ValVal
      --------------------------------------------------
3651  TCTTACTTAG CCGCAAACAA CTTGAAGATC GAGAAGACGG CGGTCGTGAG
      AGAATGAATC GGCGTTTGTT GAACTTCTAG CTCTTCTGCC GCCAGCACTC

   +3 rPheIleGly PheIleValSer GlnAlaAla ValThrAla AspValVal
      --------------------------------------------------
3701  CTTTATAGGG TTTATCGTTT CCCAAGCCGC CGTCACAGCT GATGTCGTGC
      GAAATATCCC AAATAGCAAA GGGTTCGGCG GCAGTGTCGA CTACAGCACG

   +3 rgLeuAlaThr ArgThrTyr GlyArgSerTyr LysAsnAla AspAspLeu
      --------------------------------------------------
3751  GTCTAGCCAC CCGGACTTAC GGTCGAAGTT ATAAAAACGC TGATGATCTA
      CAGATCGGTG GGCCTGAATG CCAGCTTCAA TATTTTTGCG ACTACTAGAT
```

```
   +3 AlaLysTyrLys IleAlaIle AlaAspHis CysLysLeuPhe ArgSer
      ----------------------------------------------------
3801  GCGAAGTATA AGATAGCTAT CGCTGACCAC TGCAAGTTGT TTAGATCACC
      CGCTTCATAT TCTATCGATA GCGACTGGTG ACGTTCAACA AATCTAGTGG

   +3 oArgThrArg LeuMetThrAla IleAsnCys AlaThrLeu TyrGlyThr
      ----------------------------------------------------
3851  GAGAACTCGT CTCATGACCG CGATCAACTG CGCCACCCTT TACGGCACCT
      CTCTTGAGCA GAGTACTGGC GCTAGTTGAC GCGGTGGGAA ATGCCGTGGA

   +3 erLysGluCys IleAsnTyr LeuMetAspAla LeuAspAla PheGlyHis
      ----------------------------------------------------
3901  CGAAAGAGTG CATCAATTAT CTGATGGACG CGTTGGACGC ATTCGGACAC
      GCTTTCTCAC GTAGTTAATA GACTACCTGC GCAACCTGCG TAAGCCTGTG
                                      Start of subgenome
                                                  ▼
   +3 ThrLysMetSer AspLeuHis LeuAspPro GlyPheValMet ArgVal
      ----------------------------------------------------
3951  ACTAAAATGA GCGACCTACA CTTGGATCCC GGTTTTGTCA TGCGGGTCAC
      TGATTTTACT CGCTGGATGT GAACCTAGGG CCAAAACAGT ACGCCCAGTG

   +3 rProMetLys ValAspGluArg ValTyrSer GlyGlnAsp GlyCysGln
      ----------------------------------------------------
                           Start of Capsid gene ORF
   +1                                      Met AspAlaAsn
                                            ]------------
4001  CCCCATGAAG GTGGACGAGC GGGTTTATTC CGGACAAGAT GGATGCCAAC
      GGGGTACTTC CACCTGCTCG CCCAAATAAG GCCTGTTCTA CCTACGGTTG

   +3 rgAlaAspLys ThrArgGlu LysGlnProGlu ProGlyGln ProGlyPro
      ----------------------------------------------------
   +1 ValGlnIleArg ProAlaArg AsnAsnPro SerGlnGlyAsn GlnGly
      ----------------------------------------------------
4051  GTGCAGATAA GACCCGCGAG AAACAACCCG AGCCAGGGCA ACCAGGGCCG
      CACGTCTATT CTGGGCGCTC TTTGTTGGGC TCGGTCCCGT TGGTCCCGGC

   +3 GlnGlnGlnGln GlnAlaSer ThrGlnGlu AlaGlySerLys ThrSer
      ----------------------------------------------------
   +1 gAsnAsnAsn AsnLysArgArg ArgArgArg ArgGlyLeu LysLeuPro
      ----------------------------------------------------
4101  CAACAACAAC AACAAGCGTC GACGCAGGAG GCGGGGTCTA AAACTTCCCC
      GTTGTTGTTG TTGTTCGCAG CTGCGTCCTC CGCCCCAGAT TTTGAAGGGG

   +3 oArgSerArg ThrAspTyrGln ProArgPro AspGlyArg ThrArgGlu
      ----------------------------------------------------
   +1 roValValAla ProIleThr SerProGlyGln MetAlaGlu ProAlaAsn
      ----------------------------------------------------
4151  CCGTAGTCGC ACCGATTACC AGCCCCGGCC AGATGGCCGA ACCCGCGAAC
      GGCATCAGCG TGGCTAATGG TCGGGGCCGG TCTACCGGCT TGGGCGCTTG

   +3 roArgGluHis ProGlyGln ProArgSerAsp ThrArgGlu GlyValLys
      ----------------------------------------------------
   +1 HisAlaAsnThr ArgValAsn ArgGlyArg ThrArgValArg GlyLeu
      ----------------------------------------------------
4201  CACGCGAACA CCCGGGTCAA CCGCGGTCGG ACACGCGTGA GGGGGTTAAG
      GTGCGCTTGT GGGCCCAGTT GGCGCCAGCC TGTGCGCACT CCCCCAATTC
```

```
   +3 AlaSerAspAsp GlyGluSer HisGlySer AspIleArgGly MetAsp
      ------------------------------------------------------
   +1 gGlnAlaMet MetGluSerPro MetAlaAla ThrSerGlu AlaTrpIle
      ------------------------------------------------------
4251  GCAAGCGATG ATGGAGAGTC CCATGGCAGC GACATCAGAG GCATGGATTC
      CGTTCGCTAC TACCTCTCAG GGTACCGTCG CTGTAGTCTC CGTACCTAAG

   +3 rArgLeuSer ArgProGlyArg ArgIleGln AspGluPro GlyArgArg
      ------------------------------------------------------
   +1 isAspTyrLeu AspProAsp GlyGluTyrLys ThrSerLeu AspAspGly
      ------------------------------------------------------
4301  ACGACTATCT AGACCCGGAC GGAGAATACA AGACGAGCCT GGACGACGGG
      TGCTGATAGA TCTGGGCCTG CCTCTTATGT TCTGCTCGGA CCTGCTGCCC

   +3 luAsnSerArg ArgArgAsp ThrSerValAsn MetArgSer IleSerArg
      ------------------------------------------------------
   +1 LysIleProAsp GlyAlaIle ProGlnSer ThrCysGlyGln PheArg
      ------------------------------------------------------
4351  AAAATTCCCG ACGGCGCGAT ACCTCAGTCA ACATGCGGTC AATTTCGAGG
      TTTTAAGGGC TGCCGCGCTA TGGAGTCAGT TGTACGCCAG TTAAAGCTCC

   +3 AspArgGlyArg GlnIlePro GlyThrGlu PheTyrAspAla ThrAla
      ------------------------------------------------------
   +1 yThrValGly AlaArgTyrPro GlyLeuAsn SerThrThr LeuProLeu
      ------------------------------------------------------
4401  GACCGTGGGC GCCAGATACC CGGGACTGAA TTCTACGACG CTACCGCTGG
      CTGGCACCCG CGGTCTATGG GCCCTGACTT AAGATGCTGC GATGGCGACC

   +3 yTrpArgAsp LeuAlaSerThr SerAspAla SerProVal LeuGlnAla
      ------------------------------------------------------
   +1 spGlyGlyThr TrpProLeu LeuValMetHis LeuProPhe PheArgHis
      ------------------------------------------------------
4451  ATGGCGGGAC CTGGCCTCTA CTAGTGATGC ATCTCCCGTT CTTCAGGCAT
      TACCGCCCTG GACCGGAGAT GATCACTACG TAGAGGGCAA GAAGTCCGTA

   +3 erValValVal HisHisHis HisGlnGlnHis GlySerArg SerAspGlu
      ------------------------------------------------------
   +1 ProLeuLeuPhe IleThrThr ThrSerAsn ThrGluValGlu ValThr
      ------------------------------------------------------
4501  CCGTTGTTGT TCATCACCAC CACCAGCAAC ACGGAAGTCG AAGTGACGAA
      GGCAACAACA AGTAGTGGTG GTGGTCGTTG TGCCTTCAGC TTCACTGCTT

   +3 ArgArgSerGly CysValArg GluArgLeu GluGlnGlnAsp GlyLeu
      ------------------------------------------------------
   +1 nAlaAspLeu AspAlaPheAla AsnAspTrp AsnAsnArg ThrAspTrp
      ------------------------------------------------------
4551  CGCCGATCTG GATGCGTTCG CGAACGATTG GAACAACAGG ACGGACTGGA
      GCGGCTAGAC CTACGCAAGC GCTTGCTAAC CTTGTTGTCC TGCCTGACCT

   +3 pArgSerAsp ValProLysLeu GlyAlaSer ArgGluArg ValLeuHis
      ------------------------------------------------------
   +1 hrGluAlaThr TyrProSer TrpAlaGlnVal GlyAsnVal PheTyrMet
      ------------------------------------------------------
4601  CCGAAGCGAC GTACCCAAGT TGGGCGCAAG TCGGGAACGT GTTTTACATG
      GGCTTCGCTG CATGGGTTCA ACCCGCGTTC AGCCCTTGCA CAAAATGTAC
```

```
   +3 lyArgProAsp ArgSerAla AspGlyArgThr ThrProAsp SerThrGly
      ----------------------------------------------------
   +1 ValValProThr GluAlaLeu ThrAspVal ProProProThr GlnLeu
      ----------------------------------------------------
4651  GTCGTCCCGA CCGAAGCGCT GACGGACGTA CCACCCCCGA CTCAACTGGG
      CAGCAGGGCT GGCTTCGCGA CTGCCTGCAT GGTGGGGGCT GAGTTGACCC

   +3 CysIleArgVal ThrArgGlu LeuProSer AspIleGluArg ArgHis
      ----------------------------------------------------
   +1 yValSerGly LeuLeuGluSer TyrArgLeu ThrSerSer GlyValThr
      ----------------------------------------------------
4701  TGTATCAGGG TTACTCGAGA GTTACCGTCT GACATCGAGC GGCGTCACAG
      ACATAGTCCC AATGAGCTCT CAATGGCAGA CTGTAGCTCG CCGCAGTGTC

   +3 rValLeuGln ArgThrHisSer ArgGluSer GlySerGly GlyAspArg
      ----------------------------------------------------
   +1 laTyrPheAsn AlaProThr LeuValAsnGln GlyValAla ValIleAla
      ----------------------------------------------------
4751  CGTACTTCAA CGCACCCACT CTCGTGAATC AGGGAGTGGC GGTGATCGCG
      GCATGAAGTT GCGTGGGTGA GAGCACTTAG TCCCTCACCG CCACTAGCGC

   +3 laValProThr GlyGlnArg ThrProGluGly GluProGly HisSerSer
      ----------------------------------------------------
   +1 GlnPheGlnPro AspLysGlu HisGlnLys GluAsnProAsp IleVal
      ----------------------------------------------------
4801  CAGTTCCAAC CGGACAAAGA ACACCAGAAG GAGAACCCGG ACATAGTAGC
      GTCAAGGTTG GCCTGTTTCT TGTGGTCTTC CTCTTGGGCC TGTATCATCG

   +3 ArgAspHisPro AsnGlyArg AsnValThr AlaArgArgPhe ArgAla
      ----------------------------------------------------
   +1 aGlyThrThr GlnThrGlyGly ThrLeuGln LeuGlyGly SerGlyPro
      ----------------------------------------------------
4851  CGGGACCACC CAAACGGGCG GAACGTTACA GCTCGGCGGT TCAGGGCCGA
      GCCCTGGTGG GTTTGCCCGC CTTGCAATGT CGAGCCGCCA AGTCCCGGCT

   +3 uLeuHisIle AspAspAspAsp ArgGlyPro GlyArgVal ArgGlyArg
      ----------------------------------------------------
   +1 snTyrThrLeu ThrMetThr IleGlyAspGln ValGluPhe GlyGlyAla
      ----------------------------------------------------
4901  ACTACACATT GACGATGACG ATCGGGGACC AGGTCGAGTT CGGGGGCGCA
      TGATGTGTAA CTGCTACTGC TAGCCCCTGG TCCAGCTCAA GCCCCCGCGT

   +3 erAsnProAla ThrHisGly ValAspGlyAla AspAlaGly ValGlyAla
      ----------------------------------------------------
   +1 AlaIleProLeu ProThrVal SerMetGly ProMetProGlu SerGly
      ----------------------------------------------------
4951  GCAATCCCGC TACCCACGGT GTCGATGGGG CCGATGCCGG AGTCGGGGCA
      CGTTAGGGCG ATGGGTGCCA CAGCTACCCC GGCTACGGCC TCAGCCCCGT

   +3 AlaGlyValPro AspCysGlu ProAspIle ArgArgArgLys HisAsn
      ----------------------------------------------------
   +1 nLeuValPhe GlnThrAlaAsn LeuThrPhe AspValGly AsnThrIle
      ----------------------------------------------------
5001  GCTGGTGTTC CAGACTGCGA ACCTGACATT CGACGTCGGA AACACAATCA
      CGACCACAAG GTCTGACGCT TGGACTGTAA GCTGCAGCCT TTGTGTTAGT
```

```
   +3 sHisHisAsp HisAlaAlaThr ArgValGly AspGlyAsn ValAlaIle
      ----------------------------------------------------
   +1 hrIleThrThr ThrLeuPro ProGlySerVal ThrGlyMet TrpGlnPhe
      ----------------------------------------------------
5051  CCATCACGAC CACGCTGCCA CCAGGGTCGG TGACGGGAAT GTGGCAATTC
      GGTAGTGCTG GTGCGACGGT GGTCCCAGCC ACTGCCCTTA CACCGTTAAG

   +3 isSerGlnGln ArgAspGly HisArgAspArg GlyArgGly SerAlaThr
      ----------------------------------------------------
   +1 ThrAlaSerAsn GlyThrAsp ThrValThr ValAspAlaGly AlaArg
      ----------------------------------------------------
5101  ACAGCCAGCA ACGGGACGGA CACCGTGACC GTGGACGCGG GAGCGCGACT
      TGTCGGTCGT TGCCCTGCCT GTGGCACTGG CACCTGCGCC CTCGCGCTGA

   +3 ValArgValArg SerGluPhe GlyArgLeu GlyThrGluSer AlaGly
      ----------------------------------------------------
   +1 uTyrAlaPhe GlyAlaAsnLeu AspAlaSer GluLeuAsn LeuGlnAsp
      ----------------------------------------------------
5151  GTACGCGTTC GGAGCGAATT TGGACGCCTC GGAACTGAAT CTGCAGGACA
      CATGCGCAAG CCTCGCTTAA ACCTGCGGAG CCTTGACTTA GACGTCCTGT

   +3 sGlnLeuAsn GlnAspSerThr AsnGluHis GluProAsn AspAlaGly
      ----------------------------------------------------
   +1 leAsnSerIle LysIlePro ProThrAsnMet AsnGlnMet MetGlnAla
      ----------------------------------------------------
5201  TCAACTCAAT CAAGATTCCA CCAACGAACA TGAACCAAAT GATGCAGGCA
      AGTTGAGTTA GTTCTAAGGT GGTTGCTTGT ACTTGGTTTA CTACGTCCGT

   +3 snAlaGlnAsp HisSerVal ProThrGlnArg AsnGluGly LeuLeuTyr
      ----------------------------------------------------
   +1 ThrProLysThr IleGlnPhe GlnLeuAsn GluThrLysGly PheTyr
      ----------------------------------------------------
5251  ACGCCCAAGA CCATTCAGTT CCAACTCAAC GAAACGAAGG GCTTTTATAT
      TGCGGGTTCT GGTAAGTCAA GGTTGAGTTG CTTTGCTTCC CGAAAATATA

   +3 AlaProGluGly ValProThr ArgValArg AsnAspAsnGly AspVal
      ----------------------------------------------------
   +1 tProLeuArg AlaPheGlnPro ValPheGlu MetThrMet AlaThrSer
      ----------------------------------------------------
5301  GCCCCTGAGG GCGTTCCAAC CCGTGTTCGA AATGACAATG GCGACGTCTT
      CGGGGACTCC CGCAAGGTTG GGCACAAGCT TTACTGTTAC CGCTGCAGAA

   +3 uTrpThrGly AlaMetGluAsp ThrGluAsp AsnCysGly ArgLeuPro
      ----------------------------------------------------
   +1 yrGlyProVal ArgTrpLys ThrProArgThr ThrValVal AspTyrHis
      ----------------------------------------------------
5351  ATGGACCGGT GCGATGGAAG ACACCGAGGA CAACTGTGGT AGATTACCAC
      TACCTGGCCA CGCTACCTTC TGTGGCTCCT GTTGACACCA TCTAATGGTG

   +3 roGlyAsnTrp TrpThrPro GlyTyrHisArg GlnGlnLeu ArgAspArg
      ----------------------------------------------------
   +1 ArgAlaIleGly GlyLeuGln AspThrIle AspSerAsnPhe AlaIle
      ----------------------------------------------------
5401  CGGGCAATTG GTGGACTCCA GGATACCATC GACAGCAACT TCGCGATAGG
      GCCCGTTAAC CACCTGAGGT CCTATGGTAG CTGTCGTTGA AGCGCTATCC
```

```
   +3 ArgCysArgAsp AspArgTyr ValTyrIle AsnArgThrLeu LeuGln
      ------------------------------------------------------
   +1 yValAlaAla MetThrGlyMet SerThrSer ThrValPro TyrPheLys
      ------------------------------------------------------
5451  CGTTGCCGCG ATGACCGGTA TGTCTACATC AACCGTACCC TACTTCAAGG
      GCAACGGCGC TACTGGCCAT ACAGATGTAG TTGGCATGGG ATGAAGTTCC

   +3 yValProThr LeuArgSerAsp ThrGlyGly GlyGluPro LeuGlyPro
      ------------------------------------------------------
   +1 alPheArgArg PheGluAla IleProAlaGlu GlySerPro TrpGlyPro
      ------------------------------------------------------
5501  TGTTCCGACG CTTCGAAGCG ATACCGGCGG AGGGGAGCCC TTGGGGCCCC
      ACAAGGCTGC GAAGCTTCGC TATGGCCGCC TCCCCTCGGG AACCCCGGGG

   +3 euArgGlnCys AspThrSer GluGlyArgArg GlyAlaAsn SerGlySer
      ------------------------------------------------------
   +1 PheAlaSerAla ThrProPro LysAspAsp ValAlaLeuThr ValAla
      ------------------------------------------------------
5551  TTCGCCAGTG CGACACCTCC GAAGGACGAC GTGGCGCTAA CAGTGGCTCG
      AAGCGGTCAC GCTGTGGAGG CTTCCTGCTG CACCGCGATT GTCACCGAGC

   +3 AsnLeuAspArg SerAlaPro IleArgIle ProGlyThrIle GlnArg
      ------------------------------------------------------
   +1 gThrTrpThr AspLeuHisPro PheAlaTyr ProGluArg TyrAsnGly
      ------------------------------------------------------
5601  AACTTGGACC GATCTGCACC CATTCGCATA CCCGGAACGA TACAACGGAT
      TTGAACCTGG CTAGACGTGG GTAAGCGTAT GGGCCTTGCT ATGTTGCCTA

   +3 eArgGlyPro IleArgAspGly GlyGlnAsp HisSerPro AspThrSer
      ------------------------------------------------------
   +1 heGlyAlaLeu PheAlaMet ValAlaLysThr IleAlaGln IleProArg
      ------------------------------------------------------
5651  TCGGGGCCCT ATTCGCGATG GTGGCCAAGA CCATAGCCCA GATACCTCGC
      AGCCCCGGGA TAAGCGCTAC CACCGGTTCT GGTATCGGGT CTATGGAGCG

   +3 euCysAlaIle SerSerArg SerGlyGluCys GlyAspGly LeuHisArg
      ------------------------------------------------------
   +1 TyrValArgSer AlaAlaGly ValAlaAsn AlaValThrAsp CysIle
      ------------------------------------------------------
5701  TATGTGCGAT CAGCAGCCGG AGTGGCGAAT GCGGTGACGG ACTGCATAGA
      ATACACGCTA GTCGTCGGCC TCACCGCTTA CGCCACTGCC TGACGTATCT

   +3 GluArgAspArg GluCysSer LeuGluPhe HisLeuGlyGlu AlaAla
      ------------------------------------------------------
   +1 uSerAlaThr GluSerValAla SerAsnSer ThrSerGlu ArgArgGln
      ------------------------------------------------------
5751  GAGCGCGACC GAGAGTGTAG CCTCGAATTC CACCTCGGAG AGGCGGCAAC
      CTCGCGCTGG CTCTCACATC GGAGCTTAAG GTGGAGCCTC TCCGCCGTTG

   +3 rLysSerGlu ThrCysTrpArg AsnArgSer ArgSerPro GlnSerCys
      ------------------------------------------------------
   +1 rgArgAlaArg ArgValGly GlyIleAlaArg GlyAlaArg AsnLeuVal
      ------------------------------------------------------
5801  GAAGAGCGAG ACGTGTTGGC GGAATCGCTC GAGGAGCCCG CAATCTTGTG
      CTTCTCGCTC TGCACAACCG CCTTAGCGAG CTCCTCGGGC GTTAGAACAC
```

End of Replicase and Capsid gene ORFs

```
      +3 lyProHisArg GluPro
         ----------------->
      +1 GlyArgIleGly AsnLeuSer Leu
         ------------------------>
   5851  GGCCGCATAG GGAACCTTAG CTTGTAGGTT CATTGCGACA TGGGATGTTC
         CCGGCGTATC CCTTGGAATC GAACATCCAA GTAACGCTGT ACCCTACAAG

   5901  TTCCAGTCAG CAGCTCTTCG GTTTCATCTC CACTGACGAC CCTCTGCACG
         AAGGTCAGTC GTCGAGAAGC CAAAGTAGAG GTGACTGCTG GGAGACGTGC

   5951  TATTTTCCAT CATCGTGGTG CTGATCATCA TTTGCGGTCT GGTTTGGCTC
         ATAAAAGGTA GTAGCACCAC GACTAGTAGT AAACGCCAGA CCAAACCGAG

   6001  GCTTCCTATT GTTTTCATGG CAGATCAACT CGACCCAATC GAAATCTTCA
         CGAAGGATAA CAAAAGTACC GTCTAGTTGA GCTGGGTTAG CTTTAGAAGT

   6051  TCGAATTCCT CGGATTTCGA AAGATTTCCG GAGCCATTCG CTGCTTCCAG
         AGCTTAAGGA GCCTAAAGCT TTCTAAAGGC CTCGGTAAGC GACGAAGGTC

   6101  TGTAAGCAAG TCTTGGGACC TGCAGGACCC GTGGATTCCC ACCTCACTGA
         ACATTCGTTC AGAACCCTGG ACGTCCTGGG CACCTAAGGG TGGAGTGACT

   6151  TTGTCAGTGG AAGGCCGCGG TGTTAAATCT TATTAACAAT CAGCTTTACG
         AACAGTCACC TTCCGGCGCC ACAATTTAGA ATAATTGTTA GTCGAAATGC

   6201  ACGTCGATAT CGATGAGACG AATCCGTTTC TTTACGGACC TCACCGCGAC
         TGCAGCTATA GCTACTCTGC TTAGGCAAAG AAATGCCTGG AGTGGCGCTG

   6251  TGAGATGTGG AAGACCACAT TGTTCCTTCA CACATGCCCA GAGTGCGGTT
         ACTCTACACC TTCTGGTGTA ACAAGGAAGT GTGTACGGGT CTCACGCCAA

   6301  ATTCCACCAG GGACACAGAA ACTACGAGAT CGTGTCCCCG AGATTGCCAA
         TAAGGTGGTC CCTGTGTCTT TGATGCTCTA GCACAGGGGC TCTAACGGTT

   6351  GACGGCAATC TTATGCACGC ATCTTCGGTC GGCTATATTT GTCACAAATG
         CTGCCGTTAG AATACGTGCG TAGAAGCCAG CCGATATAAA CAGTGTTTAC

   6401  CCGGTTAGAA GCAAACACAT TTTACCACGG TTTATGCTCT CAGTGTCGCG
         GGCCAATCTT CGTTTGTGTA AAATGGTGCC AAATACGAGA GTCACAGCGC

   6451  ACCGTGATAA TAAAAAACGA CGCTGAAGAG AGGACTCACA ACTACCTCGA
         TGGCACTATT ATTTTTTGCT GCGACTTCTC TCCTGAGTGT TGATGGAGCT

   6501  TCTCGTTTAT CGGACGAGTG ATACAATTGA CCCAGGGTCA TCCTGCAAAA
         AGAGCAAATA GCCTGCTCAC TATGTTAACT GGGTCCCAGT AGGACGTTTT

   6551  CACGCAGGTT TCCGATAGTG GTGCAAATCC ACCCGCCAGT CGTCGGTGGT
         GTGCGTCCAA AGGCTATCAC CACGTTTAGG TGGGCGGTCA GCAGCCACCA

   6601  CCCTTGCGGG ACCTATACGG TACCA
         GGGAACGCCC TGGATATGCC ATGGT
```

REFERENCES:

Agrawal, D.K. and J.E. Johnson. 1992. Sequence and analysis of the capsid protein of *Nudaurelia capensis* ω virus, an insect virus with T=4 icosahedral symmetry. Virology **190**:806-814.

Agrawal, D.K. and J.E. Johnson. 1995. Assembly of the T=4 Nudaurelia capensis ω virus capsid protein, post-translational cleavage, and specific encapsidation of its mRNA in a baculovirus expression system. Virology **207**:89-97.

Ahlquist, P. 1986. In vitro transcription of infectious viral RNA from cloned cDNA. Methods in Enzymology **118**:704-716.

Baneyx, F. 1999. Recombinant protein expression in *Escherichia coli*. Current Opinion in Biotechnology **10**:411-421.

Berghammer, H. and B. Auer. 1993. "Easypreps": Fast and Easy Plasmid Minipreparation for Analysis of Recombinant Clones in *E. coli*. BioTechniques **14**:527-528.

Boyer, J.C. and A.L. Haenni. 1994. Infectious transcripts and cDNA clones of RNA viruses. Virology **198**:415-426.

Brown, F. 1989. The classification and nomenclature of viruses: Summary of results of meetings of the International Committee on Taxonomy of Viruses in Edmonton, Canada 1987. Intervirology **30**:181-186.

Canady, M.A., M. Tihova, T.N. Hanzlik, J.E. Johnson, and M. Yeager. 2000. Large Confirmational Changes in the Maturation of a Simple RNA Virus, *Nudaurelia capensisω Virus (NωV)* . Journal of Molecular Biology **299**:573-584.

Chao, Y.C., H.A. Scott, and S.Y. Young. 1983. An icosahedral RNA virus of the soybean looper (*Pseudoplusia includens*). Journal of General Virology **64**:1835-1838.

Christian, P.D., T.N. Hanzlik, D.J. Dall, and K.H.J. Gordon. 1993. Insect viruses: new strategies for pest control. In Molecular Approaches to Fundamental and Applied Entomology 128-163.

Domier, L.L., K.M. Franklin, A.G. Hunt, R.E. Rhoads, and J.G. Shaw. 1989. Infectious in vitro transcripts from cloned cDNA of a potyvirus, tobacco vein mottling virus. Proc.Natl.Acad.Sci.USA **86**:3509-3513.

du Plessis, D.H., G. Mokhosi, and D.A. Hendry. 1991. Cell-free translation and identification of the replicative form of *Nudaurelia β virus* RNA. Journal of General Virology **72**:267-273.

Finch, J.T., R.A. Crowther, D.A. Hendry, and J.K. Struthers. 1974. The structure of *Nudaurelia capensis* β Virus: the First Example of a Capsid with Icosahedral Surface Symmetry T=4. Journal of General Virology **24**:191-200.

Francki, R.I.B., C. Fauquet, D.L. Knudson, and F. Brown. 1991. Classification and Nomenclature of Viruses Fifth Report of the International Committee on Taxonomy of Viruses. Archives of Virology **Supplementum 2**:80-85.

Geertsema, H. 1970. Die dennepouoogmot, Nudaurelia cytheria (Fabri.) (Lepidoptera Saturniidae). Plae en siektes van Suid-Afrikaanse bosse en hout **Pamflet 273**:1-11.

Geertsema, H. and J.H. Giliomee. 1972. Aspects of the Biology of the Pine Tree Emperor Moth, Nudaurelia cytherea (Fabr.) (Lepidoptera: Saturniidae). Forestry in South Africa **13**:1-13.

Geertsema, H. 1975. Studies on the Biology, Ecology and Control of the Pine Tree Emperor Moth, *Nudaurelia cyntherea cyntherea* (Fabr.)(Lepidoptera: Saturniidae). Annals of University of Stellenbosch **50 A (1)**:1-170.

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

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

Grace, T.D.C. and E.H. Mercer. 1965. A new virus of the saturniid *Antheraea encalypti* Scott. Journal of Invertebrate Pathology **7**:241-244.

Greenwood, L.K. and N.F. Moore. 1981. The Nudaurelia-β group of small RNA viruses of insects: Serological comparison of four members. Microbiologica **4**:271-280.

Greenwood, L.K. and N.F. Moore. 1981. A single protein *Nudaurelia* β–like virus of the Pale Tussock Moth, Dasychira pudibunda. Journal of Invertebrate Pathology **38**:305-306.

Greenwood, L.K. and N.F. Moore. 1982. The Nudaurelia β group of small RNA-containing viruses of insects: Serological identification of several new isolates. Journal of Invertebrate Pathology **39**:407-409.

Hanzlik, T.N., S.J. Dorrian, K.H.J. Gordon, and P.D. Christian. 1993. A novel small RNA virus isolated from the cotton bollworm, *Helicoverpa armigera*. Journal of General Virology **74**:1805-1810.

Hanzlik, T.N. and K.H.J. Gordon. 1997. The Tetraviridae. Advances in Virus Research **48**:101-168.

Hanzlik, T.N., K.H.J. Gordon, P.D. Christian, and J.-L. Zeddam. 1999. New views of insect small RNA viruses for the pesticide industry. (Generic)

Hendry, D.A., M.F. Bekker, and M.H.V. van Regenmortel. 1967. A non-inclusion virus of the Pine Emperor Moth, *Nudaurelia Cyntherea Capensis* Stoll. S.A.Medical Journal **42**:117-117.

Hendry, D.A., V. Hodgson, R. Clark, and J. Newman. 1985. Small RNA viruses co-infecting the Pine Emperor Moth (*Nudaurelia cytherea capensis*). Journal of General Virology **66**:627-632.

Hendry, D.A. and D.K. Agrawal. 1994. Tetraviruses. Encyclopedia of Virology 1416-1422.

Juckes, I.R.M., J.F. Longworth, and C. Reinganum. 1973. A serological comparison of some nonoccluded insect viruses. Journal of Invertebrate Pathology **21**:119-120.

Juckes, I.R.M. 1979. Comparison of some biophysical properties of the *Nudaurelia* β and ε viruses. Journal of General Virology **42**:89-94.

King, L.A., A.T. Merryweather, and N.F. Moore. 1984. Proteins expressed in vitro by two members of the Nudaurelia β family of viruses, *Trichoplusia ni* and *Dasychira pudibunda* viruses. Annals of Virology (Inst.Pasteur) **135 E**:335-342.

MacCallum, F., G. Brown, and T.W. Tinsley. 1979. Antibodies in human sera reacting with an insect pathogenic virus. Intervirology **11**:234-237.

Mathews, R.E.F. 1982. *Nudaurelia* β virus group. Intervirology **17**:135-135.

Moore, N.F., L.K. Greenwood, and K.R. Rixon. 1981. Studies on the capsid protein of several members of the Nudaurelia β group of small RNA viruses of insects. Microbiologica **4**:59-71.

Moore, N.F. and T.W. Tinsley. 1982. The small RNA-viruses of insects. Archives of Virology **72**:229-245.

Moore, N.F., B. Reavy, and L.A. King. 1985. General characteristics, gene organization and expression of small RNA viruses of insects. Journal of General Virology **66**:647-659.

Moore, N.F. 1991. The Nudaurelia β family of insect viruses. In Viruses of Invertebrates 277-285.

Morris, T.J., R.T. Hess, and D.E. Pinnock. 1979. Physicochemical characterization of a small RNA virus associated with Baculovirus infection in *Trichoplusia ni*. Intervirology **11**:238-247.

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

Reavy, B. and N.F. Moore. 1984. In vitro translation of the genome of a small RNA virus from *Trichoplusia ni*. Journal of Invertebrate Pathology **44**:244-248.

Reinganum, C., J.S. Robertson, and T.W. Tinsley. 1978. A new group of RNA viruses from insects. Journal of General Virology **40**:195-202.

Reinganum, C. 1991. Tetraviridae. In Atlas of Invertebrate Viruses 553-592.

Sambrook, J., E.F. Fritsch, and T. Maniatis. 1989. Molecular Cloning: A Laboratory Manuel. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY,

Schenk, P.M., S. Baumann, R. Mattes, and H.-H. Steinbiβ. 1995. Improved high-level expression system for eukaryotic genes in *Escherichia coli* using T7 RNA polymerase and rare ^{Arg}tRNAs. BioTechniques **19**:196-200.

Sheibani, N. 1999. Prokaryotic Gene Fusion Expression Systems and their use in Structural and Functional Studies of Proteins. Preparative Biochemistry & Biotechnology **29**:77-90.

Smith, D.B. and K.S. Johnson. 1988. Single-step purification of polypeptide expressed in *Escherichia coli* as fusion with glutathione S-transferase. Gene **67**:31-40.

Smith, D.B. and L.M. Corcoran. 1994. Expression and Purification of Glutathione-S-Transferase Fusion Proteins. Current Protocols in Molecular Biology **28**:16.7.1-16.7.7

Stemmer, W.P.C. 1994. Rapid evolution of a protein in vitro by DNA shuffling. Nature **370**:389-391.

Stemmer, W.P.C. 1994. DNA shuffling by random fragmentation and reassembly: In vitro recombination for molecular evolution. Proc.Natl.Acad.Sci.USA **91**:10747-10751.

STRATAGENE. Epicurian Coli® BL21-CodonPlus™(DE3)-RIL Competent Cells. 059001(230245). 1999. Stratagene Distributors. (GENERIC)
Ref Type: Catalog

Struthers, J.K. and D.A. Hendry. 1974. Studies of the Protein and Nucleic Acid Components of Nudaurelia capensis β Virus. Journal of General Virology **22**:355-362.

Tellinghuisen, T.L., A.E. Hamburger, R.O. Fisher, and R.J. Kuhn. 1999. In vitro assembly of alphavirus cores by using nucleocapsid protein expressed in *Escherichia coli.* Journal of Virology **73**:5309-5319.

Tooke, F.G.C. and C.S. Hubbard. 1941. The pine tree emperor moth *Nudaurelia cytherea capensis* Stoll. Dept.of Agric.and Forestry (Ent.Ser.) Sc.Bull. **210**:1-57.

Tripconey, D. 1970. Studies on a nonoccluded virus of the pine tree emperor moth. Journal of Invertebrate Pathology **15**:268-275.

Vail, P.V., T.J. Morris, and S.S. Collier. 1983. An RNA virus in *Autographa californica* nuclear polyhedrosis virus preparations: gross pathology and symptoms. Journal of Invertebrate Pathology **41**:183