MOLECULAR, FUNCTIONAL AND BIOLOGICAL CHARACTERIZATION OF A SINGLE ENVELOPED NUCLEOPOLYHEDROVIRUS INFECTING THE CROP PEST *Helicoverpa armigera*

by

Sehaam Khan

A thesis submitted in fulfillment of the requirements for the degree of Doctor Philosophiae in the Faculty of Sciences, University of the Western Cape.

June 2002

Supervisor:
Dr. S. Davison

KEYWORDS

Baculovirus

Helicoverpa armigera

Single-capsid nucleopolyhedrovirus

In vitro

Baculovirus repeat open reading frames (*bro*) genes

Ecdysteroid UDP-glucosyltranferase (*egt*) gene

Helicase gene

Chitinase gene

Late expression factor (*lef*) 8 gene

Phylogeny.

ABSTRACT

Molecular, functional and biological characterization of a single enveloped nucleopolyhedrovirus infecting the crop pest *Helicoverpa armigera*

S. Khan

PhD Thesis, Faculty of Sciences, University of the Western Cape

My study involves a detailed molecular and cellular characterisation of a South African isolate of a *nucleopolyhedrovirus* (NPV) isolated from the Lepidopteran insect *Helicoverpa armigera* (HaSNPV-SA). Chapter two entails the *in vitro* infection of *Heliothis zea* cell lines with HaSNPV-SA. Within this chapter I optimise the propagation of HaSNPV-SA within the cell-line HzAMI, I obtain a pure genotype (HaSNPV-Pl3) and compare the cell-lines HzAMI and Hz2E5 to determine which allows for faster propagation. Within chapter three, I discuss the identification of the ecdysteroid UDP-glucosyltranferase (*egt*) gene of HaSNPV-SA using a Hz-SNPV gene-specific probe. Studies have shown that the enzyme disrupts the ecdysteroid balance of the host larva, causing a delay in the onset of molting. That a secreted and active EGT is encoded by HaSNPV-SA was confirmed by assay of infected cell culture medium. Chapter four discusses the diversity of baculovirus repeated open reading frames (*bro*) in three nucleopolyhedroviruses of *Helicoverpa spp.* BROs of some NPV's have been shown to be involved in nucleosome organization that could block cellular replication and/or transcription and switch host machinery to viral DNA or RNA synthesis. I show that the three *bro*-genes present in four *Helicoverpa* isolates are not conserved. Furthermore, Northern analyses and/or RT-PCR indicated that all Ha-G4 3 *bro*-genes were transcribed at either 4h to 24h post infection (p.i). The final chapter, chapter five, identifies three essential baculovirus genes [*helicase*, *chitinase* and late expression (*lef*) 8] of HaSNPV-SA and uses them to confirm the placement of HaSNPV-SA in baculovirus phylogeny. All three trees confirm the placement of HaSNPV's as a group II type NPV. Based on sequence similarities, gene placement and conservation and phylogeny it is tempting to say the HaSNPV-SA isolate is a variant of HzSNPV-ELCAR.

June 2002

DECLARATION

I, declare that the work contained in "Molecular, functional and biological characterization of a single enveloped nucleopolyhedrovirus infecting the crop pest *Helicoverpa armigera*", is my own work and that I has not been submitted for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged by complete references.

Full name ..... **Date** .....

Signed .....

This thesis is dedicated to my Mother, Father and in memory of Amma.

BIOGRAPHICAL SKETCH

Sehaam Khan was born in Cape Town, South Africa, on the 18 February 1975. She attended Accordian Primary School and matriculated at Excelsior Secondary School in 1992. Sehaam enrolled at the University of the Western Cape in 1993 and obtained a B.Sc. degree in Biochemistry and Microbiology in 1995. In 1996 she completed a B.Sc. Hons. degree in Microbiology at the same university. In 1999 she obtained her MSc. Degree in Microbiology at UWC with distinction. She is presently a full-time employer of Cape Technikon in the capacity of lecturer.

ACKNOWLEDGEMENTS

I wish to express my sincere gratitude and appreciation to the following persons and institutions:

PROF. S. DAVISON, Department of Microbiology, University of the Western Cape, who acted as supervisor, for accepting me as a student, for his enthusiasm, devotion and guidance throughout this project;

DR. BURTRAM. FIELDING AND MR. CLIFFORD. JACOBS, for invaluable discussions, technical advice and contributions to this project but, especially to Bertie for being a FRIEND;

PROF JUST VLAK, Laboratory of Virology, Wageningen University, who acted as co-supervisor for the *bro* chapter, for providing me with the opportunity of enriching my professional and personal life by allowing me to work in his laboratory in Wageningen and for his invaluable scientific input.

STAFF AND STUDENTS, Department of Biotechnology at the University of the Western Cape, especially "Club Viro", for their encouragement and assistance;

THE NATIONAL RESEARCH FOUNDATION, THE MELLON SCHOLARSHIP AND THE SOLM YACH FELLOWSHIP, for financial assistance;

MONIQUE VAN OERS, XINWEN CHEN AND MARCEL WESTENBERG, for helpful discussions and invaluable guidance;

MAGDA USMANY, ELS ROODE AND THEA VAN BEMMEL, for amazing assistance and advice;

THE "MENSA" GROUP IN WAGENINGEN, how boring life would have been without our famous talks and discussions;

SIKANDER, PRIYA AND MICHAEL, truly good friends;

WESAAL for her support, encouragement, enthusiasm and assistance whenever and wherever I needed it most, but mostly just for being "Laalie";

MY MOTHER, if I was offered all the money in the world to exchange her, IT WOULD NOT BE ENOUGH;

MY FATHER, WALEED, UNCLE RONNIE AND AMMA^{+}, for their support and encouragement;

THE LORD, WOW where to begin... for the many blessings you bestowed upon me, I recognize and appreciate each one, and for being the best friend any one person could ask for.

CONTENT PAGE

This thesis is presented as a compilation of six chapters. Each chapter is introduced separately and is written according to the *Journal of Virology* to which chapter four will be submitted for publication. Chapter three is to be submitted for publication to *Virus Genes*. Chapter five is to be submitted to the *South African Journal of Science*.

CHAPTER ONE

MOLECULAR, FUNCTIONAL AND BIOLOGICAL CHARACTERIZATION OF A SINGLE ENVELOPED NUCLEOPOLYHEDROVIRUS INFECTING THE CROP PEST *HELICOVERPA ARMIGERA*

Helicoverpa armigera:

Helicoverpa (*Heliothis*) *armigera* (Lepidoptera: Noctuidae) commonly referred to as the bollworm, is a highly polyphagous agricultural pest. Host species for *H. armigera* include important agricultural crops such as cotton, maize, chickpea, sunflower, soybean, sorghum and including ornamentals such as carnations and geraniums (Fitt, 1989). The female moth only lives for approximately two weeks, but during this period she is capable of laying up to 5000 eggs (Taylor, 1982). Once the eggs are hatched, the larvae feed veraciously on the flowing and fruiting structures of their crops, and this leads to substantial economic losses.

H. armigera is present in mainland Europe, Asia, Africa and Australia. It has been found that few pests cause as much damage as the bollworm. For many years, this pest has been successfully controlled using chemical insecticides, but the indiscriminate use of insecticides during the 1980's and 1990's, has led to the resistance of this pest to most of them. Moderate to high levels of resistance to cypermethrin, monocrotophos, endosulphan, fenvalerate and quinalphos have been recorded. Since it has been shown, that this pest has a strong ability to develop resistance to chemical insecticides, integrated pest management (IPM), which involves

the use of many techniques, including biological control, to provide effective control of crop pests, was needed.

General:

Disease-causing pathogens such as fungi, bacteria, viruses and protozoa have great potential for successful use as microbial pesticides. An important advantage of successful microbial pesticides is that it is less likely that pests would develop resistance to biological control agents. Also, because biological control is a natural phenomenon, fears with regard to environmental safety and health risks to humans and animals are less likely to arise. Beneficial insects and other non-target organisms such as parasites and predators are also not affected, due to the target specificity of the microbes.

Baculoviruses:

Baculoviruses are one of the largest and most diverse groups of insect pathogenic viruses and have great potential as biological control agents for successful use in pest control programmes. Baculovirus diseases have been described in more than 800 species of insects, which include mainly the orders Lepidoptera, Hymenoptera, Diptera and Coleoptera, but also Neuroptera, Trichoptera, Thysanoptera, Siphonophtera as well as in crustaceans (Decopoda) (Adams and McClintock, 1991). Some baculoviruses as capable of infecting few related insect species, but the majority of baculoviruses are extremely host specific. They are able to cause epizootic in nature, which appear to play a role in controlling insect populations. At present, there are numerous baculoviruses used worldwide as ecologically sound biological pesticides and there

are more than 32 registered for pest control. Table 1. mentions only some of the baculoviruses registered at present (baculovirus.com).

Table 1. Some baculoviruses registered for pest control (baculovirus.com)

COMMODITY	INSECT PEST	VIRUS USED	VIRUS PRODUCT
Apple, pear, walnut and plum	Codling moth	Codling moth granulovirus	Cyd-X (Thermo Trilogy Corp)
Cabbage, tomatoes, cotton, (and see pests in next column)	Cabbage moth, American bollworm, diamondback moth, potato tuber moth, and grape berry moth	Cabbage army worm nucleopolyhedrosis virus	Mamestrin (Natural Plant Protection)
Cotton, corn, tomatoes	*Spodoptera littoralis*	*Spodoptera littoralis* nucleopolyhedrosis virus	Spodopterin (Natural Plant Protection)
Cotton and vegetables	Tobacco budworm *Helicoverpa zea*, and Cotton bollworm Heliothis virescens	*Helicoverpa zea* nucleopolyhedrosis virus	Gemstar LC, Biotrol, Elcar (Thermo Trilogy Corp)
Vegetable crops, greenhouse flowers	Beet armyworm (*Spodoptera exigua*)	*Spodoptera exigua* nucleopolyhedrosis virus	Spod-X (Thermo Trilogy Corp)
Alfalfa and other crops	Alfalfa looper (*Autographa californica*)	*Autographa californica* nucleopolyhedrosis virus	Gusano Biological Pesticide (Thermo Trilogy Corp)
Forest Habitat, Lumber	Douglas fir tussock moth *(Orgyia psuedotsugata)*	*Orgyia psuedotsugata* nucleopolyhedrosis virus	TM Biocontrol (USDA Forest Service)
Forest Habitat, Lumber	Gypsy moth (Lymantria dispar)	Lymantria dispar nucleopolyhedrosis virus	Gypchek

Taxonomy:

Baculovirus taxonomy is based on the number of virions occluded within an occlusion body (OB) (Fig. 1), and is divided into two taxonomic genera: *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) (Blissard *et al.*, 2000). In the case of NPVs, each OB or polyhedra (ranging in size from 0.15-1.5 :M) can contain as many as 200 virions while GV OB's or granule's (ranging in size from 300-500 nM), in contrast, contain a single virion and at most a few. NPV's have two morphotypes, SNPV and MNPV, depending on the single or multiple packaging of the nucleocapsids in the virion. Phylogenetic analysis based on various genes, divides the GV's and NPV's into separate groups, but also divides the NPV's into two groups, group I and II NPV's (Herniou *et al.*, 2001). *H. armigera* SNPV is a member of group II NPV.

Figure 1. Electron micrographs of the two genera of baculoviruses (from www.virology.net/garryfavweb11.html and www.mioti.com/virology/bac.htm). A. *Nucleopolyhedrovirus* occlusion body (polyhedron). B. *Granulovirus* occlusion body (granule).

Two phenotypic forms of the NPV baculovirus virions exist; those virions that have been released from the OB's (polyhedra) are termed occlusion-derived virus (ODV) and the budded virion (BV), which bud from the cell and are not occluded (Fig. 2). ODV is responsible for the primary infection of the insect, whereas BV is responsible for the secondary infection within the insect. The virions are comprised of rod-shaped nucleocapsids that contain the DNA-protein

complex of the virus. The NPV nucleopcapsids are approximately 40 to 60 nm (diameter) x 250 to 300 nm (length) and comprise a 39 kDa capsid protein (Pearson *et al.*, 1988; Blissard *et al.*, 1989; Thiem and Miller, 1989 and Guarino and Smith, 1990).

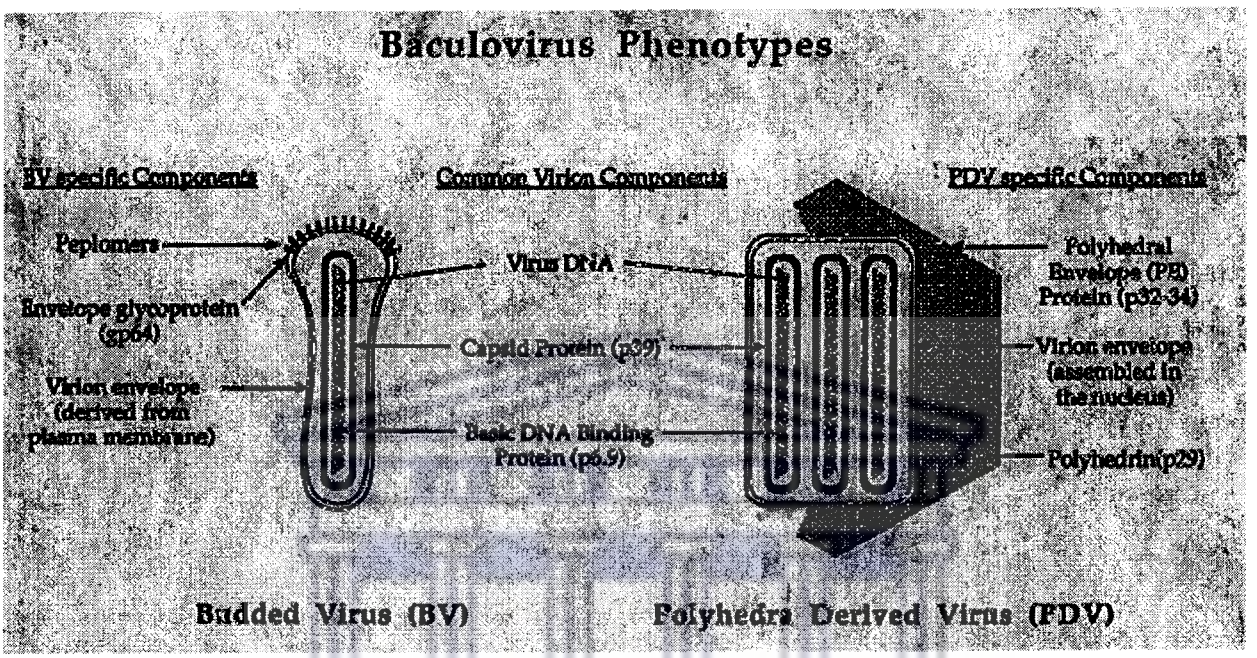


Figure 2. Schematic diagram of the two Baculovirus phenotypes, the budded virion (BV) and occluded-derived virions (ODV) (Blissard, 1996). Proteins, which are shared by both phenotypes, are indicated in the centre, and phenotype-specific components are indicated on the left for BV and on the right for ODV.

Infection cycle:

NPV's are commonly found on plant surfaces and in the soil. As is the case with the GV infection, the infection is initiated when NPV's are ingested by a susceptible during feeding of the larva. Once ingested, the occlusion body dissolves in the presence of the high alkaline pH of the insect midgut, thus releasing virions. The ODV then passes through the peritrophic membrane and fusion with the midgut epithelial cell plasma membrane commences. The pathway of NPV infection is outlined in Fig. 3. Once nucleocapsids have been synthesised within the nucleus, they pass through the plasma membrane, bud from the cell and acquire their

envelope from the plasmalemma to become budded virions. BV's are then responsible for secondary infection of surrounding cells. ODV are only formed later in the infection cycle. Studies of the *Heliothis zea* (Hz) NPV, showed virus fusion and entry into the cytoplasm, presence of a nucleocapsid in the cell cytoplasm, uncoating in the cell nucleus, BV progeny formulation and PDV progeny formation, occurred 1-4 h p.i., 2-4 h p.i., 2-4 h p.i., 8-12 h p.i., 8-48 h p.i., respectively (Granados & Williams, 1987).

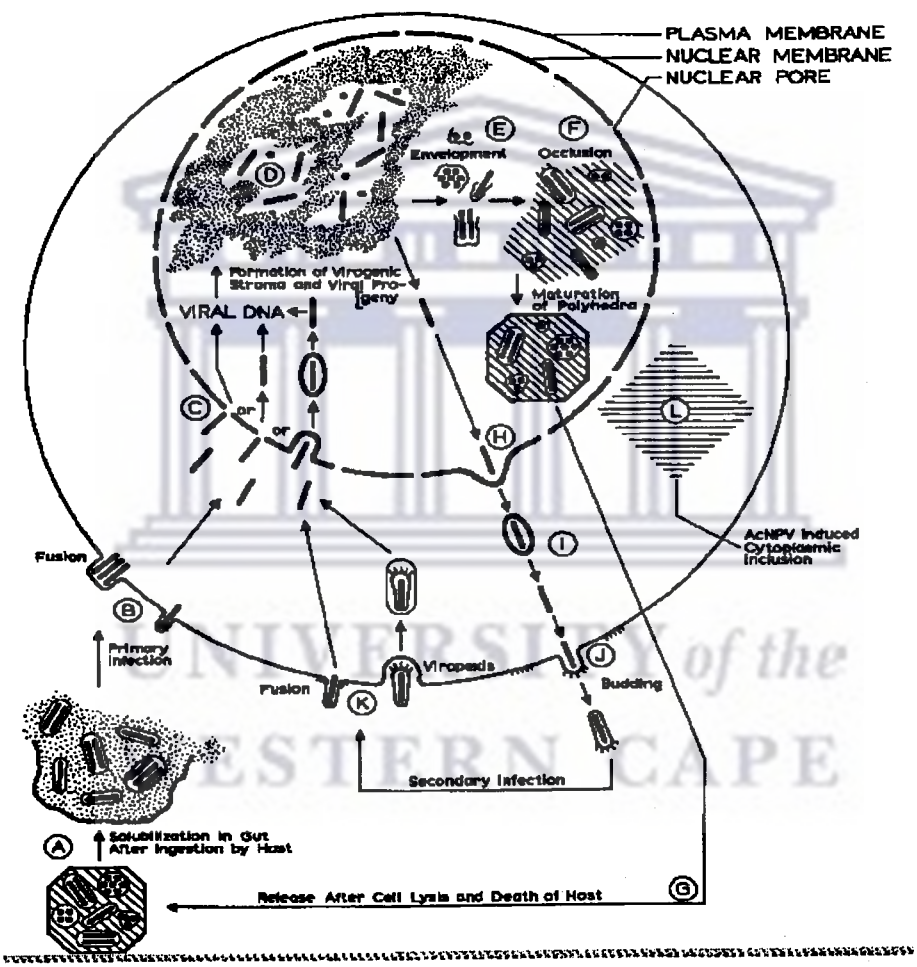


Figure 3. Schematic representation if the baculovirus infection cycle (van der Beek, 1980; van Strien, 1997). A and B. After ingestion the ODV's are dissolved within the midgut and released virions fuse with the plasma membrane. C. Virions enter the nucleus, uncoat and release viral DNA. D. Nucleocapsids are synthesised in the virogenic stroma. H, I, and J. Virions produced within the early stages of infection are released by budding and infect adjacent cells by endocytosis-K. E, F and G. During the later stages of infection, enveloped virions are occluded within the nucleus and released by lysis of the infected cell.

Symptoms:

As is the case with GV infection, typical symptoms of NPV viral infection include lethargy, loss of appetite, negative geotropism, increasing whitish appearance, and liquefaction of internal tissues. After a while complete cessation of feeding occurs, the cuticle darkens, skin becomes fragile and ruptures easily, and cadavers of infected larvae hang up side down attached by the posterior prolegs in an inverted V position known as wilting (Adams & McClintock, 1991).

Genome organisation:

As mentioned earlier, the infectious units of NPV are the nucleocapsids containing the DNA-protein complex. The supercoiled, circular, double-stranded DNA genomes are between 80 and 180 kb. Thus far 12 NPV and 3 GV genomes have been totally sequenced. These include, HaSNPV-G4 (Chen *et al.*, 2001) (NC002654); HzSNPV-ELCAR (Chen *et al.*, 2002) (NC003349); HaSNPV-N (NP 203683) (Zhang and Jin, 2000); *Orgyia pseudotsugata* NPV (OpMNPV) (NC001875) (Ahrens *et al.*, 1997); *Mamestra configurata* NPV (MacoNPV) (NC003529) (Erlandson *et al.*, 2002); *Spodoptera exigua* NPV (SeMNPV) (NC002169) (Ijkel *et al.*, 1999); *Culex nigripalpus* (CuniSNPV) (AF403738) (Afonso *et al.*, 2001); *Epiphyas postvittana* NPV (EppoNPV) (NC003083) (Hyink *et al.*, 2001); *Cydia pomonella* GV (CpGV) (NC002816) (Luque *et al.*, 2001); *Bombyx mori* NPV (BmMNPV) (NC001962) (Park, 2001); *Spodoptera litura* NPV (SlMNPV) (NC003102) (Pang *et al.*, 2001); *Plutella xylostella* GV (PxGV) (NC002593) (Hashimoto *et al.*, 2000); *Lymantria dispar* NPV (LdMNPV) (NC001973) (Kuzio *et al.*, 1999); *Autographa californica* NPV (AcMNPV) (NC001623) (Ayres *et al.*, 1994); and *Xestia c-nigrum* GV (XcGV) (NC002331) (Hayakawa *et al.*, 1999). The sizes of these

genomes range from 101 (PxGV) to 179-kb (XcGV) and contain 120 (PxGV) to 181 (XcGV) open reading frames (ORF) in the different genomes. G+C content usually ranges from 40-57%. Both strands of the genomes contain coding regions, which are functional.

Gene expression:

Baculovirus gene expression is an ordered sequence of events in which each successive phase is dependant on the previous phase (Blissard and Rohrmann, 1990) and allows for the expression of viral genes and DNA replication. The study of baculovirus replication *in vitro* has greatly simplified experiments to understand virus gene expression and replication. Studies have shown that gene expression is divided into two phases: early, late expression (Friesen & Miller, 1986). Early genes may be subdivided into immediate-early and delayed-early, whereas late genes may be distinguished as late and very late genes. Usually the expression level in each succeeding phase is higher than that of the preceding one.

Early gene transcription utilises the host RNA polymerase II complex (Friesen, 1997) and in many cases is initiated with the binding of the polymerase to a TATA box. This transcript is usually initiated within a conserved CAGT motif. Examples of the former stage include genes such as, IE-1 which is one of the first genes transcribed and enhances gene expression of other genes (Guarino & Summers, 1986), IE-N which increases IE-1 expression (Carson *et al.*, 1988), and polyhedrin envelop 38 kDa protein (PE38), a major component of the polyhedrin membrane (Krappa *et al.*, 1992), Examples of the delayed stage include a 39 kD gene which is usually detected in infected cells between 3 and 6 h p.i..

Late genes are those genes that are first transcribed after or concurrently with, the start of viral DNA synthesis and are transcribed within the late baculovirus $^{A}/_{G}$TAAG promoter. In contrast to early genes, late genes utilise an α-amanitin-resistant RNA polymerase of the virus. Examples of this phase of gene expression include: a DNA polymerase gene where activity is enhanced during baculovirus replication and is sensitive to aphidicolin (Tomalski *et al.*, 1988) and the 39- K capsid protein which is a major structural protein (Blissard *et al.*, 1989). The very late gene products include the polyhedrin protein (sizes range from 29-33 kDa), which is a major component of polyhedra, and the p10 protein, which probably has a role in polyhedra formation (Vlak *et al.*, 1988). These genes are hyper –expressed during the infection process. The immediate early transregulators 0, 1, 2 (IE0, IE1 and IE2) and PE38 regulate early gene transcription, while 19 late expression factors (LEF's) are implicated in the transcription of the late genes (Friesen, 1997 and Miller, 1997).

in vitro:

The use of Lepidopteran cell lines for the study of virus-cell interaction allows basic information with regard to the metabolism of insect cells, free of complications from the specialised functions of the whole insect. The possibility of latent virus contamination present in insects is also greatly reduced. Compared with viral preparations from the insect that usually include microbial contaminants, insect debris and require a large number of person hours to prepare, cell culture are operated under sterile conditions; consequently a far cleaner product is obtained. Mass rearing via this technique is also far less labour intensive. There are however a few drawbacks, which include high capital cost, continuos sterile conditions and a relative slow production of growth of the virus. Also, although numerous studies show that NPV produced from cell culture is as virulent and effective as those produced by mass rearing of larva, extended

passage (10-40 passages) through a cell-line will eventually reduce virulence, and mutant polyhedra might form (Chakraborty and Reid, 1999). Research suggests that viral inocula are passed a maximum of 4 times in a cell culture (Hink, 1982).

Phylogeny:

With the aid of molecular phylogeny, the Lepidopteran baculoviruses were separated into NPV's and GV's with the further subdivision of the NPV's into groups I and II. Even though previous baculovirus phylogeny have been based on the OB protein (Zanotto *et al.*, 1993), studies show, that this gene is not the ideal gene for phylogenetic analysis because of its small size and relatively high sequence conservation which provides limited phylogenetic information. Also, phylogenies based on this protein are often in conflict with other gene phylogenies (Clarke *et al.*, 1996 and Kang *et al.*, 1998). Baculovirus phylogeny based on a single gene may not accurately represent baculovirus relatedness. As times passes, the availability of complete genome sequence data allows for the comparison of combination genes, which may reduce sampling errors in phylogenetic analysis. Herniou *et al.* (2001) compiled a comprehensive study where phylogenies were generated based on three independent character sets: the individual sequences of genes shared by viruses; gene order, and gene content. Their study included the entire genome sequence of 9 baculoviruses (AcMNPV, BmMNPV, OpMNPV, HaSNPB-G4, SeMNPV, LdMNPV, CpGV, PxGV and XcGV) whose data was available as that time. With time however, 4 more complete genome data have now become available (HzSNPV, SlMNPV, MacoNPV, EppoNPV, HaSNPV-N and CuniSNPV). As suggested by Herniou *et al.* (2001), baculovirus phylogenies should be based on a combined analysis of all genes conserved among baculoviruses. Combining genes may reduce sampling errors in phylogenetic analysis and allow phylogenies to converge towards correct solutions with good support.

Single Nucleotide polymorphisms

Single nucleotide polymorphisms (SNPs) are common DNA sequence variations among individual isolates and may have great significance. The idea was first related to the human genome, since there are up to 300,000 SNPs distributed evenly throughout the human genome. Analysis of DNA sequence variation is becoming an increasingly important source of information for identifying the genes involved in both disease and in normal biological processes, such as development, aging, and reproduction. There are several types of DNA sequence variation, including insertions and deletions, differences in the copy number of repeated sequences, and single base pair differences. The latter are the most frequent and are termed single nucleotide polymorphisms (SNPs). Since the genomes of three *Helicoverpa* isolates sequenced to date (HaSNPV-G4 (NP 075195) [Chen *et al.*, 1997]; HaSNPV-N (NP 203683) [Zhang and Jin, 2000] (origin –China) and HzSNPV-ELCAR (NP 542753) [Chen *et al.*, 2002] are so similar (> 98%) it is hoped that SNP analyses of the genomes and in particular various functional or structural genes, (including *Helicoverpa armigera*-South Africa) could shed some light as to the variations if any between these isolates.

Functional and structural genes of importance

Ecdysteroid UDP-Glucosyltranferase (e*gt*) gene

O'Reilly and Miller (1989) first identified this gene in AcMNPV. They showed, that the enzyme ecdysteroid UDP-glucosyltranferase (e*gt*) disrupts the ecdysteroid balance of the host larva, causing a delay in the onset of moulting. This gene encodes an enzyme, ecdysteroid UDP-glucosyltransferase (EGT), which catalyses the conjugation of sugars (glucose or galactose) from UDP-sugars to ecdysteroid molting hormones, which then renders the hormone inactive in

infected larvae (O'Reilly, 1995; O'Reilly and Miller, 1989). An active EGT disrupts the hormonal balance of the host larvae, prevents insect larvae from moulting, and ensures the continuation of insect feeding. Deletion of the *egt* gene from the viral genome has been shown to reduce the time required to kill the target pest and also reducers the amount of crop damage as food consumption by the infected larvae is decreased (O'Reilly and Miller, 1991; Slavicek *et al.*, 1999 and Chen *et al.*, 2000). Chen *et al.*, (2000) constructed two recombinant HaSNPV's by deleting the ecdysteroid UDP-glucosyltransferase gene. Bioassay data showed, that the LT_{50} of the *egt*-deletion recombinants was about 27% faster than that of HaSNPV wild type. Thus proving that the deletion of the *egt* gene plays an important role in enhancing baculoviruses in biocontrol strategies. Ten conserved regions are usually found dispersed throughout the *egt* gene. O'Reilly (1995) suggested, that these conserved aa might play a role in the enzymatic activity and UDP-sugar binding of glucosyltransferases. Thus far, the *egt* gene has been reported in 20 NPV's and 7 GV's. However, the genome of *Culex nigripalpus* (Afonso *et al.*, 2001) does not contain an *egt* gene or a homolog thereof.

Baculovirus repeat open reading frames (*bro*) gene

It is believed that genes, which occur commonly within various genomes, are likely to be essential for the virus replication and survival. The occurrence of *bro*-genes is a common feature in baculoviruses and represents a remarkable example of duplication of a single gene, found dispersed throughout the genomes. All these *bro*-genes seem to be related to ORF 2 of the prototype baculovirus *Autographa californica* MNPV (AcMNPV) (Ayres *et al.*, 1994). It is important to note, that not only do the number of *bro*'s present on the genomes differ, AcMNPV-1; *Bombyx mori* NPV (BmNPV)-5 (Gomi *et al.*, 1999); *Orgyia pseudotsugata* MNPV (OpMNPV)-2 (Ahrens *et al.*, 1997); *Lymantria dispar* MNPV (LdMNPV)-16 (Kuzio *et al.*, 1999) and *Xestia c-nigrum* granulovirus (XcGV)-7 (Hashimoto *et al.*, 2000), but their location

and sequence composition varies too. Kuzio *et al.*, (1999) proposed that the *bro*-genes have a related sequence in their N-terminal regions but shared differing degrees of similarity in other regions. They divided the *bro*-genes into four groups based on the relationship of the different domains. Furthermore, Zemskov *et al.*, (2000) showed, that most *bro*-genes had a single-stranded DNA binding motif composed of a pattern of basic and aromatic amino acids with a consensus sequence of $K/RX_{2-5}K/RX_{4-12}F/YX_{2-14}F/YX_{6-13}F/YX_{1-19}K/RX_{3-26}F/Y/W/X_{6-12}K/R$ present in this region of the *bro*-genes. Zemskov *et al.* (2000) showed that 3 of the 5 *Bombyx mori* (Bm) BRO's (BRO-A, BRO-C and BRO-D) have nucleic acid binding activity and are involved in nucleoprotein complexes in the nuclei of infected cells. These three BROs were shown to be involved in nucleosome organization that could block cellular replication and/or transcription and switch host machinery to viral DNA or RNA synthesis. They detected the BmNPV proteins BRO-A, BRO-C and BRO-D in the histone H1 fraction using anti-BRO antibodies. Chromatographic fractions showed that only BRO-A and/or BRO-C interacted with core histones. *Bro*-a of Ha-G4 was also shown to contain a motif [Ha-G4ORF59 (5-60) $K\text{-}X_5$ $K\text{-}X_4$ $Y\text{-}X_3$ $F\text{-}X_{10}$ $F\text{-}X_7$ $R\text{-}X_{16}$ $F\text{-}X_3$ K] found in single strand binding (SSB) proteins from prokaryotic and eukaryotic organisms (Wang and Hall, 1990). This motif is found in the highly conserved N-motif of most of the BRO proteins. This finding suggests that these proteins could be interacting with nucleic acids. Even though all five of the Bm BRO's contained this motif but only three were shown to positively bind proteins. It is possible that different BRO's have different functions since deleting the *bro*-b and –e genes of BmNPV did not alter the infectivity (Kang *et al.*, 1999). Although the genomes of Ha-G4 and Hz-USA are very similar in nucleotide (nt) (97% identity) and amino acid (aa) (99% identity) sequence, a major difference is found in sequence and organisation of the *bro*-genes. *Bro*-a and *bro*-b in Hz-USA have an amino acid identity of 28% and 68% respectively, to Ha-G4 *bro*-a and -b genes. *Bro*-c is more highly conserved between the two genomes with an amino acid identity of 99.4%. Ferber *et al.*, (2000)

compared sequences and localisation of the *bro*-genes of two plaque purified BmNPV strains and showed that the 2 strains displayed similar variations with regard to their *bro*-gene composition, location and number as with the *Helicoverpa* isolates.

chitinase

The *chitinase* gene is expressed in the late stage of virus replication and its product has both endo- and exo-chitinase activity (Hawtin *et al.*, 1995). Of the 16-baculoviral genomes sequenced thus far, the *chitinase* gene only fifteen contains a *chitinase* gene. *Plutella xylostella* GV (PxGV) does not contain a *chitinase* gene. AcMNPV *chiA* is expressed as a late gene and we expect that this HaSNPV-SA homologue is also a late gene. As is the case with *polyhedrin* phylogenetic analysis based on this gene is not ideal, because of its small size and relatively high sequence conservation, which provides limited phylogenetic information. Also, phylogenies based on this protein are often in conflict with other gene phylogenies The conflict in placement could also be due to erroneous phylogenetic inferences caused by unequal rates of evolution or the lack of an unambiguous phylogenetic signal in the sequences. Furthermore, the baculoviral *chitinase* gene has a remarkable homology with that of the *Serratia marcescens* chitinase A (Hawtin et al., 1995). If the *chitinase* gene did originate from *S. marcescens*, this could possibly explain the anomalous behaviour of this gene.

helicase

Helicase genes, which have been found to play a role in replication or expression, have been found in all baculovirus genomes sequenced to date. Studies of putative baculoviral *helicases*, suggest that these enzymes are one of the key determinants of host range (Heldens *et al.*, 1997). For example, changes in a few amino acids or even in only one, in the AcMNPV *helicase* (*p143*) enable AcMNPV to replicate in cells and larvae of BmMNPV, a host which does not support

AcMNPV replication normally (Maeda *et al.*, 1993). Analysis of *helicase* gene of AcMNPV revealed a potential leucine zipper motif, a putative nuclear localization signal, and seven amino acid motifs (I, Ia, II-VI) previously identified in a number of proteins involved in NTP binding and DNA/RNA unwinding (Lu and Carstens, 1991).

lef-8

The *lef-8* gene has been targeted as a possible candidate for a viral gene involved in late expression and has been found in all baculovirus genomes sequenced to date. This gene has been shown to be necessary for the efficient expression of late and very late gene promoters in transient gene expression assays (Passarelli *et al.*, 1994). This gene sequence is one of a few genes that yielded the best trees within the comprehensive phylogenetic study compiled by Herniou *et al.*, (2001).

Scope of thesis:

When this study was undertaken, little was known with regard to the molecular characteristics that dictate the biological properties of *Helicoverpa armigera* single-enveloped nucleopolyhedrovirus (HaSNPV). For this reason, this study was initiated to broaden the molecular knowledge (which include, identification of essential genes, *in vitro* analysis, gene expression and replication etc) of HaSNPV, with specific reference to a South African isolate (HaSNPV-SA). The major aim of the present study, was to investigate and identify essential genes, found throughout previously characterised baculovirus genomes, in our HaSNPV isolate. Furthermore, transcriptional analyses of relevant genes as well as phylogenetic studies were done.

A granulovirus (GV) and a single envelope nucleopolyhedrovirus (SNPV) were found infecting a colony of *Helicoverpa armigera* in the Western Cape region of South Africa. Since studies have shown HaSNPV to be specific and highly virulent to its host (Hughes *et al.*, 1983; Teackle *et al.*, 1985), the formation of *in vitro* production systems of HaSNPV-SA was important. For further analysis, it was decided to concentrate all studies on the genus. Numerous homologous *in vivo* infection of HaSNPV-SA consistently resulted in a viral inoculum containing HaGV. We were never able to eradicate the HaGV and obtain a pure HaSNPV-SA inoculum. We also at various times had problems with latent infection. Furthermore, obtaining the larvae was problematic. To obtain a HaSNPV virus stock and to eradicate the possibility of latent virus contamination, the *Heliothis* cell line HzAM1 was infected with HaSNPV-SA (**Chapter 2**). To study the HaSNPV-SA virus cell interactions, it was imperative that we infect and successfully propagate our virus *in vitro* using permissive cell lines

As our study progressed, two *Helicoverpa* SNPV genomes were completely sequenced (Chen *et al.*, 2001 and 2002). Although a hiccup, it provided a platform for our research and allowed us the opportunity to compare the essential genes we were able to identify as well as characterise. One such gene, was the ecdysteroid UDP-glucosyltranferase (*egt*) gene that was located using an Hz-SNPV gene-specific probe (**Chapter 3**). The ecdysteroid UDP-glucosyltranferase gene (*egt*) has been found in several baculoviruses and encodes an enzyme, ecdysteroid UDP-glucosyltransferase (EGT), which catalyses the conjugation of sugars (glucose or galactose) from UDP-sugars to ecdysteroid molting hormones. Numerous studies show, that *egt* deletion mutants are enhanced viral insecticides. Four other *egt* genes of *Helicoverpa* (HaSNPV-G4 (NP 075195) [Chen *et al.*, 1997]; HaSNPV-N (NP 203683) [Zhang and Jin, 2000] (origin –China); HzSNPV-ELCAR (NP 542753) [Chen *et al.*, 2002] and HzSNPV-Popham (AAB 81201) [Popham *et al.*,

1997]) are present in GenBank. Single nucleotide polymorphisms (SNPs) showing DNA sequence variations among the individual isolates were tabulated.

A comparison of the genomes of *Helicoverpa armigera* (Ha-G4) and *Helicoverpa zea* (strain Elcar, Hz-USA) nucleopolyhedroviruses showed that the most divergent region among these isolates contained two of the three baculovirus-related open reading frames (*bro* genes) named *bro*-a and -b respectively. Although the genomes of Ha-G4 and Hz-USA are very similar in nucleotide (nt) (97% identity) and amino acid (aa) (99% identity) sequence, a major difference is found in sequence and organisation of their three *bro*-genes. Since sequence data strongly suggest that these two strains are variants of the same virus species, the low homology between *bro*-a and *bro*-b was interesting to note (Chen *et al.*, 2002). To study the evolutionary relatedness between *bro*-a and *bro*-b present in *Helicoverpa* SNPV isolates, our South African isolate HaSNPV-SA was included in this study (**Chapter 4**). Studies have shown, that some BRO's have nucleic acid binding activity and are involved in nucleoprotein complexes in the nuclei of infected cells. These BROs were shown to be involved in nucleosome organization that could block cellular replication and/or transcription and switch host machinery to viral DNA or RNA synthesis. For this reason, all three *bro*- genes of HaSNPV were analysed at the transcriptional level.

Baculovirus phylogeny based on a single gene may not accurately represent baculovirus relatedness. It is therefore of interest to check if various genes could produce trees, which are similar or different. Three important genes late expression factor (*lef*) 8, *chitinase* and *helicase* were identified during random sequence analysis of two clones (*Eco*RI-H and *Eco*RI-J), which formed part of the *Eco*RI plasmid library of the viral genome (**Chapter 5**). Single nucleotide polymorphisms (SNPs) showing DNA sequence variations among the individual isolates for

each gene of the *Helicoverpa* SNPV's were tabulated. The study supports the theory that there is danger in using phylogeny based on one or a small number of genes to infer evolution.

CHAPTER TWO

IN VITRO INFECTION OF *HELIOTHIS ZEA* CELL LINES WITH *HELICOVERPA ARMIGERA* SINGLE ENVELOPE NUCLEOPOLYHEDROVIRUS.

Abstract

A granulovirus (GV) and a single envelope nucleopolyhedrovirus (SNPV) were found infecting a colony of *Helicoverpa armigera* in the Western Cape region of South Africa. To obtain a HaSNPV virus stock and to eradicate the possibility of latent virus contamination, the *Heliothis* cell line HzAM1 was infected with HaSNPV-SA. Polyhedra were first detected 2-3 d p.i. (post infection) Further purification of the HaSNPV stock by plaque purification was completed to obtain various phenotypes of our isolate, and to select and further analyse one isolate. HaSNPV infection at various morphological stages was observed. Plaques with the morphologically distinct ring zone could be seen as well as cells completely filled with mature polyhedra. Various plaques were chosen and analysed by SDS-PAGE. Once a distinct isolate was chosen (HaSNPV-PL3), it was further assayed until it was certain the virus was pure. HaSNPV-PL3 was amplified in HzAM1 cells and titred by end point dilution assays. Titres in the range of 10^6 pfu/ml were frequently obtained. Isolation of cellular (viral) DNA was extracted for further molecular analysis and digested with the restriction enzyme *Eco*RV. The replication of HaSNPV-PL3 in two *Heliothis* cell lines, HzAM1 and Hz2E5 was also studied. HaSNPV-PL3 seems to replicate more rapidly in Hz2E5 rather than HzAM1, although high polyhedra yields were obtained for both cell lines at 4 d p.i. with an multiplicity of infection (MOI) of 1.

Introduction:

Helicoverpa spp. is a major focus of pest control programmes world wide because theses pests attack more than 60 crops (Christian, 1994). Since chemical insecticides have failed to control these pests in some areas, alternative control measures are sought. Baculoviruses are large, circular, double-stranded viruses occupying mainly in insects and are very effective as pest control agents. They contain two genera, the *Nucleopolyhedroviruses* (NPV) and the *Granuloviruses* (GV) (Blissard *et al.*, 2000). The family is characterised by the occlusion of virions into large proteinaceous capsules or occlusion bodies (OB's).

A South African isolate of *Helicoverpa armigera* single envelope nucleopolyhedrovirus (HaSNPV-SA), a wild-type (wt) baculovirus, and a granulovirus (HaGV) was isolated. Since studies have shown HaSNPV to be specific and highly virulent to its host (Hughes *et al.*, 1983; Teackle *et al.*, 1985), the formation of *in vitro* production systems of HaSNPV-SA was important.

Although *in vivo* replication of NPV's allow for a more virulent virus stock, preparations from the insect usually include microbial contaminants, insect debris and latent virus contamination. Since cell culture is operated under sterile conditions, a far cleaner product is obtained. *In vivo* isolation is also rather labour intensive and with *in vitro* isolation mass rearing of a virus is possible with fewer person hours. There are however a few drawbacks, which included high capitol cost and continuous sterile conditions. Also, after repeated passage in cell culture, the virus will eventually become less virulent, and baculovirus may suffer insertions of host cell

DNA and transposable elements with its genome (Blissard and Rohrmann, 1990). A consequence of one such insertion within a gene encoding a 25kDa protein is the production of viruses, which yield few polyhedra (FP phenotype) (Beames and Summers, 1988; 1989; 1990). Research suggest, that viral inoculum should be passaged a maximum of 4 times in a cell culture system, by which time, the inoculum should be passed through an insect once more (Chakraborty and Reid, 1999).

Homologous *in vivo* infection of HaSNPV-SA consistently resulted in a viral inoculum containing HaGV (Fig. 1). We were never able to eradicate the HaGV and obtain a pure HaSNPV-SA inoculum. We also at various times had problems with latent infection. Furthermore, obtaining the larvae was problematic. To study the HaSNPV-SA virus cell interactions, it was imperative that we infect and successfully propagate our virus *in vitro* using permissive cell lines. Our virus was shown to successfully replicate within the *Heliothis* cell lines HzAM1 (McIntosh and Ignoffo, 1983) and Hz2E5. After plaque purification of the virus HaSNPV-PL3 was chosen for future studies.

Materials and Methods:

Viral origin

A single-enveloped nucleopolyhedrovirus was used in this study.

Maintenance of cells

HzAM1 and Hz2E5 cells were maintained at 27°C in Grace's supplemented medium with 10% foetal bovine serum, and passaged when 80% of a confluent layer was

observed (split 1:10), with one medium change after 2 days. A maintenance culture was also stored at 18°C.

Hemolymph extraction and BV purification

Fifty HaSNPV-SA infected 4th instar *H. armigera* larvae were bled by amputation of the second proleg on day 3 post infection (p. i.). The hemolymph was collected in 5ml of medium in the presence of phenylthiocarbamide to inhibit prophenoloxidase activity. End-point dilution assays were used to titre the virus (1 x 10^8 median infection dose [$TCID_{50}$] units/ml). The diluted hemolymph was filtered through a 0.45 uM filter and stored in small aliquots at -70°C.. Repeated thawing was found to reduce the titre of samples.

Infection of cells

Medium was removed from HzAM1 cells in the exponential growth phase (1 x 10^6 cells/ml). Virus inoculum (Multiplicity of infection [MOI] of 1 $TCID_{50}$ units/cell) (100ul of diluted hemolymph plus 900ul medium for a 25cm^2 flask) was added and the cells were incubated for 1h at 27°C with rocking. The inocula were removed and the cells were washed once with fresh medium. Four millilitres fresh medium was then added and the incubation continued at 27°C.

Plaque assay of HaSNPV-SA

Infected HzAM1 cells were harvested in the growing medium. Thirty-five millilitre dishes were seeded with 1 x 10^6 cells/ml in Grace's supplemented medium + 10% fetal bovine serum (FBS; Gibco/BRL) (henceforth, referred to as complete medium) and incubated overnight at 27°C. The supernatant of the infected cells were diluted to

concentrations of 10^{-3}, 10^{-4} and 10^{-5} in 500ul complete medium. Medium was removed from the cells and the 500ul viral suspension was transferred to the 35 mm dishes and incubated for 90 min at 27°C (the dishes were gently rocked every 15min). After the incubation period the virus suspension was removed from the cells (1 x 10^5 cells/ml), and the cells were washed carefully with 500ul complete medium. All medium was removed carefully after the last wash step. Two millilitres of a Grace's/agarose mixture (for each dish 1ml 2 x Grace's + 20% FBS with 1 ml hot sterilised 3% Sea plaque agarose was mixed) was overlaid on the cells without loosening them. The dishes were then incubated for 4 days at 27°C in a moisten box.

Plaque collection

A sterile Pasteur pipette was used and suction applied on top of the agarose plaque. The agarose block was placed in 500ul complete medium, which was then vortexed and incubated at 4°C to allow the budded virus (BV) to diffuse from the agarose into the medium.

SDS-PAGE analysis and BV DNA extraction

Once the BV was allowed to diffuse into the medium, HzAM1 cells were infected with different plaques (as previously described). Infected cells were sloughed into the growing medium and transferred to a sterile greiner 10ml tube. Tubes were then centrifuged at 2000 rpm for 10min.

a) DNA extraction from supernatant

The supernatant (containing BV) was removed and filtered with a 0.45uM filter. The suspension was then placed in a 36ml polycarbonate tube and placed in an ultracentrifuge at 20000 rpm at 8°C using the SW28 swing rotor for 1h. The

supernatant was then discarded and the BV pellet was resuspended in 100ul 0.1 x TE (Tris/EDTA; 10mM Tris, pH 7.5; 1 mM EDTA) and placed at 4°C overnight. The resuspended pellet was transferred to a 1.5ml Eppendorf tube and proteinase K was added to a final concentration of 0.1mg/ml and incubated at 45°C for 2.5h. Sodium dodecyl sulphate (SDS) was added to a final concentration of 1% and this suspension was incubated for a further 30min. The mixture was extracted twice with phenol:chloroform:isoamylalcohol (PCI; 25:24:1) and after the final extraction the top aqueous layer was transferred to a clean Eppendorf. DNA was allowed to precipitate in the presence of 0.1V 3M NaAc (pH 5.2) and 2.5V 100% ethanol (ETOH) for 1h at -20°C. After the precipitated pellet was washed in 70% ETOH, an air-dried pellet was resuspended in 100ul 1 x TE and placed in a 37°C water-bath overnight. The resuspended genomic DNA was stored at 4°C until further use.

b) SDS-PAGE analysis of pellet

The pellet of mock-infected HzAM1 cells, HaSNPV-G4 wt (Chen *et al.*, 2001), HaSNPV-SA wt (Khan *et al.*, 2000) and 6 collected plaques (designated HaSNPV-PL1 – HaSNPV-PL6) were resuspended in 5ml phosphate buffered saline (PBS) and centrifuged at 1000rpm for 10min. The supernatant was discarded and the pellet was resuspended in 100ul PBS + 40ul sample lysis buffer + 10ul ß-mercaptoethanol. The mixture was boiled at 95°C for 10min and samples were electrophoresed on a 7.5% SDS-PAGE using a Bio-Rad mini protein 11 apparatus according to Laemmli (1970) and stained with Coomassie brilliant blue.

c) Restriction endonuclease analysis of HaSNPV-PL3

By comparing protein profiles we choose a plaque (HaSNPV-PL3) with dominant protein bands that we assumed were polyhedron and p10. Once DNA was extracted this DNA was digested with the restriction enzyme *Eco*RV (GIBCO/BRL). The

fragments were separated by electrophoresis in a 0.6% agarose gel at 45V (1.5V/am) for 16h. Lambda DNA digested with *Bam*HI-*Eco*RI-*Hin*dIII was used as the molecular size standards.

Determination of the $TCID_{50}$ titres of HaSNPV-PL3

Sixty well plates were seeded with 1 x 10^6 cells/ml in fresh complete medium. After cell attachment, the medium was replaced with 100ul inoculum (serial dilutions in complete medium) and plates were incubated at 27°C in a moistened container. On day 4 up till day 10 the wells were scored for the presence of polyhedra. The median infection dose ($TCID_{50}$) was determined as proposed by Reed and Muench (1938).

Infection of the *Heliothis* cell line HZ2E5 and HZAM1 with HaSNPV-PL3

Medium was removed with a sterile Pasteur pipette from 25cm^2 flasks containing HzAM1 and Hz2E5 cells in the exponential growth phase (1 x 10^6 cells/ml). Virus inoculum was added at MOI of 1 and the cells were incubated for 1h at 27°C with gentle rocking. The inoculum was removed and the cells were washed once with fresh medium. Four millilitres of fresh medium was then added and the incubation continued at 27°C.

Results:

Infected cells

Upon infection of the 1 x 10^5 cells/ml of HzAM1 cells with the wt HaSNPV-SA (MOI of 1 $TCID_{50}$ units/cell), polyhedra were first detected 2-3 d p.i. (Fig. 2a and 2b).

Morphology of plaques

At 6 d p.i., a lawn of viral plaques which were well separated had formed on the 35 mm dish with a diluted virus stock of 10^{-4} (Fig. 3a). Many plaques had varying and distinct features. Some contained as few as 22 infected cells, while others had too many infected cells to count. Some plaques were clustered and formed clumps of cells, while others had cells close together but separated. When using higher magnification, it was possible to distinguish between two distinct morphological stages of infection. As seen in Fig. 3b, the morphologically distinct peristomal compartment called the ring zone is clearly shown. Enlargement of the cell nucleus reduced the "cytoplasm space" between the plasma membrane and the nuclear membrane resulting in a wider ring zone. Early virion occlusion occurs within this zone. Polyhedrin proteins are deposited around the enveloped virions, thus the virions are occluded progressively. Plaques were also observed which contained a nucleus filled with mature polyhedra (Fig. 3c). This morphological stage is normally observed by 96 h p.i.

Protein and DNA analysis

Upon collection two of each variant – plaques containing many polyhedra (Fig 4. lane 4 and 5), plaques, which were clumped (Fig 4. lane 6 and 7), and plaques with scattered infected cells (Fig 4. lane 8 and 9) - were obtained. Lysed proteins of mock-infected HzAM1 cells, HaSNPV-G4 wt (Chen *et al.*, 2001), HaSNPV-SA wt (Khan *et al.*, 2000) and 6 collected plaques (designated HaSNPV-PL1 – HaSNPV-PL6) were analysed by SDS-PAGE (Fig. 4). Of the plaques collected, HaSNPV-PL3 had two clear distinct protein bands, which were correctly positioned to be polyhedron and P10. For further analysis, we decided to use this plaque. All other plaques

irrespective of their morphology yielded similar profiles. DNA extracted from HaSNPV-PL3 and digested with the restriction enzyme *Eco*RV was analysed electrophoretically on a 0.6% agarose gel (Fig. 5b). A comparison of this plaque with the wt isolate showed that they were very similar (Fig. 5a).

Infection of two permissible cells lines with HaSNPV-PL3

$TCID_{50}$ values for BV extracted from HzAM1 infected cells were = 10^5 /ml. Both 1 x 10^5 cells/ml of Hz2E5 (Fig. 6a) and HzAM1 (Fig. 6b) were infected at a MOI of 1 with HaSNPV-PL3. Although high yields of infected cells were observed for both cell lines, HaSNPV-PL3 appeared to replicate more rapidly in Hz2E5. Both micrographs were viewed at 4 d p.i.

Discussion and Conclusion:

This is the first report of a South African SNPV propagated *in vitro*. Although many SNPV's and more recently GV's have been successfully propagated within cell culture, applied baculovirus research in South Africa has been in the pioneering stages with its launch only 6-7 years earlier. The study of baculovirus replication *in vitro* has greatly simplified experiments to understand the kinetics of virus gene expression and replication. It was thus imperative that we infect permissive cell lines and that our virus could successfully propagate within cell culture.

With *in vitro* replication, we successfully eradicated the latent infection or contamination with HaGV (Fig. 1) and obtained a pure HaSNPV-SA virus stock (Fig. 2a and 2b). At present there are only a few cases of successful *in vitro* replication of

GV's. The resulting cell-lines however have low susceptibility and/or were unstable, losing their virus-permissive character after serial passage. The properties of the cell line populations are most important for the in vitro replication of baculoviruses and are the cause for limited success with GV's. It is thus highly unlikely that we were able to successfully propagate HaGV by default. Plaque purification provided the opportunity to attain a stock starting from a single infectious unit (Fig. 3a). As seen in Fig 3b and 3c, two distinct morphological stages of *in vitro* replication were observed.

The morphogenesis of HaSNPV is generally similar to previous reports on multiple enveloped nucleopolyhedroviruses (MNPV) such as *Autographa califonica* (Ac)MNPV (Williams and Faulkner, 1997), *Trichoplusia ni* (Tn)MNPV (Mackinnon *et al.*, 1974) and *Spodoptera frugiperda* (Sf)MNPV (Knudson and Harrap, 1976). However, Lua and Reid (2000) noted two significant differences of HaSNPV morphogenesis *in vitro* compared to other baculoviruses. One, early polyhedron envelope (PE) morphogenesis in HaSNPV seems to be marked by the appearance of calyx precursor structures in the virogenic stroma. In HaSNPV, these structures appear to emerge and develop randomly in the nuclei, especially in the virogenic stroma, and to have no definite form. Secondly, fibrillar structures enriched in P10 proteins are commonly found in the nuclei and cytoplasm of AcMNPV-infected cells (Lee *et al.*, 1996; Rohrmann, 1992; and Williams and Faulkner, 1997). This prominent fibrillar structure was not observed in the nuclei of HaSNPV-infected cells. Instead, fibrillar structures (presumptive homologues of AcMNPV fibrillar structures), resembling to some extent the AcMNPV fibrillar structures, were clearly detected only in the cytoplasm of HaSNPV-infected cells (Lua and Reid, 2000). It may be possible that P10 proteins of SNPV are located differently to those seen in

MNPV, such that in SNPV fibrillar structures may be exclusively located in the cytoplasm. As Lua and Reid (2000) suggested, P10 proteins could be present in the nuclei of HaSNPV-infected cells but are possibly undetected, as they do not form fibrillar structures in the nucleus. Since we did not carry out extensive electron microscope studies of our viral morphogenesis upon infection, we cannot confirm or deny these reports.

Serial passaging of HaSNPV in *H. zea* cells has led to lower yields of OB's, changes in plaque morphology and lower potency of OB's produced as a result of infection (Chakraborty and Reid, 1999). For this reason, when SDS-PAGE analysis was completed on six plaques collected, a genotype with a distinct polyhedron band was chosen (Fig. 4). Further DNA analysis of this isolate (HaSNPV-PL3) (Fig. 5a) showed that it had a similar almost identical profile to the wt HaSNPV-SA (Fig. 5b).

Infection of the two *H. zea* cell lines HzAM1 (Fig. 6a) and Hz2E5 (Fig. 6b) showed that both cell lines produced high yields of HaSNPV-PL3. This is not surprising, as viruses are generally able to multiply fast and produce large batches within cell lines, which originate from its primary host. The virus however did appear to replicate more rapidly within Hz2E5. Some viruses are rarely specific to single species, although most viruses generally do not replicate in non-permissive cell lines.

The feasibility of producing viruses in cell culture has far surpassed any expectations. Most technical problems have been overcome, and the future and viability of large-scale production using cell culture has vastly improved.

Acknowledgements:

Magda Usmany, Xinwen Chen and Marcel Westenberg are thanked for supplying HaSNPV-G4 BV and for scientific input. The research was funded in part by the National Research Foundation and the Royal Netherlands Academy of Sciences.

Figure 1. Transmission electron micrograph at 26 000 x magnification of a single enveloped nucleopolyhedrovirus (SNPV) and granulovirus (GV) isolated from *Helicoverpa armigera*.

A

B

Figure 2. Micrographs of passage 1 infected HzAM1 cells with HaSNPV-SA. **A.** Magnification of x 10. **B.** Magnification of x 40. **IC,** infected cells.

Figure 3. Micrographs of plaques formed by HzAMI cells infected with HaSNPV-SA. **A**. Magnification of x 10 depicting a number of plaques. **B**. Plaque clearly depicting the morphological "ring zone" stage (x 40 magnification). **C**. Plaque of a HzAMI cells filled with mature polyhedra (x 40 magnification). **RZ**, ring zone.

Figure 4. SDS-PAGE analysis of plaques collected at 10^{-4} dilution. Uninfected HzAMI cells (lane 1), HzAMI cells infected with wt HaSNPV-G4 (lane 2), HzAMI cells infected with wt HaSNPV-SA (lane 3), plaques collected; HaSNPV-PL1 (lane 4), HaSNPV-PL2 (lane 5), HaSNPV-PL3 (lane 6), HaSNPV-PL4 (lane 7), HaSNPV-PL5 (lane 8), HaSNPV-PL6 (lane 9). **LMW**, low molecular weight marker; **PH**, polyhedrin.

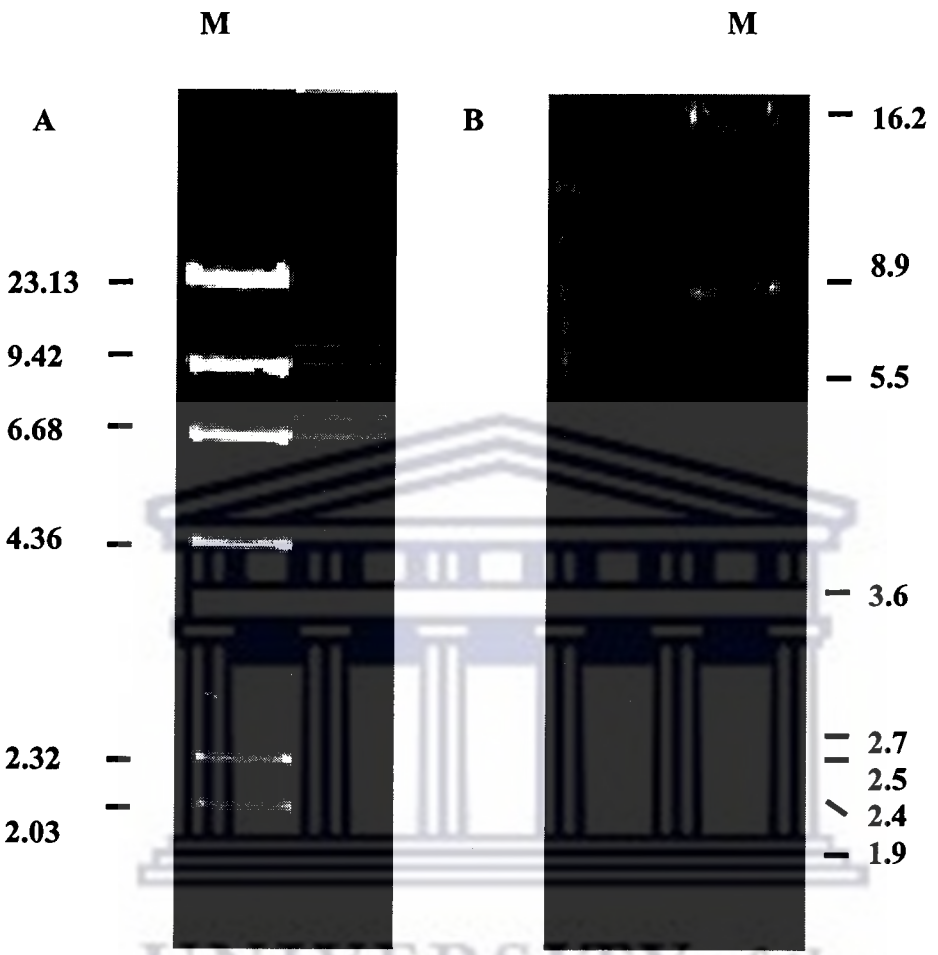


Figure 5. Genomic profiles. A. Lane 2. HaSNPV-SA digested with *Eco*RV and electrophoretically separated in 0.8% agarose (Lambda DNA digested with *Hin*dIII was used as molecular size markers with the size indicated in kilobase pairs). B. Lane 1. HaSNPV-PL3 DNA digested with *Eco*RV and electrophoretically separated in 0.6% agarose. Lambda DNA digested with *Bam*HI-*Eco*RI-*Hin*dIII was used as molecular size markers with the size indicated in kilobase pairs. M, ? marker.

A

B

Figure 6. Micrograph of cells infected with HaSNPV-PL3 with a moi = 1. **A**. Hz2E5 infected cells at 4 d p.i. (x 10 magnification) **B**. HzAMI infected cells at 4 d p.i. (x 40 magnification).

CHAPTER THREE

FUNCTIONAL CHARACTERISATION OF THE ECDYSTEROID UDP-GLUCOSYL TRANSFERASE GENE OF *HELICOVERPA ARMIGERA* SINGLE-ENVELOPED NUCLEOPOLYHEDROVIRUS ISOLATED IN SOUTH AFRICA.

Abstract:

The ecdysteroid UDP-glucosyltranferase (*egt*) gene of a single enveloped nucleopolyhedrovirus was located using an Hz-SNPV gene-specific probe. This SNPV was found infecting a colony of *Helicoverpa armigera* (HaSNPV) in the Western Cape region of South Africa. Studies have shown that the enzyme disrupts the ecdysteroid balance of the host larva, causing a delay in the onset of moulting. The open reading frame of the HaSNPV-SA *egt* is 1548 nucleotides long and encodes a predicted protein of 516 amino acids with a M_r of 58,897-Da. The 5'-noncoding region contained an early transcription initiation motif (CAGT) and a baculovirus late transcription motif (ATAAG). A transcription enhancer sequence (GATA) was also identified. Two possible TATA boxes together with am AT rich region were also recognized. A putative signal peptide of 20 residues was present at the N-terminus of the predicted EGT sequence. A polyadenylation signal (AATAAA) was found downstream of the translation stop codon. Five *Helicoverpa* NPV EGT's that have an extremely high degree of nucleotide and amino acid sequence homology were used in this study. Single nucleotide polymorphisms (SNPs) within the gene were tabulated. The *Helicoverpa* NPV *egt*'s seem to be closely related to the *egt* genes of *Mamestra configurata* NPV (MacoNPV), *Buzura suppressaria* NPV (BusuSNPV) and *Spodoptera exigua* NPV (SeMNPV) with amino acid identities of approximately 50%. The *Helicoverpa* NPV EGT's show ten conserved motifs with other EGT's. A phylogenetic tree of twenty-seven baculovirus EGT's and a human UDP-glucoronosyltransferase was constructed using Neighbour-joining within CLUSTAL X. That a secreted and active EGT is encoded by HaSNPV-SA was confirmed by assay of infected cell culture medium.

Introduction:

Baculoviruses are invertebrate-specific pathogens that have been described in more than 800 species of insects. The *Baculoviridae* family is taxonomically divided into the *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) genera (Blissard *et al.*, 2000). The proliferating interest in baculovirus research has occurred due to the realization that these relatively virulent insect pathogens have great potential as biological control agents. Baculoviruses are also important as expression vectors of foreign genes (Luckow *et al.*, 1993) and as gene delivery vectors (Kost and Condreay, 2002).

The ecdysteroid UDP-glucosyltranferase gene (*egt*) has been found in several baculoviruses. This gene encodes an enzyme, ecdysteroid UDP-glucosyltransferase (EGT), which catalyses the conjugation of sugars (glucose or galactose) from UDP-sugars to ecdysteroid molting hormones, which then renders the hormone inactive in infected larvae (O'Reilly, 1995; O'Reilly and Miller, 1989). An active EGT disrupts the hormonal balance of the host larvae, prevents insect larvae from moulting, and ensures the continuation of insect feeding. As the infected insect continues to grow and feed, high amounts of occlusion bodies (polyhedra) are produced. Deletion of the *egt* gene from the viral genome has been shown to reduce the time required to kill the target pest and also reducers the amount of crop damage as food consumption by the infected larvae is decreased (O'Reilly and Miller, 1991; Slavicek *et al.*, 1999 and Chen *et al.*, 2000 (a)). In essence, *egt* deletion mutants are enhanced viral insecticides.

Helicoverpa spp. is a major agricultural pest and is widely distributed in Australia, Southeast and southern Asia, the Middle East, Southern Europe and Africa. To study the evolutionary relatedness of *egt* genes present in *Helicoverpa* SNPV isolates, a South African isolate of *H. armigera* SNPV

(Ha-SA) was studied. We also searched for an *egt* homologue in HaSNPV-SA since this gene has been shown to play an important role in the regulation of host development. Thus far, the *egt* gene has been reported in 20 NPV's and 7 GV's. However, the genomes of *Culex nigripalpus* (CuniSNPV) (Afonso *et al.*, 2001) and *Xestia c-nigrum* GV (XcGV) (Hayakawa *et al.*, 1999). do not contain *egt* gene or a homologue thereof. Four other *egt* genes of *Helicoverpa* (HaSNPV-G4 (NP 075195) [Chen *et al.*, 1997]; HaSNPV-N (NP 203683) [Zhang and Jin, 2000] (origin –China); HzSNPV-ELCAR (NP 542753) [Chen *et al.*, 2002] and HzSNPV-Popham (AAB 81201) [Popham *et al.*, 1997]) are present in GenBank. In this study, we report the identification and sequencing of the *egt* gene from HaSNPV-SA. Single nucleotide polymorphisms (SNPs) showing DNA sequence variations among the individual isolates were tabulated. The presence of the *egt* gene in 27 isolates in GenBank supports the view that most baculoviruses may carry an *egt* gene. Although the gene of *Helicoverpa armigera* has been previously reported, this study includes the phylogenetic analysis of all 27 *egt*'s present in GenBank as well as the confirmation that our virus encodes an active EGT by assay of infected cell culture supernatant. Phylogenetic analysis of *egt* supported the baculovirus evolution scheme suggested by Herniou *et al.* (2001).

Materials And Methods:

Virus and DNA isolation

The HaSNPV-SA was isolated from a colony of diseased *H. armigera* larvae in the Western Cape region of South Africa (Khan *et al.*, 2000). The occlusion bodies were purified by continuous and discontinuous sucrose density gradient centrifugation. Viral DNA was extracted using the method prescribed by Crook *et al.*, (1985). Viral DNA was digested with various restriction enzymes according to the specifications of the manufacturer.

Cloning of DNA fragments

HaSNPV-SA genomic DNA digested with *Eco*RI was shotgun-cloned into the plasmid vector pBluescript SK^+ by using standard methods (Sambrook *et al.*, 1989).

Southern Hybridisation

Electrophoretically separated restriction fragments of the HaSNPV-SA *Eco*RI library were transferred using the capillary transfer of DNA to Hybond N^+ nylon membrane (Amersham) under alkaline conditions (Southern, 1975). The blot was hybridised with a DIG-Nick translated fragment of the HzSNPV-ZIM *egt* gene (clone *Eco*RV-I9) ("C. Jacobs, personal communication") at 55°C using standard procedures as specified by Von Miltenberg *et al.*, (1995).

Sequencing

The termini of the positive *Eco*RI clone (*Eco*RI-K), was sequenced by automated sequencing using the pUC/M13 universal forward and reverse primers. Subclones and deletions of *Eco*RI-K were constructed using the restriction enzymes *Sal*I, *Eco*RV, *Hin*dIII, *Bst*XI and *Sac*I. These subclones and deletions were further sequenced using the universal primers and a specific oligonucleotide primer. All sequencing was done with a Pharmacia ALF/Express automated sequencer. The sequences were screened using the database (GenBank) of the National Centre of BioInformatics advanced BLAST program (Altshul *et al.*, 1997) to search for sequence homology. All cloning procedures were as described in Sambrook *et al.*, (1989). The nucleotide sequence of the complete open reading frame has been deposited in GenBank and was assigned the following accession number: AF503939.

Phylogenetic analysis

The baculovirus EGT's: HaSNPV-SA; HaSNPV-G4; HaSNPV-N; HzSNPV-ELCAR; HzSNPV-Popham; MacoNPV NPV (NP613122) (Erlandson *et al.*, 2002); BusuSNPV NPV (AAB58353) (Hu *et al.*, 1997); SeMNPV NPV (NP037787) (Ijkel *et al.*, 1999); *Epiphyas postvittana* NPV (EppoNPV) (AAK85576) (Hyink *et al.*, 2001); *Cydia pomonella* GV (CpGV) (NP148925) (Luque *et al.*, 2001); *Adoxophyes honmai* GV (AhGV) (BAB69953) (Goto and Nakai, 2001); *Bombyx mori* NPV (BmMNPV) (AAL13149) (Park, 2001); *Spodoptera litura* NPV (SlMNPV) (NP258389) (Pang *et al.*, 2001); *Adoxophyes orana* GV (AoGV) (AAL02080) (Wormleaton and Winstanley, 2001); *Plutella xylostella* GV (PxGV) (NP068337) (Hashimoto *et al.*, 2000); *Lymantria dispar* NPV (LdMNPV) (NP047762) (Kuzio *et al.*, 1999); *Epinotia aporema* GV (EaGV) (AAK82354) (Manzan *et al.*, 2001); *Autographa californica* NPV (AcMNPV) (NP05044) (Ayres *et al.*, 1994); *Anticarsia gemmatalis* NPV (AgNPV) (AAK16408) (Rodrigues *et al.*, 2001); *Choristoneura fumiferana* GV (CfGV) (AAG50437) (Bah *et al.*, 1998); *Amsacta albistriga* NPV (AaNPV) (AAF15297) (Premkumar and Mathaven, 1999); *Ecotropis oblique* NPV (EoNPV) (AAD22774) (Zhang *et al.*, 1998); *Choristoneura fumiferana* NPV (CfMNPV) (AAC54601) (Barrett *et al.*, 1995); *Lacanobia oleracea* GV (LoGV) (CAA69602) (Smith and Goodale, 1998); *Choristoneura fumiferana* NPV (CfMNPV-DEF) (AAC09376) (Barrett *et al.*, 1995); *Mamestra brassicae* NPV (MbMNPV) (AAB03658) (Clarke *et al.*, 1996); *Orgyia pseudotsugata* NPV (OpMNPV) (Q65363) (Ahrens *et al.*, 1997) and one human 3,4-catechol oestrogen UDP-glucoronosyltransferase (Hs) (Ritter *et al.*, 1990) were aligned with CLUSTAL X (1.81) (Thompson *et al.*, 1997) using default parameters and the Blosum matrix. GENEDOC software (Nicholas and Nicholas, 1997) was used for homology shading of the aligned amino acid sequence. A tree of EGT's was constructed using the Neighbour-joining algorithm of CLUSTAL X. The human EGT (Hs) was chosen as the out-group. The resulting phylogenetic trees were visualised in Treeview (Page, 1996).

EGT expression assay

A 50ul mixture containing: 5ul sample; 5ul 10x Tris Malate; 5ul 10mM UDP-Glucose; 5ul 10mM UDP-Galactose; 2.5ul [^{3}H] ecdysone and 27.5ul dH_2O was incubated at 37°C for 3 hours. Two volumes of 100% ethanol were added to stop the reaction. After the addition of ethanol, the entire solution was placed in a speed vac for 2 hours. The resulting pellet was resuspended in 25ul of 60% ethanol. The reaction products were separated by thin-layer chromatography, and the plate was allowed to dry overnight. The dried plate was exposed to an IP plate and visualized by a PhosphorImager (FUJIX Bas 1500). Negative controls included SF900II medium from mock infected SF9 cells as well as Grace's supplemented medium from uninfected HzAMI cells. The supernatant of AcMNPV (wild type) was used as a positive control for conjugation. [^{3}H] E (tritiated ecdysone) was also spotted onto the TLC plate. The supernatant of HaSNPV-SA infected HzAMI cells was tested for EGT activity.

Single nucleotide polymorphisms

SNP's were determined by aligning the four *Helicoverpa egt* nucleotide sequences with CLUSTAL X (1.81) (Thompson *et al.*, 1997) using default parameters and the Blosum matrix. GENEDOC software (Nicholas and Nicholas, 1997) was used for homology shading of the aligned nucleotide sequence. SPN's were tabulated (Table 1).

Results:

Identification and sequencing

To identify the HaSNPV-SA *egt* homologue, an *Eco*RI genomic library of HaSNPV-SA was constructed using the plasmid vector pBluescript SK$^+$. After digestion with the appropriate restriction enzyme, the resulting fragments were separated by agarose gel electrophoresis and

transferred to nitrocellulose membranes. The blots were probed with a DIG-Nick labelled HzSNPV-ZIM *egt* clone (*Eco*RV-I9) at 55°C. Hybridisation was detected to the *Eco*RI-K clone with the HzSNPV-ZIM *egt* gene (data not shown). This *Eco*RI-K clone contained an inserted fragment of 4.3-kb. A restriction map of this fragment localized the *egt* gene to the region between the *Eco*RV and *Sac* I sites at 382-nucleotides (nt) and 2484-nt respectively (Fig. 1). Other restriction enzymes, which were used to map this fragment, included, *Hin*dIII, *Sal*I and *Bst*XI. The *egt* gene and its flanking regions were sequenced from both strands of subclones of the *Eco*RI-K fragment using universal and specific primers (Fig. 1).

Sequence of the *egt* gene and the EGT protein

Sequence data showed that the HaSNPV-SA *egt* gene had an open reading frame (ORF) of 1548-nt encoding a predicted protein of 516 amino acids (aa) with a M_r of 58.897-Da (Fig. 2) The 5' noncoding region of the HaSNPV-SA *egt* gene, as seen with the other *Helicoverpa's*, contained various control elements for early and late transcription (Fig. 2). In HaSNPV-SA two possible TATA boxes were located 96-nt and 309-nt upstream of the translation start codon. A Putative enhancer elements, GATA was observed at nt position, -60 while an early transcription initiation motif, CAGT was observed at –133 with respect to the ATG translation start of HaSNPV-SA a TAAG motif characteristic for baculovirus late transcription was found at nt position –94. A potential polyadenylation signal, AATAAA, was found 39-nt downstream of the TAA translation termination codon.

The HaSNPV-SA EGT contained six potential N-linked glycosylation consensus sequences (N-X-T/S), however it is not known whether these sites are indeed glycosylated (Fig. 2). A putative signal sequence of 21 amino acids is present at the N-terminus of HaSNPV-SA using the proteomics tools of the ExPASy Molecular Biology server (Fig. 2).

Comparative analyses of EGT's

The HaSNPV-SA ORF exhibits a 46% aa identity with the AcMNPV *egt* gene products, but seem to be more closely related to BusuSNPV, MacoNPV and SeMNPV with aa identities of approximately 50% to each of these EGT's. All five *Helicoverpa's* were 99% identical while 12 SNP's were shown to occur throughout the gene (Table 1). Ten conserved regions (I-X) with an identity of at least 45% have been identified (Riegel *et al.*, 1994; Barrett *et al.*, 1995; O'Reilly, 1995 and Hu *et al.*, 1997), an increased number of EGT's, however decreases this value. All of the GV's lack region X, together with AaNPV, which lacks region IX as well. Eight amino acids (HaSNPV-SA EGT residues Ser33, Trp353, Gln356, Gly373, Glu379, Pro386, Pro391 and Gln396) were shown to be absolutely conserved in all tranferases irrespective of the origin. Chen *et al.*, (1997) included one extra aa residue at position Gly304 but this residue is not conserved in AhGV or AoGV respectively (Fig. 3).

Phylogenetic analysis

The phylogeny of all 27 baculovirus EGT's present in GenBank as well as the human EGT, Hs were analysed using the Neighbour-joining algorithm of Clustal X (Thompson *et al.*, 1997). A tree using human EGT as an outgroup is shown in Fig 4. As shown by previous studies, the baculovirus EGT's from AcMNPV, BmMNPV, CfMNPV, OpMNPV and CfMNPV-DEF once again belongs to one clade. This clade is however now expanded to include the EGT's from AaNPV, EppoNPV and AgNPV. Using Hs as an outgroup, the EGT's of SeMNPV, MacoNPV, MbMNPV, EoNPV and BusuSNPV form one clade, while LdMNPV, SlMNPV and all the *Helicoverpa* SNPV's are rooted from this clade but belong to no particular group. The 5 GV's of PxGV, CpGV, CfGV, LoGV and EaGV all formed one group, while the two closely related GV's, AhGv and AoGV formed another.

Glucosyltranferase activity

The revealed that the EGT protein from the culture medium of cells infected by HaSNPV-SA and the positive control AcMNPV, were both functionally active (Fig. 5, lanes 3 and 4, respectively). Extracts from uninfected SF9 and HzAMI cells did not exhibit any enzymatic activity (Fig. 5, lanes 1 and 2, respectively). Both AcMNPV and HaSNPV-SA conjugates 78.8% of the ecdysone (data not shown).

Discussion and Conclusion:

Most identified baculovirus *egt's* have TATA boxes 5' to the coding region (Hu *et al.*, 1997), the exception being CFMNPV *egt* which has an AT rich region, rather than a canonical TATA box (Barrett *et al.*, 1995). For most baculoviruses the *egt* genes are transcribed from a start site downstream from their TATA box [eg. AcMNPV at 28-nt downstream; LdMNPV at 12-nt downstream, SlMNPV at 22-nt downstream (Toister-Achituv and Factor, 1997) and EppoNPV with three positioned at 29, 31 and 33-nt downstream (Caradoc-Davies *et al.*, 2001]. For this reason it is unlikely that the latter TATA box situated 96-nt downstream of the CAGT motif is functional (Fig. 2). It is however possible, that an imperfect CAGT motif (CAGC) located 86-nt upstream of the ATG start codon and 10-nt downstream from this putative TATA box (Fig. 2), could be functional. Blissard *et al.*, (1992) noted that with modification of the CAGT element the correct transcript was synthesized, however with significant loss of transcriptional activity. We have not yet compared the levels of *egt* activity in infected cells to confirm whether it correlates with upstream CAGT elements. Normal transcription initiation is started at a point 12-28-nt downstream of their TATA box. Based on this model, the TATA box situated 309-nt upstream of the ATG start signal (Fig. 2) is thus too far upstream to be functional. As with CFMNPV *egt* which has an AT rich region, it is

possible that HaSNPV *egt*'s follow a similar pattern. Transcriptional analysis should clarify this point.

The EGT ORF's of the five *Helicoverpa* NPV isolates are all the same size (516-aa) and have an aa identity of greater than 99% when compared to each other (Fig. 3). SNP analysis showed that single nucleotide differences occurred throughout the genes of the various isolates (Table 1). There is no evidence that these polymorphisms in any way hinder the function of *egt*, since both Ha-G4 and Ha-SA EGT's have been shown to be functionally active. When analysing various genes of HaSNPV-SA, it was found that certain genes (polyhedrin) would show a high degree of homology to HaSNPV-G4 while others were more homologous to HaSNPV-ELCAR (*egt*). This is not unusual, since comparative analysis of most these two genomes of HaSNPV-G4 (Chen *et al.*, 2001) and HzSNPV-ELCAR (Chen *et al.*, 2002) show them to have a 98% amino acid identity and a 99% identity for nucleotides. Their *egt* genes are 99% similar and 100% identical. It is therefore highly likely, that most of these isolates could be divergent strains from the same isolate (Chen *et al.*, 2002). The only group of genes not sharing a similar degree of high conservation is the baculovirus repeat ORF's (*bro*-genes) (Chapter 4, this thesis).

The alignment of the 28 EGT's (Fig. 3) showed that they displayed similarity throughout their sequence and included the UDP-glucosyltransferase signature: [FVA]-[LIVMF]-[TS]-[HQ]-[SGAC]-G-x (2)-[STG]-x (2)-[DE]-x (6)-P- [LIVMFA]-[LIVMFA]-x (2)-P- [LMVFIQ]-x (2)-[DE]-Q (Fig. 3) (Mackenzie *et al.*, 1997). Of the ten conserved regions found (Fig. 3) the largest and most conserved is region IV. Hu *et al.*, (1997) reported, that the identity of these ten regions was at least 45%. As expected however, an increased number of EGT's decreases this value. Furthermore, region X is not conserved to a high degree within the EGT's used, and is not found in any of the GV's or AaNPV, which also lacks region IX. The overall identity among the EGT's

ranges from 7% to 99%. Of the eight aa which are absolutely conserved in all known tranferases, all but one (Ser33) is located in region VII. O'Reilly (1995) suggested, that these conserved aa might play a role in the enzymatic activity and UDP-sugar binding of glucosyltransferases. Furthermore, there are 33 conserved aa which are restricted to baculovirus EGT's. Chen *et al.*, (1997) suggested that these could be involved in recognition of ecdysteroids in insects. When baculovirus EGT's were compared with UDP-glucosyltransferases from mammalian, plant, nematode and bacterial origin, only regions V and VII was conserved in all those investigated. This could be attributed to the suggested function of these conserved regions namely, carbohydrate binding.

The phylogenetic tree constructed (Fig. 4) is in close agreement with those proposed by Barrett *et al.*, (1995) (7 EGT's); Clarke *et al.*, (1996) (9 EGT's); Hu *et al.*, (1997) (11 EGT's) and Chen *et al.*, (1997) (12 EGT's), except that the present one includes much more EGT's (28 ORF's). All of the EGT's seem to root from the clade containing AcMNPV, BmMNPV, CfMNPV, OpMNPV, CfMNPV-DEF, AaNPV, EppoNPV, OpSNPV and AgNPV. It is possible that one of the EGT's in this group provided the common ancestor *egt* gene from which all other baculovirus EGT's evolved. When one compares this tree to the many proposed by Herniou *et al.* (2001), the tree is in agreement on the separation of the NPV's and GV's and the division of the NPV's into groups I and II. It important, when doing phylogenetic analysis, that one investigates the possibility of different genes producing similar or different trees. Phylogeny based on a single gene may not accurately represent baculovirus relatedness.

As seen in Fig. 5, the results obtained strongly suggest that the EGT activity detected in medium of virus-infected cells resulted from a virus-encoded enzyme. O'Reilly and Miller, (1989) confirmed

by genetic and biochemical studies that this gene encodes an ecdysteroid UDP-glucosyltransferase (*egt*).

The presence of the *egt* gene in 27 isolates in GenBank supports the view that most baculoviruses may carry an *egt* gene even though the genome of CuniSNPV and XecnGV do not. Chen *et al.*, (2000a) constructed two recombinant HaSNPV's by deleting the ecdysteroid UDP-glucosyltransferase gene. Bioassay data showed, that the LT_{50} of the *egt*-deletion recombinants was about 27% faster than that of HaSNPV wild type. Thus proving that the deletion of the *egt* gene plays an important role in enhancing baculoviruses in biocontrol strategies.

ACKNOWLEDGEMENTS

We thank Clifford Jacobs for his HzSNPV EcoRV-I9 clone. Kate Caradoc-Davies is thanked for investigating the activity of the HaSNPV-SA *egt*. The National Research Foundation supported the research. SK received grants from the National Research Foundation and the Solm Yach Fellowship.

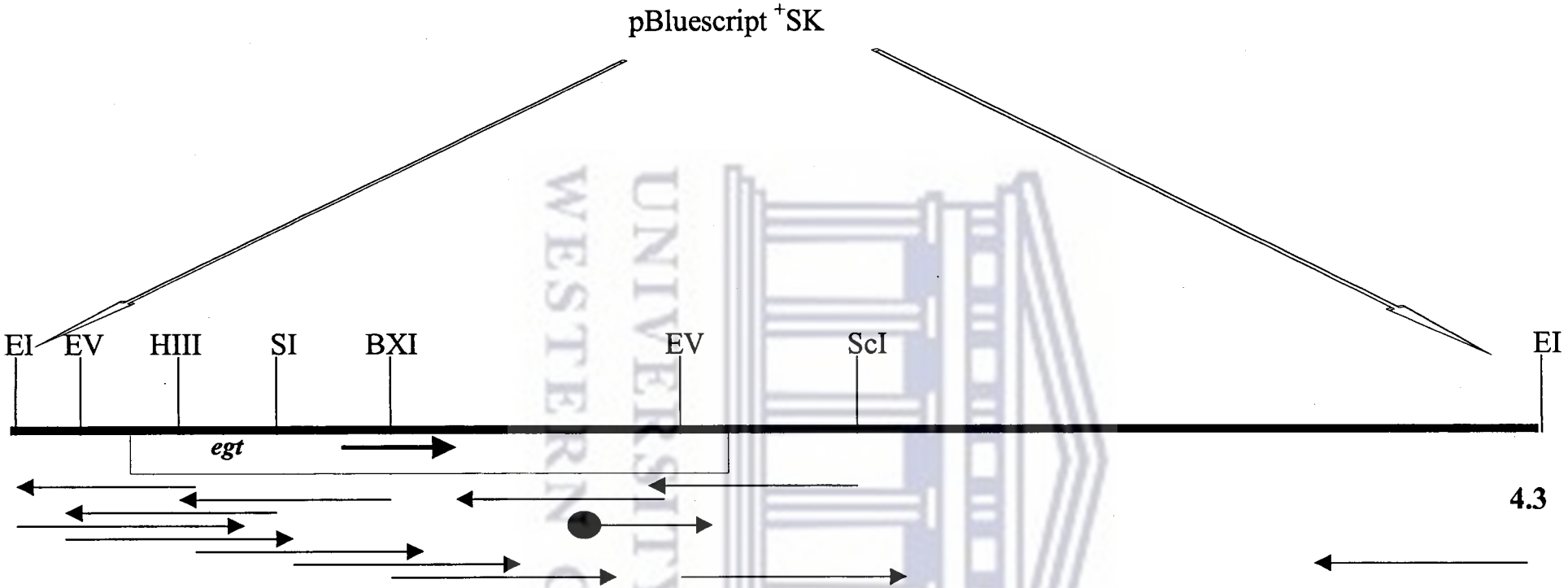


Figure 1. Location of the *egt* gene on the cloned 4.3-kb fragment of the HaSNPV-SA genome into pBluescript. For the sequencing strategy the arrows indicate the direction of sequencing and the dot at the proximal end of the arrow indicates the use of a specific primer. EI = *Eco*RI; EV = *Eco*RV; HIII = *Hin*dIII; SI = *Sal*I; BXI = *Bst*XI; ScI = *Sac*I.

```
                                                                  TCATGTAT
AGTTATCGTGTTATCAGCAGCAGCAGTCGTTGTGGTAGGCGCATCGTTAGCGATACGACGATTGCGTATAAAGGTTTT
ACAGGCGCTAGTTTTAAACAACATGACACATAACAGCAATATTATTATGAGGACAGATTTTAAAAGTATGCGGTCATT
GTTTTCTGTGTCTTTGATGTTGTCTAGTTCGTCGAACAGCGCGTCGATTACGCTATCTATGCCATCGTGATTCGACAG
AGTTATAATTATTTATTAATGTGTTGCTGTTGTTGCGTATGACATCATAGGTATTAGTACTGTCATTGCTAAATATAGA
CATGAAATAATTATCTTCGTTGAGGTTATCACGAACACTAGTAGTCATCGTGACGATAGATATCTGTAATACACACAT
CAAAGTAAACATGTTTACTTAAACAGTAACTGAATAATAATTTCAACATAGCGACGCCACTATAAGATGCAGCATCCC
GTCCGTTGGGCATCTTTCGATAAACGCTCTGACCCATAAACGGACGTGCGCTAATTTTTTTTTATTGCTAAATTCAAA
ATGTACAAACAAATAATAACTATGTTATTGTTGGTGTTGTTTCTGTCGGTTCTGGATGGAGCGCGTATCCTGTGCGTT
M  Y  K  Q  I  I  T  M  L  L  L  V  L  F  L  S  V  L  D  G  A  R  I  L  C  V
TTTCCTGTTCCTTCGTACAGTCATCATGCAGTGTTCGAAGCTTACACCAATGCTCTAGCGTCGCGTGGCCATACAATA
F  P  V  P  S  Y  S  H  H  A  V  F  E  A  Y  T  N  A  L  A  S  R  G  H  T  I
GTCAGAATCACACCGTTTCCCACTAAGAAAAACGATTCATCCAACGTGACAGATGTCGACGTTAGCTTGTCGAAAGAT
V  R  I  T  P  F  P  T  K  K  N  D  S  S  N  V  T  D  V  D  V  S  L  S  K  D
TATTTTAAAAGTCTTGTGGACCGATCTAGACTGTTCAAGAAACGAGGCGTTATTTCGGAAACGTCCAGCGTGACCGCT
Y  F  K  S  L  V  D  R  S  R  L  F  K  K  R  G  V  I  S  E  T  S  S  V  T  A
CGCAATTACATCAGTCTAGTACACATGTTGATTGATCAATTCTCTATGGAGAGTGTACGACAATTGATCGAATCCAAC
R  N  Y  I  S  L  V  H  M  L  I  D  Q  F  S  M  E  S  V  R  Q  L  I  E  S  N
AATGTTTTCGATTTGTTGGTGACCGAAGCCTTTCTAGATTATCCTCTGGTGTTTTCGCATTTGTTTGGCGATGTGCCT
N  V  F  D  L  L  V  T  E  A  F  L  D  Y  P  L  V  F  S  H  L  F  G  D  V  P
GTCATACAAATTTCGTCGGGTCACGCTTTGGCCGAAAATTTTGAGACAATGGGAGCCGTGAGCCGACATCCCATTTAC
V  I  Q  I  S  S  G  H  A  L  A  E  N  F  E  T  M  G  A  V  S  R  H  P  I  Y
TATCCAAATTTGTGGCGCAACAAATTTCAAAATTTAAACGTTTGGGAGATAATAACGGAAATCTATACAGAACTGGTG
Y  P  N  L  W  R  N  K  F  Q  N  L  N  V  W  E  I  I  T  E  I  Y  T  E  L  V
CTGTACTTGGAATTTGCTCGTTTAGCCGACGAACAAACTAAAATGCTTCGCCATCAATTCGGACCAAACACGCCCAGC
L  Y  L  E  F  A  R  L  A  D  E  Q  T  K  M  L  R  H  Q  F  G  P  N  T  P  S
GTGGAAGAACTGCGACAACGCGTTCAATTATTGTTTGTGAATACGCATCCGCTGTTTGATAATAACAGACCAGTACCG
V  E  E  L  R  Q  R  V  Q  L  L  F  V  N  T  H  P  L  F  D  N  N  R  P  V  P
CCGAGTGTACAATATTTGGGAAGTCTACATCTTGATCGAAACAATGATGTCAACGAACAGCAAACGATGGACTATAAT
P  S  V  Q  Y  L  G  S  L  H  L  D  R  N  N  D  V  N  E  Q  Q  T  M  D  Y  N
TTGATGCAATTTTTAAATAATTCTACAAACGGTGTGGTGTACGTGAGCTTCGGTACGTCTATACGAGTTTCAGACATG
L  M  Q  F  L  N  N  S  T  N  G  V  V  Y  V  S  F  G  T  S  I  R  V  S  D  M
GACGACGAATTTCTGTTTGAATTTATAACAGCTTTCAAGCAATTACCCTATAATATATTGTGGAAGACCGATGGACTG
D  D  E  F  L  F  E  F  I  T  A  F  K  Q  L  P  Y  N  I  L  W  K  T  D  G  L
CCCATGGAACACGTACTGCCTAAAAATGTGTTGACACAAACTTGGCTGCCGCAACACCATGTATTGAAACACAGCAAT
P  M  E  H  V  L  P  K  N  V  L  T  Q  T  W  L  P  Q  H  H  V  L  K  H  S  N
GTAGTTGCTTTTGTTACTCAAGGCGGAATGCAGTCAACGGACGAAGCCATCGACGCTTGTGTACCACTAATCGGAATC
V  V  A  F  V  T  Q  G  G  M  Q  S  T  D  E  A  I  D  A  C  V  P  L  I  G  I
CCGTTTATGGGCGACCAAGCATACAATACCAATAAATACGAAGAACTCGGAATCGGACGCAACCTCGATCCCGTAACG
P  F  M  G  D  Q  A  Y  N  T  N  K  Y  E  E  L  G  I  G  R  N  L  D  P  V  T
CTCACAAGTCATATTTTGGTGTCTGCCGTTTTAGATGTGACCGTCAACAACAAGAGTCGCTACACAGATAATATTAAA
L  T  S  H  I  L  V  S  A  V  L  D  V  T  V  N  N  K  S  R  Y  T  D  N  I  K
GCATTGAATCGTTCCACTAATTATCGAACACGGAAACCTATGGAAAAGGCCATCTGGTACACAGAACATGTAATTGAT
A  L  N  R  S  T  N  Y  R  T  R  K  P  M  E  K  A  I  W  Y  T  E  H  V  I  D
AATGGTAAAAATCCCATTTTAAAAACGAAGGCCGCCAACGTATCGTATAGCAAATATTATATGAGTGATATCATCGTT
N  G  K  N  P  I  L  K  T  K  A  A  N  V  S  Y  S  K  Y  Y  M  S  D  I  I  V
CCTGTTATAACGTTTTTGGTAATGACTCATTTGGGTCAGGCTATTCGGCGGTTGGTTGTTATTTAATACTGTATGACA
P  V  I  T  F  L  V  M  T  H  L  G  Q  A  I  R  R  L  V  V  I  *
ATGTACACATGTGTTAATAAAAAAGGCATTACTAATATTTAGATTGTTTCAAATTATTTACGCATGACTACCCGTCTC
CTATTGCGCAGCTACGCTAGCTTTA  AATACAGCCGATGGCGTAGTAAAGTTCATTTAAATATCTAAATTGGTTAGTTC
AACATCGCGGTGCGAGCGCACGACTTATACCATGCATCGTTCCAATAGTAACAGCAGCAAATACAAACAATCGCTGAT
AAATCGCTTTGAACTGGAATACAAAAGTGTGTCTGTGCGCGATTTGCAAAAATTGTCAGCGGCCATGTATCGTTTGTT
GGCTGTGAACGATAAACTTATGGAAAATTTACAAACTCTACCGATGCATTATAGAGCTCCAGCTTTGTCCCTTTAGTGA
```

Figure 2. Nucleotide sequence of the HaSNPV-SA *egt* gene and its flanking sequence. The predicted amino acids are indicated by one-letter code designations below the nucleotide sequence. Putative transcription initiation and termination signals are in bold (see text). The putatitive signal sequence is in bold italics and possible N-linked glycosylation sequence (N-X-S/T) is underlined. The putative enhancer element (GAT) is underlined, the possible TATA boxes are in bold, the TAAG motif is in italics and the polyadenylation signal is in bold.

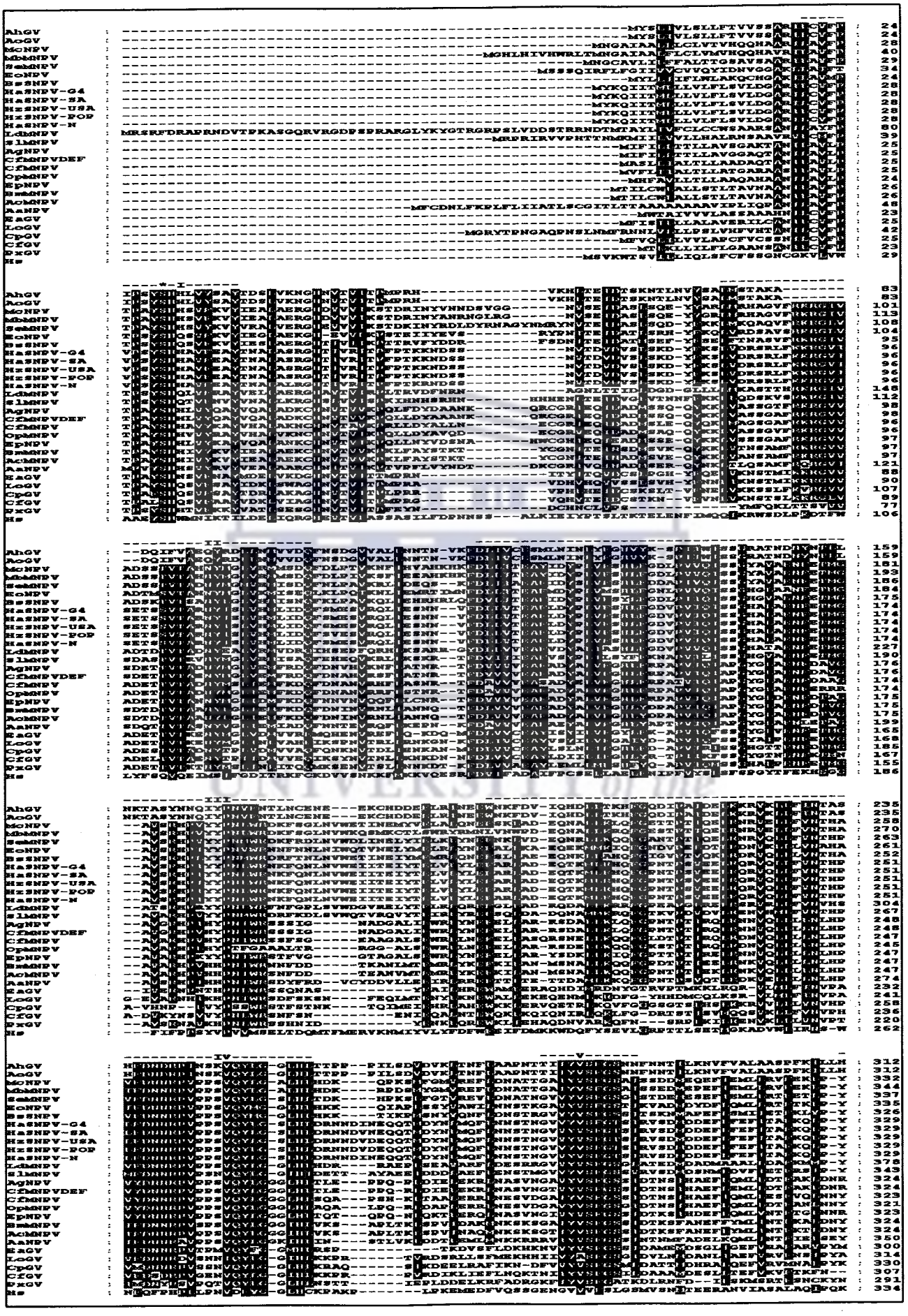


Fig. 3 continued

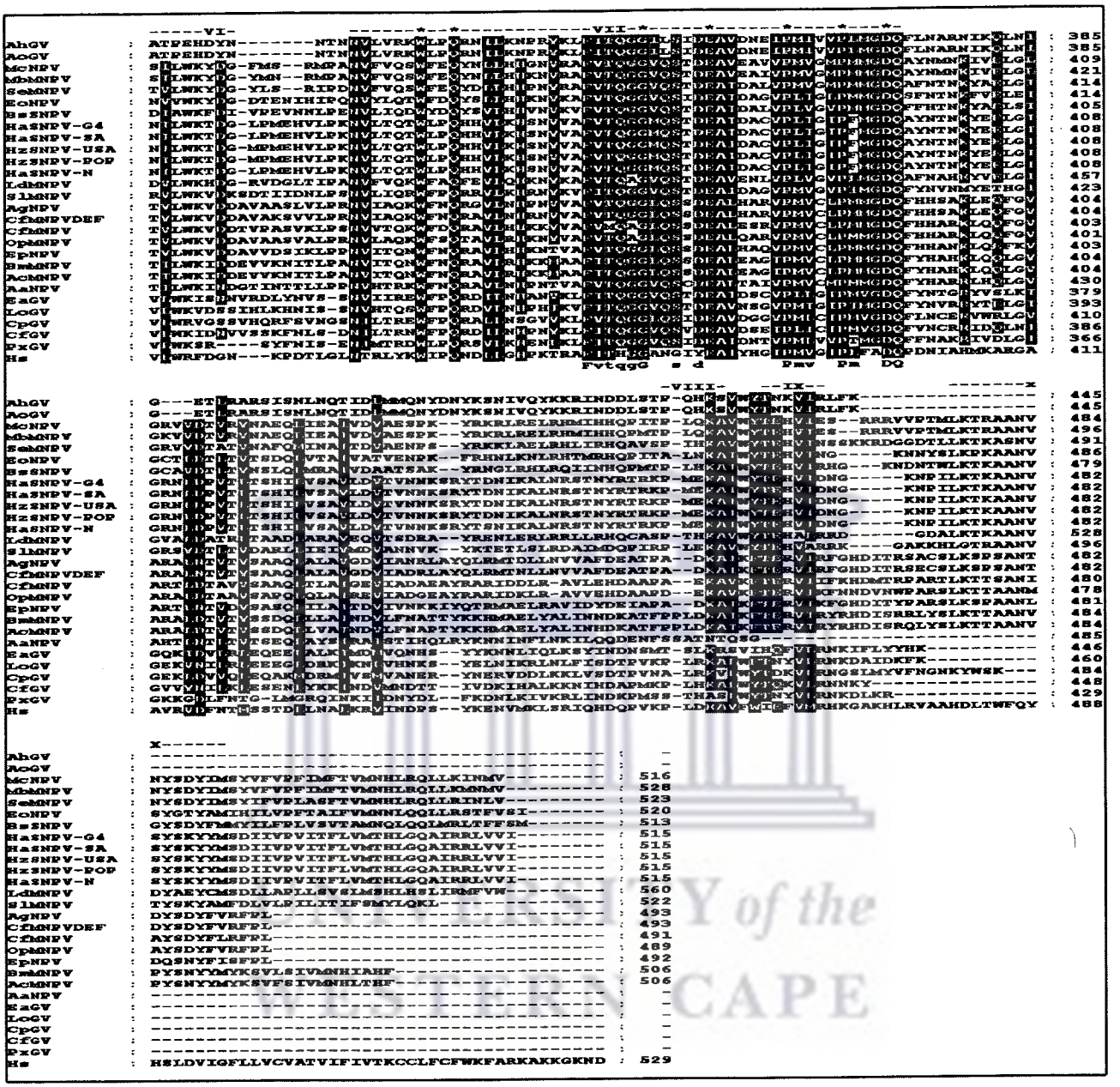


Figure 3. Alignment of the amino acid sequence of twenty-seven baculovirus *egt* genes one human 3,4-catechol oestrogen UDP-glucoronosyltransferase. The alignment was carried out by the Multiple Protein Alignment of Clustal X (1.81). GeneDoc software was used for homology shading. The conserved regions are indicated with Roman numerals. A consensus is indicated in the region of the EGT signature sequence. The sources of sequence and the abbreviations of viruses are given in the text. The seven absolutely conserved amino acids among all the UDP-glucosyltransferases investigated are indicated by asterisk (*). Gaps introduced into the alignment are indicated with dashes. Two shading levels were set: black for 100% identity and grey for 70% identity.

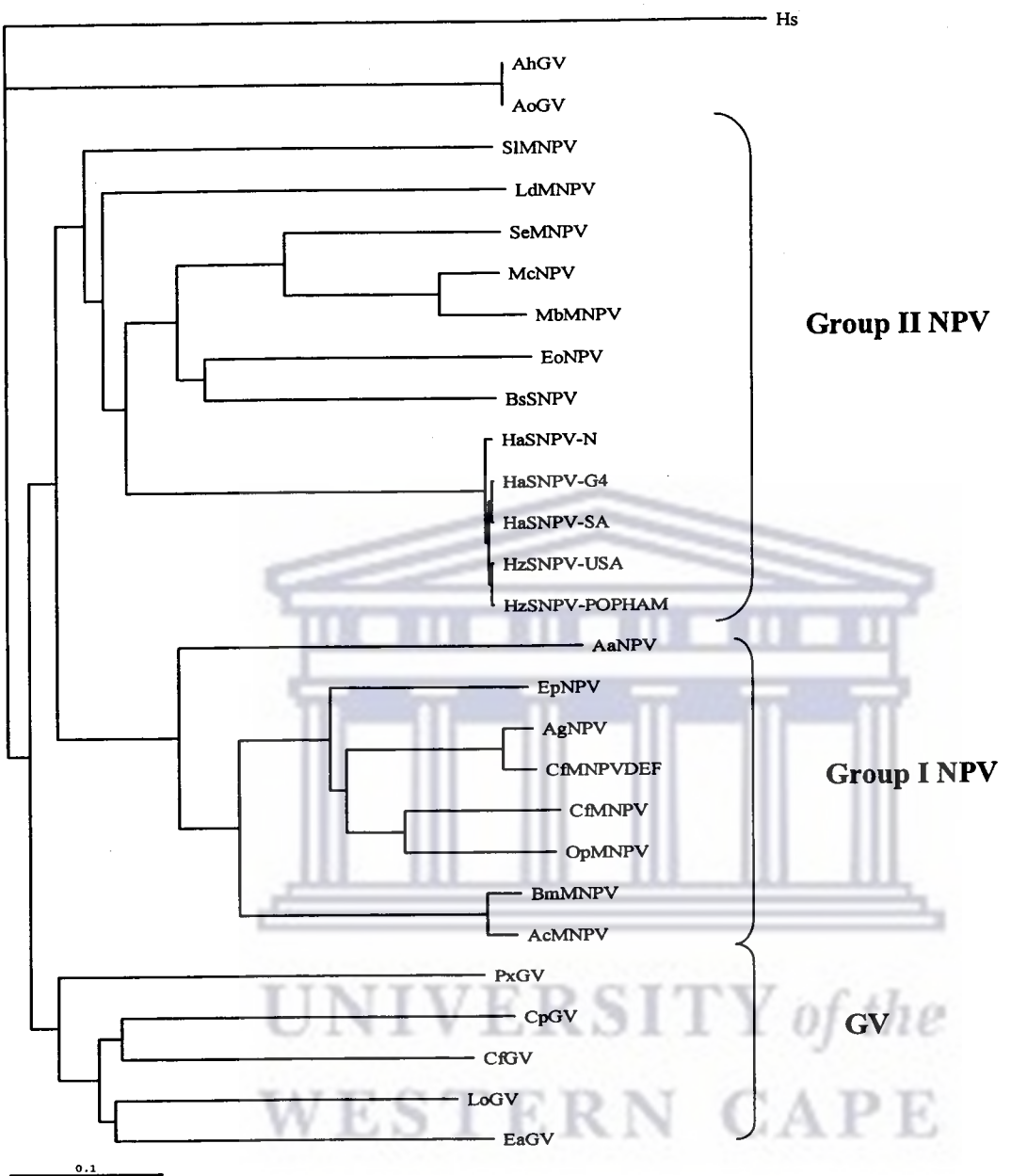


Figure 4. A phylogenetic tree of *egt* genes. A tree for twenty-seven baculovirus EGT's and one human 3,4-catechol oestrogen UDP-glucoronosyltransferase was constructed by using the Neighbour-joining algorithm of Clustal X. Hs was chosen as an outgroup.

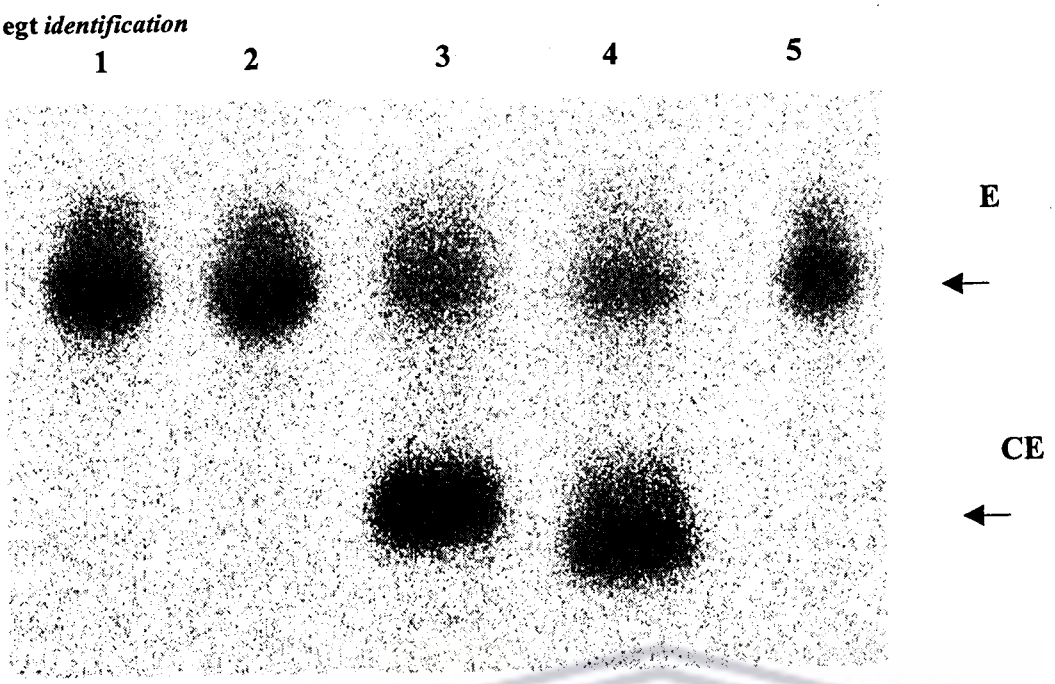


Figure 5. EGT assay of HaSNPV-SA infected cells. A PhosphorImager picture showing the separation of EGT reaction products on a TLC plate. E, free [^{3}H] ecdysone; CE, [^{3}H] ecdysone-sugar conjugates. Lane 1. SF900III medium from uninfected SF9 cells; lane 2. Grace's supplemented medium from uninfected HzAMI cells; lane 3. a sample from HaSNPV-SA-infected HzAMI cells; lane 4. a sample from AcMNPV-infected SF9 cells and lane 5. tritriated free [^{3}H] ecdysone.

Table 1. Tabulation of SNP variations within the nucleotide sequence of the *egt* gene

SNP	Ha-G4				Hz-ELCAR				Ha-SA				Ha-N				Ha-Popham			
position[a]	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T	A	C	G	T
12			+			+				+					+		+			
140		+					+				+					+		+		
165				+					+		+			+						+
185	+					+				+					+		+			
358	+				+				+				+						+	
387				+			+				+					+		+		
829	+							+				+	+						+	
831				+			+				+					+		+		
832	+							+		+			+						+	
1012		+				+					+			+				+		
1308		+					+				+					+		+		
1315			+					+				+				+			+	
1316	+					+				+				+				+		

[a]Nucleotide position as from the start codon

CHAPTER FOUR

DIVERSITY OF BACULOVIRUS REPEATED OPEN READING FRAMES (*BRO*) IN THREE NUCLEOPOLYHEDROVIRUSES OF *HELICOVERPA SPP.*

Abstract:

A comparison of the genomes of *Helicoverpa armigera* (Ha-G4) and *Helicoverpa zea* (strain Elcar, Hz-USA) nucleopolyhedroviruses showed that the most divergent region among these isolates contained two of the three baculovirus-related open reading frames (*bro* genes) named *bro*-a and -b respectively. The three *Helicoverpa* SNPV *bro*-genes were distributed in two regions for both virus strains. *Bro*-c the third of the three *bro*-genes was more highly conserved than its fellow *bro*-genes with an amino acid sequence similarity of 99.4%. *Bro*-a and -b were found juxtaposed and flanked by two homologous repeat (*hr*) regions while *bro*-c was found several ORF's downstream of *bro*-b in the genetic map (Chen *et al.*, 2002). Sequence comparison of both regions showed *bro*-a and *bro*-b to be the most divergent and the least conserved. For this reason, the same region was investigated in a South African isolate of HaSNPV (Ha-SA). Gene specific primers were designed to the flanking regions of *bro*-a and *bro*-b to study the polymorphism in this region. These primers (*Hr1F* and *Hr4R*) were used to amplify a 5.5-kb segment by polymerase chain reaction for all three isolates (Ha-G4; Hz-USA and Ha-SA). RFLP analyses of this segment of the isolates showed definite polymorphisms when digested with *Eco*RI, *Eco*RV, *Spe*I and *Sal*I. Primer walking sequenced the entire bro-a/bro-b region of Ha-SA. Phylogenetic analysis divided baculovirus BROs into four groups, group I, II, III and IV. The group II was further subdivided into group II-A and II-B. The *Helicoverpa* SNPV *bro*-a and *bro*-b genes were placed into the group II-B *bro*-type genes, while *Bro*-c consistently fell into the group II-A *bro*-group. . Northern analyses and/or RT-PCR indicated that all Ha-G4 3 *bro*-genes were transcribed at either 4h to 24h post infection (p.i).

Introduction:

Both *Helicoverpa armigera* and *Helicoverpa zea* single nucleocapsid nucleopolyhedrovirus (HaSNPV and HzSNPV) are members of the large family of viruses called *Baculoviridae*. Baculoviruses have large, circular, covalently closed, double-stranded DNA genomes that are mainly pathogenic for Lepidopteran insects.

A prominent feature of most baculovirus genomes sequenced thus far is the presence of multiple repeat open reading frames (ORF's) termed baculovirus repeat ORF's (*bro*-genes) by Kuzio *et al.* (1999), found dispersed throughout the genomes. All these *bro*-genes seem to be related to ORF 2 of the prototype baculovirus *Autographa californica* MNPV (AcMNPV) (Ayres *et al.*, 1994). It is important to note, that not only do the number of *bro*'s present on the genomes differ, AcMNPV-1; *Bombyx mori* NPV (BmNPV)-5 (Gomi *et al.*, 1999); *Orgyia pseudotsugata* MNPV (OpMNPV)-2 (Ahrens *et al.*, 1997); *Lymantria dispar* MNPV (LdMNPV)-16 (Kuzio *et al.*, 1999) and *Xestia c-nigrum* granulovirus (XcGV)-7 (Hashimoto *et al.*, 2000), but their location and sequence composition varies too. Three related *bro*-genes named *bro*-a, -b and -c were identified in both *H. armigera* (Ha-G4) (Chen *et al.*, 2001) and *H. zea* (Hz-USA) (Chen *et al.*, 2002) and found at similar locations on their genomes.

Kuzio *et al.*, (1999) proposed that the *bro*-genes have a related sequence in their N-terminal regions but shared differing degrees of similarity in other regions. They divided the *bro*-genes into four groups based on the relationship of the different domains. When analysing the *bro*-family on the basis of phylogeny, it is difficult to

establish orthology (orthologous genes are those genes that are present in two taxa because both taxa shared a common ancestor) between *bro*-genes of different genomes. For this reason when Herniou *et al.*, (2001) used the entire genome sequence data to infer baculovirus phylogeny, the *bro*-genes were omitted from the study. Since the N-terminal end of *bro*-genes seems more related, by paralogy (relationship between two genes related through gene duplication) it might be possible to phylogenetically compare the *bro*-family using this end only. Less ambiguous phylogenitc signals could be present here. Furthermore, Zemskov *et al.*, (2000) showed, that most *bro*-genes had a single-stranded DNA binding motif composed of a pattern of basic and aromatic amino acids with a consensus sequence of $K/RX_{2-5}K/RX_{4-12}F/YX_{2-14}F/YX_{6-13}F/YX_{1-19}K/RX_{3-26}F/Y/W/X_{6-12}K/R$ present in this region of the *bro*-genes.

Although the genomes of Ha-G4 and Hz-USA are very similar in nucleotide (nt) (97% identity) and amino acid (aa) (99% identity) sequence, a major difference is found in sequence and organisation of the *bro*-genes. *Bro*-a and *bro*-b in Hz-USA have an amino acid identity of 28% and 68% respectively, to Ha-G4 *bro*-a and -b genes. *Bro*-c is more highly conserved between the two genomes with an amino acid identity of 99.4%. Since sequence data strongly suggest that these two strains are variants of the same virus species, the low homology between *bro*-a and *bro*-b was interesting to note (Chen *et al.*, 2002). Ferber *et al.*, (2000) compared sequences and localisation of the *bro*-genes of two plaque purified BmNPV strains and showed that the 2 strains displayed similar variations with regard to their *bro*-gene composition, location and number as with the *Helicoverpa* isolates.

To study the evolutionary relatedness between *bro*-a and *bro*-b present in *Helicoverpa* SNPV isolates, a South African isolate of *H. armigera* SNPV (Ha-SA) was included in this study. Gene specific primers *(Hr1F* and *Hr4R)* were designed to the flanking regions of *bro*-a and *bro*-b of both Ha-G4 and Hz-USA (Fig. 1). These primers were used to amplify by PCR a 5.5-kb region, which contained both *bro*-genes -a and -b. Consistent restriction fragment length polymorphisms (RFLP) and direct sequence comparison was used to show the degree of conservation throughout. A fourth *Helicoverpa* isolate found in GenBank was included in this study. This *Helicoverpa armigera* (Ha-N) (Zhang and Jin, 2000) isolate was also from Chinese origin and had a sequence identity of >99% to Ha-G4. This isolate was thus also used in our sequence comparisons. All three *bro-* genes of Ha-G4 were analysed at the transcriptional level and appeared to be active. A phylogeny study was generated using the first 100 amino acids of the *bro*-gene or the entire gene where genes were shorter than 100 amino acids.

Materials and Methods:

Insect and Virus:

Ha-G4 BV's were kindly provided by X. Chen, while Hz-USA BV's were kindly provided by Dr. Albert Lu [Chen *et al.,* 2001 (b), Chen *et al.*, 2002]. Ha-SA (Khan *et al.*, 2000) was originally isolated from diseased *H. armigera* larvae in the Western Cape Province of South Africa in 1997. The bollworm *H. armigera* was cultured as a laboratory colony and reared on artificial diet as described by Zhang *et al.,* (1981). Hemolymph was extracted 4 days post infection (p.i). Both BV and hemolymph samples were propagated in *H. zea* cell line Hz-AM1 cells (McIntosh and Ignoffo,

1983), which were maintained at 28°C in Grace's medium supplemented with 10% fetal bovine serum. Viral DNA was isolated from insect derived polyhedra as described previously by Miller *et al.*, (1988).

Restriction endonuclease analysis of *Helicoverpa* SNPVs' genomic DNA:

Viral DNA of Ha-G4, Hz-USA and Ha-SA was digested with the restriction enzyme *Hin*dIII (GIBCO/BRL) and the fragments were separated by electrophoresis in a 0.6% agarose gel at 45 V (1.5V/am) for 16h. Lambda DNA digested with *Bam*HI-*Eco*RI-*Hin*dIII was used as molecular size standards.

Primer design:

The *Hr1F* (forward) and *Hr4R* (reverse) primers were designed using an alignment of Ha-G4 and Hz-USA to amplify a 5.5-kb fragment from ORF57/58 (Ha-G4 and Hz-USA respectively) to *he*65 of both *Helicoverpa* isolates. *Hr2HaR/Hr2HzR* and *Hr3Ha/Hr3HzF* were designed in a similar fashion to amplify smaller truncated fragments of the 5.5-kb region to allow for easier sequencing of the Ha-SA isolate. *BroA-FOR1*, *BroB-FOR1* and *BroC-RT* were each designed for the specific *bro*-genes of Ha-G4 and used in RT-PCR analysis. *BroB-REV1*, *BroB-REV2* and *BroB-FOR2* were designed homologous to the Ha-SA genome as part of a primer walk strategy to completely sequence this region of the South African isolate. The different primers are positioned as shown in Figure 1. The primer sequences, reaction conditions and the expected size of the amplification products for each primer set of PCR primers are listed in Table 1.

Polymerase Chain Reaction:

Amplification of target DNA by PCR was performed in a total reaction volume of 50 µl containing 200 µM of each deoxynucleoside triphosphate, 200 µM of each primer, 0.2 units of heat stable polymerase and polymerase buffer (Promega). For reactions using *Hr1F* and *Hr4R*, the Expand Long Template PCR system and buffer 1 was used (Roche Diagnostics) as per the manufacturer's instructions. Reaction conditions are outlined in Table 1. Ten microlitres of the amplified DNA fragments from PCR reactions were electrophoresed in 1% agarose containing EtBr (0.5 µg/ml) in TAE (Tris/Acetic acid/EDTA) buffer.

Restriction fragment length polymorphism:

Five microlitres of the PCR product of *Hr1F* and *Hr4R* was digested with *Eco*RI, *Eco*RV, *Spe*I and *Sal*I (GIBCO/BRL) in a total volume of 20 µl, according to the manufacturer's instructions. DNA fragments were separated by electrophoresis through 1.5% agarose gels.

DNA Sequencing and computer analysis of the Ha-SA Hr1F + Hr4R PCR fragment:

Plasmid DNA for sequencing was purified using the Qiagen columns (Qiagen. Inc.). Sequencing of the cloned fragments was conducted using either the M13 universal set of primers from both ends or the custom synthesis primers designed to homologous regions of sequenced areas as part of a "primer walk" strategy (Figure 1, Table 1). PCR fragments Ha-SA *Hr1F* + *Hr4R* (6-kb), Ha-SA *Hr2HzR* + *Hr1F* (3,2-kb) and Ha-SA *Hr3HaF* + *Hr4R* (2,4-kb) (Table 1.) were cloned into the plasmid vector, pGEM-T Easy, according to the instruction manual (Promega). Automated

sequencing was carried out at the Sequencing Facility in the Department of Molecular Biology of the Wageningen University. The generated sequences were analysed with the University of Wisconsin Genetics Computer Group programs (Devereux *et al.*, 1984) and DNASTAR™ (Lasergene). The DNA and the deduced amino acid sequences were compared with the updated GenBank/EMBL and a contiguous sequence formed. The nucleotide sequences of the complete ORF of Ha-SA *bro*-a and *bro*-b were deposited in GenBank and assigned the following accession number: AY092771.

Total RNA isolation, 3' RACE analysis and Northern Blot:

Total RNA was isolated from 2 x 10^6 mock infected and Ha-G4 infected HZ-AM1 cells (MOI of 5 TCID 50 units/cell) at 0, 4, 8, 12 and 24 p.i. as described by IJkel *et al.*, (2001).

RT-PCR was performed using the 5'/3' RACE kit (Roche) employing 2 µg purified total RNA as template per time point. First-strand cDNA synthesis was performed using AMV reverse transcriptase and the oligo-DT anchor primer according to the manufacturer's instructions. The cDNA mixtures were amplified by PCR using the PCR anchor primer and the gene specific primers shown to bind to each *bro*-gene specifically. For each primer extension the specific primer *BroA-FOR1*, *BroB-FOR1* and *BroC-RT* were synthesised (Table 1). The obtained PCR-products were analysed on 1.2% agarose gels.

For northern analysis 6 µg of total RNA was electrophoresed in 1.5% agarose in the presence of glyoxal and DMSO (Ausubel *et al.*, 1994) in 10mM sodium phosphate pH

6.5-7.0. The RNA was transferred to the Hybond N Nylon membrane (Amersham) in 25 mM sodium phosphate buffer and hybridised using standard techniques (Sambrook *et al.*, 1989). The northern blot was hybridised with *bro*-a, *bro*-b and *bro*-c random primed α^{33}[P]dATP-labelled probes (Van Oers *et al.*, 1993). The probes were the fragments obtained by either RT-PCR for *bro*-b and -c or the 2-GSP PCR for *bro*-a.

Phylogeny:

Forty-three *bro*-genes were used from 9 different baculovirus genomes. These genes included AcMNPV: AcORF2 (Ayres *et al.*, 1994); BmNPV: Bm*bro*-a to– *bro*-e (Gomi *et al.*, 1999); OpMNPV: OpORF67, ORF68 and ORF116 (Ahrens *et al.*, 1997); LdMNPV: Ld*bro*-a to *bro*-p (Kuzio *et al.*, 1999); XcGV: Xc*bro*-a to *bro*-g (Hashimoto *et al.*, 2000); HaSNPV (Ha-G4): HaG4ORF59, ORF60 (divided into a N-terminal and C-terminal end) and ORF105 (Chen *et al.*, 2001); HzSNPV (Hz-USA): HzORF60, ORF61 and ORF108 (Chen *et al.*, 2002); HaSNPV (Ha-N): HaNORF56, ORF57 and ORF103 (Zhang and Jin, 2000) and HaSNPV (Ha-SA): HaSA*bro*-a and *bro*-b. For each gene the first 100 amino acids from the N-terminal end (if genes were shorter than 100 amino acids the entire gene was used) were aligned with Clustal X (1.81) (Thompson *et al.*, 1994) using default parameters and the Blosum matrix. An unrooted tree was constructed using the Neighbour program from PHYLIP (Felsentein, 2000). One hundred replicates were performed. The resulting phylogenetic tree was visualized in TreeView (Page, 1996). GenDoc software was used for homology shading and scoring among the aligned sequences.

Results:

Size Variation Between Three *Helicoverpa* SNPVs Genomic DNA:

PCR amplification of the three *Helicoverpa* isolates containing *bro*-a and *bro*-b, using the primer pair *Hr1F* and *Hr4R* showed that all 3 isolates yielded a band of approximately 5.5-kb (Fig 2. lanes 1, 4 and 7). When the 3 isolates were amplified using the primer pair combination of *Hr2HaR/Hr2HzR* and *Hr1F* a variation in size was observed (Fig 2. lanes 2, 5 and 8). For Ha-G4, Ha-SA and Hz-USA the fragment lengths obtained were approximately 3.0-kb, 2.7-kb and 2.7-kb respectively. A 2.0-kb fragment was obtained for both Ha-G4 and Hz-USA using the primer pair combination of *Hr3HaF/Hr3HzF* and *Hr4R* (Fig 2. lanes 3, 6 and 9). A smaller fragment of approximately 1.8-kb was obtained for Ha-SA.

The PCR-RFLP screen was used to differentiate the *Helicoverpa* SNPV isolates. PCR products generated using the *Hr1F* and *Hr4R* primers were digested using four different restriction enzymes (*Eco*RI, *Eco*RV, *Sal*I and *Spe*I). These 4 enzymes produced restriction profiles that showed variation for the three isolates used. Although *Eco*RI did not digest the PCR fragment of Ha-SA and Hz-USA (Fig 3. lanes 2 and 3), there was one restriction site present in the 5.5-kb PCR fragment of Ha-G4 (Fig 3. lane 1). Ha-G4, Ha-SA and Hz-USA each had one *Eco*RV restriction site within the region, which appeared to be situated at a similar position (Fig 3. lanes 4, 5 and 6) while, Ha-SA had an additional *Eco*RV restriction site. *Sal*I had one restriction site present in the PCR amplified fragment of Ha-G4 and Ha-SA, and as seen by the size variation of the resultant bands this site was in a similarly located position (Fig 3. lanes 7 and 8). The amplified product of Hz-USA had 2 internal *Sal* I restriction sites

(Fig 3. lane 9). The restriction profile of the PCR amplified fragment digested with *Spe*I, showed that all 3 *Helicoverpa* isolates had 2 sites for this enzyme. Based on the sizes of the resultant bands, only one site appeared to be placed at a similar position (Fig 3. lanes 10, 11 and 12). Direct sequence analysis confirms the positions of these restriction enzymes.

Transcriptional Analysis:

To assess whether all three *bro*-genes found in Ha-G4 were expressed, total RNA was extracted from Ha-G4 purified from mock-infected HZ-AM1 cells at 1, 4, 8, 12 and 24 h p.i. Total RNA was also extracted from HZ-AM1 infected with Ha-SA at 24 h p.i. The temporal regulation of the *bro*-gene transcripts was examined by Northern Blot analysis and RT-PCR. Since the N-termini regions of most *bro*-genes are highly conserved (Kuzio *et al.*, 1999) primers were designed for each *Helicoverpa bro*-gene's specifically using the variation in sequence in the C-termini region. RT-PCR was performed with the total RNAs using the internal primer for each specific *bro*-gene and the anchor primer.

After numerous attempts, the RT-PCR analysis of *bro*-a yielded no positive result (using the internal primer *BroA-FOR1* (Table 1.) and the PCR oligo-dT anchor primer). It was decided to amplify the cDNA of Ha-G4 using two gene specific primers [*BroA-FOR1* and *Hr2HaR* (Table 1)] specific for the *bro*-a gene. As a negative control, total RNA of the corresponding cDNA was also amplified. As shown in Figure 4a no product was obtained for all of the Ha-G4 total RNA's (lanes 1-6) when the two gene specific primers were used. These two primers did however successfully amplify a 450-bp fragment for the corresponding cDNA at 24 h p.i. (Fig.

4a, lane 12). Ha-G4 genomic DNA was amplified as a positive control. Northern analysis of *bro*-a revealed a faint band of 0.8-kb that could represent the *bro*-a transcript (Fig. 4b). This transcript was only detected at 24 h p.i. The transcript size corresponds well with the predicted ORF size of 732-nt for *bro*-a of Ha-G4. The size of the *bro*-a transcript suggests that it is most likely not polyadenylated. Furthermore, since RT-PCR using the oligo-dT anchor primer and the gene specific primer yielded no positive result, this might confirm the absence of a poly(A) tail.

The RT-PCR technique allowed for information with regard to the temporal regulation of the Ha-G4 *bro*-b gene. A fragment of 520-bp was amplified (Fig. 5a) using the internal primer *BroB-FOR1* (Table 1.) and the PCR oligo-dT anchor primer. The 520-bp fragment was detected from 4 h p.i. up till and including 24 h p.i. Northern analysis of *bro*-b showed a band of 1.8-kb that could possibly represent the *bro*-b transcript (Fig. 5b). This transcript was detected at 4 h p.i. up till 24 h p.i that corresponds well with the RT-PCR result. Since the predicted ORF size of Ha-G4 *bro*-b is 1581 nt, the size of the *bro*-b transcript suggests that it is most likely polyadenylated.

RT-PCR analysis of Ha-G4 *bro*-c amplified a fragment of 480-bp at 24 h p.i. only (Fig. 6a, lane 7). The internal primer *BroC-RT* (Table 1) and the oligo-dT anchor primer were used to amplify this fragment. As previously shown by Chen *et al.*, [2001 and 2002], *bro*-c is highly conserved with a nt sequence homology of 99% between Ha-G4 and Hz-USA. Thus as expected, the internal primer *BroC-RT* which was specifically designed from the Ha-G4 sequence was also able to amplify a fragment when the cDNA of Ha-SA at 24 h p.i. was used as a template (Fig. 6a, lane

8). Northern analysis of Ha-G4 RNA using the radio labelled *bro*-c probe revealed a 1.7-kb band for both Ha-G4 and Ha-SA RNA at 24 h p.i. Since the predicted size of *bro*-c is 1503-nt, the size of the *bro*-b transcript suggests that it is most likely polyadynelated.

Comparative Sequence Analysis:

The sequence divergence between the homologous *bro*-genes of Ha-G4, Ha-SA, Ha-N and Hz-USA were investigated. The *bro*-a and *bro*-b genes of Ha-SA were sequenced and compared at the amino acid level to three known *Helicoverpa bro*-a and -b genes (HaG4, Hz-USA and Ha-N). *Bro*-c of Ha-SA was not sequenced as the percentage homology of the HaG4, Hz-USA and Ha-N was >99% for this gene among *Helicoverpa* isolates. *Bro*-a Ha-SA is a homologue of Ha-G4ORF59, Hz-USAORF60 and Ha-NORF56 proteins but lack 103 a.a. at the C-termini end. The Ha-SA *bro*-a has a relatively low amino acid identity to the proteins of HaG4, Hz-USA and Ha-N with 40%, 14% and 37% respectively while amino acid similarity is 45%, 22% and 43% respectively (Fig. 7a). The Ha-G4ORF59 had an 89% and 28% amino acid identity to Ha-NORF56 and Hz-USAORF60 respectively (Fig. 7 a).

As shown by Chen et *al.*, 2002 the Ha-G4ORF60 had a N-terminal extension of 183 amino acids when compared to Ha-SA *bro*-b, Hz-USAORF61 and Ha-NORF57 proteins. For the purpose of this analysis, Ha-G4ORF60 was divided into the N-terminal 183 amino acid portion and the 344 C-terminal ends. The additional 183 N-terminal amino acids of Ha-G4ORF60 showed homology to Ha-SA *bro*-b (25%), Ha-NORF57 (48%) and Hz-USAORF61 (28%) for aa identity and 35%, 48% and 36%

for amino acid similarity respectively (Fig. 7b). The C-terminal end of Ha-G4ORF60 showed significant homology for amino acid identity to Ha-SA *bro*-b (98%), Ha-NORF57 (64%) and Hz-USAORF61 (68%) proteins and an amino acid similarity of 98%, 77% and 79% respectively (Fig. 7b).

The *Helicoverpa* BROs were compared to other baculovirus BROs by phylogenetic analysis. Amino acid homologies allowed for the BROs to be divided into four groups: group I, II, III and IV. An unrooted tree was constructed which clearly depicted the four groups (Fig. 8). This grouping is in close agreement with the BRO groups proposed by Kuzio *et al.*, (1999), except the present one includes more *bro*-genes. The addition of more BRO's also accounts for the division of group II, which is further subdivided into group II-A and II-B. The present tree does however place two *bro*-genes Ld*bro*-g and Ld*bro*-e into Group III rather than group IV as previously suggested by Kuzio. Ha-G4ORF59, Ha-Sa*bro*-a, Ha-NORF56, Hz-USAORF60, Ha-G4ORF60-C, Ha-G4ORF60-N, Ha-NORF57, Ha-Sa*bro*-b and Hz-USAORF61 all belong to the group II-B genes. Ha-G4ORF105, Ha-NORF108, and Hz-USAORF103 all belong to group II-A.

Discussion and Conclusion:

Genome sequencing provides a view of genes retained in all genomes and the location of the respective genes. It is believed that these common genes are likely to be essential for the virus replication and survival. The occurrence of *bro*-genes is a common feature in baculoviruses and represents a remarkable example of duplication of a single gene.

Helicoverpa species commonly contain three *bro*-genes. Although the *Helicoverpa* genomes sequenced to date are highly conserved and appear to be extremely homologous, the most divergent region among *Helicoverpa* species is the region containing *bro*-a and –b. The third *bro*-gene, *bro*-c is more than 99% conserved among these isolates.

Previous studies have shown that restriction patterns of HzSNPV and HaSNPV were very similar (Knell and Summers, 1984 and Chen *et al.*, 2000). A detailed physical map of the viral genome of Ha-G4 (Chen *et al.*, 2000) placed the two *bro*-genes, *bro*-a and –b, on the *Hin*dIII fragment B. Comparative analysis of the restriction profiles of the three genomes digested with *Hin*d III revealed that the corresponding fragment containing *bro*-a and –b showed a degree of variation (data not shown). Different sized PCR fragments were also obtained when similarly located primers were used to amplify various parts within this variable region between the genomes (Fig. 2). Size variation could denote a possible insertion or deletion within a particular region or a site-specific sequence difference. RFLP analysis of the PCR fragment of all 3 *Helicoverpa* SNPV's spanning the entire *bro*-a and –b region showed distinct patterns for each isolate (Fig. 3). Variation in restriction patterns further signifies a variation in sequence. Christian *et al.*, (2000) used the PCR-RFLP approach for the identification and differentiation of *Helicoverpa* nucleopolyhedroviruses isolated from the environment. PCR primers were designed to amplify a 400-bp fragment from the polyhedrin gene region and DNA fragments from the putative protein kinase ORF. This approach was used not only to generate diagnostic DNA profiles that can be used to identify heterologous NPV's, but also to differentiate genotypic variants of *Helicoverpa* SNPV's. Since the *bro*-a and –b region of the three *Helicoverpa* isolates

used in this study and the fourth isolated found in GenBank were shown to be variable, this area within *Helicoverpa* genomes could be a suitable one for a simple distinction between these isolates.

All three *bro*-genes have a TAGT sequence instead of the CAGT early gene start motif. TAGT has previously been shown to also function as an early gene start site for baculovirus transcription (Kang *et al.*, 1999). The *bro*-b transcript could be detected as early as 4 h p.i. up till 24 h p.i. by both RT-PCR and northern analysis. However, both RT-PCR and northern analysis yielded a positive transcript for *bro*-c only at 24 h p.i. RT-PCR failed to produce any positive result for *bro*-a. The two-gene specific primer (2GSP) approach however yielded a positive fragment at 24 h p.i. and this fragment was also detected by northern hybridisation. The size obtained for both the *bro*-a and –b transcript of Ha-G4 by northern analysis could be indicative of the presence of a poly(A) tail or be indicative of 5'and 3' UTR's. Sequencing of the 3' end of the PCR fragments would clarify this point. A possible explanation for the failed RT-PCR attempts for *bro*-a could be a lack of a poly(A) tail, which is supported by the size of the *bro*-a transcript. Even though almost all cellular mRNA's posses poly(A), a significant exception is provided by the mRNA's that code for the histone proteins (a major structural component of chromosomal material) (Lewin, 1994). These mRNA's comprise most or all of the poly (A) fraction. Zemskov *et al.* (2000) showed that 3 of the 5 *Bombyx mori* (Bm) BRO's (BRO-A, BRO-C and BRO-D) have nucleic acid binding activity and are involved in nucleoprotein complexes in the nuclei of infected cells. These three BROs were shown to be involved in nucleosome organization that could block cellular replication and/or transcription and switch host machinery to viral DNA or RNA synthesis. They detected the BmNPV proteins

BRO-A, BRO-C and BRO-D in the histone H1 fraction using anti-BRO antibodies. Chromatographic fractions showed that only BRO-A and/or BRO-C interacted with core histones.

Bro-a of Ha-G4 was also shown to contain a motif [Ha-G4ORF59 (5-60) K-X_5 K-X_4 Y-X_3 F-X_{10} F-X_7 R-X_{16} F-X_3 K] found in single strand binding (SSB) proteins from prokaryotic and eukaryotic organisms (Wang and Hall, 1990). This motif is found in the highly conserved N-motif of most of the BRO proteins. This finding suggests that these proteins could be interacting with nucleic acids. Both *bro*-a of Ha-SA and Ha-N have variations of this conserved motif present in its N-terminal portion. This motif is however not present in the N-terminal end of *bro*-a of Hz-USA. For Ha-G4, the 183-N-terminal end of BRO-B also contained a similar motif but no motif was found in the C-terminal portion. None of the BRO-B proteins of any of the other three isolates contain this motif. Furthermore, none of the BRO-C proteins of all four isolates contained this motif. All five of the Bm BRO's contained this motif but only three were shown to positively bind proteins. It is possible that different BRO's have different functions since deleting the *bro*-b and –e genes of BmNPV did not alter the infectivity (Kang *et al.*, 1999).

The *Helicoverpa* BRO's were compared to the other baculovirus BRO's by phylogenetic analysis. Based on the sequence homologies of the first 100 amino acids the BRO's were divided into four groups while group II was further subdivided into subgroup II-A and II-B (if the bro-gene used was shorter than 100 amino acids, the entire BRO protein was used). Most of the BROs were in the same groupings proposed by Kuzio *et al.* (1999). Group I appears to encompass Ac*re* (AcORF2), the

five BmNPV BRO's and LdMNPV BRO-B, -J, -K, -N and –P. This group is in total agreement with the group I proposed by Kuzio *et al.* (1999). Since this tree contains more BRO's than the ones used by Kuzio, *et al.* (1999), group II is divided into two clades. Group II-A is in its entirety a new sub-group containing all three the BRO-C proteins of the *Helicoverpa* isolates and XcGV BRO-A, -E and –F. XcGV BRO-E however seems to be more closely related to the *Helicoverpa* BROs than to the other two XcGV BROs found in the same sub-group. Group II-B is in total agreement with the group II proposed by Kuzio *et al.* (1999) except for the addition of the *Helicoverpa* BRO-A and BRO-Bs. Of the *Helicoverpa* BRO-A's, the Ha-G4, Ha-N and Ha-SA seem to be more closely related than the Hz-USA BRO-A. Furthermore, the Hz-USA BRO-A seems to be closely related to the Hz-USA and Ha-SA BRO-B proteins and the C-terminal end of Ha-G4 BRO-B. The N-terminal end of Ha-G4 BRO-B and the entire BRO-B protein of Ha-N seem strongly related to the BRO-As of the *Helicoverpa* isolates. The only variation of the groups which deviate from that suggested by Kuzio *et al.* (1999), is that the two BRO's of LdMNPV, LdBRO-G and BRO-E falls in group III rather than group IV. This is however not surprising, since the shading patterns of the N-terminal end of these two proteins proposed by Kuzio *et al.* (1999) seem to be more similar to those in group III rather than group IV. It is notable that the BROs from the same virus do not always have a close relationship, suggesting that they might have a different evolutionary history. In a particular virus the BRO's could be, as a result of either gene duplication or acquisition from different or common sources, such as the insect host or another baculovirus occupying the same ecological niche.

Even though all four the *Helicoverpa* SNPVs compared contained the *bro*-a and –b genes in similar positions, their sequence conservation had different degrees of similarity. It is significant that theses two *bro*'s are flanked by homologous repeat (*hr*) regions within the genome and it is possible that they were duplicated along with these *hr* regions. Majima *et al.* (1993), showed that duplication and deletions occur most commonly in *hrs*. Furthermore, when Gracia-Maruniak (1996) studied the *hr* regions of *Anticarsia gemmatalis*, they reported that *hrs* were regions of high genotypic variation. As suggested by Munoz *et al.* (1999), various genotypes of *Spodoptera exigua* nucleopolyhedrovirus could possibly have evolved due to re-arrangement of an *hr* region. It is feasible that a similar argument holds for the variation observed between the *Helicoverpa* isolates with regard to re-arrangement within the *bro*-a and –b region of their genomes.

Acknowledgements:

We thank Magda Usmany and Els Rhoode for technical assistance. Monique Van Oers and Xinwen Chen are thanked for invaluable assistance and discussions. The Royal Netherlands Academy of Sciences supported the research. SK received grants from the National Research Foundation, the Solm Yach Fellowship.

Table 1. Details of Primers and reaction conditions used for the amplification of NPV DNA fragments

Primers	Sequences	Approximate size of amplified product	Intended purpose of primer	DNA Source	PCR Reaction Conditions				
					Step	Denature	Anneal	Extend	Number of cycles
Hr1F[A]	5'-GCTGTCGTACTAACATTGGTAAC-3'	5,5-kb	Amplification of *bro*-a	All 3 *Helicoverpa*	1	93℃, 2 min	N/A	N/A	1
Hr4R	5'-CTCAGTACATAATCGGCATAAGAC-3'		and *bro*-b genomic	SNPV's	2	93℃, 10 s	60℃, 30 s	68℃, 4 min	10
			region		3	93℃, 10 s	60℃, 30 s	68℃, 4 min?	20
					4	N/A	N/A	68℃, 7 min	1
Hr1F	5'-GCTGTCGTACTAACATTGGTAAC-3'	3,0-kb	Amplification of *bro*-a	Ha-G4	2	95℃, 30 s	60℃, 30 s	72℃, 2 min	30
Hr2HaR	5'-TGATGCGAGACTCGTGGAAGTG-3'		genomic region						
Hr1F	5'-GCTGTCGTACTAACATTGGTAAC-3'	2,7-kb	Amplification of *bro*-a	Ha-SA	2	95℃, 30 s	60℃, 30 s	72℃, 2 min	30
Hr2HzR	5'-GACTTGGCTTGAACGTAAGATGAC-3'	2.7-kb	genomic region	Hz-USA					
Hr3HaF	5'-TACTGATATTAACGTCCGTAACTAC-3'	2,0-kb	Amplification of *bro*-b	Ha-G4	2	95℃, 30 s	60℃, 30 s	72℃, 2 min	30
Hr4R	5'-CTCAGTACATAATCGGCATAAGAC-3'	1.8-kb	genomic region	Ha-SA					
Hr3HzF	5'-TGATTGAGTTTTGTAGCGATTTTGTG-3'	2,0-kb	Amplification of *bro*-b	Hz-USA	2	95℃, 30 s	60℃, 30 s	72℃, 2 min	30
Hr4R	5'-CTCAGTACATAATCGGCATAAGAC-3'		genomic region						
BroA-FOR1[B]	5'-TAAAGCTATTTTAGAAAAAGTGTCTGATGG-3'	0,6-kb	3'-RACE	Ha-G4	2	95℃, 30 s	60℃, 25 s	72℃, 1 min	35
Anchor	5'-GACCACGCGTATCGATGTCGAC-3'								
BroB-FOR1	5'-ACTTGGCCGTAGCAAATCAGGG-3'	0,45-kb	2GSP	Ha-G4	2	95℃, 30 s	60℃, 25 s	72℃, 45 s	35
Hr2Ha	5'- TGATGCGAGACTCGTGGAAGTG -3'								
BroC-RT	5'-CAGAAATGGAGGTGCTCGACGAAACTGAA-3'	0,5-kb	3'-RACE	Ha-G4	2	95℃, 30 s	60℃, 25 s	72℃, 45 s	35
Anchor	5'-GACCACGCGTATCGATGTCGAC-3'								
BroB-REV1	5'-GGTTGATAAACTTCGTTTTCGCTTGGATG-3'	-	Primer walking of	Ha-SA					
BroBREV 2	5'-GACGTCGTTTAAAATATTCGTAGTTAC-3'	-	Ha-SA	Ha-SA					
BroB-FOR2	5'-GTGTTCGAGTTGATCAACGCGAGCGACATG-3'	-		Ha-SA					

[A]Expand Long Template System Used
[B]Used in Primer walking strategy as well
?Accumulative 20 s added to each cycle

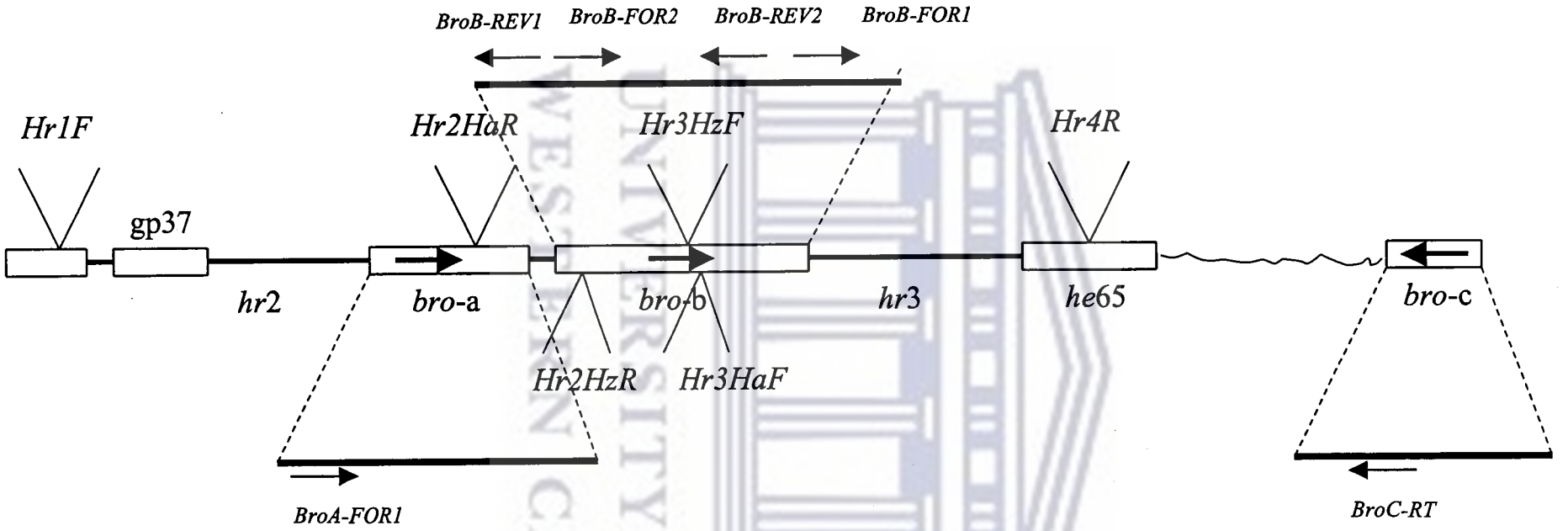


Figure 1. Diagram showing the positions of the various primers (*see also Materials and Methods*). ORF's [*bro*-a, -b and –c, Chen *et al* 2001 (a) (b)] are represented by boxes, ⟶ represent 3'-RACE and Primer walking primer positions, ➡ represent the transcription direction.

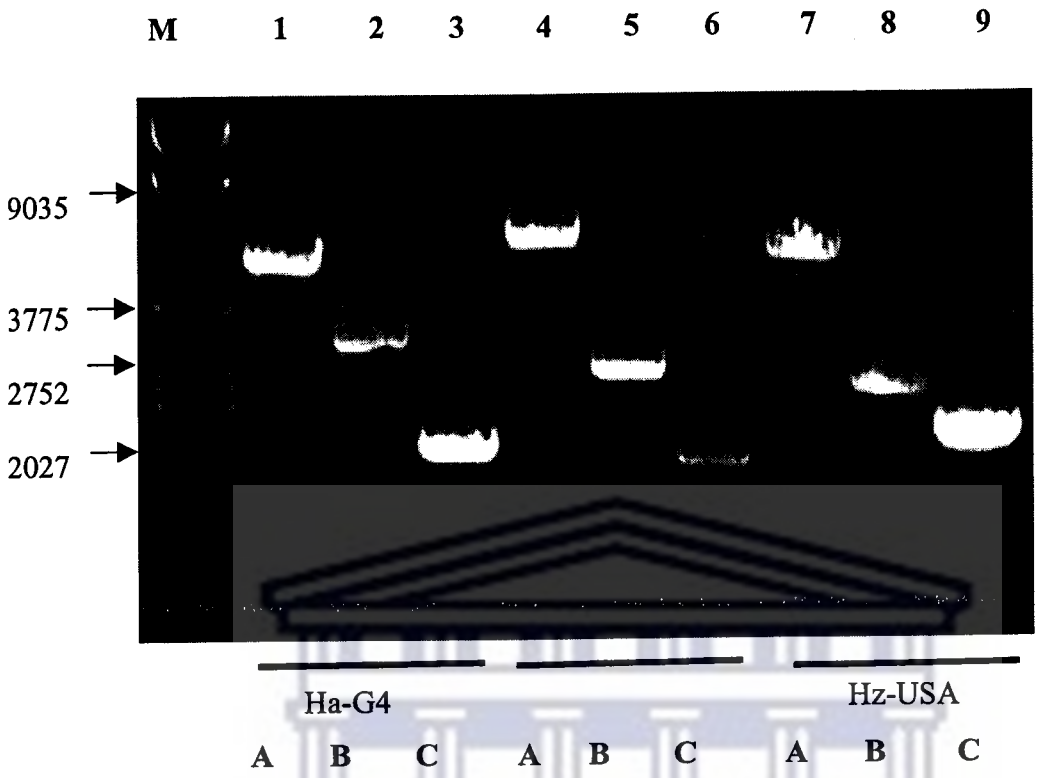


Figure 2. PCR amplification of the three *Helicoverpa* genomes using various primer pair combinations. **A.** Primer pair combination of *Hr1F* and *Hr4R*. **B.** Primer pair combination of *Hr1F* and either *Hr2HaR* or *Hr2HzR*. **C.** Primer pair combination of *Hr4R* and either *Hr3HaF* or *Hr3HzF*. Lambda DNA digested with *Bam* HI-*Eco* RI-*Hind* III was used as molecular size markers (**M**) with the sizes indicated in base pairs.

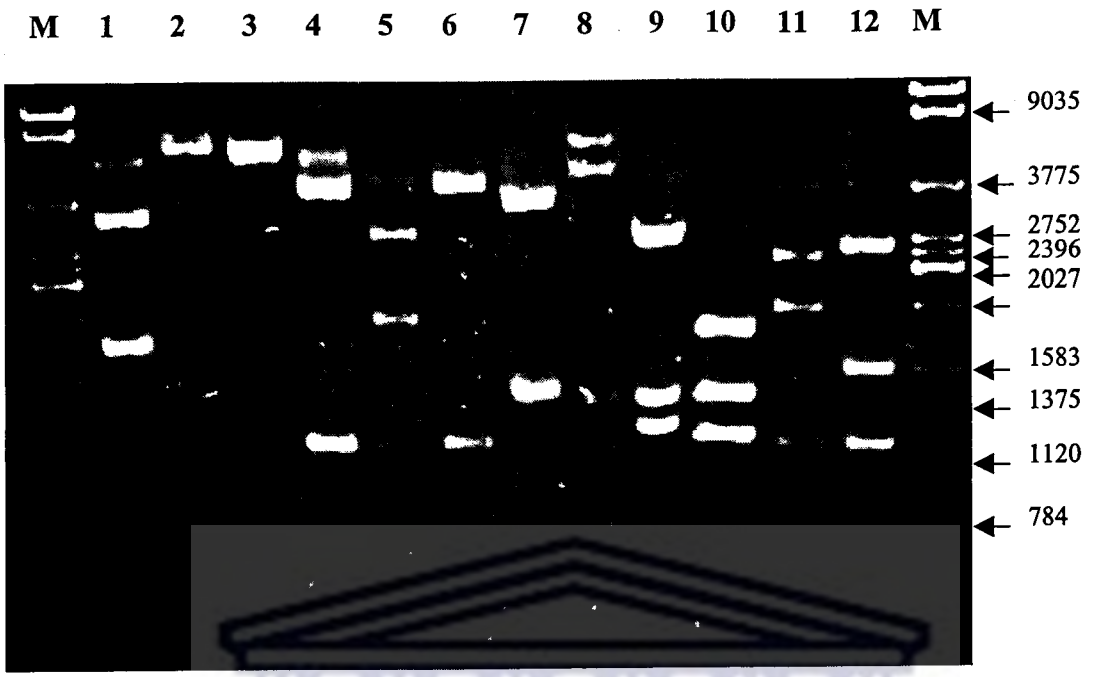


Figure 3. PCR-RFLP analysis of the 5.5-kb fragment yielded by the primer pair of *Hr1F* and *Hr4R*. Lane 1, 4, 7, 10: Ha-G4; lane 2, 5, 8, 11, Ha-SA; lane 3, 6, 9, 12, Hz-USA; 1, 2, 3, digested with *Eco*RI; 4, 5 and 6 digested with *Eco*RV; 7, 8 and 9 digested with *Sal*I; 10, 11 and 12 digested with *Spe*I. Lambda DNA digested with *Bam* HI-*Eco* RI-*Hind* III was used as molecular size markers (M) with the sizes indicated in base pairs.

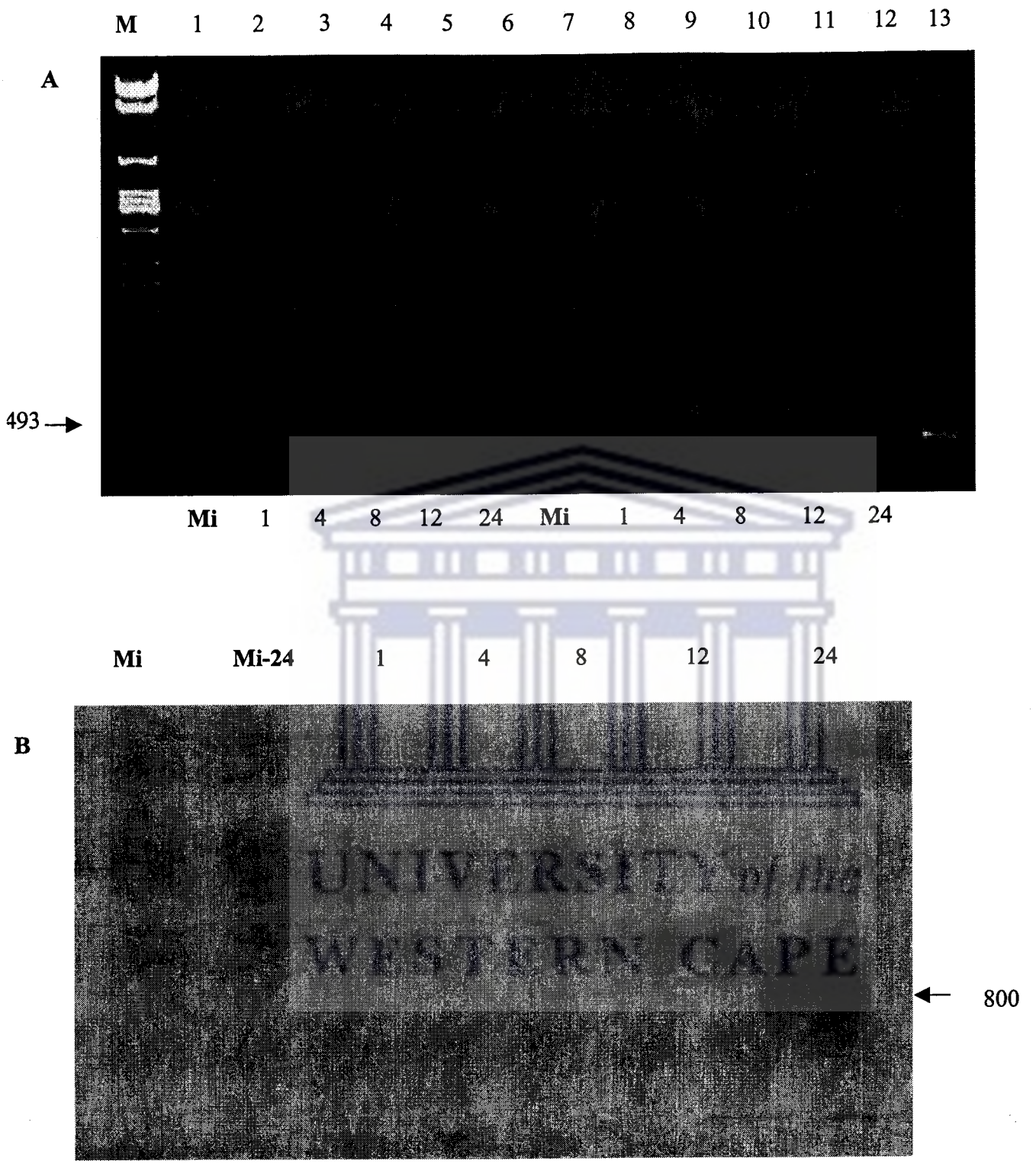


Figure 4. Expression of *bro*-a transcripts of Ha-G4. **A.** 2 Gene specific primers-2GSP (*BroA-FOR1* and *Hr2HaR*) were used to analyse the cDNA of Ha-G4. Lanes 1, 2, 3 4, 5 and 6; 2-GSP used with RNA of Ha-G4 as a template; Lanes 7, 8, 9, 10, 11 and 12; 2-GSP used with cDNA of Ha-G4 as a template; Lane 13; 2-GSP used with genomic DNA of Ha-G4 as a template. Lambda DNA digested with *Bam* HI-*Eco* RI-*Hind* III was used as molecular size markers (**M**) with the sizes indicated in base pairs. **B.** Northern analysis of *bro*-a. The approximate size of the hybridisation band is indicated on the right. The 2-GSP fragment was used as a probe.RNA used for the analysis in **B.** was total RNA extracted from Hz-AM1 cells obtained under the same conditions as for **A** (see text). Times p.i. are indicated above the lanes (**Mi**, mock infected).

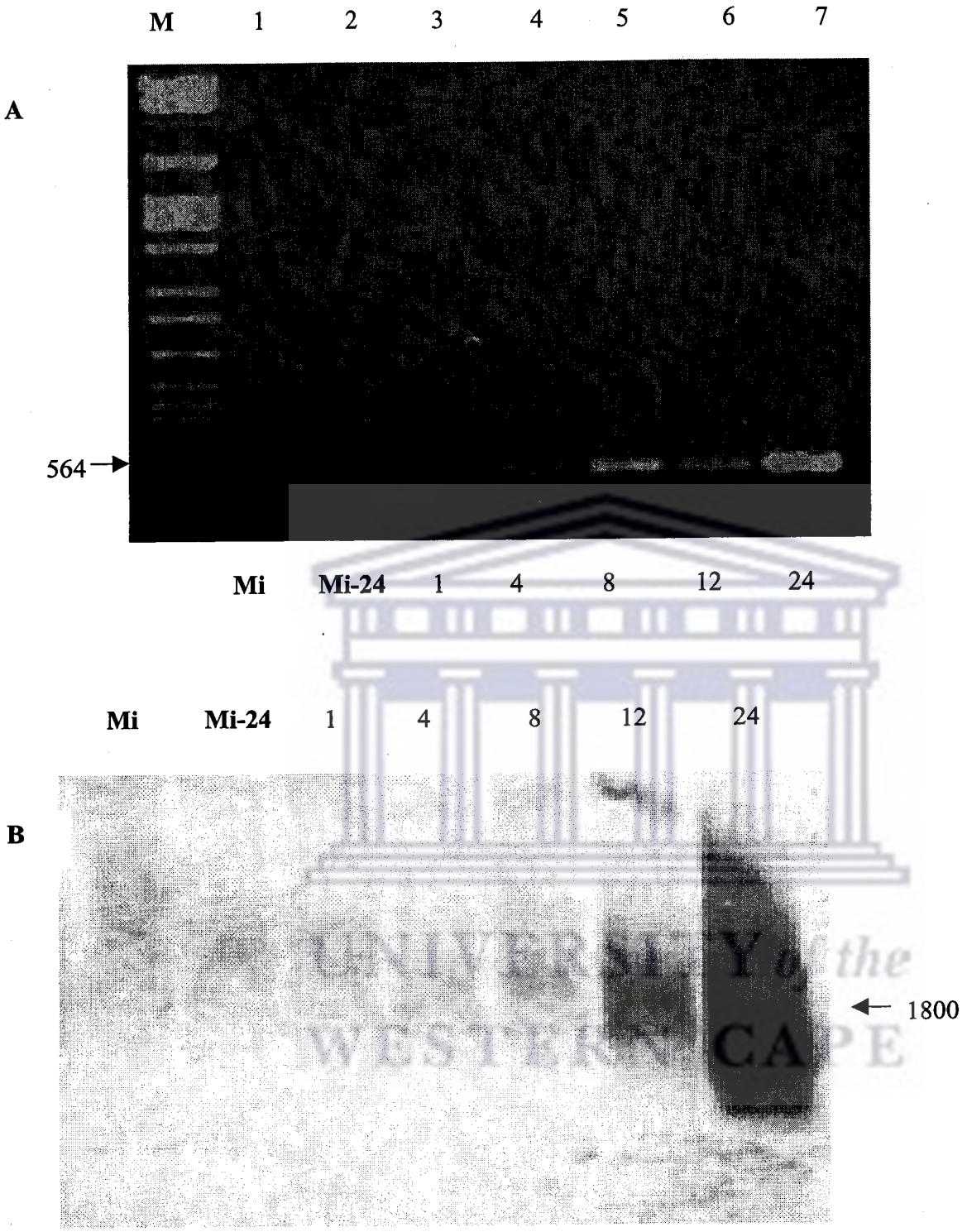


Figure 5. Expression of *bro*-b transcripts of Ha-G4. **A.** RT-PCR analysis of *bro*-b performed on total RNA extracted from Hz-AM1 cells(see text). Lambda DNA digested with *Bam* HI-*Eco* RI-*Hind* III was used as molecular size markers (**M**) with the sizes indicated in base pairs. Times p.i. are indicated below the lanes (**Mi**, mock infected). **B.** Northern analysis of *bro*-b. The approximate size of the hybridisation band is indicated on the right. RT-PCR fragments were used as probes. Times p.i. are indicated above the lanes (**Mi**, mock infected).

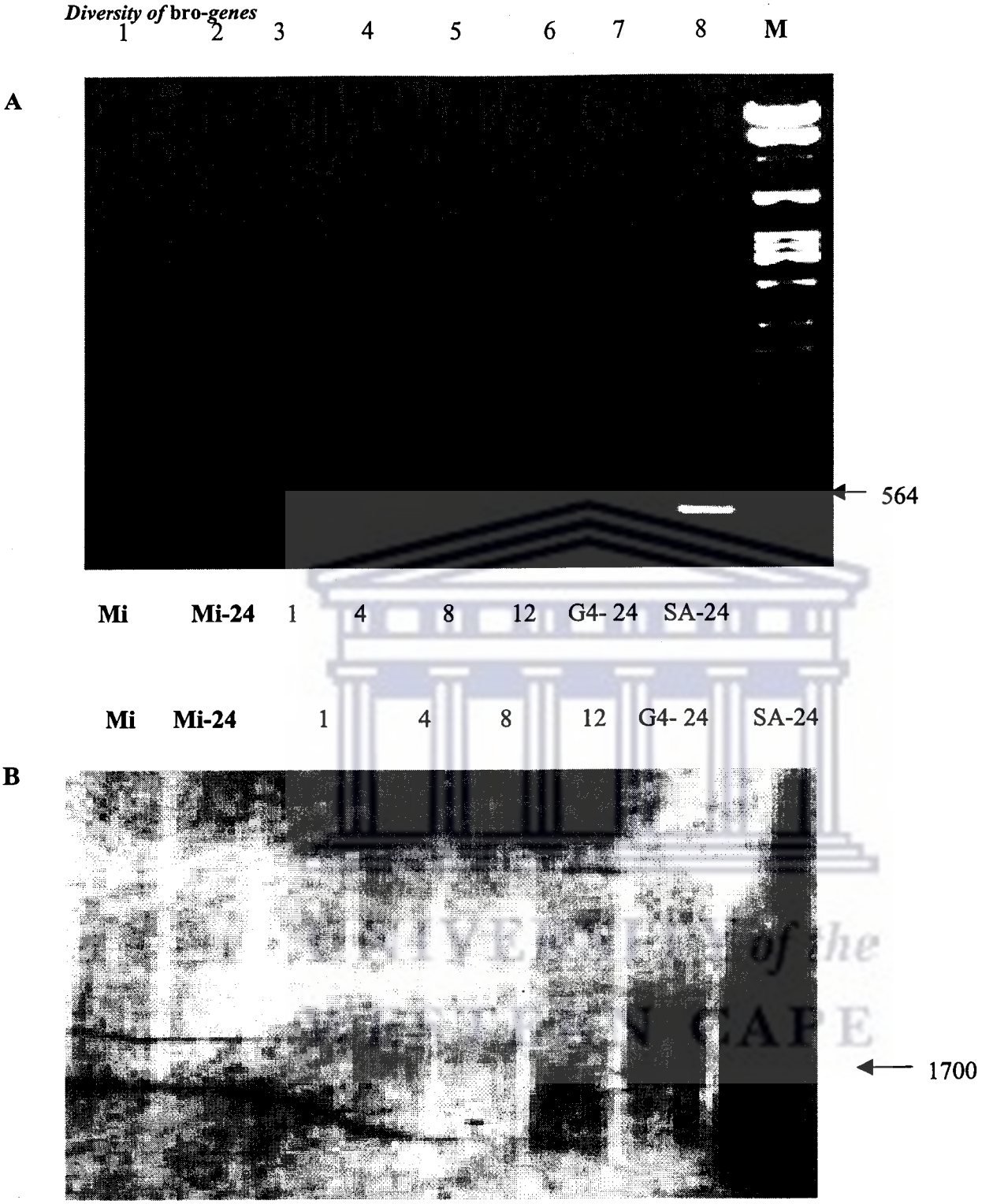
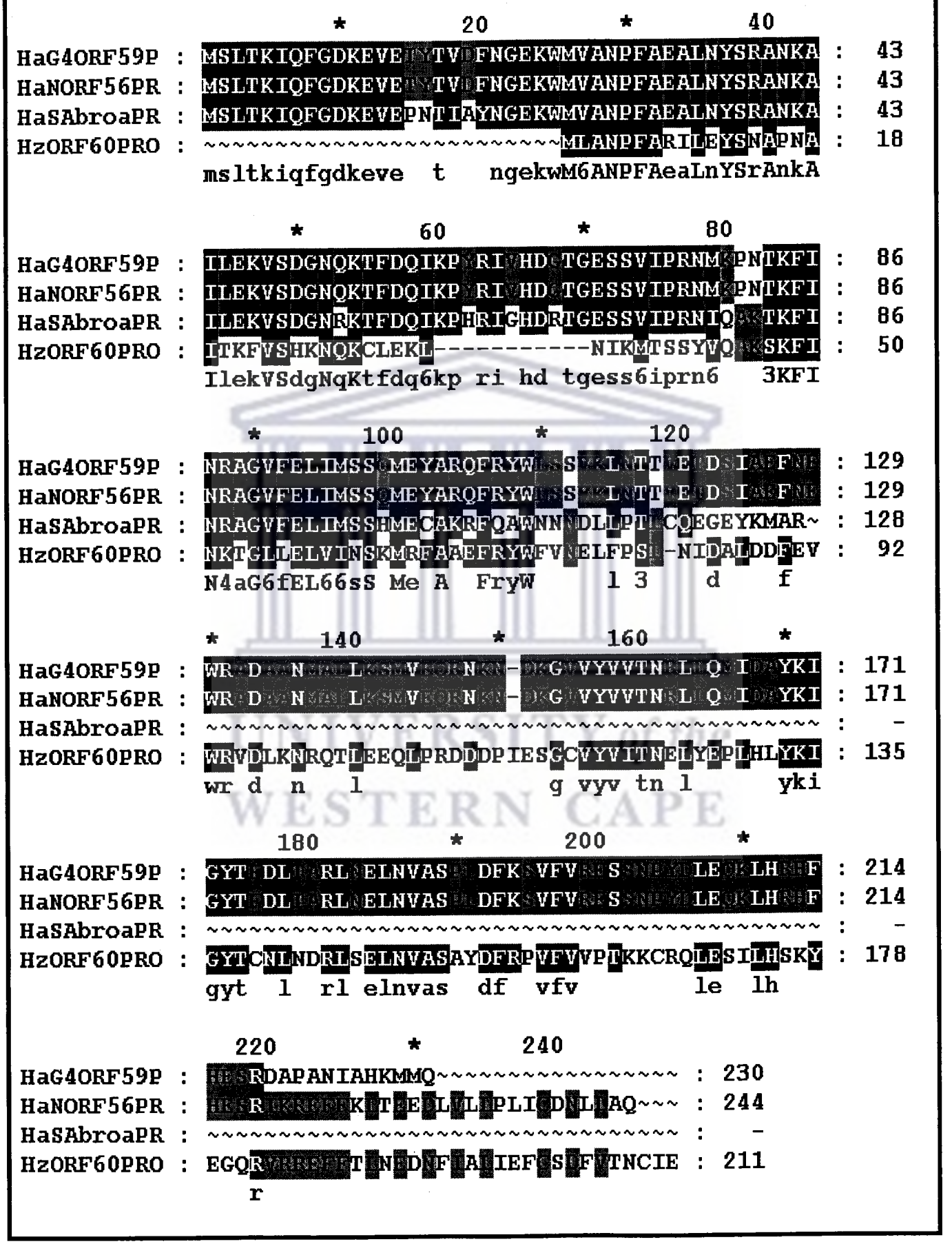


Figure 6. Expression of *bro*-c transcripts of Ha-G4. **A.** RT-PCR analysis of *bro*-c performed on total RNA extracted from Hz-AM1 cells(see text). Lambda DNA digested with *Bam* HI-*Eco* RI-*Hind* III was used as molecular size markers (**M**) with the sizes indicated in base pairs. Times p.i. are indicated below the lanes (**Mi**, mock infected). **B.** Northern analysis of *bro*-c. The approximate size of the hybridisation band is indicated on the right. RT-PCR fragments were used as riboprobes. Times p.i. are indicated above the lanes (**Mi**, mock infected).

A



B

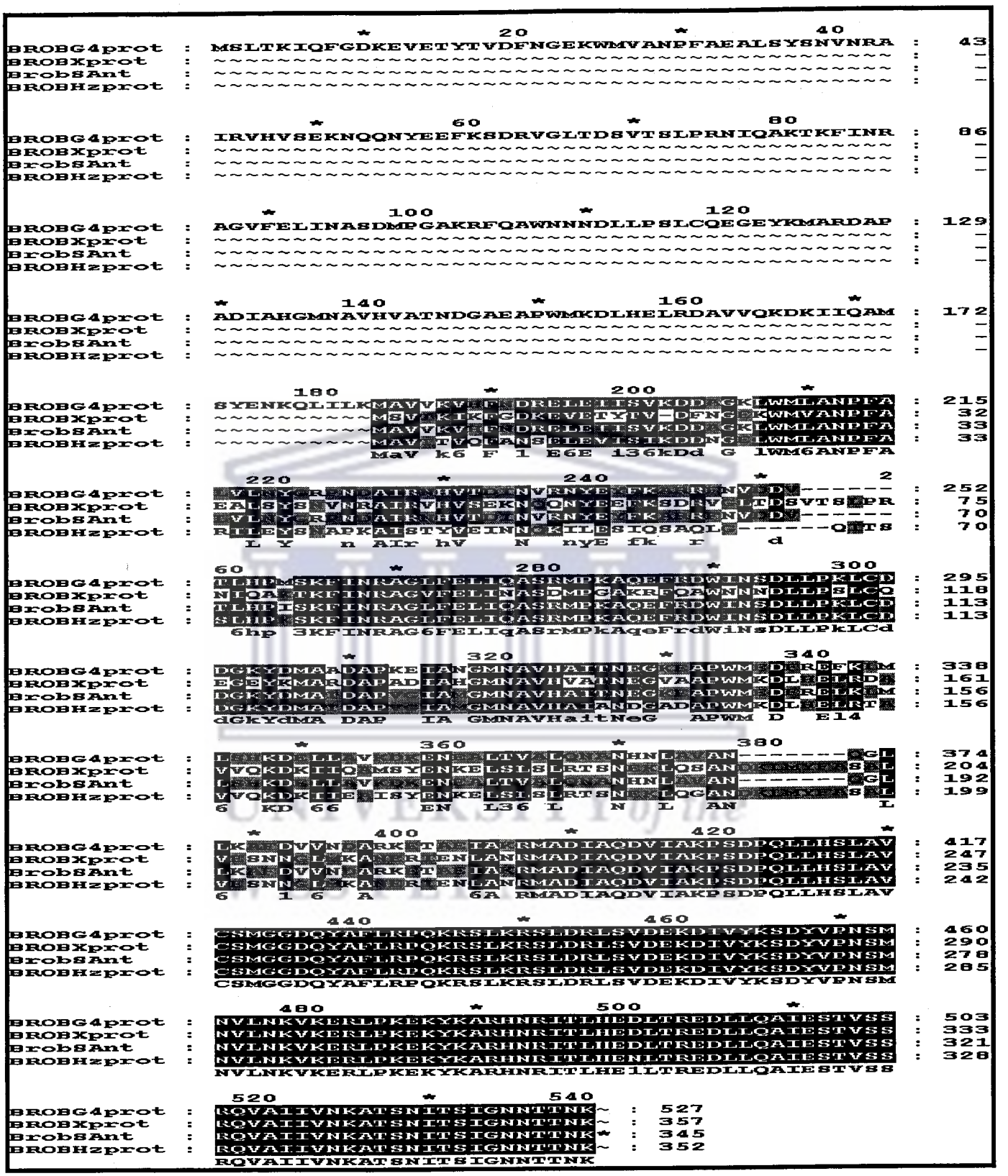


Figure 7. A and B. Alignment of the predicted amino acid sequences of the *bro*-a and -b proteins of all four *Helicoverpa* isolates, respectively. Gaps introduced to optimize the alignment are indicated with dashes. Shading is used to indicate the occurence of identical (black) or substitutional (gray) amino acids.

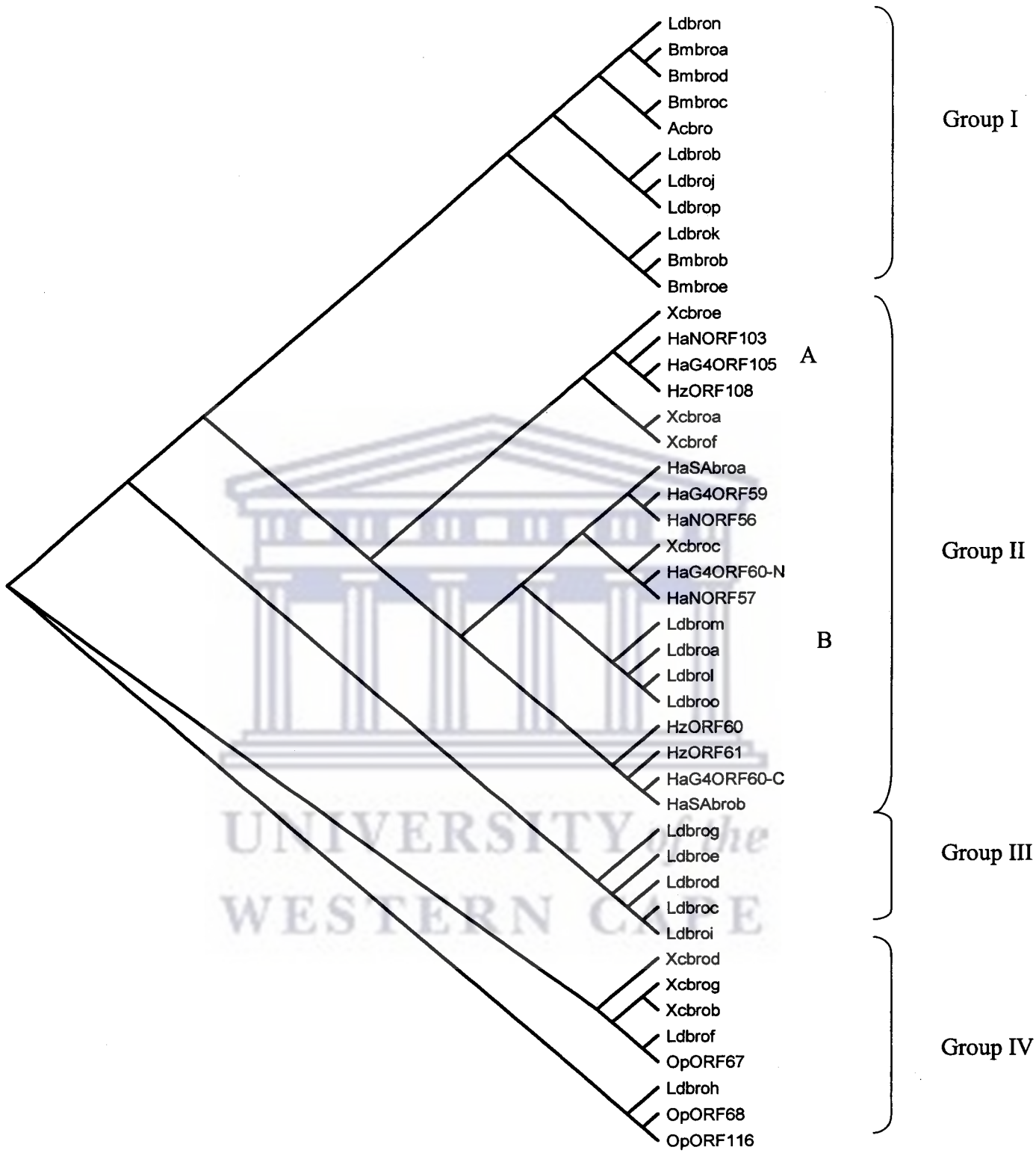


Figure 8. Phylogenetic tree of baculovirus BRO's. The unrooted tree was constructed using the Neighbour program of PHYLIP. Ha-G4 ORF60 was divided into two parts: N- terminal part (Ha-G4ORF-N) and C-terminal part (Ha-G4ORF60-C). The details are described in the text.

CHAPTER FIVE

IDENTIFICATION OF THREE ESSENTIAL BACULOVIRUS GENES OF *HELICOVERPA ARMIGERA* SINGLE-ENVELOPED NUCLEOPOLYHEDROVIRUS AND THEIR USE TO CONFIRM THE PLACEMENT OF HaSNPV-SA IN BACULOVIRUS PHYLOGENY.

Abstract:

Baculovirus phylogeny based on a single gene may not accurately represent baculovirus relatedness. It is therefore of interest to check if various genes could produce trees, which are similar or different. An *Eco*RI genomic library of a single-capsid nucleopolyhedrovirus of *Helicoverpa armigera* was constructed with the plasmid SK pBluescript. Various clones ranging from 14 kilobasepairs (kb) to 0.5-kb were constructed. Three important genes late expression factor (*lef*) 8, *chitinase* and *helicase* were identified during random sequence analysis of two clones (*Eco*RI-H and *Eco*RI-J), which formed part of the *Eco*RI plasmid library of the viral genome. Two of these genes *lef-8* and *helicase,* which have been found to play a role in replication or expression, has been found in all baculovirus genomes sequenced to date. The genome sequence of *Plutella xylostella* granulovirus (PxGV) does not contain a *chitinase* gene. For each of the three genes, amino acid sequences were aligned with CLUSTAL X using default parameters and the Blosum matrix. Single nucleotide polymorphisms (SNPs) within each gene of the *Helicoverpa* SNPV's were tabulated. Trees were constructed using the Neighbour-joining algorithm of CLUSTAL X and was rooted on *Cydia pomonella* GV. The resulting phylogenetic trees were visualised in Treeview. The two data sets of *helicase* and *lef-8* yielded phylogenies supporting the separation of the *Nucleopolyhedrovirus* (NPV) and GV genera and further division of the NPV's into group I and II. The data set of *chitinase* incorrectly grouped the *Helicoverpa chitinase* proteins as part of the group I NPV's. All three trees confirm the placement of HaSNPV's as a group II type NPV.

Introduction:

Baculoviruses are insect pathogenic viruses, which contain a circular, supercoiled double-stranded genome. These genomes have been found to range in size from 90-180- kilobasepairs (kb) with G + C contents of 40-48%, and contain 120 to 181 potential open reading frames (ORF). This class of virus is traditionally classified based on the number of virions occluded within the occlusion body (OB). Baculoviruses are thus divided into two taxonomic genera: *Nucleopolyhedrovirus* (NPV) and *Granulovirus* (GV) (Blissard *et al.*, 2000). The genera NPV contain many occluded virions, while GV's rarely contain more than two. With the aid of molecular phylogeny, the Lepidopteran NPV's have been subdivided into group I and II (Zanotto *et al.*, 1993).

Even though previous baculovirus phylogeny have been based on the OB protein (Zanotto *et al.*, 1993), studies show, that this gene is not the ideal gene for phylogenetic analysis because of its small size and relatively high sequence conservation which provides limited phylogenetic information. Also, phylogenies based on this protein are often in conflict with other gene phylogenies (Clarke *et al.*, 1996 and Kang *et al.*, 1998). As times passes, the availability of complete genome sequence data allows for the comparison of combination genes, which may reduce sampling errors in phylogenetic analysis.

Random sequence analysis of two fragments which were present in an *Eco*RI genomic library we constructed of a South African isolate of *Helicoverpa armigera* single-capsid nucleopolyhedrovirus (HaSNPV-SA) yielded three baculovirus genes [late

expression factor (*lef*)-8; *helicase* and *chitinase*]. All three genes are important baculoviral genes if present on a genome. Studies of putative baculoviral *helicases*, suggest that these enzymes are one of the key determinants of host range (Heldens *et al.*, 1997). The entire *helicase* ORF was present in a 4584-kb *Eco*RI fragment of HaSNPV-SA. The *lef-8* gene has been targeted as a possible candidate for a viral gene involved in late expression. This gene has been shown to be necessary for the efficient expression of late and very late gene promoters in transient gene expression assays (Passarelli *et al.*, 1994). The *chitinase* gene is expressed in the late stage of virus replication and its product has both endo- and exo-chitinase activity (Hawtin *et al.*, 1995). The latter two genes (*lef-8* and *chitinase*) were found on a 5610-kb fragment of HaSNPV-SA. This chapter contains a phylogenetic study based on data sets of three single genes. For the *helicase* and *lef-8* ORF's, fourteen baculoviral genes were compared, while for the *chitinase* gene only thirteen ORF's were used, as *Plutella xylostella* GV (PxGV) does not contain a *chitinase* gene. Amino acid (aa) sequences were aligned with CLUSTAL X using default parameters and the Blosum matrix. Single nucleotide polymorphisms (SNPs) showing DNA sequence variations among the individual isolates for each gene of the *Helicoverpa* SNPV's were tabulated. Trees were constructed using the Neighbour-joining algorithm of CLUSTAL X and was rooted on *Cydia pomonella* GV. The resulting phylogenetic trees were visualised in Treeview.

Materials and Methods:

Virus and DNA isolation

The HaSNPV-SA was isolated from a colony of diseased *H. armigera* larvae in the Western Cape region of South Africa (Khan *et al.*, 2000). The occlusion bodies were purified by continuous and discontinuous sucrose density gradient centrifugation. Viral DNA was extracted using the method prescribed by Crook *et al.*, (1985). Viral DNA was digested with various restriction enzymes according to the specifications of the manufacturer.

Cloning and sequencing

HaSNPV-SA genomic DNA digested with *Eco*RI was shotgun-cloned into the plasmid vector pBluescript SK^+ by using standard methods (Sambrook *et al.*, 1989). A 4584- and a 5610-kb *Eco*RI fragment of HaSNPV-SA were selected for sequencing. Automated sequencing of the end-termini of revealed the presence of the *lef-8* and *chitinase* genes on the 5610-kb fragment and the *helicase* ORF on the 4584-kb fragment. Subsequent subclones were constructed and sequenced using the M13 universal set of primers with a Pharmacia ALF/Express automated sequencer. The nucleotide sequence of the entire fragments together with the complete open reading frames has been deposited in GenBank and were assigned the following accession numbers: 4584-kb fragment- AY118079 and 5610-kb fragment- AY118080.

Phylogenetic analysis

Sequence analysis was performed using the GCG computer program. Conceptual amino acid sequences were compared with the sequence homologues at

GenBank/EMBL using the Advanced Blast Search Server (Altshul *et al.*, 1997). Sequence alignments were done by CLUSTAL X (1.81) (Thompson *et al.*, 1997) using default parameters and the Blosum matrix. GENEDOC software (Nicholas and Nicholas, 1997) was used for homology shading of the aligned amino acid sequence. The baculoviruses used for comparative analysis and phylogeny studies were (GenBank accession numbers): HaSNPV-SA (AY118079 and AY118080); HaSNPV-G4 (Chen *et al.*, 2001) (NC002654); HzSNPV-ELCAR (Chen *et al.*, 2002) (NC003349); *Mamestra configurata* NPV (MacoNPV) (NC003529) (Erlandson *et al.*, 2002); *Spodoptera exigua* NPV (SeMNPV) (NC002169) (Ijkel *et al.*, 1999); *Epiphyas postvittana* NPV (EppoNPV) (NC003083) (Hyink *et al.*, 2001); *Cydia pomonella* GV (CpGV) (NC002816) (Luque *et al.*, 2001); *Bombyx mori* NPV (BmMNPV) (NC001962) (Park, 2001); *Spodoptera litura* NPV (SlMNPV) (NC003102) (Pang *et al.*, 2001); *Plutella xylostella* GV (PxGV) (NC002593) (Hashimoto *et al.*, 2000); *Lymantria dispar* NPV (LdMNPV) (NC001973) (Kuzio *et al.*, 1999); *Autographa californica* NPV (AcMNPV) (NC001623) (Ayres *et al.*, 1994); *Orgyia pseudotsugata* NPV (OpMNPV) (NC001875) (Ahrens *et al.*, 1997) and *Xestia c-nigrum* GV (XcGV) (NC002331) (Hayakawa *et al.*, 1999). Only 13 of the above were used for the analysis of the *chitinase* ORF, since PxGV does not contain this gene. Trees were constructed using the Neighbour-joining algorithm of CLUSTAL X and was rooted on *Cydia pomonella* GV. The resulting phylogenetic trees were visualised in Treeview (Page, 1996).

Single nucleotide polymorphisms

SNP's were determined by aligning the three *Helicoverpa* SNPV *chitinase, helicase* and *lef-8* nucleotide sequences with CLUSTAL X (1.81) (Thompson *et al.*, 1997)

using default parameters and the Blosum matrix. GENEDOC software (Nicholas and Nicholas, 1997) was used for homology shading of the aligned nucleotide sequence. SPN's were tabulated (Table 1, 2 and 3, respectively).

Results:

Sequence analysis of the genes and their predicted proteins

chitinase

Sequence data showed that the HaSNPV-SA *chitinase* gene had an open reading frame (ORF) of 1728-nt encoding a predicted protein of 576 aa with a M_r of 64.116-kDa. The 5' noncoding region of the HaSNPV-SA *chitinase* gene, as seen with the other *Helicoverpa's*, contained various control elements late transcription. In HaSNPV-SA a possible TATA box was located 33-nt upstream of the translation start codon. A TAAG motif characteristic for baculovirus late transcription was found at nt position –29. No potential polyadenylation signal, AATAAA, was found downstream of the TAA translation termination codon.

helicase

Sequence data showed that the HaSNPV-SA *helicase* gene had an open reading frame (ORF) of 3762 nucleotides (nt) encoding a predicted protein of 1254 aa with a M_r of 145.962-kDa. The 5' noncoding region of the HaSNPV-SA *helicase* gene, as seen with the other *Helicoverpa's*, contained various control elements for early and late transcription. In HaSNPV-SA a possible TATA box was located 112-nt upstream of the translation start codon. The putative enhancer element, GATA, was observed at positions, -50. A TAAG motif characteristic for baculovirus late transcription was

found at nt position –48. A potential polyadenylation signal, AATAAA, was found 11-nt downstream of the TAA translation termination codon.

lef-8

Sequence data showed that the HaSNPV-SA *lef-8* gene had an open reading frame (ORF) of 2706-nt encoding a predicted protein of 902 aa with a M_r of 105.099-kDa. A search for regulatory transcription motifs in the upstream sequences of the new gene did not reveal the usual TATA box, or TAAG motifs. Like the AcMNPV *lef-8* gene, a transcription initiator motif, CAGT was observed at 89-nt with respect to the ATG translation start of HaSNPV-SA. The HaSNPV-SA also contains the 12-residue motif $GXKX_4HGQ/NKG$ conserved in DNA-dependant RNA polymerases. No potential polyadenylation signal, AATAAA, was found downstream of the TAA translation termination codon.

Comparative analyses of the three predicted proteins

CHITINASE

The HaSNPV-SA ORF exhibits a 61% aa identity with the AcMNPV *chitinase* gene product, but is slightly more related to BmMNPV and OpMNPV with aa identities of 62% to each of these CHITINASE's (Fig. 1). HaSNPV-SA was 98% identical to HaSNPV-G4 and 99% identical to HzSNPV. A number of amino acids were shown to be absolutely conserved in all the baculovirus CHITINASE's. This ORF seems to be highly conserved for the GV's had a percentage identity of greater than 45% to the NPV genera. When the nucleotide sequence of the *chitinase* gene of all three *Helicoverpa* SNPV's used in this study were compared, 11 SNP's were shown to

occur throughout (Table 1). A 15-nt insertion positioned from nt 33 uptil and including position 51 in the N-terminal end of both HzSNPV and HaSNPV-SA was also found (Table 1).

HELICASE

The HaSNPV-SA ORF exhibits a 37% aa identity with the AcMNPV *helicase* gene product, but is more closely related to MacoNPV and SlMNPV with aa identities of ≥ 45% to each of these HELICASE's (Fig. 2). All three *Helicoverpa's* were 99% identical. This ORF contains the consensus NTP-binding motif (Fig. 2, Motif I) with the consensus sequence G/Axx(G)xGKS/T (Gorbalenya *et al.*, 1989). This sequence has a hydrophobic region preceding it. In addition to this, five additional motifs have been identified in a large number of replicative proteins, some of which have ATP-dependant *helicase* activity (II-VI). Another motif (Fig. 2, Motif Ia) is a one, which is commonly identified within DNA genomes, and has a predicted ß sheet-turn-a helix secondary structure surrounding an invariant threonine residue (Hodgeman, 1988). Within the baculovirus NPV's and GV's used for this analysis, only one of these seven motifs had a high homology when compared with the consensus sequence (Fig. 3, Motif I). The other six motifs were not conserved at all when compared with the consensus sequence. When the nucleotide sequence of the *helicase* gene of all three *Helicoverpa* SNPV's used in this study were compared, 21 SNP's were shown to occur throughout (Table 2).

LEF-8

The HaSNPV-SA ORF exhibits a 60% aa identity with the AcMNPV *lef-8* gene product, but seem to be more closely related to MacoNPV and SeMNPV with aa

identities of 66% and 67% to each of these LEF-8's respectively (Fig. 3). All three *Helicoverpa's* were 99% identical. A number of amino acids were shown to be absolutely conserved in all the baculovirus LEF-8's. This ORF seems to be fairly conserved for the GV's had a percentage identity of greater than 40% to the NPV genera. The conserved motif found in all DNA dependant RNA polymerases had percentage identities ranging from 61%-100% while percentage similarities between all the baculoviruses used ranged from between 91% to 100%. When the nucleotide sequence of the *lef-8* gene of all three *Helicoverpa* SNPV's used in this study were compared, 26 SNP's were shown to occur throughout (Table 1).

Phylogenetic analysis

The phylogeny of the *chitinase, helicase* and *lef-8* genes of all 13 baculovirus genomes (if they contained the particular gene) present in GenBank as well as the HaSNPV-SA, were analysed using the Neighbour-joining algorithm of CLUSTAL X (Thompson *et al.*, 1997). Trees using CpGV as an outgroup are shown in Fig. 4 (*chitinase*), Fig. 5 (*helicase*) and Fig. 6 (*lef-8*). The trees based on the baculovirus HELICASE and LEF-8 proteins confirm the organisation shown by previous studies (Herniou *et al.*, 2001). AcMNPV, BmMNPV, OpMNPV and EppoNPV belong to one clade forming the group I type NPV. SeMNPV, MacoNPV, LdMNPV and all three *Helicoverpa* NPV's form another clade, which encompass the group II type NPV. The 3 GV's of PxGV, CpGV, and XcGV all formed one group. There is however one major difference between these two trees based on the position of SlMNPV. In Fig. 5 SlMNPV groups with the group II type NPV's, whereas in Fig 4 and 6 this virus does not group in either division but rather branches from the clades forming group I and II. The tree based on the data sequence of the CHITINASE

proteins is not in agreement with the former two trees, although there are similarities. AcMNPV, BmMNPV, OpMNPV and EppoNPV once again belong to one clade forming the group I type NPV however, the three *Helicoverpa* NPV's are now rooted from this clade. SeMNPV, MacoNPV and LdMNPV form another clade which encompass the group II type NPV. The 2 GV's of CpGV, and XcGV formed one group.

Discussion and Conclusion:

AcMNPV *chiA* is expressed as a late gene and we expect that this HaSNPV-SA homologue is also a late gene. This hypothesis is supported by the TAAG motif characteristic for baculovirus late transcription, which was found at nt position –29 upstream from the start signal. The HaSNPV-SA also shows high homology with the *chiA* genes from other baculoviral viruses. A region with the highest conservation seems to be in the central part of the protein and extends towards the C-terminus. The conserved central region encompasses the putative active site of *chitinase* (amino acids 287-331 for HaSNPV-SA; Hawtin *et al.*, 1995) and the putative catalytic residue (Glu^{316}), which is also conserved in HaSNPV-SA (Fig. 1).

Studies of putative NPV DNA *helicases* have shown that they are involved in replication and host range determination. For example, changes in a few amino acids or even in only one, in the AcMNPV *helicase* (*p143*) enable AcMNPV to replicate in cells and larvae of BmMNPV, a host which does not support AcMNPV replication normally (Maeda *et al.*, 1993). Analysis of *helicase* gene of AcMNPV revealed a potential leucine zipper motif, a putative nuclear localization signal, and seven amino

acid motifs (I, Ia, II-VI) previously identified in a number of proteins involved in NTP binding and DNA/RNA unwinding (Lu and Carstens, 1991). Within our study, we aligned the *helicase* proteins of 14 NPV's and GV's (Fig. 2). Of these *helicase* proteins, the only other baculovirus with a leucine zipper motif was BmMNPV. Only EppoNPV and BmMNPV had the putative nuclear localization signal with the consensus core signal of KxxK/R. The absence of these domains may indicate that the proteins lacking them may not be multifunctional as suggested for AcMNPV. Of the seven conserved motifs suggested by Lu and Carstens (1991), only Motif I was highly conserved within all 14 baculoviruses. The other six motifs had a fairly low homology with the consensus sequence. It is important to note, that most researchers have shown that the invariant amino acid residues in the NTP binding sequence of Motif I is rather conserved.

Unlike the *Spodoptera littoralis* NPV (*S. littoralis* NPV) *lef-8* gene, but like the AcMNPV *lef-8* gene, a transcription initiator motif, CAGT was observed in HaSNPV-SA. No regulatory elements such as a TATA box, or TAAG motif was found. This is not uncommon though, for neither is these motifs found in *S. littoralis* NPV (Faktor and Kamensky, 1997). Further experiments are needed to characterise the pattern and regulatory mechanism of *lef-8* gene expression. The HaSNPV-SA also contains the 12-residue motif $GXKX_4HGQ/NKG$ conserved in DNA-dependant RNA polymerases. This conserved motif found was found near the C-terminus and had percentage identities ranging from 61%-100% while percentage similarities between all the baculoviruses used ranged from between 91% to 100% (Fig. 3).

SNP analysis showed that single nucleotide differences occurred throughout the 3 genes of the *Helicoverpa* SNPV isolates (Table 1, 2 and 3). There is no evidence that these polymorphisms in any way hinder the function of genes, since the propagation of all 3 *Helicoverpa* SNPV's appear to be stable. All three genes used to construct the trees showed the separation of the NPV's and GV's and the further subdivision of the NPV's into groups I and II. However, the tree generated using the *chitinase* gene sequences does not group the *Helicoverpa* SNPV's in the accepted "correct" position (Fig. 4). Within this tree, the three *Helicoverpa* SNPV's are grouped as a group I type NPV. Various studies show (Chen *et al.*, 1997 and Herniou *et al.*, 2001), that these NPV's are in fact a group II type NPV. As in the case of the *polyhedrin* gene, the *chitinase* gene is relatively small and has fairly high sequence conservation. The conflict in placement could also be due to erroneous phylogenetic inferences caused by unequal rates of evolution or the lack of an unambiguous phylogenetic signal in the sequences. It is also of interest to note that CpGV does not contain a *chitinase* gene. Furthermore, the baculoviral *chitinase* gene has a remarkable homology with that of the *Serratia marcescens* chitinase A (Hawtin et al., 1995). If the *chitinase* gene did originate from *S. marcescens*, this could possibly explain the anomalous behaviour of this gene.

The two trees based on the data sets of *helicase* (Fig. 5) and *lef-8* (Fig. 6) however, were in agreement with the groupings as predicted by Herniou *et al.*, 2001. Their phylogenies were generated based on three independent character sets: the individual sequences of genes shared by viruses; gene order, and gene content. Their study included the entire genome sequence of 9 baculoviruses (AcMNPV, BmMNPV, OpMNPV, HaSNPV-G4, SeMNPV, LdMNPV, CpGV, PxGV and XcGV) whose data

was available as that time. With time however, 4 more complete genome data have now become available (HzSNPV, SlMNPV, MacoNPV and EppoNPV). A comparison of the two trees generated based on the *helicase* (Fig. 5) and *lef-8* (Fig. 6) genes are in agreement except for the placement of SlMNPV. Fig. 5 places SlMNPV as a group II type NPV, while Fig. 6 does not place this NPV in any of the two groupings for the NPV genera. This genome sequence is not included in the comprehensive study compiled by Herniou *et al.*, (2001). *lef-8* however is one of a few genes that yielded the best trees within their study. Furthermore, phylogeny based on the *egt* gene (Chapter three, this thesis) strongly places SlMNPV as a group II type NPV. All three trees confirm the placement of HaSNPV's as a group II type NPV.

The above study supports the theory that there is danger in using phylogeny based on one or a small number of genes to infer evolution. As suggested by Herniou *et al.*, (2001), we second the recommendation, that the baculovirus phylogenies should be based on a combined analysis of all genes conserved among baculoviruses. Combining genes may reduce sampling errors in phylogenetic analysis and allow phylogenies to converge towards correct solutions with good support.

Acknowledgements:

The National Research Foundation supported the research. SK received grants from the National Research Foundation and the Solm Yach Fellowship.

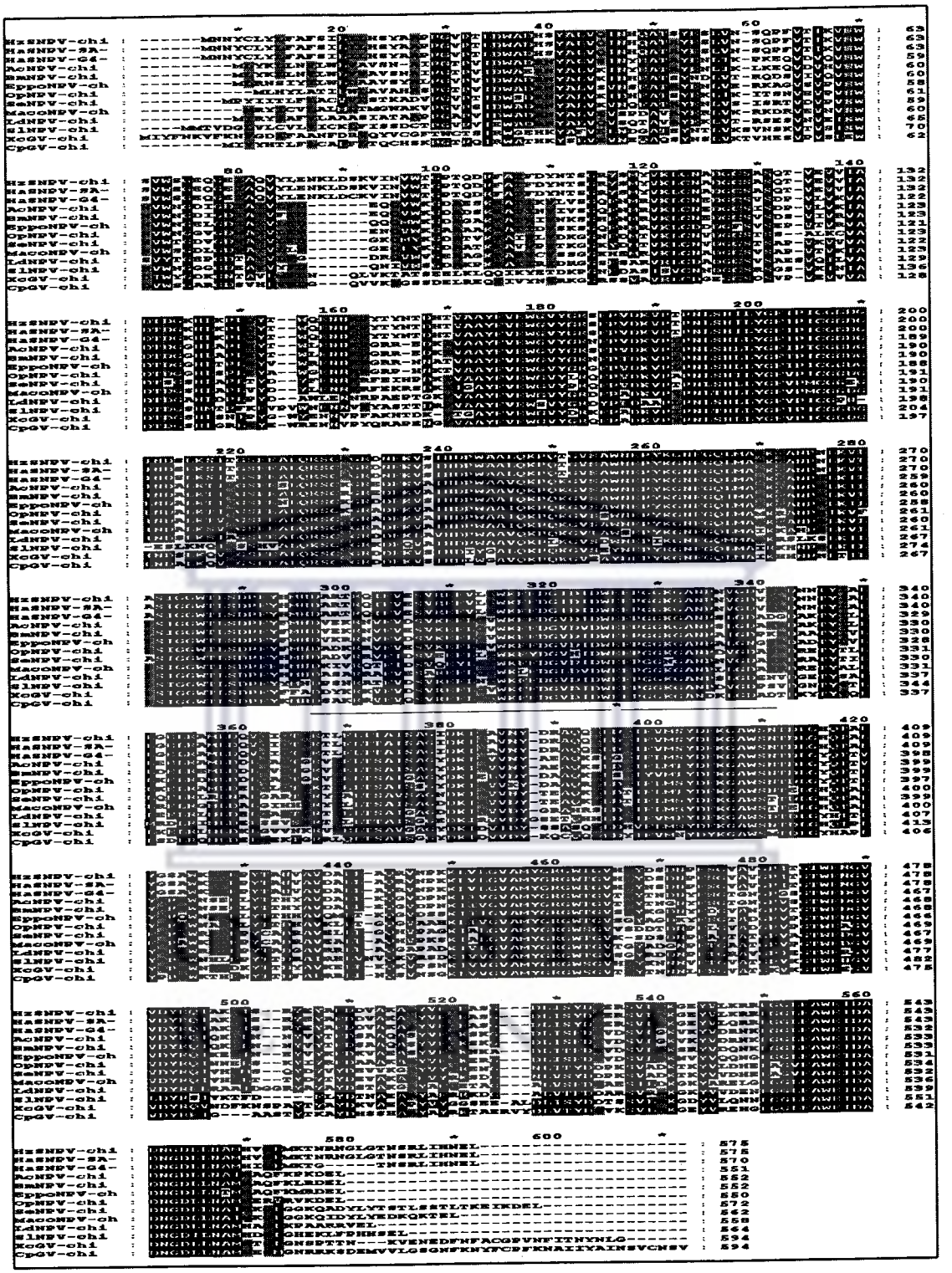


Figure 1. Alignment of deduced amino acid sequences of putative *chitinase* genes from 13 baculoviruses was done using CLUSTAL X. Dashes were introduced to align the sequences. Black shaded areas indicate 100 % homology; deep grey 80% or more; light grey 60% or more. The putative active site of chitinase is underlined and the putative catalytic residue is indicated by an asterisk (*).

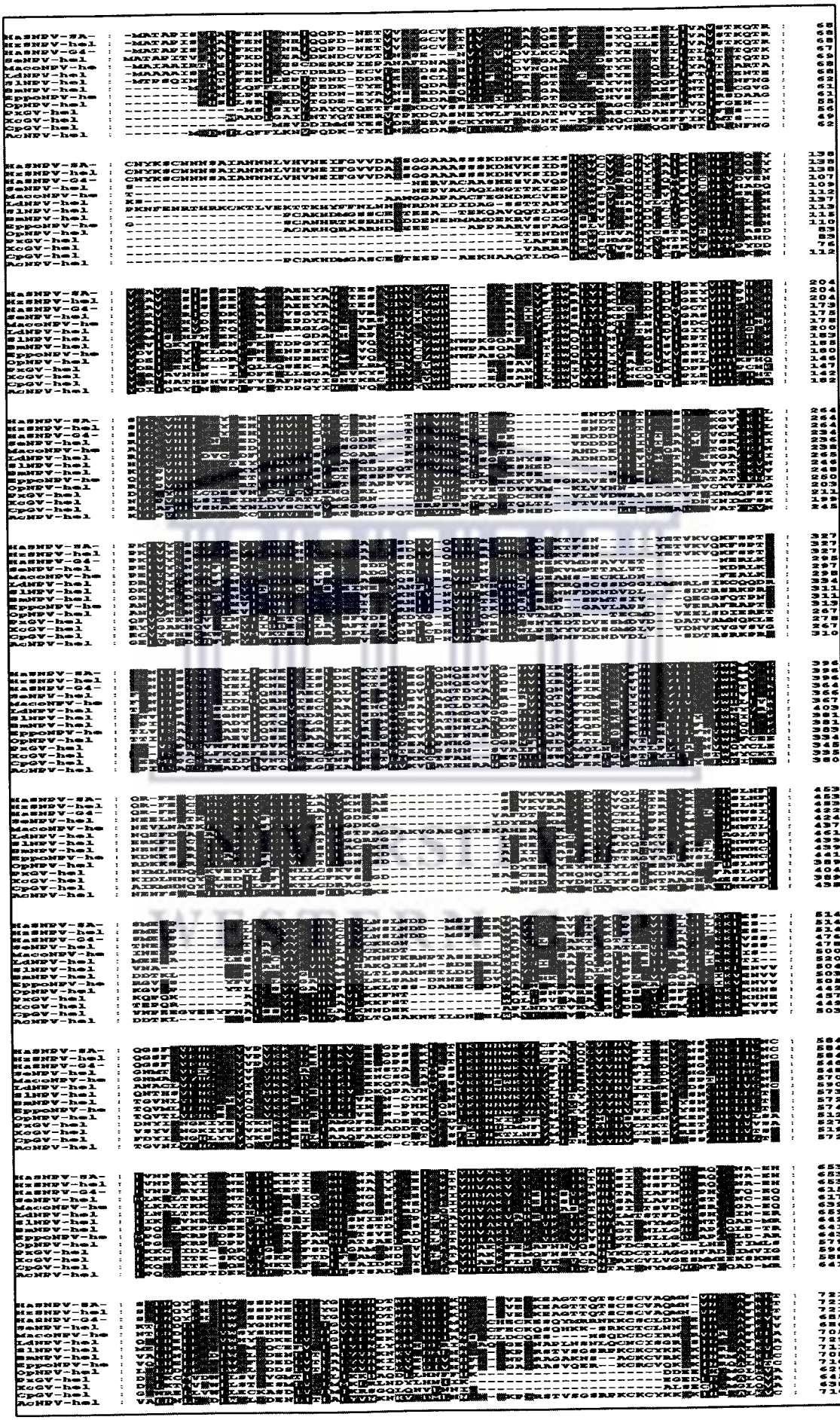


Fig. 2 continued

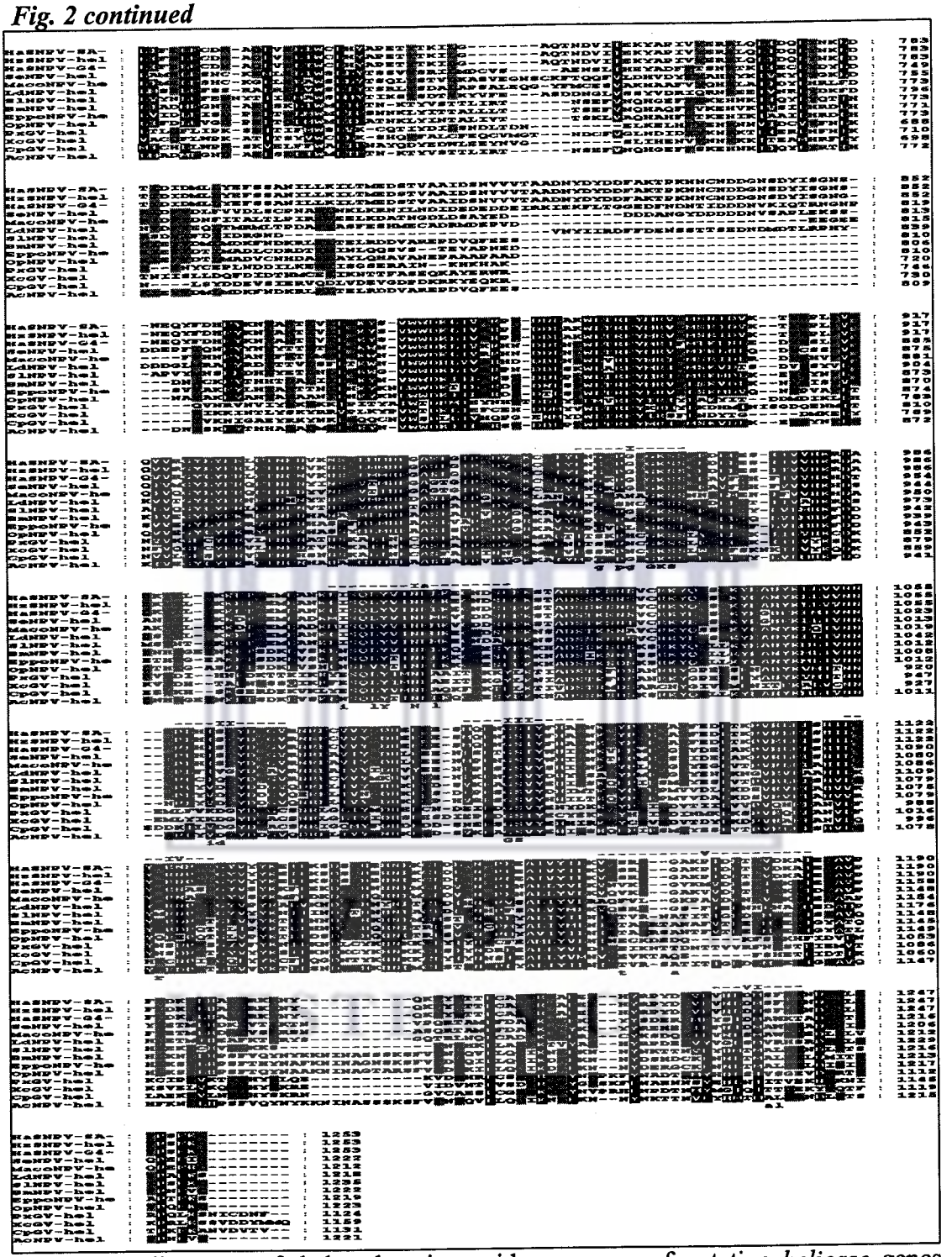


Figure 2. Alignment of deduced amino acid sequences of putative *helicase* genes from 14 baculoviruses was done using CLUSTAL X. Dashes were introduced to align the sequences. Black shaded areas indicate 100 % homology; deep grey 80% or more; light grey 60% or more. Conserved motifs are indicated by a dashed line above the sequence (----). Conserved amino acids shared among most members of the group are showed along the bottom.

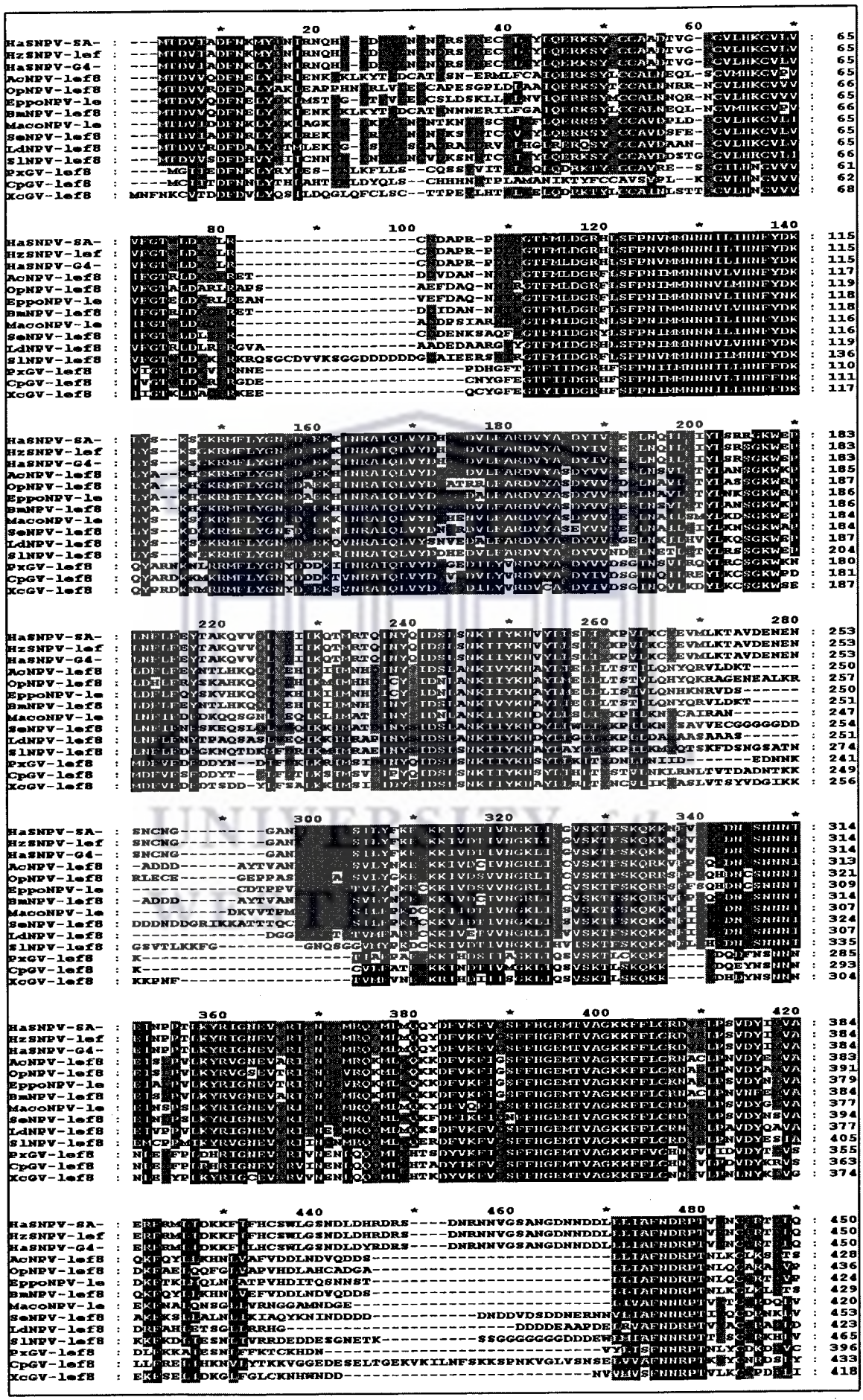


Fig. 3 continued

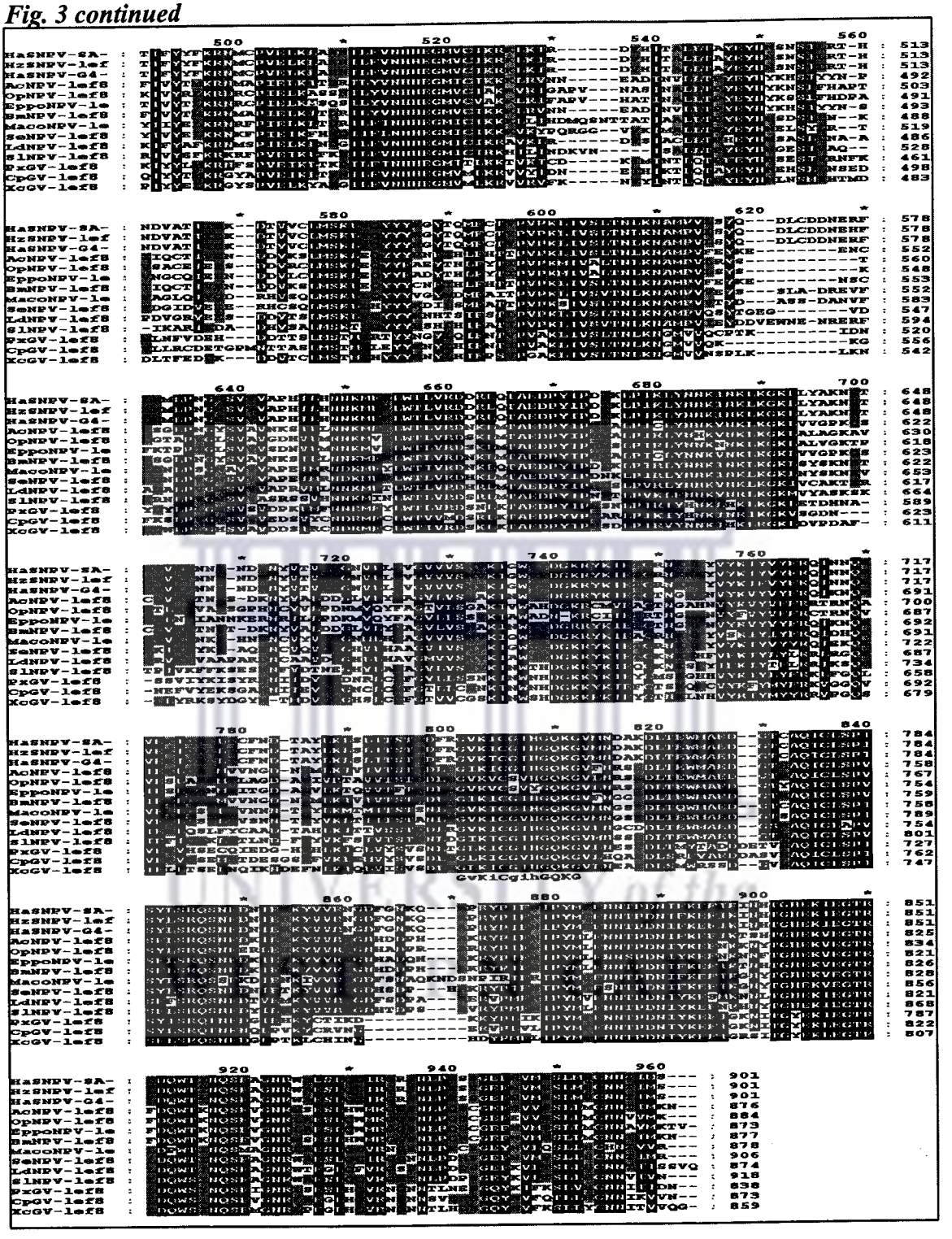


Figure 3. Alignment of deduced amino acid sequences of putative *lef-8* genes from 14 baculoviruses was done using CLUSTAL X. Dashes were introduced to align the sequences. Black shaded areas indicate 100 % homology; deep grey 80% or more; light grey 60% or more. A consensus line is indicated in the region of the RNA polymerase consensus motif $GXKX_4HGQ/NKG$.

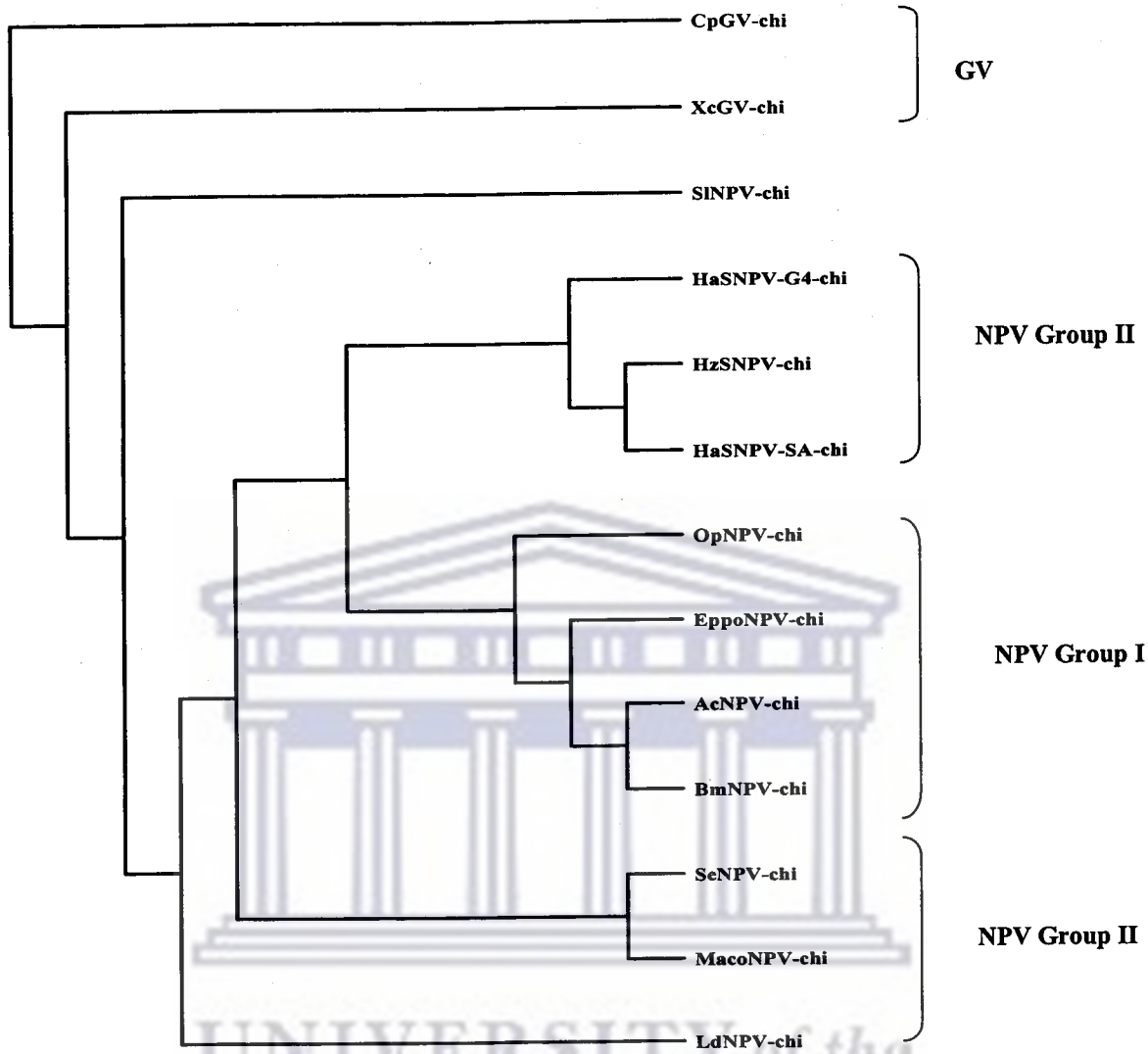


Figure 4. A phylogenetic tree of *chitinase* genes. A tree for 13 baculovirus CHITINASE's was constructed by using the Neighbour-joining algorithm of Clustal X. CpGV was chosen as an outgroup.

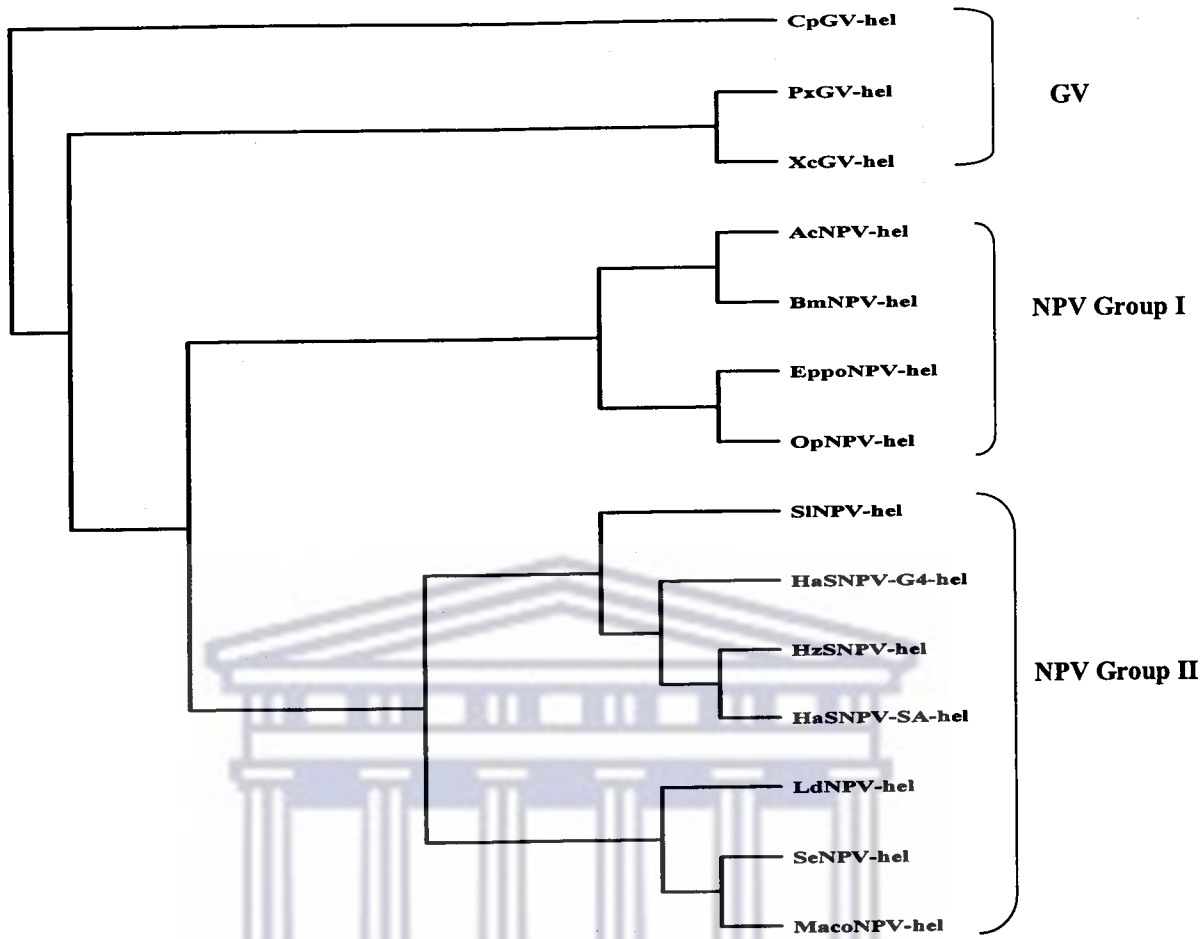


Figure 5. A phylogenetic tree of *helicase* genes. A tree for 14 baculovirus HELICASE's was constructed by using the Neighbour-joining algorithm of Clustal X. CpGV was chosen as an outgroup.

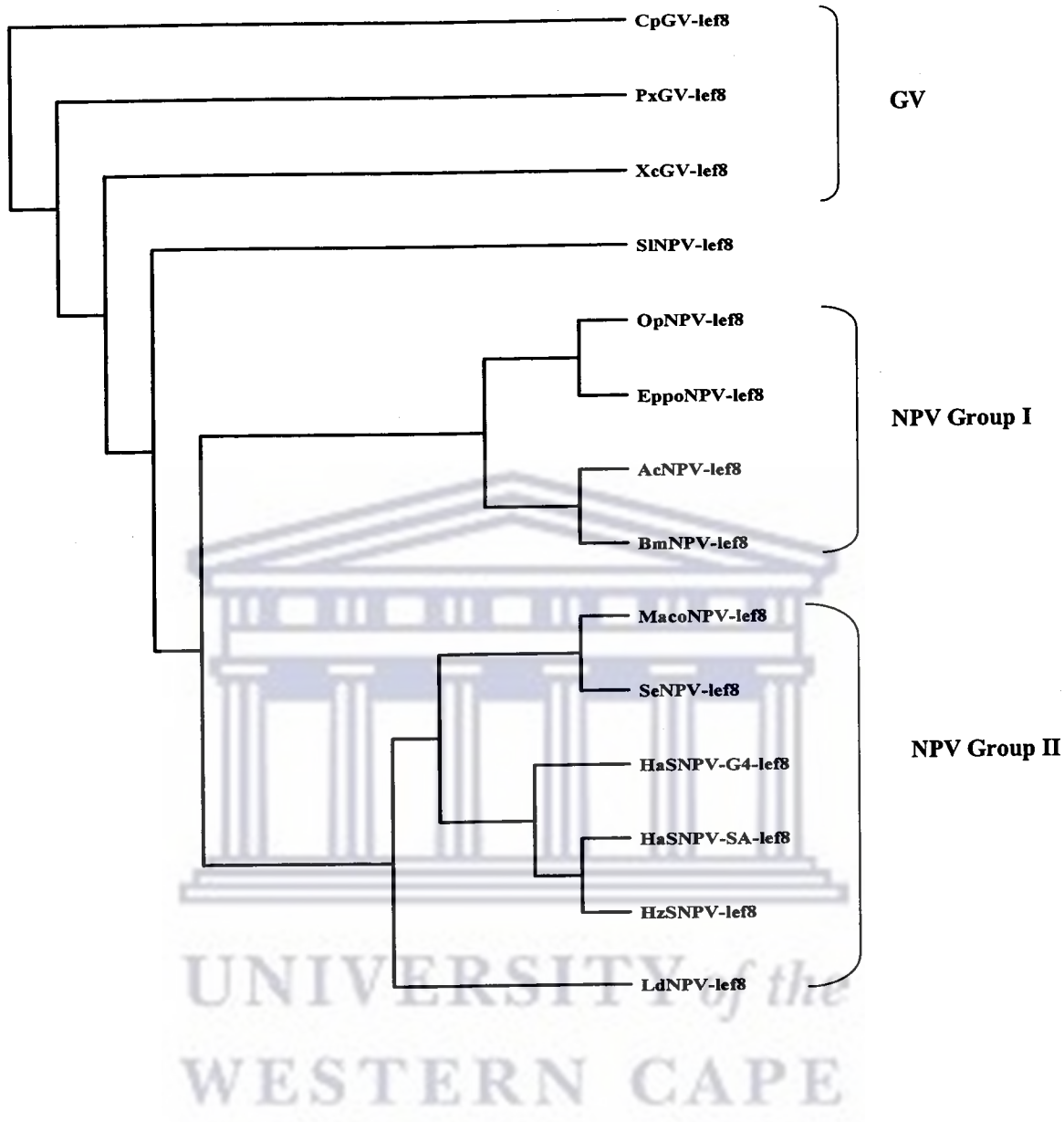


Figure 6. A phylogenetic tree of *lef-8* genes. A tree for 14 baculovirus LEF-8's was constructed by using the Neighbour-joining algorithm of Clustal X. CpGV was chosen as an outgroup.

Table 1. Tabulation of SNP variations within the nucleotide sequence of the *chitinase* gene

SNP position[a]	Ha-G4				Hz-ELCAR				Ha-SA			
	A	C	G	T	A	C	G	T	A	C	G	T
1				+		+						+
38-51	AGGCCGTTGCGATTT insertion for Hz-ELCAR and HA-SA											
69				+		+				+	+	
141			+					+				+
184	+						+				+	
360				+		+				+		
364			+		+				+			
373			+					+			+	
1393	+						+				+	
1456	+						+				+	
1479	+							+				+
1593		+				+				+		

[a]Nucleotide position as from the start codon

Table 2. Tabulation of SNP variations within the nucleotide sequence of the *helicase* gene

SNP position[a]	Ha-G4				Hz-ELCAR				Ha-SA			
	A	C	G	T	A	C	G	T	A	C	G	T
490	+						+				+	
599				+		+						+
655				+			+				+	
721				+			+				+	
1006	+						+				+	
1348				+		+				+		
1353	+						+				+	
1486	+							+				+
2065	+						+				+	
2374			+					+			+	
2488				+		+				+		
2497	+						+				+	
2660				+		+				+		
2669		+						+		+		
2670		+						+				+
2743	+						+				+	
2794	+						+				+	
2917	+						+		+			
2995		+					+				+	
3724				+		+				+		
3726				+		+				+		

[a]Nucleotide position as from the start codon

Table 3. Tabulation of SNP variations within the nucleotide sequence of the *lef-8* gene

SNP position[a]	Ha-G4				Hz-ELCAR				Ha-SA			
	A	C	G	T	A	C	G	T	A	C	G	T
61		+						+		+		
343				+	+							+
587	+							+				+
589			+		+						+	
688		+						+				+
709	+						+		+			
754				+		+				+		
957		+						+		+		
1009				+		+				+		
1066	+						+		+			
1090				+		+				+		
1234	+						+		+			
1402	+						+				+	
1516		+			+				+			
1798			+		+						+	
2014				+		+				+		
2029				+		+				+		
2062	+						+				+	
2110				+		+				+		
2146				+	+							+
2173			+				+			+		
2239	+							+				+
2341		+						+				+
2356	+						+				+	
2533	+							+				+
2676	+							+				+

[a]Nucleotide position as from the start codon

CHAPTER SIX

SUMMARY AND FUTURE STUDIES

SUMMARY

The research presented in this thesis centered on the molecular characterization of HaSNPV-SA. To acquire this information, a pure virus stock of HaSNPV-SA was needed. When the virus was first isolated, two baculoviruses were found co-infecting a colony of *Helicoverpa armigera*. This was no uncommon occurrence, since these two viruses are usually found co-infecting particular hosts in nature. Numerous attempts to isolate a pure stock of HaSNPV-SA without the HaGV *in vivo*, failed. Various factors influenced this; latent Ha-GV virus infection within the *H. armigera* colony, inability of sucrose gradients to allow for a pure viral stock without trace amounts of HaGV, and due to the smaller size of GV's they have been found to replicate more rapidly within its host than the larger NPV. To obtain a pure inoculum of HaSNPV-SA the virus was propagated *in vitro*. At present there are only a few successful cases of *in vitro* replication of HaGV (Miltenburger *et al.*, 1984; Naser *et al.*, 1984; Granados *et al.*, 1986; Dwyer *et al.*, 1988; Winstanley and Crook, 1993). The cells used for *in vitro* propagation should be very specific, since GV infect particular cells (Huger, 1963) and researchers have experienced difficulty in selecting stable permissive cell lines for GV. To thus eradicate HaGV from a virus stock containing both HaGV and HaSNPV-SA, *in vitro* replication is a possible solution, since GV replication *in vitro* is so sensitive and the chances of infecting cells *in vitro* with GV by default is slim. Within our study, we were able to successfully propagate HaSNPV-SA within permissible cell-lines and obtain pure virus stocks.

When the genomic sequences of both HaSNPV-G4 (Chen *et al.*, 2001) and HzSNPV-ELCAR (Chen *et al.*, 2002) were published, their nucleotide (97% identity) and amino acid (99% identity) sequences were not surprisingly similar. It was then concluded, that these two isolates were variants of the same species, based on genome sequence conservation and features, and other structural and biological properties (Chen *et al.*, 2002). These properties included the ability of both isolates to infect their respective heliothine host, their common morphotype (SNPV) and a similar virulence to insect host (Hughes *et al.*, 1983; Sun and Zhang, 1994). When some selected genes from the genome of HaSNPV-SA were sequenced, a similar percentage identity was observed to both HaSNPV-G4 and HzSNPV-ELCAR. Previous studies have shown that restriction patterns of HzSNPV and HaSNPV were very similar (Knell and Summers, 1984 and Chen *et al.*, 2000). Comparative analysis of the restriction profiles of the three genomes digested with *Hin*d III revealed that the fragments showed a degree of variation but basically had the same restriction profile backbone.

Single nucleotide polymorphisms (SNPs) are common DNA sequence variations among individuals and have great significance for biomedical research. Analysis of DNA sequence variation is becoming an increasingly important source of information for identifying the genes involved in both disease and in normal biological processes, such as development, aging, and reproduction. Since baculoviral Ha/HzSNPV's have great importance in biological control of pests, it was thought that a more detailed SNP comparison of essential and important genes might prove useful. One such gene was the ecdysteroid UDP-glucosyltranferase (*egt*) gene that was located using an Hz-SNPV gene-specific probe. The ecdysteroid UDP-glucosyltranferase gene (*egt*) has been found in several baculoviruses and encodes an enzyme, ecdysteroid UDP-

glucosyltransferase (EGT), which catalyses the conjugation of sugars (glucose or galactose) from UDP-sugars to ecdysteroid molting hormones. Bioassay data showed, that the LT_{50} of the *egt*-deletion recombinants was about 27% faster than that of HaSNPV wild type (Chen et al., 2000). Thus proving that the deletion of the *egt* gene plays an important role in enhancing baculoviruses in biocontrol strategies. While all the *Helicoverpa egt* genes used in this study was 99% similar, SNP analysis showed that 12 single nucleotide differences occurred throughout the genes of the various isolates. There is no evidence that these polymorphisms in any way hinder the function of *egt*, since both Ha-G4 and Ha-SA EGT's have been shown to be functionally active.

A comparison of the genomes of HaSNPV-G4 and HzSNPV-ELCAR also showed, that the most divergent region among these isolates contained two of the three baculovirus-related open reading frames (*bro* genes) named *bro*-a and -b respectively. *Bro*-a and -b were found juxtaposed and flanked by two homologous repeat (*hr*) regions while *bro*-c was found several ORF's downstream of *bro*-b in the genetic map (Chen *et al.*, 2002). Since sequence data strongly suggest that these two strains are variants of the same virus species, the low homology between *bro*-a and *bro*-b was interesting to note (Chen *et al.*, 2002). To study the evolutionary relatedness between *bro*-a and *bro*-b present in *Helicoverpa* SNPV isolates, the same region was investigated in the South African isolate of HaSNPV -SA. Gene specific primers were designed to the flanking regions of *bro*-a and *bro*-b to study the polymorphism in this region. Analysis revealed, that this region within HaSNPV-SA was also different from either of the previously sequenced *Helicoverpa* isolates.

Genetic relatedness of baculoviruses has often been based on the phylogenetic analysis of a single selected gene (Zanotto *et al.*, 1993, Hu, 1998). More often than not, this gene was the occlusion body protein (polyhedrin). While trees based on this gene correctly divided the isolates into NPV and GV and furthermore, the NPV into Group I and II, the exact placement of certain isolates were often in direct disagreement with other phylogenies based on different genes (Herniou *et al.*, 2001). It has now been concluded, that the polyhedrin gene is not the ideal gene for phylogenetic analysis because of its small size and relative high sequence conservation, which provides limited phylogenetic information. Since single gene phylogenies have proved less that accurate, combining genes may reduce sampling errors in phylogenetic analysis and allow for the correct evolutionary conclusion with good support. We then identified and used three important genes late expression factor (*lef*) 8, *chitinase* and *helicase*, for our phylogenetic studies. All three genes used to construct the trees showed the separation of the NPV's and GV's and the further subdivision of the NPV's into groups I and II. However, the tree generated using the *chitinase* gene sequences does not group the *Helicoverpa* SNPV's in the accepted "correct" position (Fig. 4). Within this tree, the three *Helicoverpa* SNPV's are grouped as a group I type NPV. Various studies show (Chen *et al.*, 1997 and Herniou *et al.*, 2001), that these NPV's are in fact a group II type NPV. As in the case of the *polyhedrin* gene, the *chitinase* gene is relatively small and has fairly high sequence conservation. Single nucleotide polymorphisms (SNPs) within each gene of the *Helicoverpa* SNPV's were tabulated and were 11, 21 and 25 SNP's for *chitinase, helicase* and *lef-8* respectively. There is no evidence that these polymorphisms in any way hinder the function of genes, since the propagation of all 3 *Helicoverpa* SNPV's appear to be stable.

CONCLUDING REMARKS

Even though *bro*-genes are so different amongst various isolates of *Helicoverpa*, evidence is still strong enough to support the theory that all these isolates are variants of the same strain. Different genotypic variants of HaSNPV-G4 and HzSNPV-ELCAR have been reported, and numerous genes of various isolated have been submitted to GenBank. Amongst these isolates, gene conservation is never below 98%. The SNP could be as a result of sequence error amongst these isolates, simply because the SNP is rarely an insertion of deletion, but rather of the substitution type. If these are indeed actual SNP's, they do see to affect the functions of the respective genes to any great extent. Furthermore, the propagation of the *Helicoverpa* SNPV's does not see to be affected. As stated before, even though all four the *Helicoverpa* SNPVs compared contained the *bro*-a and –b genes in similar positions, their sequence conservation had different degrees of similarity. It might be significant that theses two *bro*'s are flanked by homologous repeat (*hr*) regions within the genome and it is possible that they were duplicated along with these *hr* regions. Majima *et al.* (1993), showed that duplication and deletions occur most commonly in *hrs* As suggested by Munoz *et al.* (1999), various genotypes of *Spodoptera exigua* nucleopolyhedrovirus could possibly have evolved due to re-arrangement of an *hr* region. It is feasible that a similar argument holds for the variation observed between the *Helicoverpa* isolates with regard to re-arrangement within the *bro*-a and –b region of their genomes. It is notable that the BROs from the same virus do not always have a close relationship, suggesting that they might have a different evolutionary history. In a particular virus the BRO's could be, as a result of either gene duplication or acquisition from different or common sources, such as the insect host or another

baculovirus occupying the same ecological niche. If indeed these isolates are variants of the same species, a revision of baculovirus nomenclature is definitely needed to better reflect the actual genetic information rather than the host infected. Most baculoviruses can indeed cross-infect a close relative host.

FUTURE STUDIES

It would be useful to test whether the HaSNPV-SA and HaGV inoculum co-infecting a colony is more virulent or as virulent as the pure inoculums. Since these baculoviruses are often found together within the environment, a separation might alter the infectivity of the viruses. HaSNPV has been developed as a commercial biopesticide, in China, but in most cases when using naturally occurring pathogens, it might prove useful to obtain and use an indigenous strain.

Since the *bro*-a and –b genes of only *Helicoverpa* isolates were analyzed, it might prove useful to compare this region within more isolates. Transcriptional analysis proved that all three *bro*-genes were indeed active. Specific deletions of one or more of these *bro*-genes might prove useful when determining or confirming their function. The ssDNA-binding motif found within BRO-A protein also warrants further investigation and suggests that it might be involved or interacting with nucleic acids. Different *bro*'s might have different functions or be more or less important for virus replication.

3' RACE might also provide useful information with regard to the polyadenylation of specific bro's. 5' RACE may also provide useful information with regard to the transcriptional start site of the egt gene, and the presence of a TATA box. If indeed an altered CAGT motif were the start site of transcription, it would be of use to test transcription levels to see whether there is any significant decrease.

It is indeed doubtful that the SNP's are significant or provide any useful information with regard to the function or alteration of a specific gene within similar genomes. This would indeed be different only if these genes were defective. Analysis of DNA sequence variation is becoming an increasingly important source of information for identifying the genes involved in both disease and in normal biological processes. This would indeed be an interesting study if any of the *Helicoverpa* isolates were obviously less virulent.

REFERENCES

Adams, J.R. and J.T. McClintock (1991). *Baculoviridae*, Nuclear polyhedrosis viruses, Part 1. Nuclear polyhedrosis viruses of insects. pp 89-147. *in* Adams, J.R. & Bonami, J.R. (Eds.) Atlas of Invertebrate pathology. United States of America, CRC Press Inc.

Afonso, C.L., Tulman, E.R., Lu, Z., Balinsky, C.A., Moser, B.A., Becnel, J. J., Rock, D.L. and G.F. Kutish. (2001) Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. Journal of Virology. **75 (22),** 11157-11165.

Ahrens, C. H., Russel, R.L.Q., Funk, C.J., Evans, J.T., Harwood, S.H. and G.F. Rohrmann (1997). The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome. *Virology* **229,** 381-399.

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

Ausabel, F. M., Brent, R., Kingston, R.E., Moore, D.D., Seidman, J.G., Smith, J.A., and K. Struhl (1994). *In* "Current protocols in molecular biology.". New York: K. Greene and Wiley-Interscience.

Ayres, M. D., Howard, S.C., Kuzio, J., Lopez-Forber, M. and R.D. Possee (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202,** 586-605.

Bah, A., Lucarotti,J.C., Arella,M. and C. Guertin (1998). *Amsacta albistriga* nucleopolyhedrovirus. *Unpublished, Direct Submission.*

Barrett, J. W., Krell,P.J. and B. M. Arif (1995). Characterization, sequencing and phylogeny of the ecdysteroid UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis viruses isolated from *Choristoneura fumiferana. Journal of General Virology* **76,** 2447-2456.

Beames, B. and M. D. Summers. (1988). Comparisons of host cell DNA insertions and altered transcription at the site of insertions in few polyhedra baculovirus mutants. *Virology* **162,** 206-220.

Beames, B. and M. D. Summers. (1989). Location and nucleotide sequence of teh 25K protein missing from baculovirus few polyhedra (FP) mutants. *Virology* **168,** 344-353.

References

Beames, B. and M. D. Summers. (1990). Sequence comparison of cellular and viral copies of host cell DNA insertions found in *Autographa californica* nuclear polyhedrosis virus. *Viroloy* **174**, 354-363.

Blissard, G.W., Quant-Russell, R.L., Rohrmann, G.E. and G.S. Beaudreau (1989). Nucleotide sequence, transcriptional mapping, and temporal expression of the gene encoding *p39*, a major structural protein of the multicapsid nulcear polyhedrosis virus *Orgyia pseudotsogata*. *Virology* **168,** 354.

Blissard, G. W. and G. F. Rohrmann. (1990). Baculovirus diversity and molecular biology. *Annual Review of Entomology* **35**, 127-155.

Blissard, G. K., P. H., Wei, R. and G. Rohrmann (1992). A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology* **190**, 783-793.

Blissard, G. W. (1996). Baculovirus – insect cell interactions. *In* Insect cell cultures Fundamental and applied aspects (Eds J. M. Vlak, C. D. de Gooijer, J. Tramper and H. G. Miltenburger) Kluwer, Dordrecht, pp. 73-93.

Blissard, G. W., Black, B., Crook, N., Keddie, B. A., Possee, R., Rohrmann, G., Theilmann, D. A. and L. E. Volkman (2000). *In* "Virus Taxonomy. Seventh Report of the International Committee on Taxonomy of Viruses". In *Family Baculoviridae*, pp. pp. 195-202. (Eds M. H. V. van Regenmortel, C. M. Fauquet, D. H. L. Bishop, E. B. Carstens, M. K. Estes, S. M. Lemon, J. Maniloff, M. A. Mayo, D. J. McGeoch, C. R. Pringle and R. B. Wicker, Eds.). New York: Academic press.

Caradoc-Davies, K. M., Graves, S., O'Reilly, D. R., Evans, O. P. and V. K. Ward (2001). Identification and in vivo characterization of the *Epiphyas postvittana* nucleopolyhedrovirus Ecdysteroid UDP-glucosyltransferase. *Virus Genes* **22** (3): 255-264.

Chakraborty, S. and S. Ried. (1999) Serial passage of a *Helicoverpa armigera* nucleopolyhedrovirus in *Helicoverpa zea* cell cultures. *Journal if Invertebrate Pathology* **73**, 303-308.

Carson, D.D., Guarino, L.A. and M.D. Summers (1988). Functional mapping of an AcNPV immediate early gene which augments expression of the IE-1 *trans*-activated 39- k gene. *Virology* **162,** 444-451.

Chen, X., Hu, Z., Jehle, J. A., Zhang, Y. and J. M. Vlak (1997). Analysis of the ecdysteroid UDP-glucosyltransferase gene of *Heliothis armigera* single-nucleocapsid baculovirus. *Virus Genes* **15**, 219-225.

References

Chen, X., Li, M., Sun, X., Hu, Z, and J. M. Vlak (2000). Genomic organization of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus. *Archives of Virology* **145**, 2539-2555.

Chen, X., Sun, X., Hu, Z., Li, M., O'Reilly, D. R., Zuidema, D. and J. M. Vlak. (2000a) Genetic engineering of *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus as an improved pesticide. *Journal of Invertebrate Pathology* **76**, 140-146.

Chen, X., Ijkel, W. F. J., Tarchini, R., Sun, X., Sandbrink, H., Wang, H., Peters, S., Zuidema, D., Lankhorst, R. K., Vlak, J. M. and Z. Hu (2001). The sequence of the *Helicoverpa armigera* single-nucleocapsid nucleopolyhedrovirus genome. *Journal of Virology* **82**, 241-257.

Chen, X., Zhang,W.-J., Wong,J., Chun,G., Lu,A., McCutchen,B.F., Presnail,J.K., Herrmann,R., Dolan,M., Tingey,S., Hu,Z.-H. and Vlak,J.M (2002). Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *Journal General Virology* **83**, 673-684.

Christian, P. (1994). *Recombinant baculovirus insecticides: catalysts for a change of heart?* University of Queensland. Brisbane, Queensland.

Christian, P. D., Gibb, N., Kasprzak, A. B. and A. Richards (2001). A rapid method for the identification and differentiation of *Helicoverpa* nucleopolyhedroviruses (NPV *Baculoviridae*) isolated from the environment. *Journal of Virological Methods* **96**, 51-65.

Clarke, E. E. M., Cory, J. S. and D. R. O'Reilly (1996). Characterization of the ecdysteroid UDP-glucosyltransferase gene from the *Mamestra brassicae* nuclear polyhedrosis virus. *Journal of General Virlogy* **77**, 2865-2871.

Crook, N. E., Spencer, R.A., Payne, C.C. and D.J. Leisy. (1985). Variation in *Cydia polmonella* granulosis virus isolates and physical maps of the DNA from three variants. *Journal of General Virology* **66**, 2423-2430.

Devereux, J., Haeberli, P. and O. Smithies (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Research* **12**, 387-395.

Dwyer, K., Webb, S. E., Shelton, A. M. and R. R. Granados. (1988) Establishment of cells lines from *Pieris rapae* embryos: characterization and susceptibility to baculoviruses. *Journal of Invertebrate Pathology* **52**, 268-274.

References

Erlandson, M., Li, S., Gillott, C. and D. Moody (2002). Sequence and Organization of the *Mamestra configurata* Nucleopolyhedrovirus Genome. *Virology* **294**, 106-121.

Faktor, O. and B. Kamensky (1997). Genomic localization and nucleotide sequence of a *lef-8* gene of the *Spodoptera littoralis nucleoployhedovirus*. *Virus Genes* **15 (1)**, 9-15.

Ferber, M. L., Argaud, O., Croizier, L. and G. Croizier (2001). Diversity, distribution and mobility of *bro* gene sequences in *Bombyx mori* nucleopolyhedroviruses. *Virus Genes* **22**, 247-254.

Fitt, G. P. (1989). The ecology of *Heliothis* in relation to agroecosystems. *Annual Review of Entomology* **34**, 17-52.

Friesen, P.D. and L.K Miller (1986). The regulation of baculovirus gene expression. *Current Topics in Microbiology and Immunology* **131**, 31-49.

Friesen, P.D. (1997). Regulation of baculovirus early gene expression. *In* The Baculoviruses. (Ed. L.K. Miller) Plenum, New York, pp. 141-170.

Garcia-Maruniak, A., Pavan, O. H. O. and J. E. Maruniak (1996). A variable region of *Anticarsia gemmatalis* nuclear polyhedrosis virus cotnains tandemly repeated DNA sequences. *Virus Research* **41**, 123-132.

Gomi, S., Majima, K. and S. Maeda (1999). Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *Journal of General Virology* **80**, 1323-1337.

Goto, C. and M. Nakai (2001). Granulovirus prevents pupation and retards development of *Adoxophyes honmai* larvae. *Unpublished, Direct Submission*.

Granados, R. R., Derksen, A. C. G. and K. G. Dwyer. (1986) Replication of *Trichoplusia ni* granulosis virus and nuclear polyhedrosis virus in cell culture. *Virology* **152**, 472-476.

Granados, R.R. and K.A. Williams (1987). *In vivo* infection and replication of baculoviruses. *In* The biology of baculoviruses, Biological and molecular biology, Volume one. (Eds Granados, R.R. and Federici, B.A.) Boca Rota, CRC Press. pp 89.

Guarino, L.A. and M.D Summers (1986). Functional mapping of a *trans*-activating gene required for the expression of a baculovirus delayed early gene. *Journal of General Virology* **57**, 563-571.

Guarino, L.A. and M.W. Smith (1990). Nucleotide sequence and characterisation of the 39K gene region of *Autographa californica* nuclear polyhedrosis virus. *Virology* **179**, 1-8.

Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita, T., Sano, Y. and T. Matsumoto (2000). Sequence analysis of the *Plutella xylostella* granulovirus genome. *Virology* **275**, 358-372.

Hawtin, R. E., Arnold, K., Ayres, M. D., Zanotto, P. M., De, A., Howard, S. C., Gooday, G. A., Chappell, L.H. Kitts, P. A., King, L. A., and R. D. Possee. (1995) Identification and preliminary characterization of a *chitinase* gene in the *Autographa californica* nuclear polyhedrosis virus genome. *Virology* **212**, 673-685.

Hawtin, R. E., Zarkowska, T., Arnold, K., Thomas, C. J., Gooday, G. W., King, L. A., Kuzio, J. A. and R. D. Possee (1997). Liquefaction of Autographa californica Nucleoployhedrovirus-infected insects is dependent on the integrity of virus-encoded chitinase and cathepsin genes. *Virology* **238**, 243-253.

Hayakawa, T., Ko, R., Okano, K., Seong, S., Goto, C. and S. Maeda (1999). Sequence analysis of the *Xestia c-nigrum* granulovirus genome. *Virology* **262**, 277-297.

Heldens, J. G. M., Liu, Y., Zuidema, D., Goldbach, R. W. and J. Vlak (1997). Characterization of a putative *Spodoptera exigua* multicapsid nucleopolyhedrovirus helicase gene. *Journal of General Virology* **78**, 3101-3114.

Herniou, E. A., Luque, T., Chen, X., Vlak, J. M., Winstanley, D., Cory, J. S. and D. O'Reilly (2001). Use of whole genome sequence data to infer baculovirus phylogeny. *Journal of Virlogy* **75**, 8117-8126.

Hink, W.F. (1982). Production of *Autographa californica* nuclear polyhedrosis virus in cells from large scale suspension cultures. pp 493-495. *In* Microbial and viral pesticides. (Ed Kurstak, E.) New York & Basel, Marcel Dekker Inc.

Hu, Z. H., Broer, R., Westerlaken, J., Martens, J. W. M., Jin, F., Jehle, J. A., Wang, L. M. and J. M. Vlak (1997). Characterization of the ecdysteroid UDP-glucosyltransferase gene of a single nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* **47**, 91-97.

Huger, A. (1963) Granulosis of insects. *In* Insect pathology: An advanced treatise, Volume one (Ed. Steinhaus, E. A.), New York, Academic Press.

Hughes, P. R., Gettig, R. R. and W. J. McCarthy. (1983). Comparison of the time-mortality response of

Heliothis zea to 14 isolates of *Heliothis* nuclear polyhedrosis virus. *Journal of Invertebrate Pathology* **41**, 256-261.

Hyink, O., Dellow, R. A., Olsen, M., Caradoc-Davies, K. M. B., Drake, K., Herniou , E. A., Cory, J. S., O'Reilly, D. R. and V. K. Ward (2002). Whole genome analysis of the *Epiphyas postvittana* nucleopolyhedrovirus. *Journal of General Virology* **83**, 957-971.

IJkel, W. F. J., van Strien, E. A., Heldens, J. G. M., Broer, R., Zuidema, D., Goldbach, R. W. and J. M. Vlak (1999). Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *Journal of General Virology* **80**, 3289-3304.

IJkel, W. F. J., Lebbink, R. J., Op den Brouw, M. L., Goldbach, R. W., Vlak, J. M. and D. Zuidema (2001). Identification of a novel occlusion derived virus-specific protein in *Spodoptera exigua* multicapsid nucleopolyhedrovirus. *Virology* **284**, 170-181.

Kang, W., Tristem, M., Maeda, S., Crook, N. E. and D. R. O'Reilly (1998). Identification and characterization of the *Cydia pomonella* granulovirus cathepsin and chitinase genes. *Journal of General Virology* **79 (9)**, 2283-2292.

Kang, W., Suzuki, M., Zemskov, E., Okano, K. and S. Maeda (1999). Characterization of baculovirus repeated open reading frames (*bro*) in *Bombyx mori* nucleopolyhedovirus. *Journal of Virology* **73**, 10339-10345.

Khan, S., Wright, M.G. and S. Davison (2000). Polyhedrin gene sequence of a South African isolate of *Helicoverpa armigera* (Hübner) (Lepidoptera: Noctuidae) single enveloped nuclear polyhedrosis virus. *South African Journal of Science* **96**, 584-588.

Knell, J. D. and M. D. Summers (1984). A physical map of the *Heliothis zea* SNPV genome. *Journal of General Virology* **65**, 445-450.

Knudson, D. L. and K. A. Harrap. (1976). Replication of a nuclear polyhedrosis virus in a continous cell culture of *Spodoptera frugiperda*: microscopy study of the sequence of events of the virus infection. *Journal of Virology* **17**, 254-268.

Kost, T.A.,and J. P. Condreay. (2002) Recombinant baculoviruses as mammalian cell gene-delivery vectors. *Trends in Biotechnology* **20 (4),** 173-180

Krappa, R., Bohn-Krappa, A., Jahnel, F. Doerfler, W. and D. Knebel-Mörsdorf, D (1992). Differential factor binding at the promoter of early baculovirus gene PE 38 during viral infection; GATA motif is recognised by an insect protein. *Journal of Virology* **66,** 3494-3503.

Kuzio, J., Pearson, M.N. Harwood, S.H., Funk, C.J., Evans, J.T., Slavicek, J.M. and G.F. Rohrmann (1999). Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* **253,** 17-34.

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

Lee, S. Y., Poloumienko, A., Belfry, S., Chen, W., MacAfree, N., Morin, B., Lucarotti, C. and M. Krause (1996). A common pathway for p10 and calyx proteins in progressive stages of polyhedron envelope assembly in AcMNPV infected *Spodoptera frugiperda* larvae. *Archives of Virology* **141,** 1247-1258.

Lewin, B. (1994). Genes V. New York: Oxford University Press.

Lua, L. H. L. and S. Reid. (2000). Virus morphogenesis in *Helicoverpa armigera* nucleopolyhedrovirus in *Helicoverpa zea* serum-free suspension culture. *Journal of General Virology* **81,** 2531-2543.

Luckow, V.A., Lee, S.C., Barry, G.F.and P.O. Olins. (1993) Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *Journal of Virology* **67(8),** 4566-79

Luque, T., Finch, R., Crook, N., O'Reilly, D. R. and D. Winstanley (2001). The complete sequence of the *Cydia pomonella* granulovirus genome. *Journal of General Virology* **82,** 2531-2547.

Mackenzie P. I., Owens I. S., Burchell B., Bock K. W., Bairoch A., Belanger A., Fournel-Gigleux S., Green M., Hum D. W., Iyanagi T., *et al.* (1997). The UDP-glycosyltranferase gene superfamily: recommended nomenclature update based on evolutionary divergence. *Pharmocogenetics* **7,** 255-269.

MacKinnon, E. A., Henderson, J. F., Stoltz, D. B. and P. Faulkner. (1974). Morphogenesis of nuclear

polyhedrosis virus under conditions of prolonged passage *in vitro*. *Journal of Ultrastructural Research* **49**, 419-435.

Maeda, S., Kamita, S. G. and A. Kondo (1993). Host range expansion of *Autographa californica* nuclear polyhedrosis virus (NPV) following recombination of a 0.6-kilobase-pair DNA fragment originating from *Bombyx mori* NPV. *Journal of Virology* **67**, 6234-6238.

Majima, K., Kobara, R. and S. Maeda (1993). Divergence and evolution of homologous regions of *Bombym mori* nuclear polyhedrosis virus. *Journal of Virology* **67**, 7513-7521.

Manzan, M. A., Lozano, M. E., Sciocco de Cap, A., Ghiringhelli, P. D. and V. Romanowski (2001). Identification and characterization of the ecdysteroid UDP-glycosyltransferase gene of *Epinotia aporema* granulovirus. *Unpublished, Direct Submission.*

McIntosh, A. H. and. C. M. Ignoffo (1983). Characterization of 5 cell lines established from species of *Heliothis*. *Applied Entomology and Zoology* **18**, 262-269.

Miller, S. A., Dykes, D. D. and H. F. Polesky (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research* **16**, 3.

Miller, L. K. (Ed) (1997). The baculoviruses. Plenum Press, New York and London, pp. 1-32.

Miltenburger, H. G., Naser, W. L., Harvey, J. P., Huber, J. and A. M. Huger. (1984) The cellular substrate: a very important requirement for baculovirus in vitro replication. *Zeitschrift fur Naturforschung* **39c**, 993-1002.

Munoz, D., Murillo, R., Krell, P. J., Vlak, J. M. and P. Caballero (1999). Four genotypic variants of a *Spodoptera exigua* nucleopolyhedrovirus (Se-SP2) are distinguishable by a hypervariable genomic region. *Virus Research* **59**, 61-74.

Naser, W. L., Miltenburger, J. P., Harvey, J. P., Huber, J. and A. M. Huger. (1984) *In vitro* replication of the *Cydia pomonella* (codling moth) granulosis virus. *FEMS Microbiology Letters* **24**, 117-121.

Nicholas, K. B. and H. B. Nicholas (Jr) (1997). Distributed by the author.

O'Reilly, D. R. and L. K. Miller (1989). A baculovirus blocks insect molting by producing ecdysteriod UDP-glusocyl trasnferase. *Science* **24**, 1110-1112.

References

O'Reilly, D. R. and L. K. Miller (1991). Improvement of a baculovirus pesticide by deletion of the *egt* gene. *Bio-Technology* **9**, 1086-1089.

O'Reilly, D. R. (1995). Baculovirus-encoded ecdysteroid UDP-glucosyltransferases. *Insect Biochemistry and Molecular Biology* **25**, 475-485.

Page, R. D. M. (1996). TreeView: an application to display phylogenetic trees on personal computers. *Computerised Applied Biosciences* **12**, 357-358.

Pang, Y., Yu, J., Wang, L., Hu, X., Bao, W., Li, G., Chen, C., Han, H., Hu, S. and H. Yang (2001). Sequence Analysis of the *Spodoptera litura* Multicapsid Nucleopolyhedrovirus Genome. *Virology* **287**, 391-404.

Park, H. J. (2001). Molecular cloning and characterization of *egt* gene from *Bombyx mori* nuclear polyhedrosis virus K1. *Unpublished, Direct Submission.*

Passarelli, L. A., Todd, J. W. and L. K. Miller (1994). A baculovirus gene involved in late gene expression predicts a large polypeptide with a conserved motif of RNA polymerases. *Journal of Virology* **68 (7)**, 4673- 4780.

Pearson, M.N., Russell, R.L. Rohrmann, G.F. and G.S. Beaudreau (1988). *p39* a major baculovirus structural protein: immunocytochemichal characterization and genetic location. *Virology* **167**, 407-413.

Popham, H. J. R., Li, Y. and L. K. Miller (1997). Genetic Improvement of *Helicoverpa zea* Nuclear Polyhedrosis Virus as a Biopesticide. *Biological Control* **10**, 83-91.

Premkumar, A. and S. Mathavan (1999). Nucleotide sequence of *egt* gene of *Amsacta albistriga* nucleopolyhedrovirus. *Unpublished, Direct Submission.*

Reed, L. J. and H. Muench. (1938). A simple method of estimating fifty per cent endpoints. *American Journal of Hygiene* **27**, 493-497.

Riegel, C. I., Lanner-Herrera, C. and J.M. Slavicek. 1994. (1994). Identification and characterization of the ecdysteroid UDP-glucosyltransferase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *Journal of General Virology* **75**, 829-838.

References

Ritter, J. K., Sheen, Y. Y. and I. S. J. Owens. (1990) Cloning and exression of human liver UDP-glucuronosyltransferase in COS-1 cells. *Biol.ogical Chemistry* **265**, 7900-7906.

Rodrigues, J. C., De Souza, M. L., O'Reilly, D., Velloso, L. M., Pinedo, F. J., Razuck, F. B., Ribeiro, B. and B. M. Ribeiro (2001). Characterization of the ecdysteroid UDP-glucosyltransferase (*egt*) gene of *Anticarsia gemmatalis* nucleopolyhedrovirus. *Virus Genes* **22**, 103-112.

Rohrmann, G. F. (1992). Baculovirus structural proteins. *Journal of General Virology* **73**, 749-761.

Sambrook, J., Fritsch, E.F. and T. Maniatis (1989). Molecular cloning, A laboratory manual, 2nd edition number 1, 2, 3. edn. United States of America, Cold Spring Harbour: Cold Spring Harbour Press.

Slavicek, J. M., Popham, H. J. R. and C. I. Riegel (1999). Deletion of the *Lymantria dispar* multicapsid nucleopolyhedrovirus ecdysteroid UDP-glucosyltransferase gene enhances viral killing speed in the last instar of the gypsy moth. *Biological Control* **16**, 91-103.

Smith, I. and C. Goodale (1998). Sequence and in vivo transcription of *Lacanobia oleracea* granulovirus *egt*. *Journal of General Virology* **79**, 405-413.

Southern, E. (1975). Detection of specific sequences among DNA fragments separated by gel electrophoresis. *Journal of Molecular Biology* **98**, 503-517.

Sun, X. and G. Zhang. (1994) A comparison of four wild isolates of *Heliothis* nuclear polyhedrosis virus. *Virologica Sinica* **9**, 309-318.

Taylor, D. E. (1982). Entomology notes. *Zimbabwe Agricultural Journal* **79 (2)**, 45.

Teackle, R. E., Jensen, J. M. and J. C. Mulder. (1985). Susceptibility of *Heliothis armigera* (Lepidoptera: Noctuidea) on sorghum to a nuclear polyhedrosis virus. *Journal of Economic Entomology* **78**, 1373-1378.

Thiem, S.M. and L.K. Miller (1989). Identification, sequence, and transcriptional mapping of the major capsid protein gene of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology* **63 (5)**, 2008-2018.

Thompson, J. D., Higgins, D. G. and T. J. Gibson (1994). CLUSTAL W: improving the sensitivity of

progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acid Research* **22**, 4673-4680.

Toister-Achituv, M. and O. Faktor (1997). Transcriptional analysis and promoter activity of the *Spodoptera littoralis* multicapsid nucleopolyhedovirus ecdysteroid UDP-glucosyltransferase gene. *Journal of General Virology* **78**, 487-491.

Tomalski, M.D., Wu, J. and L.K. Miller (1988). The location, sequence, transcription and regulation of a baculovirus DNA polymerase. *Virology* **167**, 591-600.

van der Beek, C. P. (1980). On the origin of the polyhedral protein of the nuclear polyhedrosis virus of *Autographa californica*. Mededelingen Landbouwhogescchool Wageningen 80, 5.

van Strein, E. A. (1997). Characterisation of the Spodoptera exigua baculovirus genome: structural and functional analysis of a 20-kb fragment. PhD thesis. Wageningen Agricultural University, Wageningen, The Netherlands.

van Oers, M. M., Flipsen, J. T. M., Reusken, C. B. E. M., Sliwinsky, E. L., Goldbach, R. W. and J. M. Vlak (1993). Functionla domains of the P10 protein of *Autographa californica* nuclear polyhedrosis virus. *Journal of General Virology* 74, 563-574.

Vlak, J.M., Klinkenberg, F.A., Zaal, K.J.M., Usmany, M., Klinge-Roode, E.C., Geervliet, J.B.F., Roosien, J. and J.W.M. van Lent (1988). Functional studies on the p10 gene of *Autographa californica* nuclear polyhedrosis virus uisng a recombinant expressing a p10- -Galactosidase fusion gene. *Journal of General Virology* 69, 765-776.

von Miltenburg, R., Ruger, B., Grunewald, S., Lears, M. and C. Schroda (1995). The DIG system users guide for filter hybridization. Edited by Boehringer-Mannhein. Germany: Gmbh Biochemia.

Wang, Y. S. and J. D. Hall.. (1990). Characterization of a major DNA-binding domain in the herpes simplex virus type 1 DNA-binding protein (ICP8). *Journal of Virology* **64**, 2082-2089.

Williams, G. V. and P. Faulkner. (1997). Cytological changes and viral morphogenesis during baculovirus infection. In L. K. Miller (Ed.): *In The Baculoviruses*, Plenum, New York.

Winstanley, D. and N. E. Crook. (1993) Replication of *Cydia pomonella* granulosis virus in cell culture.

Journal of General Virology **74**, 1599-1609.

Wormleaton, S. L. and D. Winstanley (2001). Phylogenetic analysis of conserved genes within the ecdysteroid UDP-glucosyltransferase gene region of the slow-killing *Adoxophyes orana* granulovirus. *Journal of General Virology* **82**, 2295-2305.

Zanatto, P. M., De, A., Kessing, B.D., and J.E. Maruniak (1993). Phylogenetic Interrelationships among Baculoviruses: Evolutionary Rates and Host Associations. *Journal of Invertebrate. Pathology* **62**, 147-162.

Zemskov, E. A., Kang, W. and S. Maeda (2000). Evidence for nucleic acid binding ability and nucleosome association of *Bombyx mori* nucleopolyhedrovirus BRO proteins. *Journal of Virology* **74**, 6784-6789.

Zhang, G., Zhang, Y., Ge, L. and Z. Shan (1981). The production and application of the nuclear polyhedrosis virus of *Heliothis armigera* (Hubner) in biological control. *Acta Phytophylactica Sinica* **8**, 235-240.

Zhang, C., Hu, C. and X. Wu (1998). Nucleotide sequence of *Ecotropis obliqua* nuclear polyhedrosis virus (EoNPV) *egt. Unpublished, Direct Submission.*

Zhang, C. X. and W. R. Jin (2000). *Helicoverpa armigera* single-enveloped nucleopolyhedrovirus. *Submitted to Institute of Applied Entomology.*

LIST OF TABLES

Page

Chapter one:

Chapter two

Chapter four:

Chapter five

LIST OF FIGURES

Page

Chapter one:

Chapter two:

Page

Chapter five: