Molecular Characterization of the *Neodiprion abietis* Nucleopolyhedrovirus

by

Aaron Young
B.Sc., University of Victoria, 1999

A Thesis Submitted in Partial Fulfillment of the
Requirements for the Degree of

Master of Science
in the Department of Biology

We accept this thesis as conforming
to the required standard

Dr. D. Levin, Supervisor (Department of Biology)

Dr. B. Koop, Departmental Member (Department of Biology)

Dr. C. Upton, Outside Member (Department of Biochemistry and Microbiology)

Dr. R. Roper, External Examiner (Department of Biochemistry and Microbiology)

© Aaron Young, 2002
University of Victoria

All rights reserved. This thesis may not be reproduced in whole or in part, by photocopy
or other means, without the permission of the author.

Supervisor: Dr. David B. Levin

Abstract

Baculoviruses are a large, diverse family of viruses that are pathogenic for Arthropods, particularly members or the Order Insecta. These viruses are comprised of large enveloped rod-shaped capsids that contain supercoiled, double-stranded DNA genomes. Baculoviruses are of particular interest because of their use as gene expression vectors and insect biological control agents. Most of the information currently available on baculoviruses comes from studies of the *Autographa californica* nucleopolyhedrovirus and closely related species. Very little information is available on baculoviruses from other insect Orders, including Hymenoptera.

The balsam fir sawfly (*Neodiprion abietis,* Order:Hymenoptera) is a native insect species that is currently causing severe, annual damage to economically valuable balsam fir forests in eastern Canada. Due to environmental regulations, standard chemical pesticides cannot be used to control this pest and as a result, biological control agents have been proposed as a potential alternative.

A nucleopolyhedrovirus (NPV) has recently been isolated from *N. abietis* and has been shown to be able to control balsam fir sawfly population outbreaks in controlled field experiments, however, before the *N. abietis* nucleopolyhedrovirus (NeabNPV) can be registered for use as a biological control agent, research must be conducted to characterize the virus.

I have partially characterized NeabNPV at the nucleotide and protein sequence levels and have concluded that NeabNPV is highly diverged from other members of the Family *Baculoviridae.* DNA libraries were constructed for NeabNPV using sheared viral

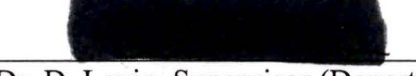
genome fragments and *Hin*dIII restriction fragments. An *Eco*RI restriction library was obtained from the Canadian Forestry Centre, Atlantic Division, and was also included in this study. Analysis of the restriction endonuclease profiles obtained for NeabNPV indicate that NeabNPV has a genome approximately 94.7 kilobase pairs in size. Nucleotide sequence data obtained from these DNA libraries was used to perform comparative sequence analyses and demonstrated that NeabNPV contains many putative baculovirus gene homologues. From the sequence data, four putative complete open reading frames (ORF) were predicted: *odv-e56, vlf-1, p74,* and *polh.*

The putative polyhedrin gene homologue was amplified from the NeabNPV genome and sequenced. Analysis of the polyhedrin gene sequence showed that it was 738 nucleotides long and encoded a protein of a predicted molecular weight of 29.5 kDa. The promoter region of the gene was determined to contain two copies of the baculovirus-specific late transcription motif, TAAG. The NeabNPV polyhedrin gene is highly diverged from Lepidopteran NPVs and GVs (Granulovirus) and shares its highest sequence identity with the polyhedrin gene of the Hymenopteran baculovirus *Neodiprion sertifer* nucleopolyhedrovirus.

Phylogenetic analyses of several putative NeabNPV gene homologues suggested that NeabNPV is highly diverged from the previously described Group I and Group II NPVs and appears in a unique evolutionary clade separate from both the Lepidopteran NPVs and GVs. Phylogenies which included the mosquito (Order: Diptera) baculovirus *Culex nigripalpus* nucleopolyhedrovirus (CuniNPV) and that used the GVs as a sister outgroup to the NPVs placed NeabNPV and CuniNPV into a unique evolutionary clade within the NPV group, separate from the Group I and Group II NPVs.

I have also investigated NeabNPV at the protein level and have putatively identified the major nucleocapsid protein, VP39. Structural proteins from budded (BV) and occlusion-derived virions (ODV) were separated on polyacrylamide gels using SDS-PAGE and several well-defined proteins were isolated and partially sequenced. The amino acid sequence data was used to perform comparative sequence analyses against our NeabNPV DNA libraries and determined that the genomic DNA library contained several putative overlapping partial *vp39* sequences. These sequences were assembled to create an 864 nucleotide long, partial NeabNPV *vp39* sequence. Analysis of the inferred amino acid sequence indicated that the NeabNPV VP39 protein shares the most identity to the GVs, but is most similar to the NPVs. The partial VP39 protein demonstrates a predicted secondary structure and hydrophilic profile that is conserved among the *Baculoviridae.*

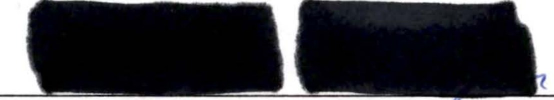
Examiners:



Dr. D. Levin, Supervisor (Department of Biology)

Dr. B. Koop, Departmental Member (Department of Biology)

Dr. C. Upton, Outside Member (Department of Biochemistry and Microbiology)



Dr. R. Roper, External Examiner (Department of Biochemistry and Microbiology)

Table of Contents

Abstract ii

Table of Contents v

List of Tables viii

List of Figures ix

Acknowledgements xi

Chapter 1. General Introduction 1

1.1. Historical Perspective 1

1.2. Prevalence and Diversity 4

1.3. Baculovirus Structure 4

1.4. Baculovirus Infection 7

1.5. Practical Application of Baculoviruses 11

1.5.1. Baculoviruses as Biological Control Agent 11

1.5.1.1. Factors Influencing the Use of Baculovirus Pesticides 12

1.5.1.2. Factors Constraining the Greater Use of Baculovirus Pesticides 13

1.5.2. Genetic Enhancement of Baculovirus Pesticides 14

1.6. The Baculovirus Genome 15

1.6.1. Overview 15

1.6.2. Common Baculovirus Genes 16

1.6.2.1. Transcription-Specific Genes 17

1.6.2.2. Replication-Specific Genes 18

1.6.2.3. Structural Genes and Proteins 19

1.6.2.4. Auxiliary Genes 20

1.6.3. Repetitive DNA Sequences 22

1.6.3.1. Coding Sequences 22

1.6.3.1.1. The VEF Gene Family 22

1.6.3.1.2. The IAP Gene Family 23

1.6.3.1.3. The CTL Gene Family 23

1.6.3.1.4. The BRO Gene Family 24

1.6.3.2. Non-Coding DNA Sequences 25

1.6.3.2.1. Homologous Regions (Hrs)....25

1.7. Genome Organization....25

1.8. Baculovirus Phylogeny....28

1.8.1. Origins and Relationships of Baculoviruses....28

1.9. The Balsam Fir Sawfly (*Neodiprion abietis)* and NeabNPV....32

1.9.1 Objectives....32

1.10. References....33

Chapter 2. Development of Partial DNA Libraries for the *Neodiprion abietis* Nucleopolyhedrovirus 39

2.1. Introduction....39

2.2. Methods and Materials....41

2.2.1. Virus Inclusion Bodies....41

2.2.2. Viral DNA Extraction....41

2.2.3. Restriction Endonuclease Analysis....42

2.2.4. Restriction Fragment Cloning and End-Terminus Sequencing....42

2.2.5. Shotgun Cloning and Sequencing....43

2.2.6. Sequence Analysis....44

2.3. Results and Discussion....45

2.4. References....60

Chapter 3. Nucleotide Sequence Analysis of the Polyhedrin Gene of the *Neodiprion abietis* Nucleopolyhedrovirus 62

3.1. Introduction....62

3.2. Methods and Materials....63

3.2.1 Virus and DNA....63

3.2.2. Restriction Endonuclease Digestion and Cloning....63

3.2.3. Sequencing Restriction Fragments....63

3.2.4. Isolation of the NeabNPV Polyhedrin Gene Sequence....63

3.2.5. Analysis of the NeabNPV Polyhedrin Gene Sequence....64

3.3. Results and Discussion....64

3.3.1. The NeabNPV Polyhedrin Nucleotide Sequence....65

3.3.2. Comparative Analysis of the NeabNPV Polyhedrin Sequence........................71
3.3.2.1. Phylogenetic Analysis..71
3.4. References..84

Chapter 4. Analysis of NeabNPV Structural Proteins....................................86
4.1. Introduction..86
4.2. Methods and Materials...88
4.2.1. Cells and Virus..88
4.2.2. Partial Virus Purification...89
4.2.3. SDS-PAGE...90
4.2.4. Protein Sequencing and Computer Analysis...91
4.3. Results and Discussion..92
4.3.1. SDS-PAGE and NeabNPV Structural Proteins...92
4.3.2. Identification of two NeabNPV Structural Proteins..................................102
4.3.3. Analysis of a Putative Partial NeabNPV *vp39* Gene.................................107
4.4. References...122

Chapter 5. Phylogenetic Analysis of Eight Putative NeabNPV Genes.....................124
5.1. Introduction...124
5.2. Materials and Methods..126
5.2.1. Baculovirus sequences...126
5.2.2. Phylogenetic inference based on baculovirus gene sequences.........................131
5.3. Results..132
5.4. Discussion...142
5.5. References...146

Conclusion..149
Appendix. Translation BLAST results for NeabNPV genomic DNA sequences.....152

List of Tables

Table 2.1. NeabNPV DNA *Eco*RI and *Hin*dIII restriction endonuclease fragment sizes (kb)....53

Table 2.2. Baculovirus Genomes....54

Table 2.3. Translation BLAST results for cloned NeabNPV DNA *Hin*dIII restriction fragments....55

Table 2.4. Translation BLAST results for cloned NeabNPV DNA *Eco*RI restriction fragments....57

Table 3.1. Translation BLAST results for the NeabNPV *Eco*RI VT3 end-terminus nucleotide sequence....74

Table 3.2. Translation BLAST results for the NeabNPV *Eco*RI MT3 end-terminus nucleotide sequence....74

Table 3.3. Amino acid sequence identities (%) of fifteen baculovirus polyhedrin (NPV) and granulin (GV) proteins....75

Table 4.1. Sizes (kDa) of protein bands in Fig. 4.1 that are less than 180 kDa....96

Table 4.2. Sizes (kDa) of protein bands less than 175 kDa seen in Fig 4.3....98

Table 4.3. Peptide sequences from NeabNPV structural proteins....105

Table 4.4. NeabNPV genomic DNA fragments which aligned with the baculovirus nucleocapsid protein VP39....105

Table 4.5. Amino acid identities (%) of fifteen baculovirus VP39 proteins....113

Table 5.1. Baculovirus genomes....127

Table 5.2. Parsimoniously-informative characters from each baculovirus amino acid sequence alignment....127

Table 5.3. Amino acid sequence identities (%) of the baculovirus genes *odv-e56* (A), *p74* (B), *polh* (C), and *vlf-1* (D) from 15 baculoviruses....128

List of Figures

Figure 1.1. Baculovirus occlusion body, virion, and nucleocapsid form....................6
Figure 1.2. Cellular infection of a nucleopolyhedrovirus....................................9
Figure 1.3. Dissected midgut from *C. nigripalpus* third-instar larvae.........................10
Figure 1.4. Comparison of Group I NPV and Group II NPV baculovirus genome clusters...27
Figure 1.5. Occlusion protein (polyhedrin and granulin) evolutionary gene tree..........30
Figure 1.6. Phylogenetic tree constructed using the major capsid protein VP-39.........31
Figure 2.1. NeabNPV DNA restriction endonuclease digests...............................50
Figure 2.2. NeabNPV *Hin*dIII restriction endonuclease digest (H) separated by pulsed field gel electrophoresis (PFGE)..51
Figure 2.3. NeabNPV *Hin*dIII and *Eco*RI DNA digests....................................52
Figure 3.1. The nucleotide sequence of NeabNPV *polh*....................................76
Figure 3.2. Amino acid alignment of two Hymenopteran (NeseNPV and NeabNPV) baculovirus polyhedrin proteins..78
Figure 3.3. Graphic representations of the collinearity of baculovirus polyhedrin proteins of AcMNPV (A), LdMNPV (B), NeseNPV (C), and MacoNPV (D) obtained by dot plot analysis...................................79
Figure 3.4. Comparison of the 5' untranslated region (UTR) from fifteen baculovirus polyhedrin and granulin genes...81
Figure 3.5. Frameshift analysis of the NeabNPV polyhedrin sequence.......................82
Figure 3.6. Phylogenetic tree of baculovirus occlusion protein sequences..................83
Figure 4.1. Putative NeabNPV structural proteins..95
Figure 4.2. Putative NeabNPV structural proteins ....,,.97
Figure 4.3. Structural proteins of AcMNPV virus and virus fractions........................99
Figure 4.4. Pairwise alignment of the inferred, partial amino acid sequence of the putative NeabNPV *vp39* gene and the SpltMNPV *vp*39 gene...............106

Figure 4.5. Partial alignment of the amino acid sequences of 15 baculovirus VP39 proteins..110

Figure 4.6. Predicted secondary structure of VP39 proteins..................................115

Figure 4.7. Partial amino acid sequence alignment of baculovirus VP39 proteins..118

Figure 4.8. Predicted hydrophilicity of a 240 amino acid region from baculovirus VP39 proteins..120

Figure 5.1. The single most parsimonious tree topologies obtained for the individual phylogenetic analyses of 4 baculovirus genes: *p143* (A), *lef-8* (B), *lef-9* (C), and *vp-91* (D)..135

Figure 5.2. The single most parsimonious tree topologies obtained for the individual phylogenetic analyses of 4 baculovirus genes: *odv-e56* (A), *p74* (B), *polh* (C), and *vlf-1* (D)..138

Figure 5.3. Phylogenetic tree of the combined sequence of *odv-e56*, *p74*, *polh*, and *vlf-1*, *p143*, *lef-8*, *lef-9*, and *vp-91*..141

Acknowledgements

I would like to thank Dr. David Levin for affording me the opportunity to practice scientific research, for providing me with guidance and helpful assistance, and for financial support throughout the duration of my research. I would also like to thank all the members of our lab, past and present, for their friendship and support during my graduate degree, particularly Simon Duffy and David Harrison, for their technical assistance, guidance, and overall camaraderie. I am also grateful for the assistance and advice I have received from many people in the Centre for Biomedical Research, particularly members of the laboratory of Dr. Ben Koop including Mike Wilson, Roberto Alberto, Linda McKinell, and Melanie Conrad. As well, I would like to thank members of Dr. Koop and Dr. Olafson's laboratories for their DNA and protein sequencing assistance, respectively.

I would like to take this opportunity to thank Eleanore Floyd, for her remarkable administrative assistance, and Pauline Tymchuk, for her assistance, kind words, and friendship. I would like to thank Dr. Imre Otvos for his assistance gaining financial support and would also like to extend my appreciation to my supervisory and examining committee, which includes Dr. Ben Koop, Dr. Chris Upton, and Dr. Rachel. Roper.

My deepest gratitude goes to my dearest and closest friend, Lara Sloan, for her support and patience with my academic endeavours, and for her understanding and compassion during the last few years; I will not forget.

Chapter 1. General Introduction

1.1. Historical Perspective

> *All at once, in the weak ones, the skin appears yellow. Then they swell up and a foul inactivity comes in the bodies of those who have fallen down. Finally they break open and everything is infected with repulsive, putrid gore; diseased blood from all sides flows from the bodies* (Marco Vida, 1527).

According to baculovirus folklore, the earliest written accounts of a baculovirus infection were found in ancient Chinese literature dating back as far as 2000 years, which describe the disease of silkworm (*Bombyx mori*) populations. However, in western culture the first record of baculovirus infection did not appear until 1527, when a description of the 'jaundice disease' of silkworms appeared in a book written by the Italian bishop Marco Vida of Cremoa (above). During this time, the cause of the disease was thought to be water, especially seawater. Around 150 years later, Maria Sibylla Merian described a 'jaundice disease' observed in silkworm larvae in her book, “The Miraculous Metamorphosis of Caterpillars and their Unique Flower Nutrition”. Merian attributed the cause of the disease to thunderstorms; for centuries the disease was not recognized as being contagious (Granados and Federici, 1986).

The earliest scientific description of baculovirus infection did not occur until the mid-19th century when, through the use of microscopes, Italian scientists Maestri and Cornalia describe “strongly refractive crystal-like corpuscules in the blood and tissue of jaundiced silkworms (Evanson, 1922). In order for true scientific research to begin on baculoviruses, the development of several important scientific techniques and

technologies, as well as advancements in the understanding of disease, had to occur. This included the development and investigation of the centrifuge, the creation of the Berkefeld and Chamberland microbial filter (to remove bacteria from liquids), the invention of dark-field and phase-contrast microscopes, and the postulate of Schleiden and Schwann (1839), that all organisms are made up of cells. With the advent of these scientific tools, research on baculoviruses began to flourish. By the beginning of World War I (1914) a huge breakthrough had occurred in the study of baculoviruses with the discovery that the cause of the jaundice disease was a "filterable virus" capable of passing through a Berkefeld and Chamberland filter (Granados and Federici, 1986).

The next advancement in baculovirus research came with the invention of the electron microscope. Although magnetic electron microscopes had been used throughout the 1930's, it was not until 1943 with the work of Gernot Bergold, that electron microscopes were used in baculovirus research. In 1947 Bergold produced and published the first electron micrographs of baculoviruses from *B. mori* and *L. dispar*. These pictures showed rod shaped virions within crystalline polyhedra and demonstrated the viral nature of the jaundice disease of silkworms. These findings proved to be very important and stimulated reflections on the nature of these viruses.

During this same period baculoviruses were discovered to be effective biological control agents. This was discovered through the accidental control of an insect infestation caused by the European spruce sawfly (*Gilpinia hercyniae*). This insect (Order: Hymenoptera) was introduced into North America in the late 19th century and became a serious pest of spruce trees in eastern Canada and northeastern US by the 1930's. Introduction of parasitoids that were unknowingly carrying baculoviruses specific to *G. hercyniae* caused populations of the European spruce sawfly to rapidly decline by the

early 1940's. Accompanying this decline was the appearance of a polyhedrosis in the sawfly larvae, which led to the discovery that the polyhedrosis of sawflies develops only in the cells of the midgut epithelium and not elsewhere in the body, unlike NPV infections observed in insect larvae from the Order Lepidoptera.

From 1970 to 1985 several important advances were made in the understanding of the pathology and genetics of baculoviruses. The development of cell culture and baculovirus purification techniques paved the way for molecular biology and the *in vitro* production of baculoviruses. *Autographa californica* nucleopolyhedrovirus (AcNPV) became the focus of much baculovirus research due to its stability and ease of propagation in cell culture and its relatively broad range of insect hosts. The realization that there were two different forms of baculoviruses, a budded form and occluded form, contributed to the understanding of the behaviour of baculoviruses in cell culture and their pathology in the insect host (Miller, 1997).

The "modern period" of baculovirus research has led to the development of baculoviruses as vectors for the expression of heterologous genes. These recombinant baculoviruses, called baculovirus expression vectors (BEVs), are used extensively throughout the molecular biology and biomedical research communities and are among the best tools for the expression of recombinant genes in eukaryotic hosts (Jarvis, 1997). Interest in baculoviruses has also focussed on their use as biological control agents with recent research focussed on sequencing baculovirus genomes in an attempt to understand these viruses at the genetic level.

1.2. Prevalence and Diversity

Baculoviridae is a large family of viruses pathogenic for Arthropods. The majority of baculoviruses infect the larval stage of insects, and baculoviruses have been isolated from around 600 insect species (Wood Granados, 1991). Although they are primarily pathogens of insects within the Order Lepidoptera (moths), they also infect species from the insect Orders Hymenoptera (sawflies), Diptera (mosquitoes), Coleoptera (beetles), Thysanura (silverfish), and Trichoptera (caddisflies) (Martignoni and Iwai, 1986). Several putative baculoviruses have also been reported in arachnids and crustaceans, but the status of these viruses as true members of the *Baculoviridae* has yet to be confirmed.

1.3. Baculovirus Structure

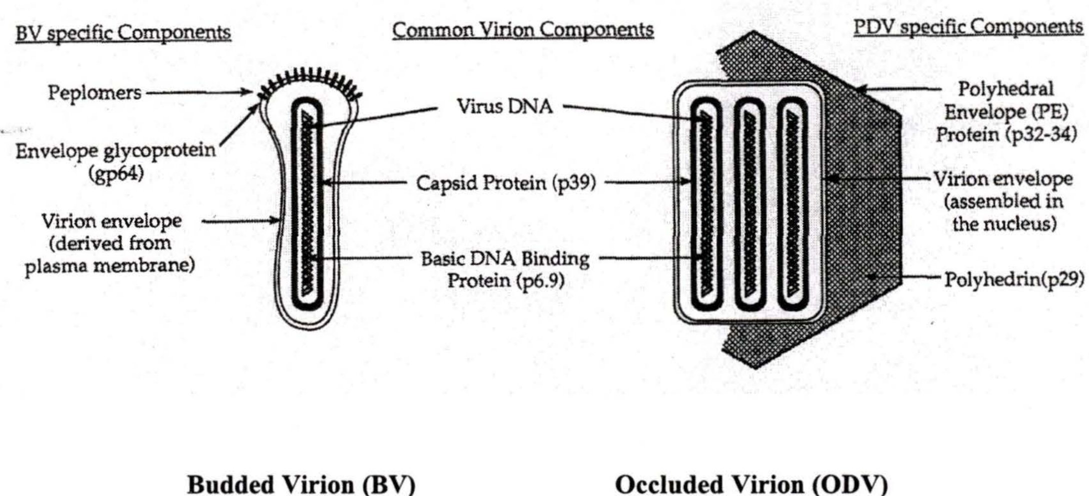
Baculoviruses are rod-shaped, enveloped viruses with large, circular, covalently closed, double-stranded DNA genomes that range in size from 80-180kb. The nucleocapsid is capable of extending to accommodate larger DNA genomes, such as recombinant viruses carrying large DNA inserts. The ends of the nucleocapsid are structurally different providing a polarity to the rods (Miller, 1997).

A distinctive feature of baculoviruses is their occlusion within a crystalline protein matrix, or occlusion body (OB). The family *Baculoviridae* has been divided into two genera, the *Nucleopolyhedroviruses* (NPVs) and the *Granuloviruses* (GVs), based on the morphology of these occlusion bodies. In NPVs, occlusion bodies are formed within the nucleus of the host cell and are comprised primarily of a protein called polyhedrin. NPV occlusion bodies are polyhedral in shape, giving rise to the additional name polyhedrin inclusion bodies (PIBs), and have been divided into two groups based on the extent of their nucleocapsid aggregation. These include the multiple nucleocapsid type, or MNPV,

which have two or more nucleocapsids present within a single membranous envelope, and the single nucleocapsid type, or SNPV, which have a single nucleocapsid present within a single envelope (Miller, 1997) (Fig. 1.1).

In GVs, occlusion bodies are formed throughout the host cell and are comprised primarily of a protein called granulin. Granulin and polyhedrin are very similar proteins (approximately 55% amino acid identity), each composed of approximately 245 amino acids and with molecular weights of 29 kilodaltons (Rohrmann, 1986). GV occlusion bodies are small and ellipsoidal in shape and only contain SNPVs. Surrounding both NPV and GV occlusion bodies is a membranous envelope called the outer calyx which is thought to provide increased stability to the occlusion body structure.

NPVs are the most common and widely distributed group of baculoviruses, having been reported from over 400 insect species distributed among several insect Orders (Lepidoptera, Hymenoptera, Diptera, Coleoptera, Arachnida, Thysanura, Trichoptera, and Crustacea) (Martignoni and Iwai, 1986). In contrast, GVs have been reported from approximately 200 insect species, all of which belong to the Order Lepidoptera.



Budded Virion (BV) **Occluded Virion (ODV)**

Figure 1.1. Baculovirus occlusion body, virion, and nucleocapsid form (Blissard and Rohrmann, 1990).

1.4. Baculovirus Infection

Baculovirus diseases result primarily from infection of host larvae, with the progression of the disease dependant on several factors including larval instar stage, infective viral dose, and host nutrition. Natural epizootics due to NPV and GV infection arise from insect larvae feeding on plant foliage contaminated with occlusion bodies. Baculovirus occlusion bodies protect the occluded virions from environmental damage and enable them to persist for years on plant foliage, plant debris, and in the soil. Once ingested, the alkaline pH (9.5 to 11.5) of the insect midgut dissolves the crystalline protein matrix of the occlusion body and releases the embedded virions. The released occlusion-derived virions (ODVs) make their way to the surface of the midgut where they bind to the peritrophic membrane, a non-cellular membrane composed primarily of chitin and proteins, which acts as a mechanical barrier to pathogenic microorganisms. After passing through the peritrophic membrane the ODVs fuse with the membrane of the midgut epithelial cells allowing the nucleocapsids to enter the microvilli (Blissard, 1996).

In Lepidopteran hosts, the virus enters the epithelial midgut cells, travels to the nucleus, uncoats the viral DNA, and initiates viral replication (Fig. 1.2). Non-occluded budded virions (BV) form after DNA replication and pass out of the infected epithelial midgut cell and into the insect hemolymph. The hemolymph then circulates the BVs throughout the body, facilitating the rapid transmission of the virus to other tissues, which spreads the infection. The formation of hundreds of occlusion bodies within the cells of these tissues results in the enormous hypertrophy of the infected nucleus, which in turn leads to hypertrophy of the infected cells. As a result the larvae will take on a swollen appearance near the end of the disease. Cellular lysis, followed by liquefaction of the

insect, occurs which releases the virus containing occlusion bodies back into the environment to reinitiate the infectious cycle (Blissard, 1996).

In contrast to Lepidopteran baculoviruses, Hymenopteran NPVs only infect the insect midgut. Infection of the midgut results in an infectious and to some extent a proliferation of midgut cells to replace those that are sloughed off during the latter stages of the disease. The infectious diarrhea contains thousands of occlusion bodies, which will ultimately contaminate plant foliage and soil, thus reinitiating the viral infectious cycle.

Recent studies of the mosquito baculovirus *Culex nigripalpus* NPV (CuniNPV) have increased our understanding of baculovirus development in the host midgut. Development of this virus is restricted to the midgut and is very localized. CuniNPV selectively infects the epithelial cells of the gastric caeca and the posterior stomach leaving the cardia and anterior stomach uninfected (Moser *et al.,* 2001) (Fig. 1.3). Only a single round of genome replication occurs within an infected cell, which leads to the production of both BVs and ODVs. The BVs disseminate the virus to other midgut cells. This pattern of infection is similar to that of other non-Lepidopteran NPVs, but the mechanisms by which these other NPVs spread from cell to cell in the insect midgut are not known (Moser *et al.,* 2001).

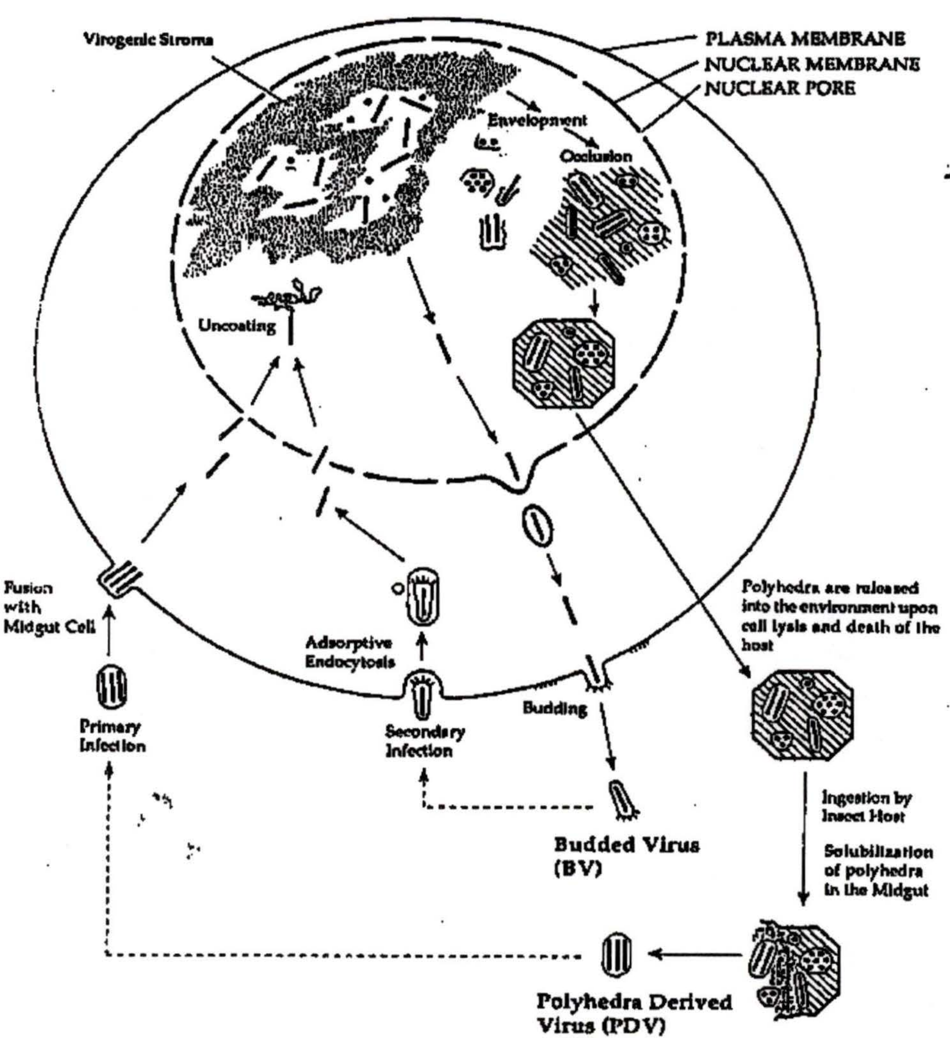


Figure 1.2. Cellular infection of a nucleopolyhedrovirus (Picture taken from Wood and Granados, 1991).

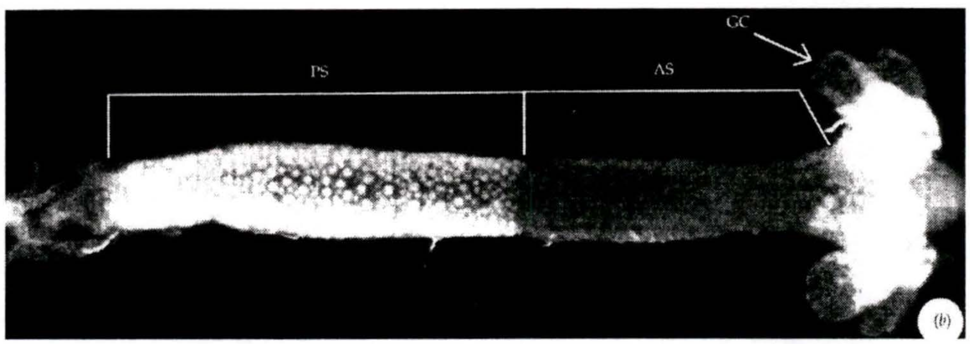


Figure 1.3. Dissected midgut from *C. nigripalpus* third-instar larvae. GC (gastric caeca), AS (anterior stomach), PS (posterior stomach). The nuclei of almost every cell of the PS are infected while cells of the AS were uninfected. Infection in the GC is primarily restricted to the proximal cells while cells in the distal portions are uninfected (Moser *et al.* 2001).

1.5. Practical Application of Baculoviruses

1.5.1. Baculoviruses as Biological Control Agents

As previously mentioned, baculoviruses are of major interest for use in many different biological control programs. Their ability to produce virions protected by a proteinaceous occlusion body enables them to be easily dispersed and survive for long periods of time in the environment. These factors make baculoviruses optimal for use as biological insecticides. Moreover, recent advances in molecular biology have provided vast amounts of sequence information on baculoviruses, which has enabled scientists to enhance baculovirus genomes in order to optimize the virus for use as biological insecticides (Miller, 1995).

For over a century baculoviruses have been recognized as useful biological control agents. Records in Germany dating back to the late 19th century recorded the collapse of a nun moth (*Lymantria monarcha*) infestation after treatment with a bacterial biological control agent. Another example, and perhaps the greatest success of any biological control agent, was the use of an NPV to control an outbreak of the European spruce sawfly (*G. hercyniae*) in the 1930's. In an effort to control the outbreak of the European spruce sawfly, Hymenopterous parasites, which inadvertently contained NPVs, were introduced into infested forests. The virus was first detected in 1936 and by 1938 it had spread to over 31,000 km^2. The sawfly populations were reduced to acceptable levels by 1943 and still remain under natural control today (Federici, 1986).

Therefore the potential of baculoviruses to control insect infestations is very great and forestry and agriculture industries will be able to benefit from their use. As of 1991, there were seven baculoviruses registered for use against forest and agricultural insect

pests in Canada and the United States, two of which were commercial products (TM-Biocontrol-1 and Gypchek). The most successful commercial baculovirus in terms of total area treated is the NPV of the velvetbean caterpillar (*Anticarsia gemmatalis*), a major pest of soybean in Brazil. In 1989-1990 enough product, called Multigen, was sold to spray approximately 200,000 acres and these numbers increased to over 4 million acres by the mid 1990's (Moscardi, 1999).

1.5.1.1. Factors Influencing the Use of Baculovirus Pesticides

Typically, chemical pesticides have been the primary means of controlling insect infestations. Unfortunately, there have been an increasing number of problems associated with the use of these broad-spectrum chemical insecticides including environmental contamination, the development of insect resistance and the elimination of non-target or beneficial insects within the ecosystem. As a result, there is an urgent need to develop alternative, environmentally compatible, biological insecticides. Biological insecticides utilizing baculoviruses have been viewed as an attractive alternative to both chemical and other biological pesticides. The primary advantages of most baculoviruses are that they are naturally occurring and extremely host-specific, thereby not affecting other non-target or beneficial insects in the ecosystem. Evidence to date also shows that insect hosts do not develop resistance against pathogenic baculoviruses. They do not accumulate in food chains, do not interact with other pesticide residues, and do not degrade into undesirable secondary products (Bohmfalk, 1986). As well, baculoviruses have the ability to replicate in their target host, enabling them to persist or spread in the environment and maintain host populations below damaging levels for many years, as seen with the *G. hercyniae* NPV.

1.5.1.2. Factors Constraining the Greater Use of Baculovirus Pesticides

The primary factor that limits interest in baculoviruses is their slow speed in stopping pest damage and killing the host, particularly when compared to chemical insecticides. It takes a virus several days to initiate disease and kill its host, during which time the insect continues to feed and damage crops. Another limiting factor for industry is the inability to produce baculovirus *in vitro* on a large-scale. The main problems are lack of low cost, serum-free media and lack of suitable cell lines and virus strains. Large-scale production of virus is currently achieved by the infection of laboratory-reared larvae. This method is cheap, but it is also labour intensive, time consuming and is commercially unattractive in developed nations. As mentioned above, baculoviruses are highly host-specific which presents a problem to industry, but is beneficial to the environment. Host-specificity is advantageous in conserving beneficial insects but it limits the potential market for their use and decreases their economic return. From the perspective of the farmer, host-specific insecticides mean having to buy several insecticides for each pest that threatens his/her crops, which is uneconomical when compared to the price of a single broad-spectrum chemical pesticide. Ideally farmer education about the use of biological insecticides and their potential is needed to overcome this obstacle along with providing information about the harmful environmental effects associated with the use of chemical pesticides.

Currently, industry interest in the use of baculoviruses is increasing due to the build-up of insect resistance to commonly used chemical pesticides, banning of various insecticides in certain countries, increasing demand from the public for safer pesticides and less chemicals in the food and water. As well, increased knowledge of baculovirus

genetics has enabled scientists to genetically modify baculoviruses and increase their infectivity.

1.5.2. Genetic Enhancement of Baculovirus Pesticides

The emergence of recombinant DNA technology provided the most significant drive for the continued growth of biological control research. The quest to understand the genes and gene products that are important to biological control has led to the ability to genetically alter baculoviruses, which may prove to be useful for pest control.

The most commonly used strategy for insertion and expression of foreign genes in baculoviruses is through the use of the baculovirus expression vector system (BEVs). The basis of this system is the replacement of the polyhedrin coding sequence with a foreign gene sequence. Polyhedrin is under the control of a strong, very late promoter and is expressed at extremely high levels during the late stages of viral infection. As a result, the engineered foreign gene sequence is highly expressed, providing copious amounts of foreign protein to the infected cell (Jarvis, 1997). Many different foreign gene inserts have been used including insect-specific toxins, insect hormones, insect hormone receptors, metabolic enzymes, and growth regulators. Expression of these foreign genes is aimed at enhancing the speed at which the baculovirus can kill the infected larvae to stop their feeding (Chen *et al.,* 2000).

The genetic improvement of HzNPV as a biological control agent demonstrates the potential benefit of genetic engineering. HzNPV has been previously registered and commercially produced as a pesticide called Elcar. The product did not compete successfully however, against several types of pesticides due to its slow speed of kill. Popham *et al.* (1997) reduced the time required for HzNPV to kill its insect host larvae by

inserting a potent insect-selective mite neurotoxin gene (*tox*34) into an HzNPV gene that encodes ecdysteroid UDP-glucosyltransferase (*egt*). EGT is a viral protein that inactivates insect larval ecdysteroids by transferring a molecule of glucose from UDP-glucose (O'Reilly and Miller, 1989). Ecdysteroids are a family of steroid hormones that are responsible for causing molting in insect larvae. Upon inactivation of the *egt* gene the ecdysteroids were able to cause larval molting and stop the larvae from feeding. In another experiment performed by Popham *et al.* (1997) the *egt* gene was replaced with a *tox*34 gene placed under the control of an early viral promoter (*ie-1* promoter). This caused the tox34 protein to be expressed early in the infectious cycle and resulted in a 50% increase in mortality within 40 hours after infection. This increased rate of mortality represented a significant improvement in kill time over the wild type virus.

1.6. The Baculovirus Genome

1.6.1. Overview

Advancements in DNA sequencing technology and computer assisted sequence analysis in the last decade have provided a powerful means of determining the genetic content of baculovirus genomes and investigating their gene organization and evolution. Eleven NPVs (AcMNPV, OpMNPV, BmMNPV, LdMNPV, EppoNPV, MacoNPV, SeMNPV, SpltMNPV, CuniNPV, HearSNPV, and HzSNPV) and three GVs (XcGV, PxGV, CpGV) have been completely sequenced and analysis of this data has given us an initial view of the content and organization of baculovirus genomes. The discovery of common and unique genes will provide us with an abundance of useful information. Genes that are retained in all baculovirus species will help provide insight into the ancient

origins of the *Baculoviridae* family, while unique genes will provide us with a view of how baculoviruses are currently changing.

The first complete baculovirus gene to be sequenced was the AcMNPV polyhedrin gene. Shortly after, the entire sequence of the AcMNPV genome was completed, making it the first baculovirus genome to be completely sequenced (Ayres *et al.,* 1994). Several properties of AcMNPV, such as the ability to propagate the virus *in vitro* and its relative stability in cell culture systems, made it an ideal choice for molecular biological studies. As a result, much of what is known about baculoviruses today comes from studies done on various strains of AcMNPV.

As mentioned previously, eleven NPVs and three GVs have been completely sequenced providing a wealth of information about the structure, organization and distinctive features of baculovirus genomes. The great diversity between baculoviruses can be seen in the size range of their genomes (~80-180kb) and number of open reading frames (ORFs) (120-180 larger than 50 amino acids). Approximately 400 different baculovirus ORFs have been discovered, 65 of which appear to be universally conserved among the *Baculoviridae* (Hayakawa *et al.,* 2000). These 65 genes are likely fundamental for the survival of the virus and constitute the "essence" of being a baculovirus.

1.6.2. Common Baculovirus Genes

Approximately 65 predicted genes are commonly shared by all of the completely sequenced NPVs and GVs. The 65 common baculovirus genes include transcription-specific genes, DNA replication genes, structural genes, and what have been called "auxiliary" genes. Auxiliary genes are not necessary for viral replication but instead control some kind of host process, which confers a selective advantage upon the virus.

Why do all baculoviruses share this set of genes and what are their functions? Below is a brief description of some of the commonly shared baculovirus genes.

1.6.2.1. Transcription-Specific Genes

Baculovirus gene expression is carried out by a variety of virus-encoded proteins, particularly genes expressed late in the infectious cycle. *Lef* (late gene expression factor) genes are a set of baculovirus genes required for the transactivation of gene expression from late (e.g. *vp-39*) and very late (e.g. *polh*) promoter sequences. Studies have shown that baculoviruses encode a novel α-amanitin resistant RNA polymerase II enzyme, which is responsible for transcription of late and very late genes. During the viral replication cycle, host RNA polymerase II is utilized for immediate early and early viral gene transcription. Transition from the early to late stages of viral infection results in a decreased level of activity from host RNA polymerase II and production of a virus encoded α-amanitin resistant RNA polymerase II (Blissard, 1996). Transcription from viral RNA polymerase II shows significantly higher levels of gene expression than that of the host RNA polymerase II (Ooi *et al.*, 1989).

LEF-4, LEF-8, and LEF-9 are conserved in all the sequenced baculoviruses and are thought to help form a complex that is involved with late RNA polymerase activity. LEF-8 and LEF-9 proteins also contain amino acid sequence motifs that are conserved in prokaryotic and eukaryotic RNA polymerases, which provides further support to the hypothesis that LEF-8 and LEF-9 are involved in late RNA polymerase activity (Guarino *et al.*, 1998).

1.6.2.2. Replication-Specific Genes

One major group of conserved baculovirus genes are the DNA replication genes. Five essential replication genes (*p143, ie-1, lef-1, lef-2,* and *lef-3*) and a virus encoded DNA polymerase (*dnapol*) are required for DNA replication and appear to be conserved among the sequenced baculovirus genomes (Hayakawa *et al.,* 2000).

Most DNA viruses replicate in the nucleus of the host cell since all the biochemical apparatus necessary for DNA replication is present within the nucleus (Cann, 1996). Normally DNA viruses do not encode a DNA polymerase and instead rely on host DNA replication proteins, but some viruses have evolved to become more independent. For instance, both herpesviruses and baculoviruses replicate in the host cell nucleus but encode their own virus-specific DNA polymerase. Herpesvirus DNA is replicated solely by this virus-specific DNA polymerase whereas baculoviruses utilize both viral and host DNA polymerase. In fact, it has been shown that baculovirus DNA polymerase is not essential to DNA replication and instead acts more to stimulate replication (Lu and Miller, 1995). This suggests that the essential baculovirus specific DNA replication genes can act in concert with the host DNA polymerase to replicate the genome. These DNA replication genes have been found to be very similar to a number of genes required for DNA replication in the herpesvirus family. As a result, the herpesvirus DNA replication system, which has been well characterized, is used as a model to describe baculovirus DNA replication.

1.6.2.3. Structural Genes and Proteins

Baculoviruses undergo a unique biphasic lifecycle involving two viral phenotypes, the budded virus form (BV) and the occlusion-derived virus form (ODV). The BV form is created when viral nucleocapsids move out of the host cell nucleus and bud through the plasma membrane, acquiring a membranous envelope in the process. This enveloped nucleocapsid, or BV, contains proteins (e.g. *gp67)*, which enable the virus to recognize and infect several different tissue types within the host insect, thereby assisting in the spread of the infection (Federici, 1997). In contrast, the ODV form is created when viral nucleocapsids are enveloped within the cell nucleus and are occluded within polyhedra (NPVs) or granules (GVs). The process of ODV envelopment within the cell nucleus is still not very well understood but is thought to involve a de novo assembled envelope. The ODV envelope contains proteins (e.g. *p74*) that are essential for infectivity of the virus within the host midgut (Faulkner *et al.,* 1997). Both viral forms are essential for the natural propagation of the occluded baculovirus and perform specialized functions within the insect host. These functions are likely regulated, in part, by different proteins present in the membranous envelopes of the two viral forms (Braunagel and Summers, 1994). For instance, experiments performed by Faulkner *et al.* (1997) demonstrated that the deletion of the *p74* gene resulted in the complete loss of infectivity of viral occlusion bodies when fed to insect larvae.

Among the 14 baculovirus genomes currently sequenced, fifteen structural genes are found conserved between them including occlusion body proteins (polyhedrin and granulin), nucleocapsid-associated proteins (*vp-39* and *vp-91*), budded virus envelope-associated proteins *(gp-41)*, occlusion-derived virion proteins *(p74* and *odv-e56)* and DNA binding proteins *(p6.9)* (Funk *et al.,* 1997). Few studies, however, have been

performed that compare the structural proteins of the BVs and ODVs and there is limited information on the protein composition and similarities/differences of BV and ODV envelope and nucleocapsid proteins.

1.6.2.4. Auxiliary Genes

The evolution of larger viral genomes has enabled viruses to carry a wider array of genes. Many of these genes are still considered essential and help to reduce the dependency of the virus to particular parts of the host cell machinery (e.g. DNA replication proteins). This can be seen in poxviruses, which have evolved (or acquired) all the necessary genes for transcription and replication of their genomes. As a result poxviruses are able to replicate in the cytoplasm of the infected cell and do not need to rely on the nucleus for transcription and replication enzymes. In addition to these essential genes, baculoviruses have also acquired non-essential genes that confer upon the virus a selective advantage against its host (O'Reilly, 1997). A few such auxiliary genes are *sod*, *ubi*, and *iap*.

Superoxide dismutases (*sod*) catalyze the dismutation of the O_2^- superoxide radical into H_2O_2 and O_2. The O_2^- anion is a by-product of oxygen utilization in aerobic organisms and is responsible for most of the cellular toxicity caused by oxygen. The function of viral *sod* genes is presently unknown but it has been hypothesized that *sod*s help to protect occluded virions in the environment from superoxide radicals generated by exposure to sunlight (Tomalski *et al.,* 1991).

Ubiquitin (*ubi*) is one of the most abundant and highly conserved proteins known, being found in all eukaryotes and having only a three amino acid residue difference between animals and yeast. Baculoviruses are the only known viruses to encode ubiquitin

and have the most divergent forms of this gene (AcMNPV ubiquitin shows 76% identity to animal ubiquitin). The principal function of ubiquitin in cells is to target other proteins for degradation, but its function in baculoviruses is still unknown. Studies have shown, however, that viral ubiquitin mutants produce five to tenfold less progeny BV than wild type strains, indicating that ubiquitin plays a major role in the development of progeny virus during the infectious cycle (Haas *et al.,* 1996).

Many viruses, including baculoviruses, induce host cells to undergo controlled cell death, or apoptosis. In most cases the resultant cell death is detrimental to the virus as the apoptotic cell environment is not optimal for viral transcription and replication. To overcome this host defence mechanism, baculoviruses have acquired *iap* genes. IAPs (inhibitors of apoptosis) are a family of genes involved in blocking apoptosis and have been discovered in all the baculoviruses sequenced to date (Hayakawa *et al.,* 2000). The exact mechanism by which *iap* genes block apoptosis is not yet known, but it has been hypothesized that the *iap* proteins interact with components of particular signal transduction pathways, such as caspases, and prevent the cell from signalling its own suicide (Clem, 1997).

Other auxiliary genes found in some baculoviruses include: 1) the ecdysteroid UDP-glucosyltransferase gene (*egt*), which suppresses host molting and helps the virus replicate (O'Reilly and Miller, 1989); 2) the cathepsin (*cath*) and chitinase (*chi*) genes, which facilitate host cuticle breakdown and help to release viral occlusion bodies into the environment (Hawtin *et al.,* 1997); and 3) the conotoxin-like (*ctl*) genes, which may act as paralytic insect toxins (Eldridge *et al.,* 1992). These genes provide fascinating examples of how well baculoviruses can manipulate and control their environment (the host insect cell) and demonstrate the high level of sophistication that viruses can achieve.

1.6.3. Repetitive DNA Sequences

1.6.3.1. Coding Sequences

The great diversity of baculoviruses is reflected in the size range of their genomes. Fifteen baculovirus genomes have been fully sequenced and range in size from 101kb (*Plutella xylostella* GV) to 180kb (*Xestia c-nigrum* GV). Smaller genome sizes have been determined, however, from restriction digests of several Hymenopteran baculoviruses. The smallest baculovirus genome was estimated to approximately 82kb for the *Neodiprion lecontei* nucleopolyhedrovirus (Arif, B. unpublished). Why do the genomes vary so much in size? What is the source of this variability? Baculoviruses contain a large variety of duplicated DNA sequences, including both gene families and non-coding intergenic sequences (called homologous regions, hrs). The largest fully sequenced baculovirus (XcGV) contains over 20 different repeated genes, which account for 37.5kb, or 20%, of the genome (Hayakawa *et al.,* 1999). In other baculoviruses repetitive DNA sequences (duplicated genes or repeating non-coding DNA sequences) contribute much less to genome size (2% in AcMNPV). Some of the repetitive gene families (two or more duplicated genes) found within baculovirus genomes include the *vef*, *iap*, *ctl* and *bro* gene families.

1.6.3.1.1. The VEF Gene Family

Viral enhancing factors (*vefs)* (also known as enhancins) are proteins that have demonstrated the ability to "enhance" viral infectivity. Genes encoding enhancins have been identified in several GVs, but are only found in one NPV (LdMNPV) In TnGV, the VEF protein forms about 5% of the occlusion body mass and seems to facilitate

baculovirus infection. This occurs by disrupting the peritrophic membrane, thereby allowing virions access to epithelial midgut cells (Kuzio *et al.,* 1999).

1.6.3.1.2. The IAP Gene Family

As mentioned above, *iap* genes encode proteins that block the host cells from undergoing apoptosis caused by baculovirus infection. There are four different baculovirus *iap* genes, three of which encode full-length proteins (*iap* 1-3) and one which encodes a truncated protein (*iap*-4). IAP proteins contain two easily identifiable sequence motifs. Near the C-terminus of the protein is a type of zinc finger called a RING finger (Clem, 1997). At the N-terminus are two imperfect repeated sequences known as baculovirus *iap* repeats (BIRs). Both motifs are capable of binding zinc and undoubtedly are important for allowing IAP proteins to assume the proper conformation (Ahrens *et al.,* 1997). The i*ap* genes are found repeated in several NPVs, but are only found in single copies in GVs.

1.6.3.1.3. The CTL Gene Family

Baculovirus conotoxin-like peptides (*ctl*) are small (53 amino acid), cysteine-rich peptides identified by their sequence similarity to the ω-conotoxins, a family of calcium channel antagonists found in predatory marine snails of the genus *Conus*. The baculovirus CTL peptides contain domains that are highly similar to those found in animal conotoxins. Multiple copies of the *ctl* gene are found in the OpMNPV and LdMNPV genomes only. The function of these toxins is currently unknown but it has been

hypothesized that they cause some form of paralysis in infected insects (Eldridge, *et al.*, 1992).

1.6.3.1.4. The BRO Gene Family

A common characteristic of baculovirus genomes is the presence of a group of related genes that have been called the baculovirus repeated ORFs, or *bro* genes. *Bro* genes are unique to the *Baculoviridae*, are present in all fully sequenced baculoviruses, except PxGV, and are the most highly repeated gene family (16 copies in LdMNPV). Many of the *bro* genes found within a single baculovirus have been shown to share a high sequence similarity to each other and are probably the result of gene duplication events. This could be due in part to their close proximity to homologous regions (hrs), the non-coding DNA sequences which have been suggested to mediate recombination.

The *bro* gene N-terminus has been shown to share sequence similarity to proteins of unknown function from several viruses including, ORF266 of bacteriophage BK5-T, ORF5 of bacteriophage rlt, ORFA of bacteriophage A2 and HI 1418 of *Haemophilus influenzae* Rd (Hayakawa *et al.*, 2000). Families of related genes have been seen in other virus species, including African swine fever virus and *Molluscum contagiosum* virus (MCV), yet no similarity exists when families of proteins or genes are compared between different virus groups (Kuzio *et al.,* 1999). The function of the *bro* genes is still unknown but recent studies with BmMNPV have demonstrated that BRO proteins are capable of binding to nucleic acids (Zemskov *et al.,* 2000).

1.6.3.2. Non-Coding DNA Sequences

1.6.3.2.1. Homologous Regions (Hrs)

Homologous regions (*hrs*) appear to be a novel feature of baculovirus genomes and are present in all the baculoviruses sequenced so far. *Hrs* are non-coding segments of DNA comprised of repeating sequences and are found interspersed throughout the viral genome. They were first characterized in AcMNPV, where 9 unique *hrs* were identified (called *hr1*, *hr1a*, *hr2*, *hr2a hr3*, *hr4a*, *hr4b*, *hr4c* and *hr5*). It was originally thought that *hrs* had conserved locations within different baculovirus genomes, but this hypothesis was later proven to be false (Ahrens *et al.,* 1997). The structure of *hrs* varies between viruses but a conserved nucleotide structure can be seen. All *hrs* are comprised of repeating segments of DNA, which contain a conserved core 30-bp imperfect palindrome flanked by direct DNA repeats, which vary in length. Studies have shown that *hrs* can act as enhancers of RNA polymerase II mediated transcription (Guarino and Summers, 1986) and as origins of DNA replication in transient replication assays (Leisy *et. al.,* 1995) and may also play a part in the evolution of baculovirus genomes. The close relatedness of the *hr* sequences within each viral genome suggests that most of the viral *hrs* descended from a common viral ancestor and that the *hr* sequences undergo coevolution, possibly evolving with a protein (such as IE-1) with which they interact.

1.7. Genome Organization

Comparative studies of baculovirus genomes have helped to improve our understanding of the similarities and differences found between different baculovirus species. Recent comparative studies using GeneParityPlots (Hu *et al.,* 1998) have shown

that NPV genomes have a high degree of genomic variability, although large regions of partially conserved genes are apparent. In contrast, GVs show a very high degree of conservation in the organization of their genomes with only 5 of approximately 100 homologous ORFs not conserved in the same order and orientation within each genome (Hayakawa *et al.,* 2000). GeneParityPlots compare the genomes of two different viruses, placing the gene order of one genome on the x-axis and the gene order of another genome on the y-axis. By comparing two genomes in this fashion one can see the relative genome position and direction of homologous gene clusters within the plot. Fourteen gene clusters (a, b, c, d, e, f, g, h, i, j, k, l, m, n) are conserved in all sequenced baculoviruses, with small variations in cluster structure occurring in different species (Fig. 1.4). These gene clusters are of particular interest from a molecular evolution standpoint as they can be used as phylogenetic markers to help study the ancestral relationship of baculoviruses, independent of the evolution of individual genes.

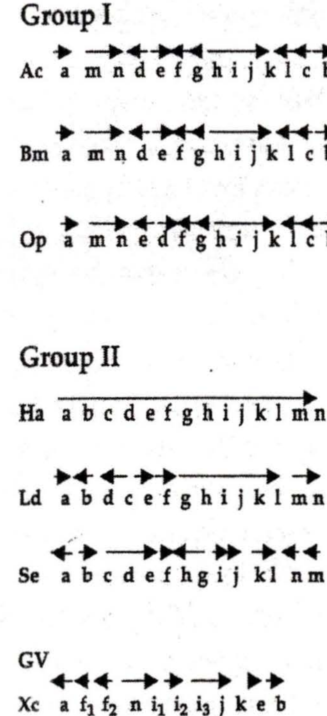


Figure 1.4. Comparison of Group I NPV and Group II NPV baculovirus genome clusters (Chen *et al.*, 2001). *A. californica* NPV (Ac), *B. mori* NPV (Bm), *O. pseudostugata* NPV (Op), *H. armigera* NPV (Ha), *L. dispar* NPV (Ld), *S. exigua* NPV (Se), *X. c-nigrum* GV (Xc)

1.8. Baculovirus Phylogeny

1.8.1. Origins and Relationships of Baculoviruses

Fundamental to understanding the origin of baculoviruses is an appreciation of their diversity. Baculoviruses have been isolated from over 600 insect species and have been reported from several species of crustaceans as well. Crustaceans are known to be a very ancient group of invertebrates, appearing almost 600 million years ago during the pre-Cambrian period. If baculoviruses were infecting species of Crustaceans 600 million years ago, then it may be possible that these viruses have evolved into a vast number of unique viral species (Granados and Federici, 1986). As previously mentioned, baculoviruses are attractive candidates for use in biological control programs due to their narrow host range. This host specificity provides non-target insects with a natural barrier to infection and will ultimately determine the use of baculoviruses in particular control programs. As well, host specificity will be relevant to safety assessments of genetically modified baculovirus genomes, should they ever be used (Bulach *et al.,* 1999). Phylogenetic studies between different baculoviruses will help improve our understanding of the relationships between baculovirus genomes and will hopefully provide some insight into particular baculovirus biological adaptations such as host-specificity and host range.

Baculovirus phylogenies are determined by comparing the similarities and differences between particular gene sequences and viral genome organization and gene content, and using the data to build phylogenetic trees. Typically gene trees have relied on occlusion protein sequences (Fig 1.5) since they are the most widely available baculovirus sequences, but other gene sequences have also been used, including the major

capsid protein VP39 (Fig. 1.6), the hormone inactivating protein EGT, and DNA polymerase.

Occlusion protein sequences are not ideal for classifying the *Baculoviridae* due to the small size of the protein sequence and the high number of invariant residues, which provide few data sites for phylogenetic estimations (Bulach *et al.,* 1999). Sequence data for a large number of viral DNA polymerases (~3000bp) have been determined and used in phylogenetic studies. Such studies have shown that AcMNPV and LdMNPV DNA polymerases are related to DNA polymerases from viruses outside the *Baculoviridae*, including entomopoxviruses (Braithewiat and Ito, 1993).

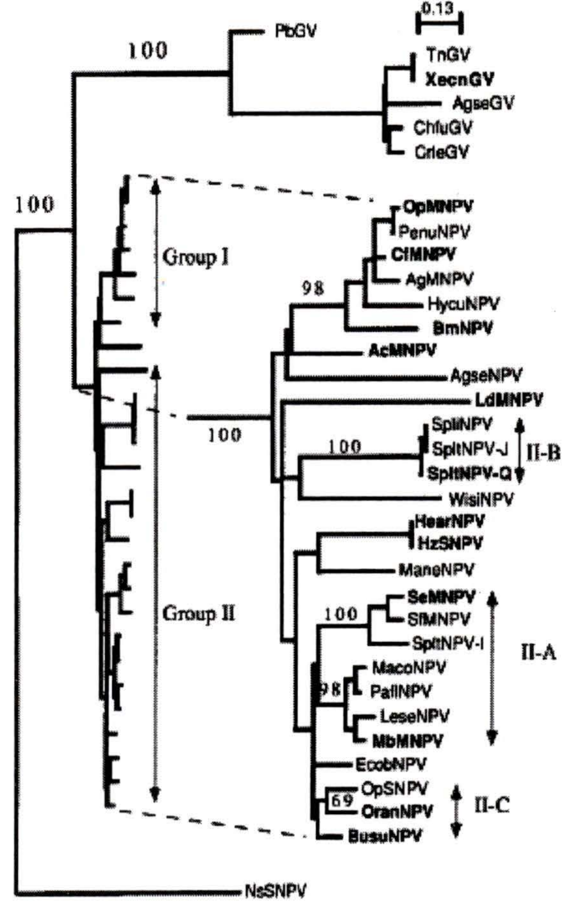


Figure 1.5. Occlusion protein (polyhedrin and granulin) evolutionary gene tree. Percentages of bootstrap support (100 replicates) greater than 50% are shown along the branches. This phylogeny is similar to that estimated using the baculovirus DNA polymerase gene (Bulach *et al.,* 1999).

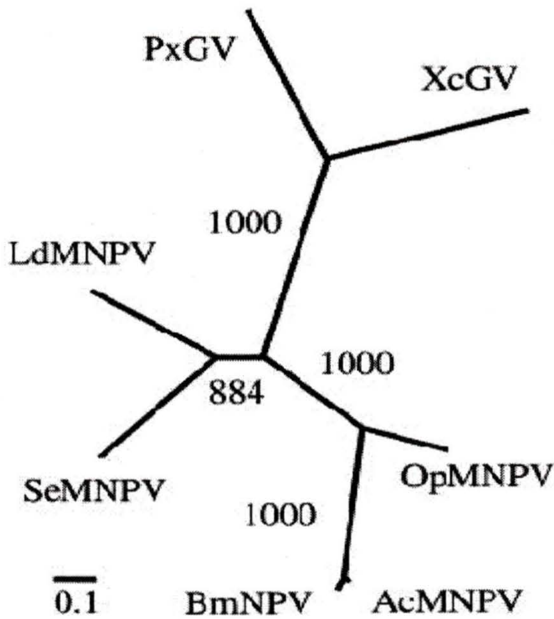


Figure 1.6. Phylogenetic tree constructed using the major capsid protein VP-39. OpMNPV, AcMNPV, and BmMNPV form part of the Group I baculoviruses. LdMNPV and SeMNPV form part of the Group II baculoviruses (Hayakawa *et al.,* 2000).

1.9. The Balsam Fir Sawfly (*Neodiprion abietis)* and NeabNPV

The balsam fir sawfly (*Neodiprion abietis*) is a native insect species (Order: Hymenoptera) whose population has reached infestation levels in the eastern Canadian province of Newfoundland. The increased populations of the balsam fir sawfly (BFS) threaten to destroy tens of thousands of hectares of forest, resulting in millions of dollars in annual damage. There is currently no acceptable control agent for use against the BFS, but studies are underway to develop an alternative, environmentally compatible, biological insecticide, which utilizes a NPV specific to *N. abietis* (NeabNPV). Tests involving aerial spraying of NeabNPV occlusion bodies onto infested balsam fir forest test plots indicate that NeabNPV is capable of controlling the insect infestations. However, virtually nothing is currently known about the genetic composition or organization of the NeabNPV genome and development of this virus as a biological control agent will require further basic research.

1.9.1 Objectives

The objectives of this study were to:

1) Develop a restriction endonuclease library for the NeabNPV *Hin*dIII digest

2) Sequence the end-termini of each cloned restriction fragment and analyse the data

3) Develop a NeabNPV genomic library by nebulization of the genome

4) Sequence the NeabNPV genome and analyse the data

5) Characterize proteins of NeabNPV using SDS-PAGE and peptide sequence data

1.10. References

Ahrens, C. H., Russell, R. L., Funk, C. J., Evans, J. T., Harwood, S. H., and Rohrmann G. F. (1997). The Sequence of the *Orgyia pseudotsugata* Multinucleocapsid Nuclear Polyhedrosis Virus Genome. *Virol.* **229**, 381-399.

Atlas, M. R. (1995). Viral Replication. Principals of Microbiology. Mosby-Year Book, Inc., St. Louise.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M., and Possee, R. D. (1994). The Complete DNA Sequence of *Autographa californica* Nuclear Polyhedrosis Virus. *Virol.* **202**, 586-605.

Beames, B., and Summers, M. D. (1988). Comparisons of host cell DNA insertions and altered transcription at the sites of insertions in few polyhedra (FP) baculovirus mutants. *Virol.* **162**, 344-353.

Blissard, G. W., and Rohrmann, G. F. (1990). Baculovirus Diversity and Molecular Biology. *Annu. Rev. Entomol.* **35**, 127-155.

Blissard, G. W. (1996). Baculovirus-insect cell interactions. *Cytotechnology.* **20**, 73-93.

Botstein, D. (1980). A Theory of Modular Evolution for Bacteriophages. *Annals of the New York Academy of Sciences.* **345**, 485-491.

Bohmfalk, G. T. (1986). Practical Factors influencing the Utilization of Baculoviruses as Pesticides. The Biology of Baculoviruses. R. R. Granados and B. A. Federici. Florida, CRC Press, 223-236.

Braithwaite, D. K., and Ito, J. 1993. Compilation, alignment, and phylogenetic relationships of DNA polymerases. *Nucleic Acids Res.* **21**, 787-802.

Braunagel, S. C. and M. D. Summers (1994). Autographa californica nuclear polyhedrosis virus, PDV, and ECV viral envelopes and nucleocapsids: structural proteins, antigens, lipid and fatty acid profiles. *Virol.* **202**, 315-28.

Bulach, D. M., Kumar, A., Zaia, A., Linag, B., and Tribe, D. E. (1999). Group II Nuclepolyhedrovirus Subgroups Revealed by Phylogenetic Analysis of Polyhedrin and DNA Polymerase Gene Sequences. *J Inverteb. Pathol.* **73**, 59-73.

Canne, A. J. (1997). Principals of Molecular Virology. 2nd Ed. San Diego, Academic Press. 115-117.

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

Chen, X., Sun, X., Hu, Z., Li, M., O'Reilly, D. R., Zuidema, D., Vlak, J. M. (2000). Genetic engineering of *Helicoverpa armigera* single nucleopolyhedrovirus as an improved bioinsecticide. *J. Inver. Pathol.* **76**, 140-146.

Clem, R. J. (1997). Regulation of programmed cell death by baculoviruses. In *The Baculoviruses*, pp. 237-266. Edited by L. K. Miller. New York: Plenum Press.

Croizier, G., Croizier, L., Argaud, O., and Poudevigne, D. (1994). Extension of *Autographa californica* Nuclear Polyhedrosis Virus Host Range by Interspecific Replacement of a Short DNA Sequence in the *p143* Helicase Gene. *PNAS.* **91**, 48-52.

Eldridge, R., Li, Y., and Miller, L. K. (1992). Characterization of a baculovirus gene encoding a small conotoxin-like polypeptide. *J. Virol.* **66**, 6563-6571.

Faulkner, P., J. Kuzio, et al. (1997). Analysis of p74, a PDV envelope protein of *Autographa californica* nucleopolyhedrovirus required for occlusion body infectivity in vivo. *J. Gen. Virol.* **78**, 3091-100.

Federici, B. A. (1997). Baculovirus Pathogenesis. In *The Baculoviruses*, pp. 33-59. Edited by L. K. Miller. New York: Plenum Press.

Funk, J., Braunagel, S.C., and Rohrmann, G.F. (1997). Baculovirus structure. In *The Baculoviruses*, pp. 33-59. Edited by L. K. Miller. New York: Plenum Press.

Gomi, S., Majima, K., and Maeda, S. (1999). Sequence of the Genome of *Bombyx mori* nucleopolyhedrovirus. *J. Gen. Virol.* **80**, 1323-1337.

Granados, R. R., and Federici, B. A. (1986). The Biology of Baculoviruses: Biological Properties and Molecular Biology, Vol. I. CRC Press, Florida.

Guarino, L. A., and Summers, M. D. (1986). Interspersed homologous DNA or *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *J. Virol.* **60**, 215-223.

Guarino, L. A., Xu, B., Jin, J., and Dong, W. (1998). A virus-encoded RNA polymerase purified from baculovirus-infected cells. *J. Virol.* **72**, 7985-7991.

Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita, T., Sano, Y., and Matsumoto, T. (2000). Sequence Analysis of the *Plutella xylostella* Granulovirus Genome. *Virol.* **275**, 358-372.

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

Hayakawa, T., Ko, R., Okano, K., Seong, S., Goto, C., and Maeda, S. (1999). Sequence Analysis of the *Xestia c-nigrum* Granulovirus Genome. *Virol.* **262**, 277-297.

Hayakawa, T., Rohrmann, G. F., and Hashimoto, Y. (2000). Patterns of Genome Organization and Content in Lepidopteran Baculoviruses. *Virol.* **278**, 1-12.

Ijkel, W., Strien, E. A., Heldens, J., Broer, R., Zuidema, D., Goldbach, R. W., and Vlak, J. M. (1999). Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J. Gen. Virol.* **80**, 3289-3304.

Haas, A. L., Katzung, D. J., Reback, P. M., and Gaurino, L. A. (1996). Functional characterization of the ubiquitin variant encoded by the baculovirus *Autographa californica* nucleopolyhedrovirus. *Biochemistry* **35**, 5385-5394.

Hu, Z. H., B. M. Arif, et al. (1998). Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome. *J Gen Virol.* **79**, 2841-51.

Jarvis, D. L. (1997). Baculovirus expression vectors. In *The Baculoviruses*, pp. 389-431. Edited by L. K. Miller. New York: Plenum Press.

Kuzio, J., Peason, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M., and Rohrmann, G. F. (1999). Sequence and Analysis of the Genome of a Baculovirus Pathogenic for *Lymantria dispar*. *Virol.* **253**, 17-34.

Levin, D. B., and Whittome, B. (2000). Codon usage in nucleopolyhedroviruses. *J. Gen. Virol.* **81**, 2313-2325.

Leisy, D. J., Rasmussen, C., Kim, H. T. and Rohrmann, G. F. (1995). The *Autographa californica* nuclear polyhedrosis virus homologous region 1a: identical sequences are essential for DNA replication activity and transcriptional enhancer function. *Virol.* **196**, 722-730.

Lewin, B. (1987). Genes III. New York, Wiley.

Lu, A. and L. K. Miller (1995). The roles of eighteen baculovirus late expression factor genes in transcription and DNA replication. *J Virol.* **69**, 975-82.

Martignoni, M. E., and Iwai, P. J. (1986). *A catalog of viral diseases of insects, mites, and ticks*, 4th edn. Portland Oregon: USDA Forest Services PNW-195.

McDougal, V. V. and L. A. Guarino (2000). The Autographa californica nuclear polyhedrosis virus *p143* gene encodes a DNA helicase. *J Virol.***74**, 5273-9.

McDougal, V. V. and L. A. Guarino (2001). DNA and ATP binding activities of the baculovirus DNA helicase P143. *J Virol.* **75**, 7206-9.

Miller, L. K. (1995). Genetically engineered insect virus pesticides: present and future. *J Invertebr Pathol* **65**, 211-6.

Miller, L. K. (1997). Introduction to the baculoviruses. In *The Baculoviruses,* pp. 1-6. Plenum Press, New York.

Moscardi, F. (1999). Assessment of the Applications of Baculoviruses for Control of Lepidoptera. *Annu. Rev. Entomol.* **44**, 257-289.

Moser, B. A., Becnel, J. J., White, S. E., Afonso, C., Kutish, G., Shanker, S., and Almira, E. (2001). Morphological and molecular evidence that *Culex nigripalpus* baculovirus is an unusual member of the family *Baculoviridae. J. Gen. Virol.* **82**, 283-297.

Ooi, B. G., Rankin, C., and Miller, L. K. (1989). Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. *J. Virol.* **210**, 721-736.

O'Reilly, D. R., and Miller, L. K. (1989). A baculovirus blocks insect molting by producing an ecdysteroid UDP-glucosyl transferase. *Science.* **245**, 1110-1112.

O'Reilly, D. R., Miller, L. K., and Lucklow, V. A. (1992). Baculovirus Expression Vectors: A Laboratory Manual. W. H. Freeman, New York.

O'Reilly, D. R. (1997). Auxiliary genes of baculoviruses. In *The Baculoviruses*. Edited by L. K. Miller. New York: Plenum Press.

Olszewski, J. and L. K. Miller (1997). A role for baculovirus GP41 in budded virus production. *Virol.* **233**, 292-301.

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

Rohrmann, G. F. (1986). Polyhedrin structure. *J Gen Virol.* **67**, 1499-513.

Russels, P. J. (1998). Transposable Elements. *In* Genetics, Fifth Ed. The Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Sadler, T. J., Glare, T. R., Ward, V. K., and Kalmakoff, J. (1999). Physical and genetic map of the *Wiseana* nucleopolyhedrovirus genome. *J. Gen. Virol.* **81**, 1127-1133.

Tomalski, M. D., Eldridge,R., and Miller, L. K. (1991). A baculovirus homolog of a Cu/Zn superoxide dismutase gene. *Virol.* **184**, 149-161.

Wood, H. A., and Granados, R. R. (1991). Genetically Engineered Baculoviruses as Agents for Pest Control. *Annu. Rev. Microbiol.* **45**, 69-87.

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

Chapter 2. Development of Partial DNA Libraries for the *Neodiprion abietis* Nucleopolyhedrovirus

2.1. Introduction

The *Baculoviridae* are a large, diverse family of more than 600 species (Murphy *et al.,* 1995), which is comprised of two genera: the *Nucleopolyhedroviruses* (NPV) and *Granuloviruses* (GV). Baculoviruses are rod-shaped virions that contain circular, double-stranded DNA genomes ranging from approximately 80 to 180 kb in size. The *Baculoviridae* are characterized by the embedding of virion particles into large, proteinaceous inclusion bodies (PIBs). PIBs have a polyhedral shape in the NPVs and a granular shape in the GVs. NPV nucleocapsids are found both singly (SNPV) and multiply (MNPV) enveloped whereas GV nucleocapsids are primarily singly enveloped. Phylogenetic studies of SNPVs and MNPVs have shown that these morphotypes are not important traits with respect to virus classification, as SNPVs were found to radiate from within what was considered to be a MNPV specific clade (Zanotto *et al.,* 1993).

Baculoviruses are pathogenic to Arthropods, in particular insects, and have been proposed as alternative control agents to commercial chemical pesticides. Because of their host specificity and ability to persist in the environment, baculoviruses are an attractive means to specifically control insect pests with minimal adverse environmental effects (Moscardi, 1999). However, when compared to chemical pesticides, baculoviruses have several drawbacks. Their slow speed of kill (due to the time needed for the virus to initiate an infection) and narrow host range as opposed to the broad range, rapid knockdown effect of chemical pesticides, is problematic to farmers, limiting their use. Knowledge of the molecular biology and genetics of baculoviruses has grown

significantly over the past decade and has enabled the genetic manipulation of certain baculoviruses. The introduction of new genes to (and deletion of genes from) baculovirus genomes has been shown to broaden the host range and greatly increase the speed of kill of baculoviruses (Wood and Granados, 1991). Studies have demonstrated that the introduction of insect-selective neurotoxic genes into the genome of HzSNPV greatly enhanced the baculoviruses speed of kill (Miller, 1995).

The balsam fir sawfly (*Neodiprion abietis*) is a native insect species whose population is currently causing severe economic damage to forests in Newfoundland. Due to environmental regulations, common broad-spectrum pesticides are banned from being used, and as a result, there is currently no acceptable means of pest control. The use of a baculovirus (NeabNPV) has been proposed as a potential means to control this insect infestation and controlled field release studies are currently underway to determine the effectiveness of NeabNPV. Knowledge of the genetic makeup and organization of the NeabNPV genome will be required for registration of NeabNPV as an insecticide.

I have attempted to characterize the NeabNPV genome at the nucleotide sequence level. Restriction fragment profiles of the NeabNPV genome have been determined using several restriction enzymes including *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sal*I, *Xho*I and partial restriction libraries were created by randomly cloning *Eco*RI and *Hin*dIII DNA restriction fragments into plasmid vectors. A partial genomic library of the NeabNPV genome was also developed by shotgun cloning 1 to 3 kb NeabNPV DNA fragments into bacteriophage M13. The partial or total nucleotide sequence of all cloned DNA fragments was determined and compared with the NCBI Genbank database in an attempt to identify potential ORFs and baculovirus gene homologues.

2.2. Methods and Materials

2.2.1. Virus Inclusion Bodies

NeabNPV polyhedral inclusion bodies (PIBs) were isolated from diseased *N. abietis* larvae collected from a Canadian Forestry Services Atlantic Division (CFS) forest test plot near Stephenville, Newfoundland. The test plot was used to conduct aerial test sprays of NeabNPV onto forest plots infested with *N. abietis* in the summer of 2001. PIBs collected by members of CFS were resuspended in ddH_2O prior to being sent to our lab at the University of Victoria, Victoria, British Columbia.

2.2.2. Partial Viral Purification and DNA Extraction

NeabNPV PIBs were semi-purified from crude matter by spinning the obtained PIB solution (in ddH_20) at 1000 g for 5 min (25°C). The pellet (large debris) was discarded and the supernatant was spun at 2500 g for 10 min (25°C) to pellet any PIBs remaining in the solution. DNA was extracted from the pellet (contains PIBs) by incubation in an alkaline dissolution buffer (0.1M Na_2CO_3, pH 11.0) for 15 min. at 4°C, followed by treatment with proteinase-K (0.5ug/ul) and SDS (0.25%) at 55°C for 15 h. Any undisrupted PIBs, polyhedrin protein, or crude matter were sedimented by spinning at 2000 g for 5 min. (25°C). The remaining supernatant was then spun at 15,000 g for 30 min. (4°C) to pellet the released polyhedra-derived virions. DNA was isolated and purified from the pellet using a phenol/chloroform extraction and ethanol precipitation as described in Sambrook et al., (1989). The purity of the DNA was determined spectrophotometrically.

2.2.3. Restriction Endonuclease Analysis

NeabNPV DNA was digested with restriction endonucleases *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sal*I, and *Xho*I (NEB) following manufacturers protocols and the fragments were separated by electrophoresis in 1x TBE buffer, through a 14cm, 0.8% agarose gel at 15 volts for 16 hours. A 1kb DNA ladder (GibcoBRL) was used as a molecular size standard. DNA fragments from a *Hin*dIII digest of the NeabNPV genome were also separated by pulsed field electrophoresis in a 1x TBE buffer, through a 14cm, 1.0% agarose gel at 30 volts for 48 hours.

2.2.4. Restriction Fragment Cloning and End-Terminus Sequencing

A partial NeabNPV DNA *Eco*RI restriction library was obtained from the Canadian Forest Services (CFS), New Brunswick. NeabNPV DNA restriction fragments were shotgun cloned into the *Eco*RI site of a T3/T7 plasmid vector multiple cloning site, transformed, amplified, purified, and sent to Dr. Levin's lab as a recombinant plasmid preparation. An *Eco*RI digest of the NeabNPV genome was also obtained from the CFS.

Viral *Hin*dIII DNA restriction fragments were randomly cloned into the *Hin*dIII restriction site of the pBluescriptII KS+ plasmid vector, using a FastLink DNA Ligation Kit (Epicentre Technologies) and following the manufacturers protocols. Restriction fragments absent from the library were isolated from agarose gels and purified using a QIAquick Gel Extraction Kit following manufacturers protocols (QIAGEN) and cloned individually into pBluescriptII KS+. After transformation into *E. coli* XL1-Blue competent cells (Stratagene), white colonies were randomly selected and recombinant plasmid DNA was isolated using a QIAprep Plasmid Extraction Kit (QIAGEN).

All cloned viral *Hind*III and *Eco*RI restriction fragments were end-sequenced using the dideoxy chain termination method of Sanger *et al.,* (1977). Universal forward and reverse M13 primers and universal T3 and T7 primers were used to sequence the cloned *Hind*III and *Eco*RI restriction fragments, respectively. Sequencing was conducted using an ABI 373 automated DNA sequencer (Applied Biosystems Inc.) and all sequencing reactions were performed using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready reaction kit with AmpliTaq DNA polymerase (Perkin Elmer) according to the manufacturers protocols. Sequencing was performed at the Centre for Environmental Health (CEH) sequencing facility (Dr. Ben Koop's lab), University of Victoria, Victoria, British Columbia.

2.2.5. Shotgun Cloning and Sequencing

NeabNPV DNA was sheared by nebulization using pressurized nitrogen gas. Total DNA was isolated by ethanol precipitation and viral DNA fragments were separated by electrophoresis in 1x TAE buffer, through a 7cm, 2.0% agarose gel at 60 volts for 2 hours.

DNA fragments ranging in size from 1.6kb to 2kb and 2kb to 3kb were gel purified using a QIAquick Gel Extraction Kit following manufacturers protocols (QIAGEN). Gel purified DNA fragments were blunt-end repaired using Mung Bean Exonuclease (50 units) incubated for 30 min. at 37°C. Blunt-ended DNA fragments were then phenol/chloroform extracted and ethanol precipitated. DNA fragments were further blunt repaired using T4 DNA Polymerase (6 units), and Klenow fragment (15 units) for 2 hours at 16°C followed by phenol/chloroform extraction and ethanol precipitation.

Blunt-ended NeabNPV DNA fragments were then randomly cloned into the *Sma*I site of the M13mp19 vector using T4 DNA ligase (200 units). Recombinant M13 DNA was transformed into competent *E. coli* DH5α cells by electroporation and transformed cells were rescued on LB agar plates for 18 hours at 37°C. Transformation was confirmed by blue/white selection and white viral plaques containing the recombinant M13 DNA vector were picked and used to infect cultures of DH5α *E. coli* cells, which were incubated for 17 hours at 37°C in 2x YT Medium. Recombinant M13 DNA was isolated and purified using a QIAGEN M13 Miniprep kit following manufacturers instructions (QIAGEN).

Cloned NeabNPV DNA fragments were sequenced (CEH sequencing facility University of Victoria, Victoria, British Columbia) using the dideoxy chain termination method of Sanger *et al.,* (1977). Standard universal M13 primers were used for all sequencing reactions. Sequencing was conducted using an ABI 373 automated DNA sequencer (Applied Biosystems Inc.) and all sequencing reactions were performed using the ABI PRISM Big Dye Terminator Cycle Sequencing Ready reaction kit with AmpliTaq DNA polymerase (Perkin Elmer) according to the manufacturers protocols.

2.2.6. Sequence Analysis

All generated sequences were analysed using Chromas (v1.62) (Technelysium Pty), Bioedit (v5.0.9) (Hall, 1999), PHRED (Ewing and Green, 1998), and DNASTAR (Lasergene). Comparisons of nucleotide and inferred amino acid sequences with the NCBI Genbank database were performed using BLAST (v2.0) (Altschul *et al.,* 1990).

2.3. Results and Discussion

Restriction endonuclease digestion of NeabNPV DNA with *Bam*HI, *Eco*RI, *Hin*dIII, *Pst*I, *Sal*I, and *Xho*I yielded the restriction fragment profiles shown in Fig.2.1. Bands smaller than 1 kb were not detected, and high molecular weight fragments of the *Bam*HI, *Pst*I, and *Xho*I digests could not be resolved. The *Hin*dIII digest appeared to be complete and was further analysed using pulsed field gel electrophoresis (PFGE) which was able to resolve the largest *Hin*dIII fragment (A) into three separate high molecular weight DNA bands: A, B, and C (Fig 2.2). A closer analysis of a *Hin*dIII and *Eco*RI digest of the NeabNPV genome showed 11 and 17 DNA fragments, respectively. The fragments were designated alphabetically with "A" being the largest fragment for each digest, as proposed by Vlak and Smith (1982). Well-resolved restriction profiles of NeabNPV DNA *Hin*dIII and *Eco*RI digests are shown in Fig. 2.3 and the approximate sizes of the fragments were summarized in Table 2.1. The sizes of the *Hin*dIII fragments were estimated from both a standard and pulsed field agarose gel while the sizes of the *Eco*RI fragments were estimated from a standard agarose gel only.

We estimated the total size of the NeabNPV genome to be approximately 94.7 kb. This is similar in size to the genomes of NeleNPV (~82 kb) (Arif, unpublished), NeseNPV (87.1 kb) (Maruniak, unpublished), PxGV (101 kb) (Hashimoto *et al.,* 2000) and CuniNPV (108.2 kb), but considerably smaller than the other known baculovirus genomes (Table 2.2). Eight fragments from a NeabNPV *Hin*dIII digest and sixteen fragments from a NeabNPV *Eco*RI digest were cloned into the pBluescriptII KS(+) and pT3/T7mp18 plasmids (performed by CFS), respectively. The cloned *Hin*dIII Y and Z fragments do not appear in the NeabNPV *Hin*dIII restriction digest (Fig 2.1) as DNA

fragments less than 1.6kb were not visualized. Both of these restriction fragments, however, were seen inconsistently in a variety of *Hin*dIII digests of the NeabNPV genome and may represent restriction fragments from a unique strain of the NeabNPV species. Different strains of particular baculovirus species are known to occur in nature and show heterogeneity at certain sites between their genomes yet are nearly 100% identical (Sadler *et al.,* 1998). End-terminus sequence data from the *Hin*dIII Y fragment appears to be 99% identical (over a 490 bp region) to a stretch of sequence within the *Hin*dIII I fragment. The 1% difference between these sequences appeared to be due to miscalled bases (N's) in the *Hin*dIII Y sequence. Translation BLAST analysis of this sequence showed significant homology (e-value less than 10^{-26}) to the baculovirus HOAR protein, indicating that this sequence putatively codes for a protein and does not represent a baculovirus-specific non-coding region (hrs, see Chapter 1). Due to the high level of sequence identity between the *Hin*dIII Y and I fragments, and due to the inconsistent appearance of both the *Hin*dIII Y and Z fragments in digests of the NeabNPV genome, both Y and Z were excluded from the estimation of the size of the NeabNPV genome.

As mentioned in the methods, the *Eco*RI library was obtained from the Canadian Forest Services, as was an *Eco*RI digest of the NeabNPV genome. A comparison of the sizes of the cloned NeabNPV *Eco*RI fragments to those present in the *Eco*RI restriction digest showed different bands. There are fewer DNA bands in the *Eco*RI digest of the genome than there are cloned *Eco*RI fragments. A possible explanation for this is that the *Eco*RI digest was an incomplete restriction digest that resulted in fewer DNA bands. Star activity (non-specific digestion of DNA due to factors such as excessive amounts of enzyme) may also be a factor and may have resulted in more DNA bands being cloned

than are represented in the digest of the NeabNPV genome. The estimated size of the genome should still be accurate as these partially digested DNA fragments were well resolved, enabling an accurate determination of their sizes. Approximately 600 base pairs of sequence data was obtained from the end-termini of each cloned *Eco*RI and *Hin*dIII fragment which, when combined, accounts for approximately 28%, or 26 kb, of the NeabNPV genome.

The end-terminus nucleotide sequence data obtained from each of the cloned NeabNPV DNA restriction fragments was compared to the NCBI Genbank database using the translation BLAST (BLASTx) tool, which indicated the presence of several putative baculovirus gene homologues (Table 2.3 and Table 2.4).

The most notable alignments (e-value less than 10^{-15}) from the *Hin*dIII library sequence data came from the $H_{Reverse}$, $I_{Forward}$, $K_{Forward}$, and $K_{Reverse}$ sequences, which aligned significantly with the baculovirus genes *lef-5, p74, p74,* and *hoar*, respectively. LEF-5 is a baculovirus late expression factor that is required for late gene expression. Studies conducted by Harwood *et al.* (1998) showed that the LEF-5 protein contains within its C-terminal 32 amino acids a region homologous to the zinc ribbon domain of RNA polymerase II elongation factor IIS (TFIIS) from a variety of taxa. P74 is a 74 kDa baculovirus protein found associated with the membranous envelope surrounding the occlusion-derived virions. Faulkner *et al.* (1997) showed that the P74 protein is essential for establishing the primary infection of midgut cells in insect larvae that ingest baculovirus occlusion bodies. HOAR is an arginine-rich protein with a RING-finger zinc-binding motif. Its function has not been demonstrated but due to the presence of a zinc finger, it is though to be involved in binding DNA (Wang *et al,* 1991).

The most notable alignments (e-value less than 10^{-15}) from the *Eco*RI library sequence data came from the K_{T3}, M_{T3}, T_{T3}, and V_{T3} sequences, which aligned significantly with *lef-9, polh, p47,* and *polh* genes, respectively. LEF-9 is a baculovirus late transcription factor that is required for late gene expression. The *lef-9* gene encodes a 59 kDa protein that contains an amino acid sequence similar to a conserved motif identified in the largest subunit of prokaryotic and eukaryotic DNA-dependant RNA polymerases (Lu and Miller, 1994). The polyhedrin protein is the major structural protein of the viral occlusion body, which protects the virion from environmental decay when the virion is found outside of its host. The P47 protein is also a late expression factor that is required for late and very late gene expression (Carstens *et al.,* 1993).

The NeabNPV genomic library constructed using nebulized genomic DNA is currently in the progress of being sequenced. Some of the sequence data obtained so far has been compared to the NCBI Genbank database, which indicated the presence of several putative baculovirus genes (Appendix). Some of the most significant alignments (e-value less than 10^{-10}) from the genomic library sequence data include the *hoar*-like protein (5e-44), *odv-e56* (7e-46), *p74* (4e-43), *lef-9* (5e-34) and *iap-3* (1e-33). IAP-3 is an anti-apoptotic protein, which is part of a larger family of proteins known as inhibitors of apoptosis (iap). Several *iap* genes have been discovered in baculoviruses, each of which is identifiable by a C-terminal zinc finger known as a RING finger and an N-terminus imperfect repeat sequence known as a baculovirus *iap* repeat (BIR) motif (Clem, 1997). The function of the HOAR-like, P74, and LEF-9 proteins was previously mentioned (see above).

Using both the *Eco*RI and *Hin*dIII restriction fragment sequence data and the NeabNPV genomic library sequence data, four putative baculovirus ORFs were assembled. These putative ORFs included the baculovirus genes *p74, polh, odv-e56,* and *vlf-1*. The *vlf-1* gene encodes a very late expression factor that is required for expression of the very late genes *polh* and *p10* (Yang and Miller, 1999). The function of *p74, polh,* and *odv-e56* was previously mentioned (see above). The nucleotide sequence data of the putative NeabNPV ORFs was compared to the NCBI Genbank database using the translation BLAST tool. The results indicated that each of the assembled NeabNPV ORFs were most similar to baculovirus DNA sequences. The *p74, polh, odv-e56* and *vlf-1* ORFs were compared individually to Genbank and yielded e-values as high as 1e-129 (EppoNPV *p74*), 1e-70 (AnfaNPV *polh*), 3e-55 (PxORF *odv-e56*), and 1e-48 (SpltMNPV *vlf-1*), respectively. A more in-depth analysis of these genes will be necessary to help better characterize the relationship of NeabNPV to the other known baculoviruses.

In summary, *Eco*RI and *Hin*dIII restriction digests of the NeabNPV genome suggest a genome size of approximately 94 kb, making NeabNPV one of the smallest baculovirus genomes known to date. The size of the NeabNPV genome is similar to the other Hymenopteran NPVs, NeleNPV (~82 kb) and NeseNPV (~87 kb), suggesting a possible close evolutionary relationship. The sequence data obtained from both the restriction libraries and genomic library aligned with a variety of baculovirus sequences found in the NCBI Genbank database. This suggests that NeabNPV may have many more genes common to known baculovirus genomes, however, further analysis of the sequence data will be necessary to determine which baculovirus genes the NeabNPV genome contains.

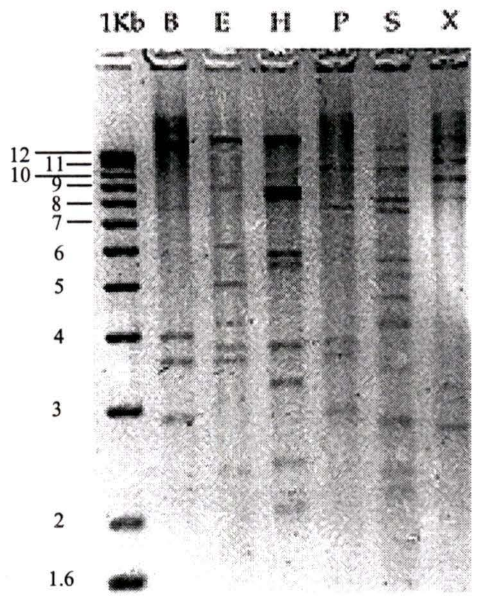


Figure 2.1. NeabNPV DNA restriction endonuclease digests. Genomic DNA was digested with *Bam*HI (B), *Eco*RI (E), *Hin*dIII (H), *Pst*I (P), *Sal*I (S), and *Xho*I (X) and electrophoretically separated in 0.8% agarose and stained with ethidium bromide. A 1 kb DNA ladder was used as a molecular size marker with sizes indicated in kilobase pairs.

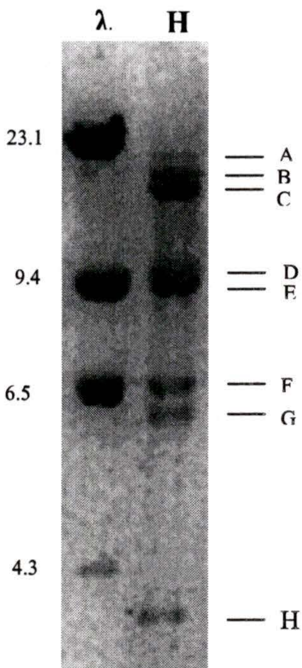


Figure 2.2. NeabNPV *Hin*dIII restriction endonuclease digest (H) separated by pulsed field gel electrophoresis (PFGE) and stained with ethidium bromide. Lambda DNA cut with *Hin*dIII (λ) was used as a molecular size marker with sizes indicated in kilobase pairs. Note – bands sizes seen here are not representative of those shown in Figure 2.3 or those listed in Table 2.1. This figure is included solely to show the separation of the high molecular weight bands A, B, C, which appear as a single band in Fig. 2.1

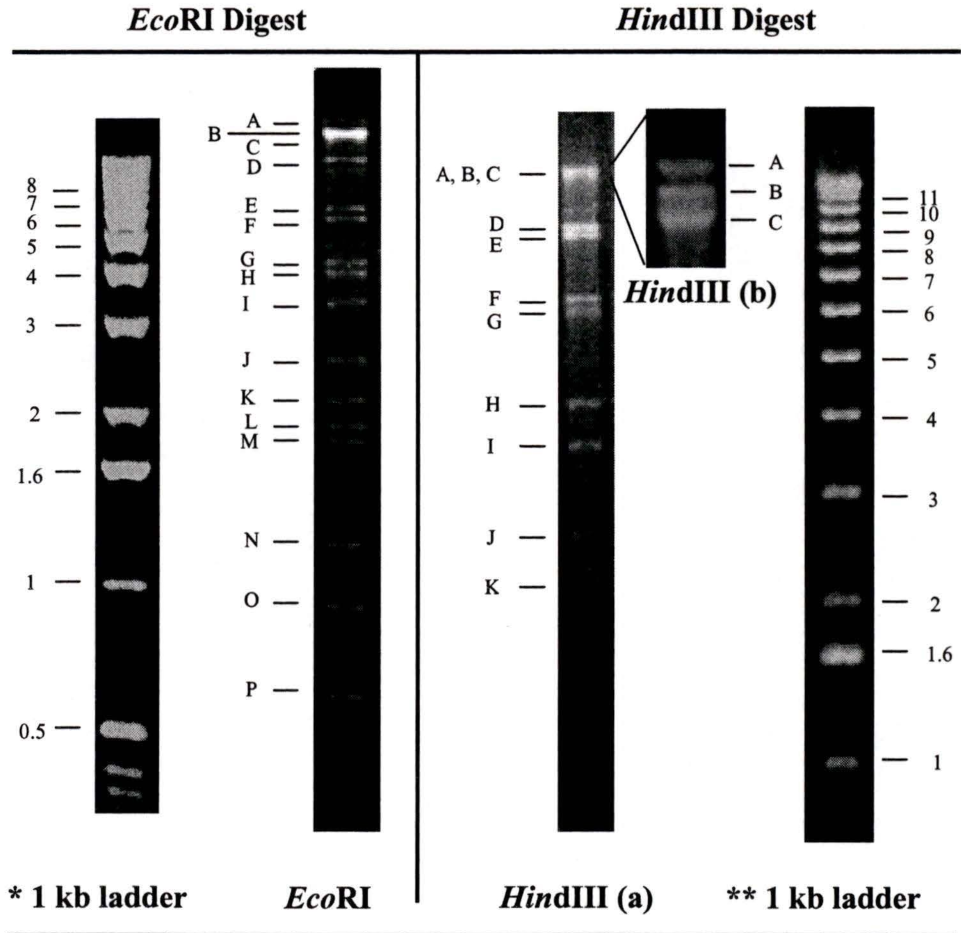


Figure 2.3. NeabNPV DNA *Hin*dIII and *Eco*RI digests. Fragments are named alphabetically, with "A" being the largest fragment. A 1 kb ladder was used as a molecular size marker for each digest with sizes indicated in kilobase pairs. A lambda/*Hin*dIII ladder was used to estimate the sizes of the *Hin*dIII A,B,C fragments and *Eco*RI A,B,C,D fragments (ladder not shown). The *Eco*RI, *Hin*dIII (a) digests, and 1 kb DNA ladders were run on standard 1.0% agarose gels. The *Hin*dIII (b) digest was run on a 1% agarose gel using pulsed field electrophoresis. All digests were stained with ethidium bromide.

Table 2.1. NeabNPV DNA *Eco*RI and *Hin*dIII restriction endonuclease fragment sizes (kb). All fragment sizes were estimated from the digests shown in Fig. 2.3. *Eco*RI fragments A,B,C,D were analyzed using a λ/*Hin*dIII molecular size marker which is not shown in Fig. 2.3. *Hin*dIII fragments A,B,C were analyzed using a λ/*Hin*dIII molecular size marker which is not shown in Fig. 2.3. The estimated size of the NeabNPV genome was determined from the sum of the restriction fragments, shown in bold at the bottom of each column.

Restriction Fragment	EcoRI	HindIII
A	17.4	19.0
B	15.6	17.3
C	14.3	16.3
D	11.2	8.8
E	6.50	8.5
F	5.30	6.6
G	4.20	6.2
H	4.10	4.3
I	3.60	3.7
J	2.80	2.6
K	2.30	2.1
L	2.10	
M	1.90	
N	1.20	
O	0.91	
P	0.61	
Sum of Fragment Sizes → (Size of Genome)	94.0	95.4

Table 2.2. Baculovirus Genomes

Baculovirus Genome	Size (bp)	Accession no.	References
AcMNPV	133,894	L22858	Ayres et al.
BmMNPV	128,413	L33180	Gomi et al.
OpMNPV	131,990	U75930	Ahrens et al.
EppoNPV	118,584	AY043256	Hyink et al.
SeMNPV	135,611	AF169823	Ijkel et al.
LdMNPV	161,046	AF081810	Kuzio et al.
HearSNPV	131,403	AF271059	Chen et al.
HzSNPV	130,869	AF334030	Chen et al.
SpltMNPV	139,342	AF325155	Pang et al.
MacoNPV	155,060	AF325155	Li et al.
NeleNPV	~82,000	None	B. Arif (Unpublished)
NeseNPV	87,100	None	J. Maruniak (Unpublished)
NeabNPV	~94,710	None	* This Study *
CuniNPV	108,252	AF403738	Alfonso et al.
XcGV	178,733	AF162221	Hayakawa et al.
PxGV	100,999	AF270937	Hashimoto et al.
CpGV	123,500	U53466	Luque et al.

Fragment	Orientation	Putative ORF	Baculovirus	E-Value
F	Forward	ORF55	*Plutella xylostella*	2e-11
		LEF-1	*Helicoverpa zea*	8e-09
		ORF82	*Xestia c-nigrum*	5e-08
		LEF-1	*Choristoneura fumiferana*	1e-05
			Lymantria dispar	3e-05
			Spodoptera exigua	2e-04
			Autographa californica	7e-03
			Bombyx mori	7e-03
			Orgyia pseudotsugata	4e-03
	Reverse	None		
G	Forward	None		
	Reverse	None		
H	Forward	HOAR-like	*Neodiprion sertifer*	2e-02
	Reverse	ORF69	*Plutella xylostella*	3e-19
		LEF-5	*Cryptophlebia leucotreta*	1e-17
			Spodoptera exigua	1e-15
			Lymantria dispar	1e-15
			Autographa californica	2e-15
			Bombyx mori	8e-15
			Heliocoverpa armigera	1e-12
			Orgyia pseudotsugata	1e-09
		ORF95	*Xestia c-nigrum*	4e-15
I	Forward	P74	*Choristoneura fumiferana*	3e-61
			Spodoptera littoralis	3e-61
			Spodoptera litura	4e-61
			Orgyia pseudotsugata	2e-60
			Lymantria dispar	2e-59
			Autographa californica	4e-58
			Bombyx mori	7e-57
			Spodoptera exigua	5e-55
			Heliocoverpa armigera	7e-55
			Xestia c-nigrum	2e-26
		ORF49	*Plutella xylostella*	3e-50
		ORF77	*Xestia c-nigrum*	9e-50
	Reverse	ORF81	*Lymantria dispar*	6e-12
			Anticarsia gemmatalis	2e-11
			Orgyia pseudotsugata	5e-11
			Bombyx mori	4e-10
			Autographa californica	4e-10
		ORF86	*Plutella xylostella*	2e-10
		ORF120	*Xestia c-nigrum*	2e-09
		ORF79	*Spodoptera exigua*	2e-08

Fragment	Orientation	Putative ORF	Baculovirus	E-Value
J	Forward	ODV-e56	*Autographa californica*	1e-05
			Bombyx mori	8e-05
			Orgyia pseudotsugata	5e-04
			Cydia pomonella	1e-03
			Choristoneura fumiferana	2e-03
		ORF16	*Plutella xylostella*	2e-04
	Reverse	None		
K	Forward	P74	*Autographa californica*	6e-16
			Bombyx mori	1e-15
			Orgyia pseudotsugata	2e-15
			Lymantria dispar	4e-15
			Choristoneura fumiferana	5e-15
			Spodoptera exigua	1e-12
			Heliocoverpa armigera	4e-12
			Spodoptera litura	4e-11
			Spodoptera littoralis	1e-10
		ORF49	*Plutella xylostella*	2e-10
		P74	*Xestia c-nigrum*	6e-10
		ORF77	*Xestia c-nigrum*	6e-10
	Reverse	HOAR-like	*Neodiprion sertifer*	7e-27
Y	Forward	None		
	Reverse	HOAR-like	*Neodiprion sertifer*	2e-26
Z	Full	Polh	*Autographa californica*	1e-02

Table 2.3. Translation BLAST results for cloned NeabNPV DNA *Hin*dIII restriction fragments. Nucleotide sequence data was obtained for the termini of each NeabNPV *Hin*dIII fragment, translated in all six frames, and compared to the NCBI Genbank database. Orientation refers to the universal M13 primer that was used for sequencing.

Fragment	Orientation	Putative ORF	Baculovirus	E-Value
E	T7	None		
F	T3	None		
H	T3	None		
	T7	None		
I	T3	Alkaline Nuclease	*Autographa californica*	3e-13
			Bombyx mori	3e-13
			Plutella xylostella	2e-12
			Xestia c-nigrum	5e-12
			Spodoptera littura	2e-11
			Choristoneura fumiferana	2e-11
			Helicoverpa zea	1e-10
			Heliocoverpa armigera	2e-10
			Orgyia pseudotsugata	2e-10
			Lymantria dispar	4e-10
			Spodoptera exigua	2e-07
	T7	LEF-9	*Heliocoverpa armigera*	1e-08
			Lymantria dispar	3e-07
			Orgyia pseudotsugata	5e-07
			Bombyx mori	1e-06
			Autographa californica	2e-06
			Spodoptera exigua	2e-05
		ORF99	*Plutella xylostella*	2e-08
		ORF139	*Xestia c-nigrum*	2e-07
J	T3	None		
	T7	None		
K	T3	LEF-9	*Heliocoverpa armigera*	4e-26
			Lymantria dispar	2e-25
			Orgyia pseudotsugata	8e-22
			Bombyx mori	5e-23
			Autographa californica	5e-23
			Spodoptera exigua	8e-22
		ORF99	*Plutella xylostella*	5e-24
		ORF139	*Xestia c-nigrum*	3e-23
L	T3	None		
	T7	None		
M	T3	Polh	*Anagrapha falcifera*	3e-38
			** Plus 67 more unlisted significant polyhedrin hits**	
	T7	None		

Fragment	Orientation	Putative ORF	Baculovirus	E-Value
N	T3	None		
	T7	ORF90	*Spodoptera exigua*	5e-06
		ORF135	*Xestia c-nigrum*	6e-06
		ORF68	*Xestia c-nigrum*	6e-06
			Bombyx mori	7e-05
			Autographa californica	3e-04
		ORF96	*Plutella xylostella*	1e-04
		ORF80	*Lymantria dispar*	4e-04
		ORF73	*Epiphyas postvittana*	9e-03
O	T7	GP-41	*Heliocoverpa armigera*	5e-06
			Spodoptera exigua	1e-05
			Spodoptera frugiperda	2e-05
			Autographa californica	4e-05
			Bombyx mori	7e-05
			Lymantria dispar	2e-04
			Orgyia pseudotsugata	3e-03
		P40	*Helicoverpa zea*	5e-06
			Heliocoverpa armigera	3e-05
P	T3	None		
	T7	None		
Q	T7	None		
R	T7	None		
T	T3	ORF51	*Plutella xylostella*	5e-18
		P47	*Choristoneura fumiferana*	1e-17
			Spodoptera exigua	2e-16
			Lymantria dispar	1e-15
			Bombyx mori	1e-15
			Autographa californica	3e-15
			Orgyia pseudotsugata	2e-14
			Heliocoverpa armigera	2e-14
		ORF78	*Xestia c-nigrum*	3e-16
T	T7	ORF51	*Plutella xylostella*	2e-11
		P47	*Heliocoverpa armigera*	3e-11
			Bombyx mori	5e-11
			Autographa californica	1e-10
			Choristoneura fumiferana	3e-10
			Spodoptera exigua	5e-10
			Lymantria dispar	7e-10
			Orgyia pseudotsugata	7e-10
			Xestia c-nigrum	6e-09
V	T3	Polh	*Wiseana signata*	4e-29
			** Plus 36 more unlisted significant polyhedrin hits**	
W	T3	None		
	T7	None		

Table 2.4. Translation BLAST results for cloned NeabNPV DNA *Eco*RI restriction fragments. Nucleotide sequence data was obtained for the termini of each NeabNPV *Eco*RI fragment, translated in all six frames, and compared to the NCBI Genbank database. Orientation refers to the universal T3/T7 primer that was used for sequencing.

2.4. References

Afonso, C. L., E. R. Tulman, et al. (2001). Genome sequence of a baculovirus pathogenic for Culex nigripalpus. *J Virol.* **75**, 11157-65.

Carstens, E. B., A. L. Lu, et al. (1993). Sequence, transcriptional mapping, and overexpression of p47, a baculovirus gene regulating late gene expression. *J Virol.* **67**, 2513-20.

Clem, R. J. (1997). Baculovirus Pathogenesis. The Baculoviruses. L. K. Miller. New York, Plenum Press, 237-266.

Faulkner, P., J. Kuzio, et al. (1997). Analysis of p74, a PDV envelope protein of Autographa californica nucleopolyhedrovirus required for occlusion body infectivity in vivo. *J Gen Virol.* **78**, 3091-100.

Ewing, B. & Green, P. (1998). Base calling of automated sequencer traces using phred II error probabilities. *Genome Res.* **8**, 186-94.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp Ser.* **41**, 95-98.

Harwood, S. H., L. Li, et al. (1998). AcMNPV late expression factor-5 interacts with itself and contains a zinc ribbon domain that is required for maximal late transcription activity and is homologous to elongation factor TFIIS. *Virol.* **250**, 118-34.

Hashimoto, Y., T. Hayakawa, et al. (2000). Sequence analysis of the Plutella xylostella granulovirus genome. *Virol.* **275**, 358-72.

Lu, A. and L. K. Miller (1994). Identification of three late expression factor genes within the 33.8-to 43.4-map-unit region of Autographa californica nuclear polyhedrosis virus. *J Virol.* **68**, 6710-8.

O'Reilly, D. R., L. K. Miller, et al. (1992). Baculovirus Expression Vectors: A Laboratory Manual. New York, W. H. Freeman and Company, 139-142.

Sadler, T., V. Ward, et al. (1998). Examination of New Zealand's endemic Wiseana nucleopolyhedrovirus by analysis of the viral polyhedrin gene. *Arch Virol.* **143**, 2273-88.

Sambrook, J., E. F. Fritsch, et al. (1989). Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbour, New York: Cold Spring Harbour Laboratory.

Vlak, J. M. and G. E. Smith (1982). Orientation of the Genome of *Autographa californica* nuclear polyhedrosis virus: a proposal. *J Virol.* **41**, 1118-1121.

Wang H., Chen X., Arif B. M., Vlak J. M. and Hu Z. (2001). Nucleotide sequence and transcriptional analysis of a putative basic DNA-binding protein of *Helicoverpa armigera* nucleopolyhedrovirus. *Virus Genes*. **22**: 113-120.

Yang, S. and L. K. Miller (1998). Expression and mutational analysis of the baculovirus very late factor 1 (vlf-1) gene. *Virol.* **245**, 99-109.

Chapter 3. Nucleotide Sequence Analysis of the Polyhedrin Gene of the *Neodiprion abietis* Nucleopolyhedrovirus

3.1. Introduction

The *Baculoviridae* are composed of two genera, the *Nucleopolyhedroviruses* and *Granuloviruses*, which are differentiated by the morphology of their occlusion bodies (PIBs). The primary function of a PIB is to protect embedded virion particles from environmental decay when outside of the host. After ingestion of PIBs by susceptible host larvae, proteins making up the PIB dissolve in the alkaline solution of the insect midgut, releasing virion particles, which initiate the viral infectious cycle.

GVs have granular shaped PIBs that contain rod-shaped virions with a single (S) enveloped nucleocapsid. In contrast, NPVs have polyhedral shaped PIBs that contain rod-shaped virions with multiple (M) enveloped nucleocapsids. The major protein component of PIBs is polyhedrin in NPVs and granulin in GVs. Both polyhedrin and granulin are approximately 29.5 kDa in size and appear to be highly conserved among baculoviruses isolated from insect hosts within the Order Lepidoptera (Vlak and Rohrmann, 1985). Synthesis of polyhedrin and granulin is temporally regulated, occurring very late in the viral infectious cycle, and is controlled by a very powerful promoter element TAAG. Many polyhedrin genes have been sequenced and studied, the majority of which have come from Lepidopteran NPVs, although GV sequences are becoming more prevalent. In this study I report the sequence and analysis of the polyhedrin gene from NeabNPV, a baculovirus pathogenic to a Hymenopteran insect, the balsam fir sawfly.

3.2. Methods and Materials

3.2.1 Virus and DNA

See Methods and Materials chapter 2.

3.2.2. Restriction Endonuclease Digestion and Cloning

See Methods and Materials chapter 2.

3.2.3. Sequencing Restriction Fragments

See Methods and Materials chapter 2.

Analysis of the restriction fragment end-sequence data from both *Eco*RI and *Hin*dIII libraries indicated the presence of putative polyhedrin gene sequences on several different cloned restriction fragments. Further analysis enabled the construction of a putative complete polyhedrin gene sequence.

3.2.4. Isolation of the NeabNPV Polyhedrin Gene Sequence

Two primers, NeabNPV Polh Left (5' CCGCAATTGTCAAGATGCCCA 3') and NeabNPV Polh Right (5' CTGGTCCATTAACAAACGGTGCT 3'), containing the flanking sequences of the putative NeabNPV polyhedrin gene were designed and used to amplify the complete polyhedrin gene using a touch down PCR protocol. The program used for the amplification consisted of cycles of 94°C for 30 sec, 65°C for 45 sec and 72°C for 35 sec; followed by 3 cycles of 94°C for 30 sec, 62°C for 45 sec and 72°C for 35 sec; followed by 3 cycles of 94°C for 30 sec, 60°C for 45 sec and 72°C for 35 sec; followed by 3 cycles of 94°C for 30 sec, 58°C for 45 sec and 72°C for 35 sec. The final

part of the program consisted of 20 cycles of 94°C for 30 sec, 55°C for 45 sec and 72°C for 35 sec. The PCR product was analysed by gel electrophoresis in a 1x TBE buffer, through a 7cm, 0.8 % agarose gel and then purified using QIAquick Gel Extraction kit (QIAGEN) following manufacturers protocols. The PCR product was sequenced using the NeabNPV polyhedrin left and right specific primers as described above.

3.2.5. Analysis of the NeabNPV Polyhedrin Gene Sequence

The NeabNPV polyhedrin nucleotide sequence was compared to the NCBI Genbank database using BLAST (v2.0) (Altschul *et al.,* 1990). ClustalX (v1.81) (Jeanmougin *et al.,* 1998) was used to create multiple sequence alignments of baculovirus polyhedrin and granulin inferred amino acid sequences. Phylogeny studies were performed using PAUP (v4.0b10) (Swafford, 1998) and phylogenetic trees were visualized and edited using TreeView (v1.6.6) (Page, 1996) and Windows Metafile Companion software (v1.11), respectively.

3.3. Results and Discussion

The NeabNPV genome was digested with the restriction enzyme *Hin*dIII and the resultant DNA fragments were randomly cloned into the pBluescriptII KS+ plasmid. A NeabNPV *Eco*RI/pT7/T3α18 plasmid library obtained from the CFS Atlantic Division was also used in this study. Cloned restriction fragments from both the NeabNPV *Eco*RI and *Hin*dIII DNA libraries were sequenced from both termini and the resultant end-sequence data was compared to the NCBI Genbank database using BLASTx. After comparison to Genbank, the NeabNPV polyhedrin gene was determined to be located on

the *Eco*RI V (~ 500bp) and *Eco*RI M (~3.5kb) fragments of the NeabNPV *Eco*RI library, as well as the *Hin*dIII Z (~500bp) fragment of the NeabNPV *Hin*dIII library. The pairwise alignment data for both the *Eco*RI V and M fragments is shown in Table 3.1 and Table 3.2, respectively. Overlapping sequences of the *Eco*RI V and *Eco*RI M fragments enabled the complete NeabNPV polyhedrin gene to be determined (Fig 3.1). Synthetic primers were designed using the end sequence data from the *Eco*RI V and *Eco*RI M fragments and the double-stranded sequence of the polyhedrin gene was generated from the NeabNPV genome. The PCR amplified polyhedrin gene was subsequently cloned into the pBluescriptII KS+ plasmid and sequenced from both ends.

3.3.1. The NeabNPV Polyhedrin Nucleotide Sequence

The NeabNPV polyhedrin gene is 741 nucleotides long and potentially encodes a 246 amino acid long protein with a predicted molecular weight of 29.5 kDa (Fig.3.1). This size agrees with the estimated size of the polyhedrin protein from NeabNPV occlusion bodies, which was determined to be approximately 28.5kDa after analysis by SDS-PAGE (unpublished, this study). The NeabNPV polyhedrin protein was similar in size to the other baculovirus occlusion proteins, which range in size from 243 amino acids (WisiNPV) to 249 amino acids (SpltMNPV). NeabNPV polyhedrin is identical in length to NeseNPV polyhedrin, 2 amino acids shorter than the GV granulins, and 1 amino acid shorter than the Group I NPV polyhedrin proteins. Both NeabNPV and NeseNPV polyhedrin sequences lacked the PAY motif found at the C-terminus end of all the other occlusion protein sequences used in this study.

The NeabNPV *polh* gene ends with a TAA translation stop codon but does not appear to contain either a polyadenylation signal (AATAAA) or the baculovirus motif

AAT_9 within the 130bp region downstream of the TAA stop codon (Van Strien *et al.,* 1992). However, the 3' untranslated region (3'UTR) of NeabNPV polyhedrin is AT rich and contains several sequences within 130bp downstream of the TAA stop codon that are similar to both polyadenylation signals (AATAGAA and AATATAA) and the baculovirus AAT_9 motif (AATTATTTTTA). It is also possible that a polyadenylation signal, or baculovirus AAT_9 motif, may exist even further downstream of the TAA stop codon than 130bp, as some baculovirus polyhedrin transcriptional stop signals are known to exist as far as 230bp downstream of the stop codon (Van Strien *et al.,* 1992). Further comparisons of polyhedrin 3'UTRs involving more Hymenopteran polyhedrin nucleotide sequences, as well as transcriptional studies, would help to determine what differences exist between Lepidopteran and Hymenopteran NPV transcriptional stop signals.

The putative NeabNPV polyhedrin gene homologue is most similar to that from NeseNPV. The most variable region between these polyhedrin proteins occurred within amino acid residues 149 to 174 (Fig. 3.2). This 25 amino acid region showed only 1 identical amino acid residue and 2 similar amino acid residues between the Hymenopteran NPVs. Pairwise comparison of the NeabNPV polyhedrin protein to those from other baculoviruses using a dot-plot showed that AcMNPV (Group I NPV), LdMNPV (Group II NPV), and MacoNPV (Group II NPV) also lack any significant amino acid identity or similarity with the NeabNPV polyhedrin protein between residues 149 to 174 (Fig. 3.3). To rule out the possibility that this region of low amino acid similarity was caused by errors in the sequence data, several analyses were performed. The chromatograms (DNA sequence trace files) for NeabNPV DNA sequences that were used to assemble the putative polyhedrin gene sequence were checked by eye and the quality of the DNA sequence was assessed using PHRED. The trace files showed no

obvious errors such as compressions, miscalled bases (N), or dye terminator peaks and the quality of the sequence was determined to be high (probability of error = 10^{-6}). Therefore the putative NeabNPV polyhedrin DNA sequence appears to be correct and the low similarity region (amino acids 149 to 174) may be the result of acquired mutations.

This low similarity region may not be crucial for the determination of the structure of polyhedrin and as a result, may lack reasonable selection pressure. Due to a lack of functional constraint, a higher rate of mutation may exist within this region of the polyhedrin gene, giving rise to a 'hot spot' for the accumulation of mutations. Since the NeabNPV polyhedrin gene is highly diverged from the Lepidopteran NPV and GV occlusion genes and since it has been hypothesized that the Hymenopteran NPVs may be ancient among the baculoviruses (Rohrmann, 1981), the NeabNPV polyhedrin gene may represent an ancestral form of the polyhedrin gene. If this is true then over the course of time the NeabNPV polyhedrin gene may have acquired multiple, acceptable point mutations within this particular region resulting in the sequence divergence seen.

It is also possible, however, that this low similarity region is the result of a frameshift in the DNA sequence. Frameshifts occur when nucleotides are inserted into, or deleted from, a sequence, which alters the normal translational reading frame of that DNA sequence. A frameshift near the beginning of the 25 amino acid region, accompanied by a second frameshift near the end of the region to restore the sequence to its original reading frame, could explain the divergence of the NeabNPV polyhedrin protein between amino acids 149 to 174. To check this hypothesis the reading frame of the DNA sequence was altered by deletion of a single nucleotide from the codon for amino acid 149 (GAA → _AA). This frameshift increased the number of identical amino acids between NeabNPV and NeseNPV polyhedrin from 1 to 11 over amino acids 149 to

174 (Fig 3.5). If the reading frame is restored by an additional frameshift downstream (involving the insertion of a nucleotide from a codon near amino acid 173), then the NeabNPV polyhedrin sequence retains its similarity to NeseNPV over the remainder of the sequence (~75 amino acids). Whether a nucleotide insertion or deletion occurred at the beginning or end of this low similarity region is unclear, however, the significant increase in identical amino acids caused by a nucleotide insertion within the codon for amino acid 149 suggests that a frameshift mutation may have created the sequence divergence seen between amino acids 149 to 174 of the NeabNPV polyhedrin.

Amino acid residues 33 to 36 of the NeabNPV polyhedrin gene contain a KKRK sequence, which is similar, but not identical, to the nuclear localization signal KxKK of most NPVs. NeseNPV has a domain identical to the NeabNPV polyhedrin KKRK. Jarvis *et al.* (1991) showed that the AcMNPV polyhedrin gene contains a domain KRKK at position 32 to 35, which is responsible for the nuclear localization and supramolecular assembly of the polyhedrin protein after translation. This variation of the nuclear localization signal from KxKK to KKRK appears to have no affect on the polyhedrin protein as studies have shown that NeabNPV replicates and forms occlusion bodies in the nucleus of infected insects cells, particularly epithelial cells of the Hymenopteran digestive tract (Lucarotti *et al.,* unpublished). Most GVs appear to lack the KxKK domain at all, instead showing sequences such as RRKK (HbGV, CpGV, ClGV), RHKK (CfGV), RHKE (XcGV, TniGV), RRKD (AsGV), or KHKK (PbGV) (Bideshi *et al.,* 2000).

This lack of a nuclear localization signal, however, is explained by differences in the cytopathology of the GVs as compared to the NPVs. NPVs form occlusion bodies exclusively within the nucleus of the cell they have infected. In contrast, GV infection results in the degradation of the infected cell nucleus, resulting in a mixture of the

cytoplasmic and nucleoplasmic contents. GV occlusion bodies are formed within this intracellular mixture (Federici, 1997). The Hymenopteran domain KKRK may be the ancestral sequence of the Lepidopteran nuclear localization signal KxKK found in most NPVs. Since it has been hypothesized that the Hymenopteran NPVs diverged from the GVs before the radiation of the Lepidopteran NPVs (Zanotto *et al.,* 1993, Rohrmann, 1981), and since the GVs lack the nuclear localization signal, it is possible that the Hymenopteran KKRK gave rise to the Lepidopteran nuclear localization signal KxKK. Studies involving oligonucleotide-directed mutagenesis would have to be performed to confirm whether or not the Hymenopteran NPV KKRK motif actually functions as a nuclear localization signal for these viruses. As well, further phylogenetic studies will have to be performed to help determine whether or not the Hymenopteran NPVs are more ancient than the GVs, and if so, are they the ancestral form of the Lepidopteran NPVs.

Analysis of the 5' untranslated region (5'UTR) of the NeabNPV polyhedrin gene showed several motifs commonly found in other baculovirus polyhedrins including CAAT and TATA-like motifs as well as the baculovirus-specific TAAG motif. Two TATA like motifs (TTAATAAAATTATA and TATTATTATTATTA) were located 98bp and 40bp respectively, upstream of the ATG start codon. As well, two CAAT motifs were found upstream of the translation initiation start site at positions -60bp and -73bp, respectively. The TATA box and CAAT motif are required for recognition of the host cell RNA polymerase II, which is necessary for transcription by early genes. The TAAG motif, in contrast, is essential for recognition of the viral encoded RNA polymerase and is required for transcription by late and very late viral genes.

Two TAAG sequence motifs were found 127bp (TTAAG) and 24bp (ATAAG) upstream of the translation initiation start codon (ATG). All late baculovirus genes

contain the consensus late promoter element TAAG which functions as both a promoter and mRNA start site (Rohrmann, 1986). The transcription of late baculovirus genes results from a viral-encoded RNA polymerase that recognizes the TAAG motif. This short promoter motif is unusual in structure when compared to promoters recognized by eukaryotic and bacterial RNA polymerases, which usually consist of multiple non-contiguous blocks of sequences (Rohrmann and Blissard, 1990). In contrast, the short promoter motif of baculoviruses appears to be more similar to promoters recognized by RNA polymerases specific to yeast mitochondrial DNA, which consist of a highly conserved nucleotide sequence ATATAAGTA (Biswas *et al.,* 1985). Experiments involving the deletion of either TAAG sequence will be necessary to determine which motif is functional.

Fifteen baculovirus polyhedrin 5'UTRs were compared including Group I NPV, Group II NPV, and GV promoter sequences. A comparative analysis of the NeabNPV 5'UTR, from 3bp upstream of the first TAAG motif to the ATG start codon, indicated that the NeabNPV polyhedrin 5'UTR is highly diverged from the GVs and NPVs (Fig. 3.4). The NeabNPV polyhedrin promoter appeared similar to the NPVs (except SeNPV, SfNPV, and MacoNPV) and GVs as it shared an ATAAG later promoter motif. However, the NeabNPV polyhedrin promoter was more similar in size to the GVs than to the NPVs having only 21 nucleotides from the end of the ATAAG motif to the ATG start codon. NPV promoters had between 45 (MacoNPV, SfNPV, SeMNPV) and 50 nucleotides (LdMNPV) from the ATAAG motif to the start codon. The sequence of the NeabNPV 5'UTR showed a higher degree of sequence identity to the GVs (~45%) than to the NPVs (~32%). All of the NPV and GV polyhedrin 5'UTR conformed to Kozak's rule (PuNNATG) for efficient eukaryotic translation (Kozak, 1987).

3.3.2. Comparative Analysis of the NeabNPV Polyhedrin Sequence

The NeabNPV polyhedrin gene showed an overall low homology with the baculovirus GV and NPV occlusion proteins. The degree of amino acid identity of the inferred NeabNPV polyhedrin amino acid sequence ranged from 36% to 80% while the degree of amino acid similarity ranged from 57% to 86% (Table 3.2). The average amino acid similarity among the NPV polyhedrin proteins was 67%, whereas the average similarity among the GV granulin proteins was 58%. The highest level of sequence homology occurred between the Hymenopteran NPVs (NeabNPV and NeseNPV), which showed an amino acid identity of 80% and similarity of 86% between their polyhedrin proteins.

3.3.2.1. Phylogenetic Analysis

The availability of 43 baculovirus occlusion gene sequences allowed for an estimation of the evolutionary relationships of the baculovirus occlusion proteins (Fig 3.6). The inferred amino acid sequence of each gene was used to create a multiple sequence alignment, performed by CLUSTAL (v1.81) (Jeanmougin *et al.,* 1998) with default parameters. An unrooted parsimonious tree was calculated using the heuristic algorithm of PAUP (v4.0b10) and the estimated phylogenetic tree was viewed using Treeview (v1.6.6) (Page, 1996) and edited using Windows Metafile Companion (v1.11). Bootstrap analysis of the constructed tree could not be performed due to the large size of the data set. However, the overall branching pattern of the gene tree was identical to that of gene trees constructed in Chapter 5 (well-supported by bootstrap analysis). The purpose of constructing this phylogenetic tree was only to obtain an idea of the evolutionary relationship of the Hymenopteran NPVs within the *Baculoviridae* and not to

make any definitive conclusions (see Chapter 5). The phylogeny obtained from this analysis showed the characteristic separation of the NPVs and the GVs, as well as the further separation of the NPVs into Group I and Group II NPVs, a topology that agrees well with previous comprehensive phylogenetic studies performed by Zanotto *et al.,* (1993), Bulach *et al.,* (1999), Herniou *et al.,* (2000).

Systematic analysis using the polyhedrin protein sequence indicated that NeabNPV is not closely related to either the Group I or Group II NPVs. However, the Hymenopteran NPVs NeseNPV and NeabNPV grouped together along a unique branch between the base of the NPVs and GVs. Placement of the Hymenopteran NPVs into a clade separate from the NPVs and GVs is supported by the high level of amino acid sequence identity (80%) and similarity (86%) shared between the polyhedrin proteins of NeabNPV and NeseNPV (Table 3.3 and Fig. 3.2).

Studies performed by Bideshi *et al.,* (2000) demonstrated that the 5' UTR end of the polyhedrin gene was a phylogenetically important region. This region is the most variable part of the polyhedrin sequence within NPV and PV occlusion sequences being conserved within, but not between the Group I and II NPVs (Sadler *et al.,* 1998). The 5' UTR of the NeabNPV polyhedrin gene did not display a high degree of sequence identity to either the NPVs or GVs, but did share some of the characteristics from each of the NPV and GV (see above). Assuming that the 5' UTR of the polyhedrin gene is phylogenetically important, then the placement of NeabNPV between both the NPV and GV clades would be expected.

The Lepidopteran NPV polyhedrin proteins are more similar to the Lepidopteran GVs than they are to the Hymenopteran NPVs (Table 3.3), a finding that does not agree with the placement of NeabNPV and NeseNPV closer to the Lepidopteran NPVs than the

GVs (Fig 3.6). According to Rohrmann *et al.,* (1981) the Hymenopteran NPVs originated via a cross-infection of a Lepidopteran NPV into a Hymenopteran insect. If this hypothesis were correct, it would explain the phylogeny of the tree obtained. However, a study conducted by Zanotto *et al.,* (1993) disagrees with this hypothesis. They constructed a polyhedrin gene tree that showed the GVs branching off from NeseNPV, followed by the NPVs branching off from the GVs. This finding suggests that the Lepidopteran baculoviruses diverged from the Hymenopteran lineage before the divergence of the Lepidopteran GV from the NPV, a finding that agrees with a study conducted by Rohrmann (1992). Due to the lack of Hymenopteran NPV genes available for phylogenetic studies and the unavailability of an unambiguous outgroup it is difficult to determine the evolutionary relationship of the Hymenopteran NPVs to the other baculoviruses. Future studies involving multiple genes from several Hymenopteran NPVs, as well as genomic gene content and gene order data should provide a better understanding of *Baculoviridae* phylogeny (Chapter 5).

By comparing viruses by the relative sequence identity of their polyhedrin proteins, we assume that this protein is a realistic measure of the difference between the viral genomes. Many analyses have been made on baculovirus polyhedrin genes and their inferred amino acid sequences (Sadler *et al.,* 1998, Hyink *et al.,* 1998, Li *et al.,* 1997) and a common pattern is emerging. Phylogenetic trees constructed using the polyhedrin gene appear to conflict with phylogenies estimated using other baculovirus genes, and the robustness of the trees are relatively weak when compared to other datasets. As a result, caution needs to be exercised when attempting to determine phylogenetic relationships of baculoviruses using only single genes or proteins, particularly polyhedrin.

Table 3.1. Translation BLAST results for the NeabNPV *Eco*RI VT3 end-terminus nucleotide sequence.

Baculovirus Aligned With	E - Value	Identity (%)	Similarity (%)
Wiseana signata NPV	3e-29	50	76
Bombyx mori NPV	1e-28	50	74
Anticarsia gemmatalis NPV	1e-28	51	74
Antheraea pernyi NPV	3e-28	49	74
Orgyia anartoides NPV	4e-28	49	72
Anagrapha falcifera NPV	5e-28	48	74
Epiphyas postvittana NPV	5e-28	49	74
Choristoneura fumiferana NPV	5e-28	49	74

Table 3.2. Translation BLAST results for the NeabNPV *Eco*RI MT3 end-terminus nucleotide sequence.

Baculovirus Aligned With	E - Value	Identity (%)	Similarity (%)
Hyphantria cunea NPV	1e-37	53	69
Anagrapha falcifera NPV	3e-28	52	68
Spodoptera frugiperda NPV	1e-36	51	70
Amsacta albistriga NPV	1e-37	50	68
Panolis flammea NPV	1e-36	51	69
Spodoptera exigua NPV	1e-36	51	70
Buzura suppressaria NPV	2e-36	50	68
Autographa californica NPV	2e-36	50	68

Table 3.3. Amino acid sequence identities (%) of fifteen baculovirus polyhedrin (NPV) and granulin (GV) proteins. The amino acid sequence similarity between NeabNPV polyhedrin and the other NPV polyhedrin and GV granulin proteins is shown in brackets ( ).

	AcMNPV	OpMNPV	LdMNPV	HzSNPV	SeMNPV	SfNPV	CfMNPV	MacoNPV	SpltMNPV	WisiNPV	**NeabNPV**	NeseNPV	CpGV	XcGV	PxGV
AcMNPV	100	89	80	85	85	85	89	88	83	80	46 (66)	41	56	52	53
OpMNPV	---	100	79	83	84	83	98	85	82	80	48 (67)	42	55	53	56
LdMNPV	---	---	100	80	81	80	80	83	80	78	45 (65)	40	55	53	54
HzSNPV	---	---	---	100	87	86	82	87	86	82	46 (67)	42	56	53	55
SeMNPV	---	---	---	---	100	98	85	90	84	80	46 (67)	43	54	54	54
SfNPV	---	---	---	---	---	100	84	90	84	80	46 (67)	43	54	54	54
CfMNPV	---	---	---	---	---	---	100	85	81	80	48 (68)	43	56	54	56
MacoNPV	---	---	---	---	---	---	---	100	84	82	47 (68)	44	56	54	54
SpltMNPV	---	---	---	---	---	---	---	---	100	81	45 (66)	42	56	53	54
WisiNPV	---	---	---	---	---	---	---	---	---	100	48 (68)	44	52	51	50
NeabNPV	---	---	---	---	---	---	---	---	---	---	100 (100)	80 (86)	36 (57)	38 (59)	37 (59)
NeseNPV	---	---	---	---	---	---	---	---	---	---	---	100	32	34	33
CpGV	---	---	---	---	---	---	---	---	---	---	---	---	100	87	85
XcGV	---	---	---	---	---	---	---	---	---	---	---	---	---	100	87
PxGV	---	---	---	---	---	---	---	---	---	---	---	---	---	---	100

1 ACATCGTGAAATATGAATTCGC**TTAAG**CCGTCCAAATGTAATAATGTATATGTTAATAAAATTATATGCT

*

71 CCTGTACCAATAAACGGTATCAATGACACACGTTTGAAATTATTATTATTATTATT**ATAAG**CAAATCCGC

* * *

141 AATTGTCAAG**ATG**CCCAATCTTGCAGCAGGTTATCAAACCTCGGCCAAATCGTATATTTATGACAATAAA

M P N L A A G Y Q T S A K S Y I Y D N K

211 TATTACAGAGGTTTAGGTGATATAATCAACAGCGCTAAGAAGCGTAAACATGACCAAGATTGGGAAAAAC

Y Y R G L G D I I N S A **K K R K** H D Q D W E K

281 ACGCCGAAGAACGTCGTGCTTTGAACGGATTTATTTTGCCGTTGGACCCACGAACGGGTCCCGGTAAACA

H A E E R R A L N G F I L P L D P R T G P G K H

351 TGTAAAAATGGTCATGTTTCAGGAGGTTCGTAACATCAAAGCCAATACAATGAAATTGGCTATCAATTGG

V K M V M F Q E V R N I K A N T M K L A I N W

421 TCTGGTAGAGAGTATCTCAGGGAAGTATGGACTACATTTATCGAAGATACTTTTCCCATCAATAATTACC

S G R E Y L R E V W T T F I E D T F P I N N Y

491 AGGAATTCACTGATGTTTTTCTAGAAATACGTTGTACACCTAACAAAAGCAACAGGCATTATCGTTTTCT

Q E F T D V F L E I R C T P N K S N R H Y R F L

561 CGCACAACACGGTCTTCGTATGGATGAAGATTTTGTCCCATGTGATACTATTCGTGTTATTGAACCAGAA

A Q H G L R M D E D F V P C D T I R V I E P E

631 TATTTGCAGGGTAACACCGTGTCATTGAGTCTGTTAAAACGTGATGGGGGATGTCCTATGATGAAAATTC

Y L Q G N T V S L S L L K R D G G C P M M K I

701 GTCAACAATTCAACGAACTGGATCTCGAACAATTTGTTGACCGTATTCTGTGGTGTCATTTCCACAGACC

R Q Q F N E L D L E Q F V D R I L W C H F H R P

771 TATCGTTTACATAGGTACAGACAGCGGAGAGGAAGAAGAAGTTTTTATCGAAGCTTCTTTAACGTTTATT

I V Y I G T D S G E E E E V F I E A S L T F I

841 ATCAAAGAATTTGCACCTGAAGCACCGTTTGTTAATGGACCAGGAATG**TAA**AAAAATAGAATTTTTATTG

I K E F A P E A P F V N G P G M *

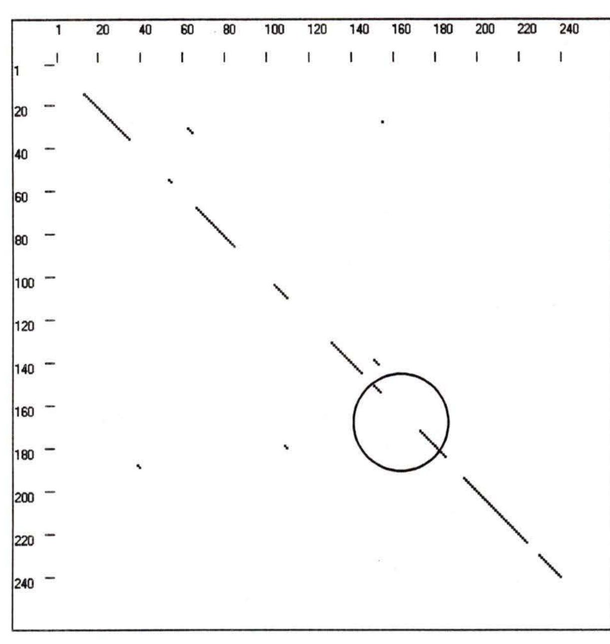
911 CGATTTTATTCAATCGATATGTTTTTCATTATTGATAAAACATATTTGTGATTTTGTATGAATATAACAA

981 TTATTTTTAATTATAATATATTCTATCATCAAAATACAGTATTAAA

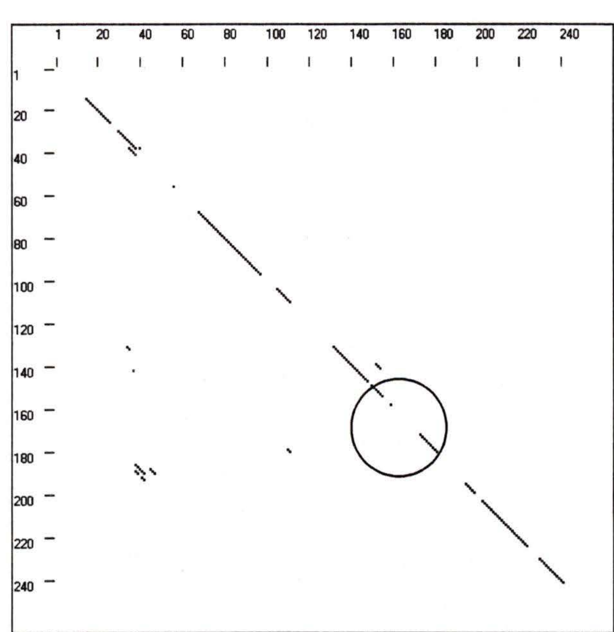
Figure 3.1. The nucleotide sequence of NeabNPV *polh*. The predicted amino acids are indicated with one-letter code designations (start codon ATG and stop codon TAA are shown in bold). The TAAG consensus sequence for late baculovirus transcription initiation is shown in bold and underlined. In the 5' UTR the eukaryotic cell RNA polymerase II promoter motifs CAAT and TATA like regions are denoted by an asterisk (*) and underlined. In the 3' UTR motifs similar to the polyadenylation signal and baculovirus AT_9-like regions are underlined. The nuclear localization-like signal (KKRK) is shown in bold and underlined.

```
                  10         20         30         40         50         60
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
NeabNPV   MPNLAAG--- YQTSATSYIY DNKYYRGLGD IINSAKKRKH DQDWEKHAEE RRALNGFILP
NeseNPV   .....Q.--- .....K.T.. .....K.... .V........ ........DQ ..........

                  70         80         90        100        110        120
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
NeabNPV   LDPRTGPGKH VKMVMFQEVR NIKANTMKLA INWSGREYLR EI-TTFIEDT FPINNYQEFT
NeseNPV   .G.......N .......... .......... ....SK.F.. .VW....... ..........

                 130        140        150        160        170        180
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
NeabNPV   DVFLEIRCTP NKSNRHYRFL AQHGLRMDED FVPCDTIRVI EPEYLQ-GNT VSLSLLKRDG
NeseNPV   .......... ......H... ........MI LCHELPFALS NLCI.R-ETP FLSV......

                 190        200        210        220        230        240
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
NeabNPV   GCPMMKIRQQ FNELDLEQFV DRILWCHFHR PIVYIGTDSG EEEEVFIEAS LTFIIKEFAP
NeseNPV   .GCV..F..E ....S.DE.I .......... .......... .....V.V.. ..........

                 250
          ....|....|
NeabNPV   EAPFVNGPGM
NeseNPV   ..........
```

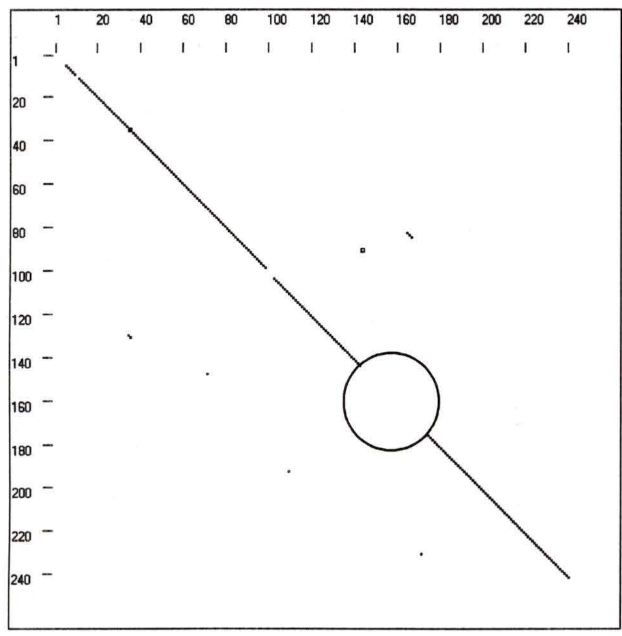
Figure 3.2. Amino acid alignment of two Hymenopteran (NeseNPV and NeabNPV) baculovirus polyhedrin proteins. Amino acids residues identical to NeabNPV are indicated by a dot (.) and amino acid residues similar to NeabNPV are shaded grey.



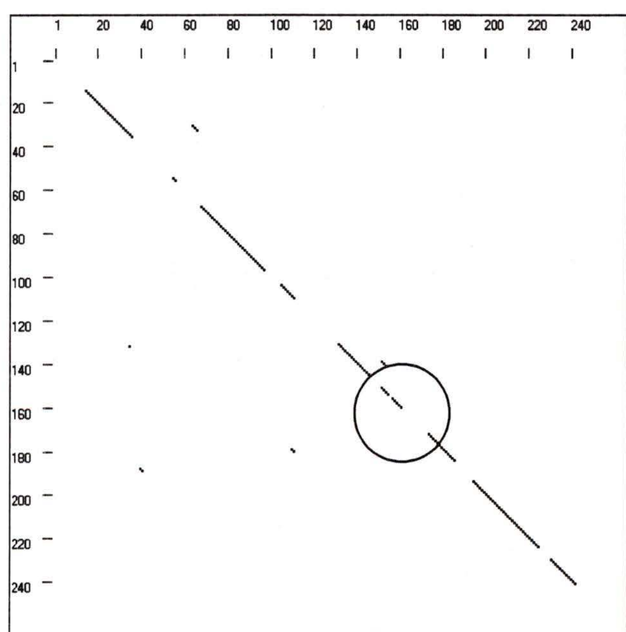
A



B



C



D

Figure 3.3. Graphic representations of the collinearity of baculovirus polyhedrin proteins of AcMNPV (A), LdMNPV (B), NeseNPV (C), and MacoNPV (D) obtained by dot plot analysis. Amino acid residues 140 to 170 of the NeabNPV polyhedrin protein show a low level of sequence identity to all other baculovirus polyhedrins in the same region (shown by a circle).

```
                  10         20         30         40         50         60
          ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|

NeabNPV   ATTATAAGCA AATCCGCAAT TGTCAAGATG
CpGV      TTTATAAGGA ATTTTAATTT TAATAACAAC ATCATG
PxGV      ATTATAAGGA ATTTTTATTT TTACAAAAAC ATG
XcGV      TTTATAAGGA ATTTATATCA AAACAATG
LdMNPV    CCAATAAGTA TTTTATTCTT TTCGTAAAGA TTTTGGAAAA ATCAAATACA CCGTAAAATG
AcMNPV    TAAATAAGTA TTTTACTGTT TTCGTAACAG TTTTGTAATA AAAAAACCTA TAAATAATG
OpSNPV    TCAATAAGTA TTTTTGTCCT TTCGTAAAAC ATTGTGAAAT TTCAAATACA CCATAATG
HearSNPV  AAAATAAGTA TTTTTTTCCT ATTGTTCAAG ATTGTGAAAA ATCAAATATC CCATAATG
HzSNPV    AAAATAAGTA TTTTTTTCCT ATTGTTCAAG ATTGTGAAAA ATCAAATATC CCATAATG
BusuNPV   TCAATAAGTA TTTTTTTCCT ATTGTAAAAC ATTGCGAAAA ATCAAATACA ACATAATG
SpltMNPV  ATGATAAGGA ATTTATTACT ATTGTTCTAG ATAGTGAAAA ATCAAATATC CCATAATG
OpMNPV    TTAATAAGTA ATTTCCTGTT ATTGTAACAA TTTTGTAAAA AAATTTCCTA TAACCATG
MacoNPV   AATGTAAGTA ATTTTCTCCT TTCGTAGAAG ATTGTGAAAA ATAAAATATA ATG
SfNPV     ATTGTAAGTA ATTTTTTCCT TTCGTAAAAC ATTGTGAAAA AATAAATATA ATG
SeNPV     ATTGTAAGTA ATTTTTTCCT TTCGTAAAAC ATTGTGAAAA AATAAATATA ATG
```

Figure 3.4. Comparison of the 5' untranslated region (UTR) from fifteen baculovirus polyhedrin and granulin genes. The baculovirus late transcription signal, TAAG, is shown in bold and the ATG translation start site is shown underlined to the right.

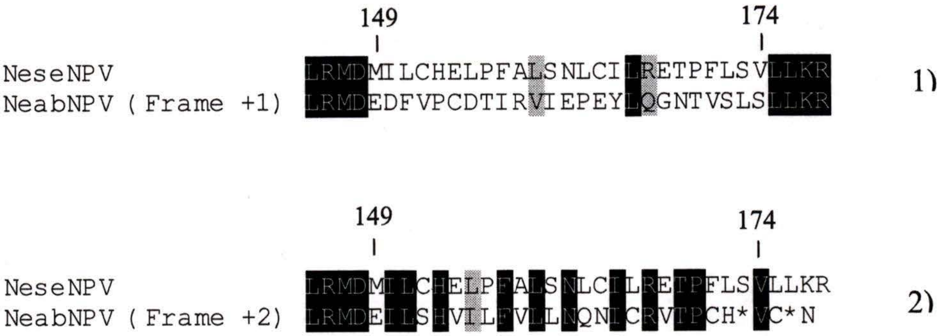


Figure 3.5. Frameshift analysis of the NeabNPV polyhedrin sequence. 1). Alignment of NeabNPV and NeseNPV polyhedrin amino acid sequences from amino acid 149 to 174. 2). Alignment of NeabNPV and NeseNPV polyhedrin amino acid sequences from amino acid 149 to 174 after the introduction of a frameshift into the NeabNPV sequence through the deletion of a single nucleotide from the codon for amino acid 149 (GAA → _AA). Similar amino acid residues are shaded in grey. Identical amino acid residues are shaded in black.

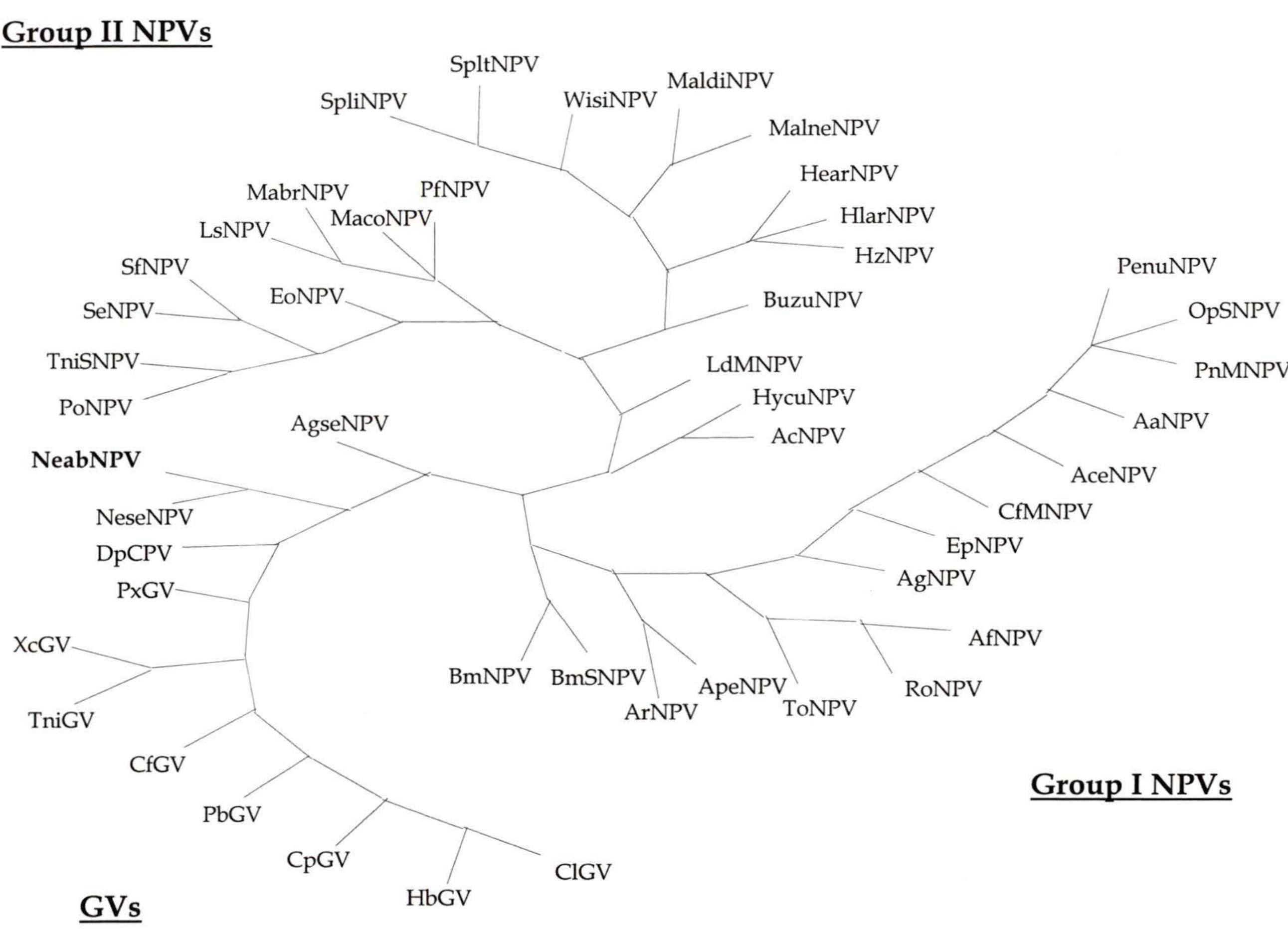


Figure 3.6. Phylogenetic tree of baculovirus occlusion protein sequences. The inferred amino acid sequence of 43 baculovirus polyhedrin and granulin genes was used. The most parsimonious tree was constructed using the 'heuristic' algorithm of PAUP (v4.0b10).

3.4. References

Afonso, C. L., E. R. Tulman, et al. (2001). Genome sequence of a baculovirus pathogenic for Culex nigripalpus. *J Virol.* **75**, 11157-65.

Biswas, T. K., J. C. Edwards, et al. (1985). Characterization of a yeast mitochondrial promoter by deletion mutagenesis. *Proc Natl Acad Sci USA.* **82**, 1954-8.

Blissard, G. W. and G. F. Rohrmann (1990). Baculovirus diversity and molecular biology. *Annu Rev Entomol.* **35**, 127-55.

Bulach, D. M., C. A. Kumar, et al. (1999). Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. *J Invertebr* Pathol. **73**, 59-73.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science.* **284**, 2124-9.

Federici, B. A. (1997). Baculovirus Pathogenesis. The Baculoviruses. L. K. Miller. New York, Plenum Press: 33-59.

Hyink, O., S. Graves, et al. (1998). Mapping and polyhedrin gene analysis of the Epiphyas postvittana nucleopolyhedrovirus genome. *J Gen Virol.* **79**, 2853-62.

Jarvis, D. L., D. A. Bohlmeyer, et al. (1991). Requirements for nuclear localization and supramolecular assembly of a baculovirus polyhedrin protein. *Virol.* **185,** 795-810.

Kozak, M. (1987). An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15(20)**, 8125-48.

Li, S., M. Erlandson, et al. (1997). A physical map of the Mamestra configurata nucleopolyhedrovirus genome and sequence analysis of the polyhedrin gene. *J Gen Virol.* **78**, 265-71.

McLachlin, J. R. and L. K. Miller (1994). Identification and characterization of vlf-1, a baculovirus gene involved in very late gene expression. *J Virol.* **68**, 7746-56.

O'Reilly, D. R., L. K. Miller, et al. (1994). Baculovirus Expression Vectors: A Laboratory Manual. New York, Oxford University Press.

Rohrmann, G. F. (1986). Polyhedrin structure. *J Gen Virol.* **67**, 1499-513.

Rohrmann, G. F. (1992). Baculovirus structural proteins. *J Gen Virol.* **73,** 749-61.

Sadler, T., V. Ward, et al. (1998). Examination of New Zealand's endemic Wiseana nucleopolyhedrovirus by analysis of the viral polyhedrin gene. *Arch Virol.* **143**, 2273-88.

van Strien, E. A., D. Zuidema, et al. (1992). Nucleotide sequence and transcriptional analysis of the polyhedrin gene of Spodoptera exigua nuclear polyhedrosis virus. *J Gen Virol.* **73**, 2813-21.

Vlak, J. M. & Rohrmann, G. F. (1985). The Nature of Polyhedrin. In *Viral Insecticides for Biological Control*, pp. 489-544. Edited by K. Maramorosh & K. E. Sherman. Orlando: Academic Press.

Yang, S. and L. K. Miller (1999). Activation of baculovirus very late promoters by interaction with very late factor 1. *J Virol.* **73**, 3404-9.

Chapter 4. Analysis of Putative NeabNPV Structural Proteins

4.1. Introduction

The complex life cycle of Lepidopteran baculoviruses involves two structurally distinct viral forms, each of which are essential for the natural propagation of the occluded baculovirus. These two unique forms are called the occlusion-derived virion (ODV) and budded virion (BV). The occlusion-derived virion is produced during the latter stages of infection where viral nucleocapsids acquire a membranous envelope within the host cell nucleus and newly formed virions are packaged into protein occlusion bodies (OBs). ODVs are protected from environmental decay by the protein occlusion body and are capable of persisting for long periods of time outside the host. ODVs are relatively non-infectious for cultured cells (Volkman and Summers, 1977) and are responsible for the initiation of viral infection in the midgut cells of susceptible insect hosts.

Following infection of the insect midgut cells by ODVs, viral DNA replication is initiated and nucleocapsids begin assembly within the nucleus. These early stage nucleocapsids travel from the nucleus to the host cell plasma membrane where they acquire a membranous envelope by budding through the plasma membrane. This budded virus phenotype spreads the infection from cell to cell throughout the insect body and is highly infectious to cultured cells (Volkman and Summers, 1977).

Although the nucleocapsids of the two viral phenotypes appear to have similar structure and protein content, the composition and structure of their membranous envelopes are different (Funk *et al.,* 1997). The difference in envelope protein

composition is most likely due to the unique functions required of each viral form. The BV envelope is adapted for movement and infection of tissues within the insect host and is suitable for an intracellular and extracellular environment. In contrast, the ODV envelope is adapted for interaction with the proteins of the viral occlusion bodies and for facilitating the infection of midgut epithelial cells in the harsh environment of the alkaline insect gut (Funk *et al.,* 1997).

Recent studies have shown that NeabNPV infects the midgut epithelial cells of *N. abietis* larvae and appears to not produce a secondary infection within other tissues types (unpublished, Lucarotti). This type of localized infection does not appear to occur in Lepidopteran insects infected by NPVs, however, some GVs are known to specifically infect only midgut epithelial cells (HbGV) (Federici, 1997). The mechanism by which the NeabNPV spreads from cell to cell in the midgut epithelium is not known and may or may not involve a budded virus phenotype.

Fifteen structural protein genes appear to be conserved among the *Baculoviridae* and include genes encoding nucleocapsid associated proteins (*vp39, vp91*), ODV envelope proteins (*odve18, odve25, odve56, odve66, p74*), and a DNA binding protein *p6.9*. VP39 is the major structural protein of baculovirus nucleocapsids and appears to be highly conserved among the Lepidopteran baculoviruses. Structural protein genes that have been found in some, but not all baculoviruses include the nucleocapsid proteins *p87* and *AcORF1629*, as well as the BV specific envelope proteins *gp64* and *ld130.*

Very little data is currently available for NeabNPV at the DNA, RNA, or protein level, and as such, the protein composition of NeabNPV ODV and BV nucleocapsids and envelopes are unknown. Members of our lab attempting to develop an insect cell culture

system that is susceptible to infection by NeabNPV have reported cytopathic effects (see below). Infected cell culture supernatants were obtained for protein analysis and putative virions present in the supernatant have been labelled 'budded virus', although the exact form of any virus particles found outside infected cells has yet to be determined.

In this study I have attempted to identify NeabNPV structural proteins by isolating and partially sequencing several proteins obtained from both putatively infected cell culture supernatant and infected insect larvae. Proteins were separated by SDS-PAGE and the partial amino acid sequences of 4 putative BV, 4 ODV, and 2 OB proteins were determined by mass spectrometry. Amino acid sequences were then compared to translated DNA sequences obtained from our NeabNPV genomic libraries. From this analysis the baculovirus specific nucleocapsid protein VP39 was putatively identified and the inferred amino acid sequence data for this protein was analysed. As well, the protein profiles of the NeabNPV 'BV' and 'ODV' forms were compared to the protein profiles of the AcMNPV BV and ODV forms. The AcMNPV structural protein profiles were obtained from Braunagel and Summers (1994).

4.2. Methods and Materials

4.2.1. Cells and Virus

NeabNPV proteins were isolated from two different sources. The ODV form was isolated from occlusion bodies obtained from infected *Neodiprion abietis* larvae. Dead *N. abietis* larvae were collected from balsam fir test plots in Newfoundland, Canada. These test plots were sprayed with preparations of NeabNPV OBs as part of a controlled aerial

spray program being conducted by the Canadian Forest Services, Atlantic Division, New Brunswick. Dead, infected larvae were homogenized and OBs were isolated and resuspended in ddH_2O according to O'Reilly *et al.* (1992). Samples of the NeabNPV OBs were then sent to our lab at the University of Victoria, Victoria, British Columbia and ODVs were isolated from these OBs (described below).

Putative NeabNPV 'budded virions' were isolated from supernatants of *Choristoneura fumiferana* (Cf70) (FMPI-299) ovarian tissue culture cells infected with NeabNPV. Infected cell culture supernatants were provided by another member of our lab, Beatrixe Whittome, who performed all the NeabNPV cell culture system research: NeabNPV DNA (2ug) extracted from viral OBs was used to transfect uninfected Cf70 cells. Several days post-infection the cells demonstrated cytopathic effects (swollen, rounded appearance, inability to adhere to tissue culture flask, lyses/death). The NeabNPV polyhedrin gene was also PCR amplified from putatively infected Cf70 cells using primers specific to the NeabNPV polyhedrin sequence (Chapter 3). Media and cells were collected after 8 passages and centrifuged at 400 g for 10 min. This infected cell culture supernatant was then used to partially purify any extracellular virus.

4.2.2. Partial Virus Purification

Extracellular virus was partially purified from the infected cell culture supernatant by centrifugation at 5000 g for 10 min (4 °C) to pellet remaining tissue culture cells and debris. The supernatant was removed, spun at 10,000 g for 15 min (4°C) and no pellet was seen. The supernatant was then spun at 15,000 g for 45 min (4 °C), the supernatant was removed, and the pellet was resuspended in a small volume of ddH_20. Centrifugation

at 15,000 g followed by washing in a small volume of ddH_20 was repeated again and the final, washed pellet was resuspended in a small volume of TE buffer (10mM Tris-HCl, 1mM EDTA, pH 7.0) and stored at 4°C. The presence of virion particles within this preparation was not confirmed by microscopy.

Samples containing NeabNPV OBs (obtained from the CFS Atlantic Division) were treated with SDS (0.5%) followed by incubation in an alkaline dissolution buffer (0.1M Na_2CO_3, 10mM Tris-HCl, 1mM EDTA, pH 11.0) to release ODVs. After 30 min at 25°C, the solution was neutralized with Tris-HCl (0.1M, pH 7.6), clarified by centrifugation (5,000 g, 5 min), and ODVs were pelleted by spinning at 15,000 g, for 1 hour (4 °C). The partially purified sample was next treated with Nonidet P40 (NP40) at a final concentration of 1% and incubated for 1 hour at room temperature with gentle agitation. Any released ODV nucleocapsids were sedimented at 80,000 g for 45 min. (7°C). Removal of the viral envelope and the presence of viral nucleocapsids in the pellet were not confirmed by microscopy. The obtained pellet was dissolved in a small volume of TE buffer (10mM, pH 7.5) and stored at 4°C. Putative membranous envelope proteins present in the supernatant were acetone-precipitated at –20°C (1 hour), pelleted by centrifugation (15,000 g, 1 hour, 4°C), and resuspended in a small volume of TE buffer (10mM, pH 7.5).

4.2.3. SDS-PAGE

Putative NeabNPV BV and ODV proteins were separated using discontinuous SDS-PAGE with a dissociating buffer system (Hames and Rickwood, 1981). Gels consisted of a stacking phase (4% acrylamide) and a resolving phase (10% acrylamide)

and were run at 60 volts and 120 volts, respectively. Protein bands were visualized by staining with the GelCode Blue Stain Reagent (Pierce), following manufacturers protocols. A broad range protein marker (NEB) was used as a molecular size standard. Polyacrylamide gels were analysed using the TotalLab 1-D gel analysis imaging software (Nonlinear Dynamics) to determine the approximate size of protein bands.

4.2.4. Protein Sequencing and Computer Analysis

Major protein bands were excised from SDS-PAGE gels and partially sequenced using mass spectrometry (Genome BC Proteomics Centre, University of Victoria). Amino a sequence determination involves several steps: 1) protein bands are destained, 2) digested with trypsin to produce smaller peptide fragments, 3) attached to an organic matrix, 4) analysed by MALDI-TOF mass spectrometry. MALDI-TOF stands for matrix-assisted-laser-desorption-ionization time-of-flight. The basic principal behind this technology involves the determination of the mass of a molecule based on its charge and time of flight through a vacuum. MALDI-TOF mass spectrometry can be used to determine the exact molecular weight of the individual amino acid molecules that make up a peptide, which in turn is used to determine the amino acid sequence of a peptide fragment. The partial amino acid sequence data obtained for each protein was compared with the NCBI Genbank database as well as partial NeabNPV restriction and genomic DNA libraries using BLAST (v2.0) (Altschul, *et al.,* 1990). Multiple sequence alignment of a partial NeabNPV VP39 protein was performed using ClustalX (v1.81) and all modifications to the sequence data were performed using Bioedit (v5.0.9).

4.3. Results and Discussion

4.3.1. SDS-PAGE and NeabNPV Structural Proteins

NeabNPV BVs were isolated from infected Cf70 cultured cells while NeabNPV ODVs were isolated from polyhedral occlusion bodies obtained from infected balsam fir sawfly larvae. Proteins from the BVs, ODVs, and viral occlusion bodies were separated by SDS-PAGE, visualized using the GelCode Blue Stain Reagent, and analysed using TotalLab 1-D Gel Imaging software. Mock-purified (uninfected) Cf70 cell culture supernatant and a 15% fetal bovine albumin solution (FBS) were included as negative controls for the BVs. Figure 4.1 shows the protein profiles for the NeabNPV BV, ODV, and OB as well as proteins present in the negative controls. Approximately 21, 9, and 5 proteins were detected in the BV, ODV, and OB, respectively (Fig. 4.1, lanes 5, 7, and 4). Some of the protein bands (indicated by a dot) estimated by the TotalLab software are artifacts and were not considered during analysis. These artifacts were dots that appeared at dark spots or creases on the polyacrylamide gel picture, which TotalLab determined to be protein bands. Gel pictures were inspected by eye and dots that did not appear to represent protein bands (artifacts) were excluded from the analysis. The FBS and uninfected Cf70 cell culture supernatant contained several proteins indicated in Fig. 4.1, (lanes 2 and 3, respectively). In particular, a prominent protein band was seen at approximately 65.3 kDa in both the FBS and uninfected cell culture supernatant. This band may represent the fetal bovine albumin protein (66.5 kDa), which was present in the Cf70 cell culture nutrient broth (a 15% solution of fetal bovine albumin was used as an ingredient for the cell culture nutrient broth). Several protein bands in the uninfected cell

culture supernatant were present at very low concentrations and were difficult to visualize. Faint protein bands appear in this lane at approximately 175, 119, 70, 65, 57, 47, 34, and 33 kDa (Figure 4.1, lane 3 - some bands not visualized). The sizes of major protein bands that were visualized are indicated in Table 4.2.

The viral envelope was putatively removed (but not confirmed) from ODV particles by treatment with NP-40 and nucleocapsid particles were isolated by centrifugation. Putative envelope and nucleocapsid fraction proteins were separated by SDS-PAGE and visualized using the GelCode Blue Stain Reagent. The sizes of major protein bands that were visualized are indicated in Table 4.2. Figure 4.2 shows the protein profile of putative ODV nucleocapsid (lane 2) and envelope fractions (lane 3 – no proteins appeared to be present in this fraction) as well as the BV form (lane 4). Protein bands were visualized from the ODV nucleocapsids and BV but not from the ODV envelope fraction, which could have occurred for several reasons. First, the envelope proteins may have been present in too low of a concentration to be visualized or may not have precipitated properly after treatment with acetone. Secondly, the envelope may have been lost earlier in the purification process during centrifugation and as a result may not have been present in the first place for removal by treatment with NP40. Third, it is possible that the membranous envelope was not properly removed from the nucleocapsids and remained attached. Since the protein profile of the ODV nucleocapsids (Fig. 4.2, lane 2) is almost identical to the protein profile of the intact ODVs (Fig. 4.1, lane 7), it seems plausible that the viral envelope was either lost during the centrifugation step or was not properly removed from the nucleocapsids after treatment with NP-40. A study conducted by Braunagel and Summers (1994) involving the isolation of viral

nucleocapsids and envelopes showed that nucleocapsids were typically contaminated with a small amount of membranous envelope after treatment with NP-40. As a result, it is also possible that the viral envelope was only partially removed from the nucleocapsids but was still present in high enough quantity to provide the protein profile seen in Fig. 4.2 (lane 2). Since the prepared viral samples were not visualized by electron microscopy it was difficult to determine if the viral envelope was removed or was left intact.

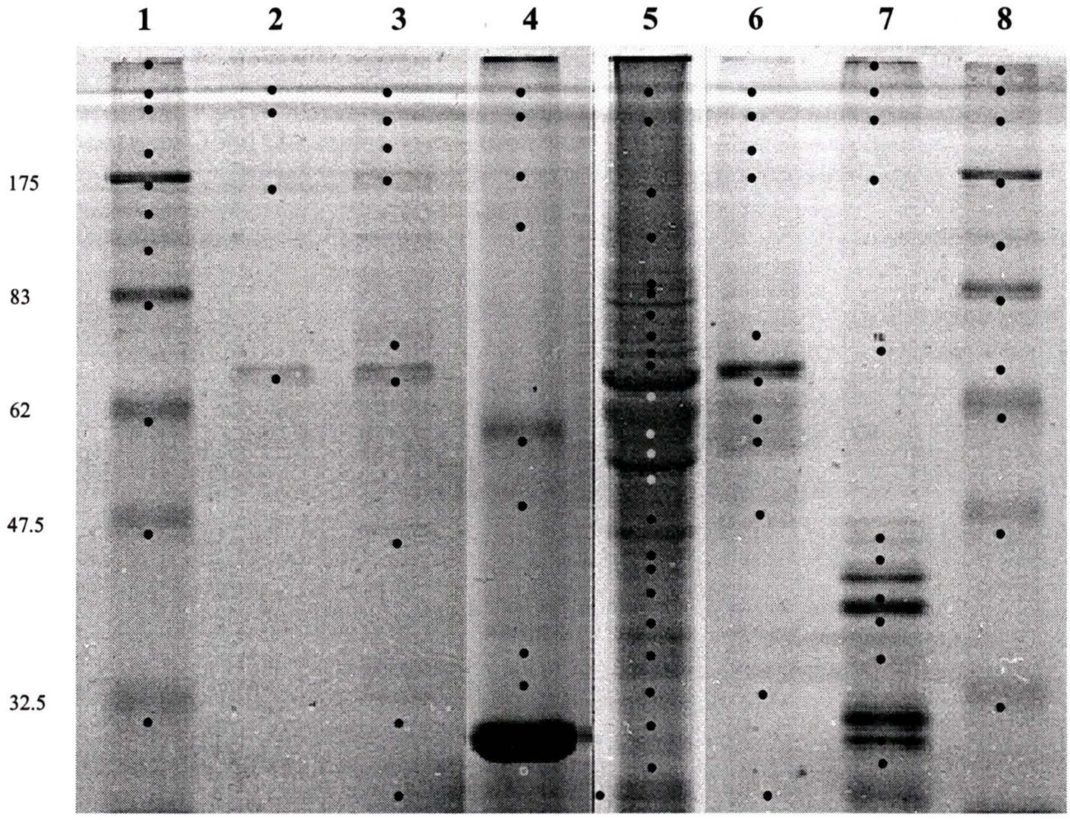


Figure 4.1. Putative NeabNPV structural proteins. Proteins were separated on a 10% polyacrylamide gel by SDS-PAGE and stained with the GelCode Blue Stain Reagent. 1 - protein marker (kDa), 2 - graces medium with 15 % FBS, 3 - Uninfected Cf70 cell culture supernatant, 4 - NeabNPV occlusion body, 5 - NeabNPV infected cell culture supernatant, 6 - NeabNPV infected cell culture supernatant (1/10th volume of lane 6), 7 - NeabNPV occlusion-derived virion protein, 8 - protein marker. Putative protein bands are indicated by a black/grey dot.

Band	Lane 1	Lane 2	Lane 3	Lane 4	Lane 5	Lane 6	Lane 7	Lane 8
1					170			
2					129.7			
3		170			96	72.7		
4		65.3		138.8	89.7	65		179.1
5	175		70.2	59.9	79.6	62.2	69.2	125.2
6	148.2		65.3	51	72.7	59.9	47.2	86.9
7	118.8		46.5	36.6	69.2	50	44.7	66.7
8	83		32.6	34.7	67.1	34.1	40.8	62.4
9	62		26.9	28.5	64	26.7	38.7	47.8
10	47.5				61		36.3	33.5
11	32.5				59		31.1	
12					55.2		29.4	
13					49.6		26.4	
14					45.3			
15					43.6			
16					41.3			
17					38.9			
18					36.3			
19					34.3			
20					32.5			
21					29.2			

Table 4.1. Sizes (kDa) of protein bands in Fig. 4.1 that are less than 180 kDa. Lane 1 - protein marker, Lane 2 - graces medium with 15 % FBS, Lane 3 - Uninfected Cf70 cell culture supernatant, Lane 4 - NeabNPV occlusion body, Lanes 5 and 6 - NeabNPV infected cell culture supernatant, Lane 7 - NeabNPV occlusion-derived virion protein, Lane 8 - protein marker.

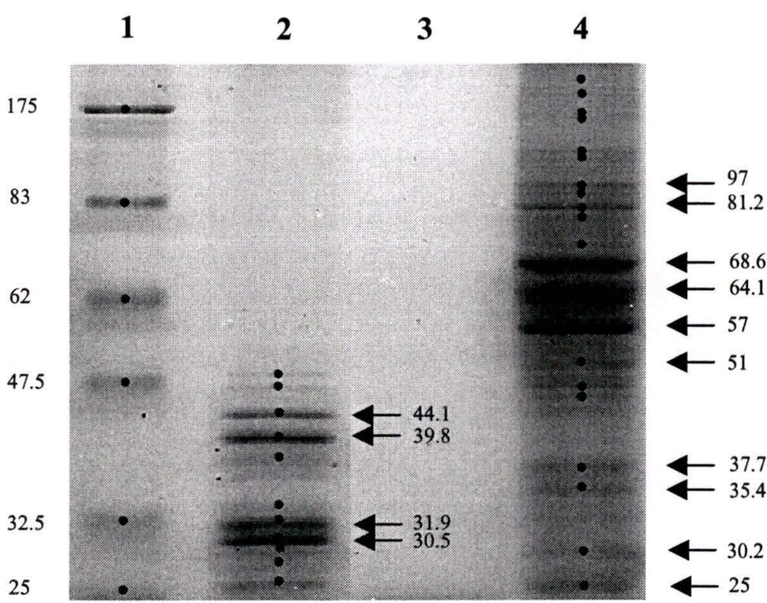


Figure 4.2. Putative NeabNPV structural proteins. Proteins were separated on a 10% polyacrylamide gel by SDS-PAGE and stained with the GelCode Blue Stain Reagent. Lane 1 - Broad range protein marker (kDa). Lane 2 - Putative occlusion-derived virion (ODV) nucleocapsid fraction. Lane 3 – Putative occlusion-derived virion envelope fraction (no proteins appeared to be present in this fraction). Lane 4 – Putative budded virion. Protein bands are indicated by a black dot. The sizes (kDa) of several well-defined protein bands are indicated by an arrow.

Band	Lane 1	Lane 2	Lane 3	Lane 4
1	175	49.9		
2	83	47.7		
3	62	44.1		169.6
4	47.5	39.8		159.5
5	32.5	38.1		124.3
6		33.7		116.8
7		31.9		97
8		30.5		88.3
9		29.5		81.2
10		27.9		78.5
11		26.4		73.4
12				68.6
13				64.1
14				57
15				51
16				47
17				45.3
18				37.7
19				35.4
20				30.2
21				25

Table 4.2. Sizes (kDa) of protein bands less than 175 kDa seen in Fig 4.3. Lane 1 - protein marker, Lane 2 – Putative NeabNPV ODV nucleocapsid proteins, Lane 3 - Putative NeabNPV ODV envelope proteins, Lane 4 - Putative NeabNPV BV proteins.

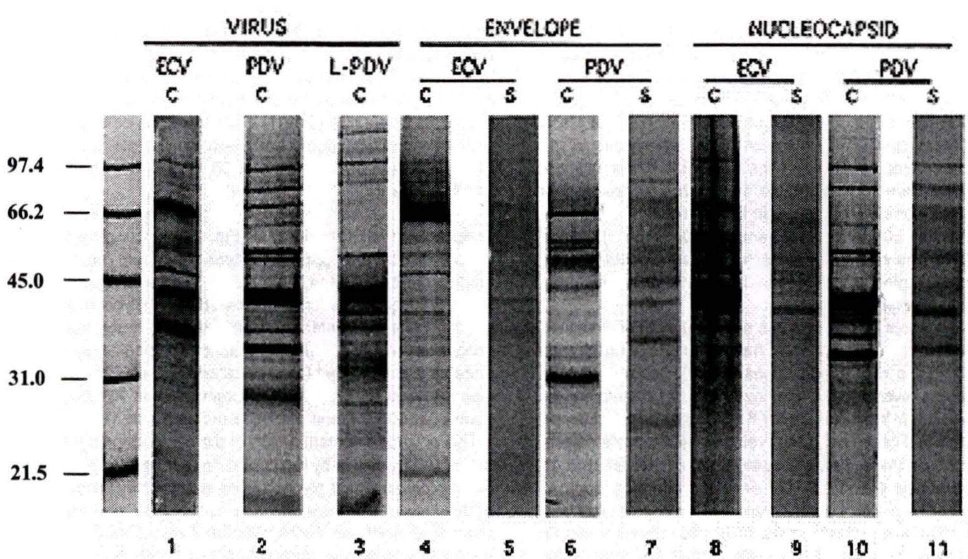


Figure 4.3. Structural proteins of AcMNPV virus and virus fractions. ECV - extracellular virus, PDV - cell culture polyhedra-derived virus, L-PDV - larval occlusion-derived virus. Different staining procedures were used including Coomassie Brilliant Blue (C), and Silver stain (S). Sizes (kDa) of protein bands are shown to the left of the protein marker. (Picture from Braunagel and Summers, 1994).

Staining with the GelCode Blue Stain Reagent revealed approximately 11 proteins in the ODV and 20 proteins in the BV (Fig. 4.2, lanes 2 and 4, respectively). Comparison of the occlusion-derived and budded virion forms revealed qualitative differences between their protein profiles. In particular are the proteins at 44.1, 39.8, 31.9, and 30.5 kDa in the ODV, which do not appear to be present in the BV. In contrast, the BV appears to have several proteins that are not present in the ODV, including proteins at approximately 97, 81.2, 68.6, 64.1, 57, and 51 kDa. Assuming there are no contaminating, non-viral proteins present in the ODV and BV protein preparations, the different ODV and BV protein bands could represent unique viral envelope proteins. Previous studies have indicated that baculovirus nucleocapsids (ODV and BV) have almost identical protein composition, however, envelope proteins may differ between the two viral forms (Funk *et al.,* 1997). Only a partial crude purification process was performed for this study and as a result it is almost impossible to conclude that the protein bands seen in the above gels are indeed NeabNPV structural proteins. The visualized protein bands may represent NeabNPV structural proteins but a more complete protein purification process involving viral purification using sucrose gradients and confirmation of viral structures (e.g. nucleocapsids) within samples would have to be performed prior to making any definite conclusions.

Limited information is available on proteins specific for BVs and ODVs, but no overall study is available that allows the direct comparison of BV and ODV structural proteins from different baculovirus species. A comprehensive study of the structural proteins of *Autographa californica* NPV was performed by Braunagel and Summers (1994) (Fig. 4.3) and included an in-depth analysis of the proteins from the BV and ODV

nucleocapsid and envelope fractions. This study showed that viral protein profiles differed depending on where the virus was isolated. For example, ODVs isolated from infected insect larvae showed different protein banding patterns than ODVs isolated from infected Cf70 cell culture. This same study also showed that different proteins were detected from the same virus preparation when different staining techniques were used. For example, staining the proteins of AcMNPV BVs with Coomassie Brilliant Blue revealed 21 proteins while silver staining revealed 29 proteins (due to its increased sensitivity – capable of detecting proteins in the nanogram (ng) range). As a result, it is difficult to directly compare proteins from different baculoviruses unless identical procedures are used to isolate virus fractions, separate proteins with SDS-PAGE, and stain the separated proteins.

A comparison of the protein profiles of AcMNPV and NeabNPV showed several similarities and differences. In AcMNPV (Fig. 4.3) approximately 25 to 30 proteins, in the range of 21 to 97 kDa, were detected by SDS-PAGE (12.5% acrylamide) and Coomassie Brilliant Blue staining (performed by Braunagel and Summers (1994), Fig. 4.3). This is similar to NeabNPV where approximately 25 proteins, in the range of 25 to 97 kDa, were detected by SDS-PAGE (10% acrylamide) and staining with the GelCode Blue Stain Reagent (Fig 4.2). Proteins at 89, 70, 60, 50, and 25 kDa appeared to be specific to AcMNPV ODVs whereas proteins at 67, 45, and 35 kDa appeared to be specific to AcMNPV BVs (Fig. 4.3). NeabNPV ODVs appeared to lack any proteins above 50 kDa but did show a protein band at 26.4 kDa (Fig. 4.2). NeabNPV BV showed protein bands at 68.6 and 35.4 kDa that were specific to this particle, but lacked the strong band near 45 kDa that is seen in the AcMNPV BV. Although some structural

proteins appear to be conserved between AcMNPV and NeabNPV (e.g. 26.4 kDa protein) their overall protein profiles appear to be quite different. Further studies involving silver staining, which is capable of detecting proteins concentrations in the nanogram range (as opposed to Coomassie blue staining which is only capable of detecting protein concentrations in the microgram range) may reveal protein bands that were not visualized in this experiment. The differences between the protein profiles of AcMNPV (Braunagel and Summers, (1994), Fig. 4.3) and NeabNPV, however, are most likely the result of non-viral proteins present in the NeabNPV sample preparations. In contrast to the experiment conducted by Braunagel and Summers (1994), NeabNPV PIBs, OB derived ODVs, and cell culture derived BVs were not purified by centrifugation through a sucrose gradient. This step is important for the specific isolation of baculovirus particles and removal of contaminating membranes and proteins, which may carry over sample preparations.

4.3.2. Identification of two NeabNPV Structural Proteins

Discrete protein bands were excised from SDS-PAGE gels and their partial amino acid sequence was determined. Proteins that were analysed include 4 ODV (44.1, 39.8, 31.9, 30.5 kDa; Fig. 4.2, lane 2), 5 BV (68.6, 57, 30.2, 25, 6.8 kDa; Fig. 4.2, lane 4) and 2 OB proteins (59.9, 28.5 kDa; Fig. 4.1, lane 4). The 6.8 kDa BV protein is not seen in Fig. 4.2 (lane 2) because the protein band was not visualized in that polyacrylamide gel (low molecular weight protein bands did not stain well). As a result, the 6.8 kDa protein was excised from a different polyacrylamide gel (picture not shown). The sequence data

obtained from the proteins was compared to translated DNA sequences from our NeabNPV genomic libraries (Chapter 2), as well as the NCBI Genbank database.

Pairwise comparison of the NeabNPV protein sequence data resulted in very few significant alignments which may have occurred for several reasons: first, the protein sequences obtained may not have been present in the NeabNPV restriction libraries or genomic library; secondly, the protein sequences that were obtained were relatively short (7 to 14 amino acids), which can make sequence alignment difficult. Shorter sequences may occur more randomly in a sequence database and produce less significant score values for local alignments. As a result, the statistical significance (e-value) of such alignments may exceed the threshold limit of acceptable alignments and go unnoticed. The sequence data obtained for the 39.8 kDa ODV protein and 28.5 kDa occlusion protein, however, aligned with several translated sequences from our NeabNPV DNA libraries. Table 4.3 shows the amino acid sequences obtained for peptides from both the 39.8 kDa ODV protein and 28.5 kDa occlusion protein.

Several of the peptide sequences from the 28.5 kDa protein aligned significantly (100% identity) with end sequence data from the *Eco*RI VT3 and MT3 restriction fragments. Both the MT3 and VT3 end sequence data aligned significantly (e-value less than 10^{-29}) with baculovirus polyhedrin sequences present in Genbank (Chapter 2), suggesting that the 28.5 kDa protein is the major occlusion protein, polyhedrin. The estimated size of this protein (28.5 kDa) also agrees well with the molecular weight of polyhedrin proteins isolated from other baculovirus species (~29.5 kDa) (Rohrmann, 1986).

Several of the peptide sequences from the 39.8 kDa ODV protein aligned significantly (100% identity) with regions of several translated DNA sequences from our NeabNPV genomic library. These alignments included sequences 863, 954, 1065, 1538, 1840, and 1888. Comparison of each of these DNA sequences was then compared to the NCBI Genbank protein database using BLASTx, which revealed that each DNA fragment was most similar to the VP39 protein from SpltMNPV (Table 4.4). NeabNPV DNA sequences 863, 1065, 1538, 1840, and 1888 aligned over the same region of the SpltMNPV VP39 protein (amino acids 15 to 141) and were assembled into a contiguous sequence. The consensus sequence from this contig was 1070 nucleotides in length (356 amino acids) and was compared to the NCBI Genbank database. The putative NeabNPV VP39 displayed an amino acid identity of 33% and a similarity of 56% over a 123 amino acid region (Fig. 4.4). Sequence 954 also aligned most significantly with the SpltMNPV VP39 protein, but in a different region (amino acid residue 160 to 246). The consensus sequence (863, 1065, 1538, 1840, and 1888) and sequence 954 were combined to create a partial NeabNPV *vp39* sequence which was then used for further sequence analyses.

39.8 kDa	28.5 kDa
LAGLLTTDLK	PNLAAGYQTSAK
LNLTNLLCPR	LALNWSGR
HLEDYFR	VAGDLLNSAK
SNTNECLFR	MVMFQEVR
SVEVTSVTR	FVFFQEVR
PLDGQLDALLSDVK	FNELDLEQFVDR
NYEGGNFK	FLAQHGLR
NTNESSLFR	

Table 4.3. Peptide sequences from NeabNPV structural proteins. The 39.8 kDa protein was isolated from NeabNPV ODVs. The 28.5 kDa protein was isolated from NeabNPV occlusion bodies.

ꟾeabNPV Sequence	Aligned With	Score	Alignment Length
863	SpltMNPV	2.00E-05	72-141
954	SpltMNPV	5.00E-06	160-358
1065	SpltMNPV	7.00E-12	15-137
1538	SpltMNPV	4.00E-12	15-137
1840	SpltMNPV	3.00E-11	15-141
1888	SpltMNPV	8.00E-12	15-141

le 4.4. NeabNPV genomic DNA fragments which aligned with the ılovirus nucleocapsid protein VP39.

```
Neab: 1   NECIFRYARENDKLNQTIDPGYNCSNYLS---GLYICQKHLEDYFRLSLRKINQLEGTDQ
          N CIF+  R  +  NQ  +     CS+  S   G+++CQ HL  +F++     I   +GT
Splt: 15  NYCIFQGVRPIE-FNQCSNYRSPCSDDASQNDGVFMCQYHLSRFFKIEKTSIAIPDGT--

Neab: 58  GRVFERLSVLSRLPHNYR--NRVIVPTRRNYEDIFKVAYLPISYQLVFHLLYQNQSAIDK
          G+    R+    S + HN R   +R+++PT+ NY+ +   V+ LP + +LV HL+Y N++A  +
Splt: 72  GQKLYRIVGKSLVSHNARANDRILIPTQENYQAVMNVSMLPPAERLVLHLIYNNRTAAAE

Neab: 116 ICQDVKISGNKFSINGGAIDALISDVKKRLAGIITTDIKCSVEVTSVTRNWESVPNIINI
          IC  ++    N  S      +   ++ +  R+  I    +  CSV  ++  R++ +  ++
Splt: 132 ICNQLRQQENFRS----DVVENVTSMAYRIIQITDPEAFCSVVASNDIRSFGNEADLTRT

Neab: 176 MDDPQEPLFIKNLITNLICPREVVIDNKTFKFETAANVIFVEEGLIVDNLYNPEAELYPL
           +     P FIKNLI  L+ P  + + N TF  +  A V      GL    LYN E E  L
Splt: 188 YN--SLPPFIKNLINTLVAPEVLTLGNNTFPLKACATVSITTNGLQAVQLYN-EVEPKYL

Neab: 236 YN
          YN
Splt: 245 YN
```

Figure 4.4. Pairwise alignment of the inferred, partial amino acid sequence of the putative NeabNPV *vp39* gene and the SpltMNPV *vp*39 gene. *Neodiprion abietis* NPV (Neab), *Spodoptera litura* MNPV (Splt). The statistical significance (e-value) for this alignment was estimated to be 9e-21.

4.3.3. Analysis of a Putative Partial NeabNPV *vp39* Gene

The putative partial NeabNPV *vp39* sequence is 864 nucleotides long (288 amino acids) and aligned most significantly with the SpltMNPV VP39 protein. Analysis of the partial sequence revealed 3 putative phosphorylation sites (S/T-X-R/K) and 3 N-linked glycosylation sites (N-X-S/T) (Fig 4.5). The inferred amino acid sequence of NeabNPV *vp39* was determined and compared to 14 other baculoviruses VP39 proteins (Fig. 4.5). A multiple sequence alignment was performed using ClustalX (v1.81) with default parameters and the alignment was trimmed using Bioedit (v5.0.9) to include only those amino acids that aligned with the partial NeabNPV *vp39*. Comparison of these sequences indicated an amino acid identity ranging from more than 90% between AcMNPV/BmMNPV and HearSNPV/HzSNPV, to less than 20% between CuniNPV and all other baculoviruses (Table 4.5). The average amino acid identity of the partial NeabNPV *vp39* was 23% with the NPVs and 22.1% with the GVs. The sequence identity of the partial NeabNPV VP39 protein ranged from 16.5% with CuniNPV to 27% with SpltMNPV (Table 4.5). A maximum parsimony phylogenetic tree was constructed (data not shown) and yielded a phylogeny with a topology almost identical to that shown in Fig. 5.3 (Chapter 5).

Alignment of the *vp39* sequences indicated the presence of 24 conserved amino acid residues (Fig. 4.5). Fifteen of these residues appeared to be universally conserved among the baculoviruses and may be necessary for the proper folding of the protein. Nine amino acid residues are conserved among all the NPVs and GVs except CuniNPV. Previous studies have shown that CuniNPV lacks any homologues of ODV envelope or occlusion body proteins, which suggests that CuniNPV ODVs and BVs may be

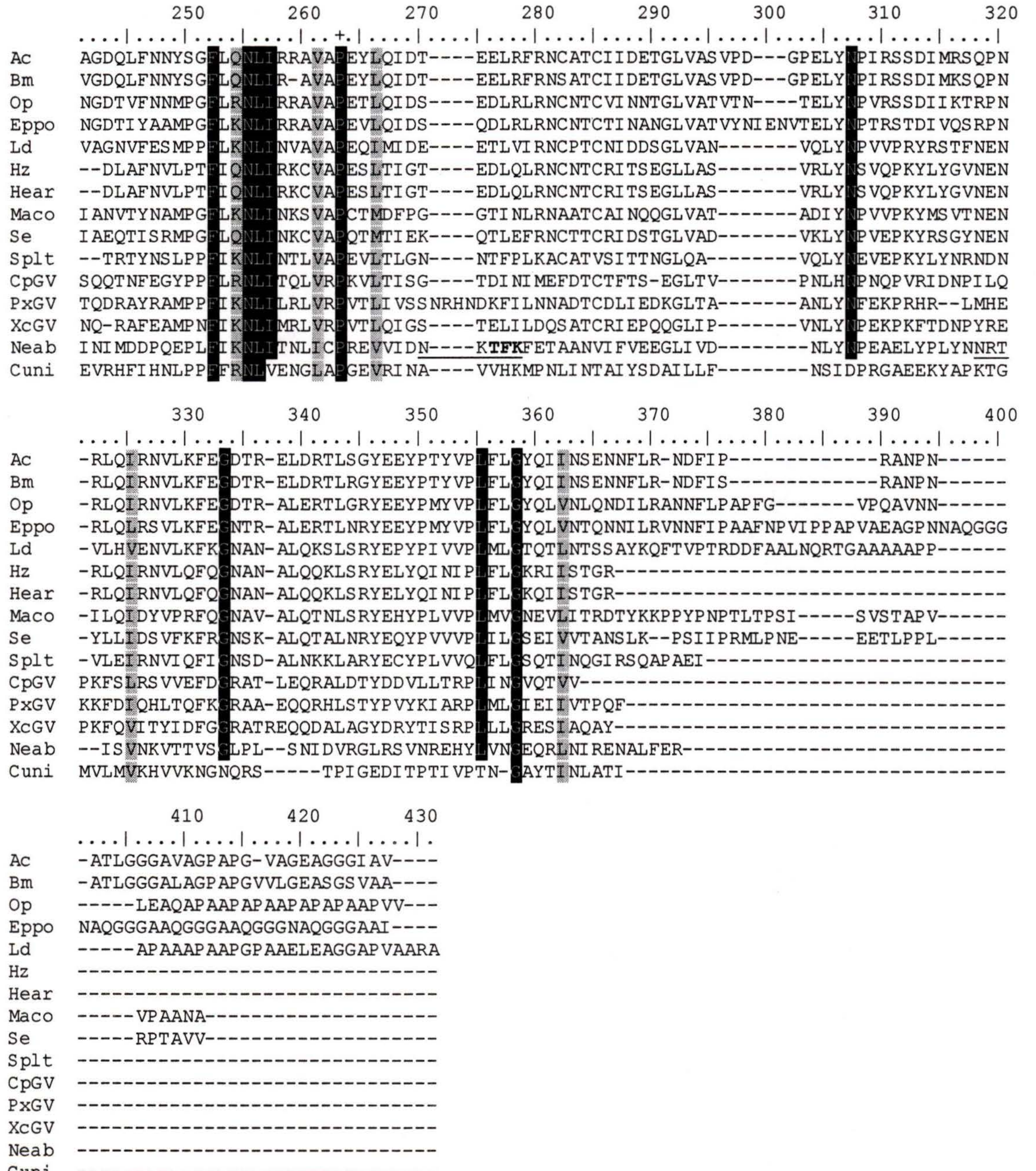
structurally and compositionally different from those of other baculoviruses (Afonso *et al.,* 2001). As well, the genome of CuniNPV shows very low levels of identity between homologous ORFs, a lack of conservation in gene order, and the absence of many genes found in all baculovirus genomes (Afonso *et al.,* 2001). These differences suggest a large evolutionary distance between CuniNPV and the other baculoviruses and may explain the absence of the 9 conserved amino acid residues. However, ORF024 of the recently sequenced CuniNPV genome shows homology (25% identity over 289 amino acids) with the Lepidopteran AcMNPV VP39 protein and is putatively considered to encode the major nucleocapsid protein VP39.

Comparison of the *vp39* sequences showed that the partial NeabNPV VP39 protein had 5 cysteine and 2 proline residues that appeared to be conserved among the baculoviruses (Fig. 4.7). Cysteine residues often form disulphide bonds that are critical for proper protein functioning. Proline residues are known to break alpha helices and beta sheets and are commonly associated with beta turns in the secondary structure of proteins (Chou and Fasman, 1977). The highly conserved nature of these particular proline and cysteine residues suggests that they may play a crucial role in determining the proper conformation of the VP39 protein.

The secondary structure of the VP39 proteins was predicted using the nnPredict computer program (Kneller and Landridge, 1990) (Fig. 4.6). The *vp39* sequences from each baculovirus showed 9 regions with potential secondary structure (A to I), including 3 regions that were primarily alpha helical (D, E, H) and 2 regions that appeared to be exclusively beta sheets (A, E) (Fig. 4.5). Five regions of major secondary structure appeared to vary between baculovirus species demonstrating both alpha helices and beta

sheets. For example, region G was primarily alpha helical in all the baculoviruses except AcMNPV, BmMNPV, and CuniNPV. Overall the predicted secondary structure of the partial NeabNPV VP39 protein appeared to be most similar to that of the NPVs, although some regions of the protein appear to be more closely related to the GVs than the NPVs.

```
                 10        20        30        40        50        60        70        80
        ....|....|....|..*.|....|....|....|...*|....|....|...*|....|....|....|....|....|
Ac      MALVPVGMAPRQMRVNRCIFASIVSFDA--CITYKSP-CSPDAYHD--DGWFICNNHLIKRFKMSKMVLPIFDEDD--NQ
Bm      MALMPVGMAPRKMRVNRCIFASIVSFDA--CITYKSP-CSPDAYHD--DGWFICNSHLIKRFKMSKMVLPIFDEDD--NQ
Op      MALVSPGVSSRRS-TNHCIFGAIEPFDS--CVTYRSP-CSSDASVD--DGWFICDYHLKLRFKMSKMVLPIYDEDD--NQ
Eppo    MALVAPGVSSRRL-TNHCIFGGADSFNV--CVTYRSP-CSSDASVD--DGFFICNYHLKMYFKMSKMVLPIFDEDD--NQ
Ld      MALVSGALSTNRL-RNYCVFGAVQPFDN--CRAYGSP-CSPDSTNN--DGWFICDYHSSIRFKIEKMVLPIPDAEG--NI
Hz      MALVTVPTATTRL-RNFCVFSSVKPLDF--CDQYSSP-CSSDATVD--DGWFVCEYHASRFFKMEKLALAIPDGTG--NN
Hear    MALVTVPTATTRL-RNFCVFSSVKPLDF--CDQYSSP-CSSDATVD--DGWFVCEYHASRFFKMEKLALAIPDGTG--NN
Maco    MALTPYGSNQPQ--SNNCIFGAIRPFDT--CRTYSSP-CSNDASQE--DGWFICEYHLSIRFRMEKMVLPIPDAEG--TI
Se      MALTPVSSTATRR-GNYCIFGAVQPFDV--CRTYTNN-CSPDASFD--DGWYICEYHASIRFKMEKMAMPIPDGEG--NT
Splt    MALVSGGNANSRM-KNYCIFQGVRPIEFNQCSNYRSP-CSDDASQN--DGWFMCQYHLSRFFKIEKTSIAIPDGTG--QK
CpGV    ---MDVVTYEPCELNNYCVFQGVMV-DIMRCNDYGTQ-CSSDAYNSRSDGTFICNYHLGKYFRIRKSRFEIPSGKDN-RS
PxGV    -----MMSLRQNRVYNNCIFQAVSYSNSSLCADPVLH-CSKDASND--DGTFICNHHLSMYFPLEKMTLEIPSGTG--TS
XcGV    MNYDQSIACLNNELKNLCIFQGVQPPEFMNCGIYTPP-CSDDCVNK--DGTFVCSYHLARYFKLKKEVFEIPSGVNN-TS
Neab    ---------------NECIFRYARENDKLNQTIDPGYNCSNYLS-----GLYICQKHLEDYFRLSLRKINQLEGTDQGRV
Cuni    -----MSNNQLSRRTG-CVFAMLGTETDNPDTKYTNA-CNDKSLHT--EKFLICPEHLSNVFKLRVCHVNPFANKN-ALS

                 90       100       110       120       130       140       150      160
        ....|....|....|....|..+.|....|....|....|....|....|....|....|....|....|....|....|
Ac      FKMTIARHLVGNKERGIKRILIPSATNYQDVFNLNSMMQAEQLIFHLIYNNEN-------------------------AV
Bm      FKMTIARHLVGNKERGIKRILIPSATNYQEVFNLNSMMQAEQLIFHLIYNNEA-------------------------AV
Op      YKRTIARHLVGHKERGVKRILVPTRANYMTVFNLPGMMLAEQLIFHLIYDNRL-------------------------EV
Eppo    YKRTICRHLVSHTDTDDKRILVPTYAKYQTVFNLNSMMLAEQLIIHMIYNNTV-------------------------EI
Ld      YNRTVGKSLVNHKTLGAARVLIPTRDNYKTVLNLNSMSLAEQLVTHMIYDNVE-------------------------EQ
Hz      YYRTVGKSLVDDKAEGIERILIPSQNNYETVLNLSLLGPAERLVFYMIYDNKE-------------------------KQ
Hear    YYRTVGKSLVDDKAEGIERILIPSQNNYETVLNLSLLGPAERLVFYMIYDNKE-------------------------KQ
Maco    YNRSVGKSLISG-TESN-RVLIPTKTNYEDVLKLPSMSLPEQLIFHMIYDEPD-------------------------KQ
Se      YFRTVGKSLVSDKAEGNERILIPTADNYETGLNINAMSLPEQLVFHMIYNNKA-------------------------KQ
Splt    LYRIVGKSLVSHNARANDRILIPTQENYQAVMNVSMLPPAERLVLHLIYNNRT-------------------------AA
CpGV    FKMLIGQSLIPQSATDR--VLIPP--THESHFNTTNRSAMEKFVIYTIYEDKE-------------------------GL
PxGV    FKLLIGKSLVQQDATRN--IIIPSKANYIDYLRVNNMSPAEKFIMYSIYGESATEPT---------------------GLI
XcGV    FKYLVGVSLIQQNVPTANRITIPAKDNYYSYLNVANMSSMEKYVFYSIYDEPDTVTTRQTDPNKAPTGTEISTGVYEGPI
Neab    FERLSVLSLPHNYRNR---VIVPTRRNYEDIFKVAYLPISYQLVFHLLYQNQS-------------------------AI
Cuni    NAVPVFHVLQPTVSVSNTGILLPLTHEAFRNVLGEYINFSTDIIVQLLYGVVT--------------------------A

                170       180       190       200       210       220       230       240
        ...*|....|....|....|....|....|....|.*..|....|....|....|....|....|....|....|....|
Ac      NTICDNLKYT-EGFTSNTQRVIHSVYATTKSILDTTNPNTFCSRVSRD-ELRFFDVTNAR---------------ALRGG
Bm      NVICDNLKYT-EGFTSGKQRVIHSVYATTRSILDTTNPNTFCSRVSRD-ELRFFDVTNAR---------------TRRGG
Op      NRICASLKNN-ENFIDNTYSVVESVYSATRNILSLTDPQAYCSRVAND-DVRFFDANVVDN--------------NYQAG
Eppo    QRICNMLKTN-DNFTNDAYSIVERVYSSTRNIIALTDPNAYCTRVAND-DTRHFDVNRQNG--------------NIVPG
Ld      EAVCKALQHN-ENFQTETYRLAEDMFNRTSAILAMTNPRRYCSQVNSN-YARIWTT-------------------DDVN-
Hz      NEICQQLRMY-ERFRP---EVVEELYNSTLRVLALTNPDAYCAQTNTN-ESRSFGL-------------------SVED-
Hear    NEICQQLRMY-ERFRP---EVVEELYNSTLRVLALTNPDAYCAQTNTN-ESRSFGL-------------------SVED-
Maco    NEICKLLQYN-ENFHSDLYKVVERVYNKTAEVLAKTNPERFCARVNIN-NFRSYGT-------------------ADETD
Se      ERVCQMLQFN-EHFQTDIYKLVENVYNNTMSVLALTDPTRYCSRVSQN-STRIYGV-------------------NDEND
Splt    AEICNQLRQQ-ENFRS---DVVENVTSMAYRIIQITDPEAFCSVVASN-DIRSFGN-------------------EADL-
CpGV    DKLCKQLVDQ-EFFQQP---TWVKLQFNINTILGLVSPSMLCKVVANNGSTRTF--------------------------
PxGV    TQLCESLRSQ-DFYTDD---MLSDLYSVVAEIMSKINPAIYCRPILNN-SSRSFGKTSLTDTELQKIVTSSTVDANDYDT
XcGV    RALCSSLASQ-EFYTED---VLSDLNSIVSSLMGAIKPSIICRPTREE-TVRTF-----------------SSNDNRQPDE
Neab    DKICQDVKISGNKFSINGGAIDALISDVKKRLAGIITTDIKCSVEVTS-VTRNWES--------------------VPNI
Cuni    EDACSSYPRYIGQIRYVS-EILRDLSDSYKSCGSQNEFHASNSERTWT-VSKDFPT---------------------QD
```



250 260 270 280 290 300 310 320
Ac AGDQLFNNYSGFLQNLIRRAVAPEYLQIDT----EELRFRNCATCIIDETGLVASVPD---GPELYNPIRSSDIMRSQPN
Bm VGDQLFNNYSGFLQNLIR-AVAPEYLQIDT----EELRFRNSATCIIDETGLVASVPD---GPELYNPIRSSDIMKSQPN
Op NGDTVFNNMPGFLRNLIRRAVAPETLQIDS----EDLRLRNCNTCVINNTGLVATVTN----TELYNPVRSSDIIKTRPN
Eppo NGDTIYAAMPGFLKNLIRRAVAPEVLQIDS----QDLRLRNCNTCTINANGLVATVYNIENVTELYNPTRSTDIVQSRPN
Ld VAGNVFESMPPFLKNLINVAVAPEQIMIDE----ETLVIRNCPTCNIDDSGLVAN-------VQLYNPVVPRYRSTFNEN
Hz --DLAFNVLPTFIQNLIRKCVAPESLTIGT----EDLQLRNCNTCRITSEGLLAS-------VRLYNSVQPKYLYGVNEN
Hear --DLAFNVLPTFIQNLIRKCVAPESLTIGT----EDLQLRNCNTCRITSEGLLAS-------VRLYNSVQPKYLYGVNEN
Maco IANVTYNAMPGFLKNLINKSVAPCTMDFPG----GTINLRNAATCAINQQGLVAT-------ADIYNPVVPKYMSVTNEN
Se IAEQTISRMPGFLQNLINKCVAPQTMTIEK----QTLEFRNCTTCRIDSTGLVAD-------VKLYNPVEPKYRSGYNEN
Splt --TRTYNSLPPFIKNLINTLVAPEVLTLGN----NTFPLKACATVSITTNGLQA--------VQLYNEVEPKYLYNRNDN
CpGV SQQTNFEGYPPFLRNLITQLVRPKVLTISG----TDINIMEFDTCTFTS-EGLTV-------PNLHNPNQPVRIDNPILQ
PxGV TQDRAYRAMPPFIKNLILRLVRPVTLIVSSNRHNDKFILNNADTCDLIEDKGLTA-------ANLYNFEKPRHR--LMHE
XcGV NQ-RAFEAMPNFIKNLIMRLVRPVTLQIGS----TELILDQSATCRIEPQQGLIP-------VNLYNPEKPKFTDNPYRE
Neab INIMDDPQEPLFIKNLITNLICPREVVIDN----KTFKFETAANVIFVEEGLIVD--------NLYNPEAELYPLYNNRT
Cuni EVRHFIHNLPPFFRNLVENGLAPGEVRINA----VVHKMPNLINTAIYSDAILLF--------NSIDPRGAEEKYAPKTG
330 340 350 360 370 380 390 400
Ac -RLQIRNVLKFEGDTR-ELDRTLSGYEEYPTYVPLFLGYQIINSENNFLR-NDFIP-------------RANPN------
Bm -RLQIRNVLKFEGDTR-ELDRTLRGYEEYPTYVPLFLGYQIINSENNFLR-NDFIS-------------RANPN------
Op -RLQIRNVLKFEGDTR-ALERTLGRYEEYPMYVPLFLGYQLVNLQNDILRANNFLPAPFG-------VPQAVNN------
Eppo -RLQLRSVLKFEGNTR-ALERTLNRYEEYPMYVPLFLGYQLVNTQNNILRVNNFIPAAFNPVIPPAPVAEAGPNNAQGGG
Ld -VLHVENVLKFKGNAN-ALQKSLSRYEPYPIVVPLMLGTQTLNTSSAYKQFTVPTRDDFAALNQRTGAAAAAPP------
Hz -RLQIRNVLQFQGNAN-ALQQKLSRYELYQINIPLFLGKRIISTGR----------------------------------
Hear -RLQIRNVLQFQGNAN-ALQQKLSRYELYQINIPLFLGKQIISTGR----------------------------------
Maco -ILQIDYVPRFQGNAV-ALQTNLSRYEHYPLVVPLMVGNEVLITRDTYKKPPYPNPTLTPSI-----SVSTAPV------
Se -YLLIDSVFKFRGNSK-ALQTALNRYEQYPVVVPLILGSEIVVTANSLK--PSIIPRMLPNE-----EETLPPL------
Splt -VLEIRNVIQFIGNSD-ALNKKLARYECYPLVVQLFLGSQTINQGIRSQAPAEI--------------------------
CpGV PKFSLRSVVEFDGRAT-LEQRALDTYDDVLLTRPLINGVQTVV-------------------------------------
PxGV KKFDIQHLTQFKGRAA-EQQRHLSTYPVYKIARPLMLGIEIIVTPQF---------------------------------
XcGV PKFQVITYIDFGGRATREQQDALAGYDRYTISRPLLLGRESIAQAY----------------------------------
Neab --ISVNKVTTVSGLPL--SNIDVRGLRSVNREHYLVNGEQRLNIRENALFER----------------------------
Cuni MVLMVKHVVKNGNQRS-----TPIGEDITPTIVPTN-GAYTINLATI---------------------------------
410 420 430
Ac -ATLGGGAVAGPAPG-VAGEAGGGIAV----
Bm -ATLGGGALAGPAPGVVLGEASGSVAA----
Op -----LEAQAPAAPAPAAPAPAPAAPVV---
Eppo NAQGGGAAQGGGAAQGGGNAQGGGAAI----
Ld -----APAAAPAAPGPAAELEAGGAPVAARA
Hz -------------------------------
Hear -------------------------------
Maco -----VPAANA--------------------
Se -----RPTAVV--------------------
Splt -------------------------------
CpGV -------------------------------
PxGV -------------------------------
XcGV -------------------------------
Neab -------------------------------
Cuni -------------------------------

Figure 4.5. Partial alignment of the amino acid sequences of 15 baculovirus VP39 proteins. AcMNPV (Ac), BmMNPV (Bm), OpMNPV (Op), EppoNPV (Eppo), LdMNPV (Ld), HzSNPV (Hz), HearSNPV (Hear), MacoNPV (Maco), SeMNPV (Se), SpltMNPV (Splt), CpGV, PxGV, XcGV, NeabNPV (Neab), and CuniNPV (Cuni). Amino acid residues conserved in greater than 90% of the sequences are shaded in black. Amino acid residues with similar biochemical properties that are conserved in 100% of the sequences are shaded in grey. Putative N-linked glycosylation sites (N-X-S/T) are shown underlined and potential phosphorylation sites (S/T-X-R/K) are shown underlined and bolded. Conserved cysteine and proline residues are indicated by a star (*) and plus (+) sign, respectively, above the sequence.

	AcMNPV	BmMNPV	OpMNPV	EppoNPV	LdMNPV	HzSNPV	HearSNPV	MacoNPV	SeMNPV	SpltMNPV	NeabNPV	CuniNPV	CpGV	PxGV	XcGV
AcMNPV	100.0	97.1	69.7	62.5	46.7	46.7	47.1	45.1	45.9	40.3	24.5	18.9	29.8	31.0	30.2
BmMNPV	---	100.0	69.7	62.5	45.9	45.9	46.3	44.3	45.9	39.9	24.1	18.5	29.8	30.6	30.6
OpMNPV	---	---	100.0	76.2	52.4	50.0	50.4	50.8	50.8	45.5	22.9	17.7	31.0	31.8	31.4
EppoNPV	---	---	---	100.0	47.9	44.7	45.1	46.7	50.0	44.7	22.9	16.9	33.4	33.4	29.8
LdMNPV	---	---	---	---	100.0	52.8	53.2	58.4	60.4	46.7	22.1	17.3	33.0	33.4	33.4
HzSNPV	---	---	---	---	---	100.0	99.5	50.8	54.4	54.0	20.1	16.5	33.4	36.6	36.2
HearSNPV	---	---	---	---	---	---	100.0	50.8	54.4	54.4	20.5	16.5	33.8	36.6	36.2
MacoNPV	---	---	---	---	---	---	---	100.0	59.2	47.1	22.5	14.1	29.0	36.6	33.0
SeMNPV	---	---	---	---	---	---	---	---	100.0	46.7	22.1	16.9	33.0	33.8	31.8
SpltMNPV	---	---	---	---	---	---	---	---	---	100.0	27.0	17.3	32.2	35.0	35.0
NeabNPV	---	---	---	---	---	---	---	---	---	---	100.0	16.5	23.3	20.9	22.1
CuniNPV	---	---	---	---	---	---	---	---	---	---	---	100.0	16.1	15.7	18.1
CpGV	---	---	---	---	---	---	---	---	---	---	---	---	100.0	39.5	41.1
PxGV	---	---	---	---	---	---	---	---	---	---	---	---	---	100.0	51.2
XcGV	---	---	---	---	---	---	---	---	---	---	---	---	---	---	100.0

Table 4.5. Amino acid identities (%) of fifteen baculovirus VP39 proteins.

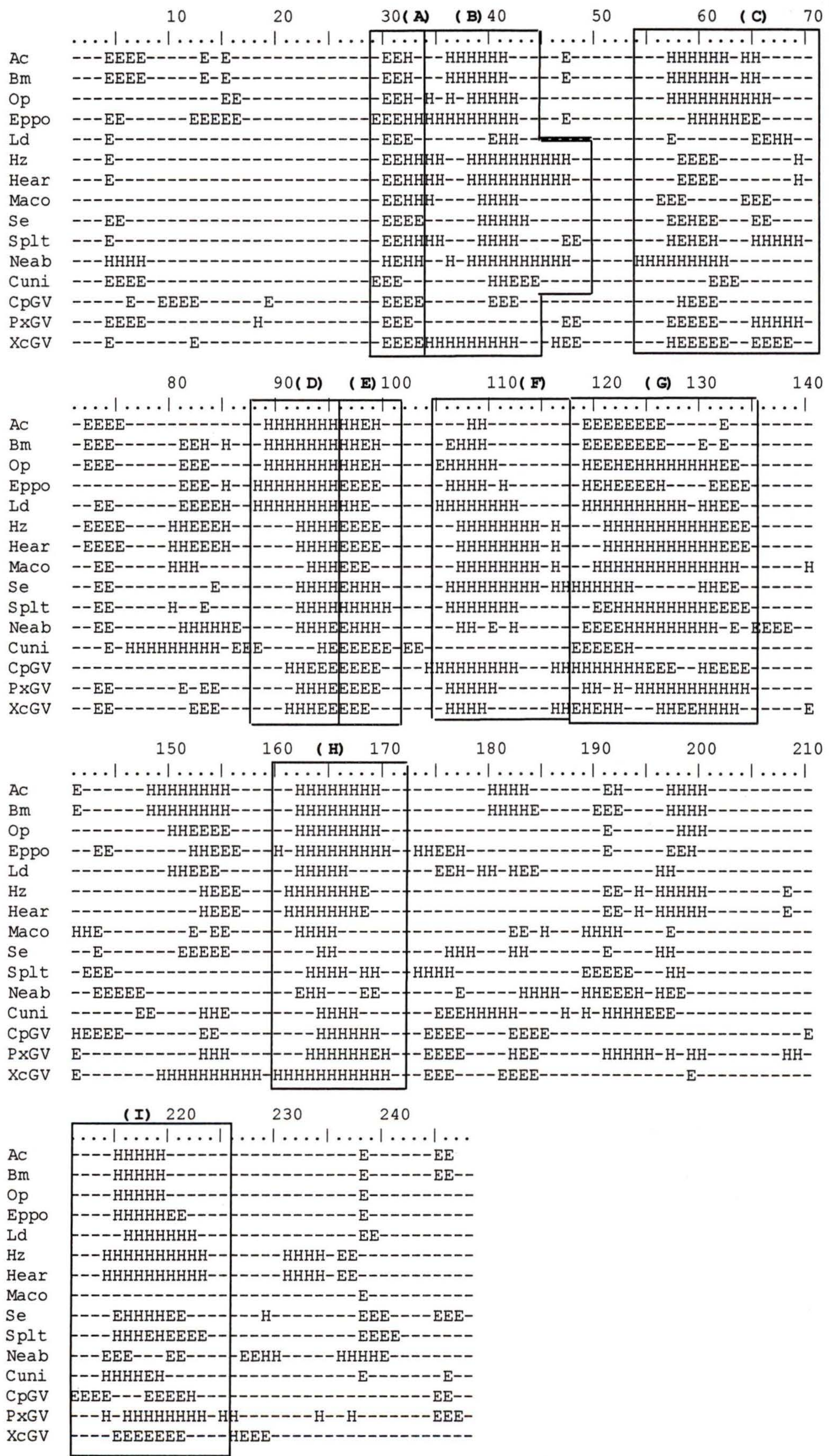

(A)
(B)
(C)
(D)
(E)
(F)
(G)
(H)
(I)
Ac
Bm
Op
Eppo
Ld
Hz
Hear
Maco
Se
Splt
Neab
Cuni
CpGV
PxGV
XcGV

Figure 4.6. Predicted secondary structure of VP39 proteins. Secondary structure was predicted from the inferred amino acid sequences shown in Fig 4.12 using the nnPredict computer program (Kneller and Landridge, 1990). Gaps in the multiple sequence alignment have been omitted. H – alpha helix element, E – beta sheet element, (-) – beta turn element. Nine areas of extensive secondary structure are labelled A to I: A and C – beta sheet structure, B, D, and E – alpha helical structure.

Alpha helices and beta strands are the major secondary structures found in proteins. These conformations form readily in proteins and are highly stable due to their optimal use of hydrogen bonding (Lehninger, 1993). Alpha helices and beta sheets create linear regions within the protein molecule whereas beta turn elements cause the molecule to fold back on itself, creating regions that are globular in nature (Rohrmann, 1986). Several regions of VP39, including those near amino acids 25, 50, 140, 165, 170 and 205 appear to have conserved beta turn elements. A multiple sequence alignment of the baculovirus VP39 proteins, which had all gaps in the alignment removed, (Fig. 4.7) showed that beta turn elements occurred near regions of the protein where proline residues were conserved. This suggests that these particular beta turn elements may be critical for the proper folding of the protein and may be necessary for correct functioning. Although the prediction of secondary structure (Chou and Fasman, 1978) is a common component of DNA sequence analysis programs, results should be taken with caution as the relationship between predicted secondary structures and the higher ordered structure of proteins can be elusive (Rohrmann, 1986).

Hydrophilicity profiles were constructed for each partial VP39 protein using the method of Hopp and Woods (1981) and the Bioedit (v5.0.9) computer program. Hydrophilicity profiles help to predict several important features of proteins, including which regions are present on a proteins surface. Regions of a protein with hydrophilic side-chains are organized so that they are exposed to the surrounding aqueous solvents while regions with hydrophobic side-chains are organized so as to minimize this contact. As a result, exterior hydrophilic portions of a protein can be distinguished from the interior hydrophobic regions by averaging the hydrophilicity or hydrophobicity of the

amino acids along an entire protein molecule (Kyte and Doolittle, 1982). Although the VP39 protein sequences vary by as much as 86% (Table 4.5), the basic pattern of hydrophilicity is conserved (Fig. 4.8). Comparison of the hydrophilicity plots for each *vp39* showed the presence of several highly conserved hydrophilic and hydrophobic regions, and revealed the presence of strong hydrophilic peaks near amino acids 50, 150, 180, and 210. In contrast, regions near amino acids 25, 100, 160, and 205 showed a very low level of hydrophilicity. These latter regions correspond with conserved areas of *vp39* that contain beta turn elements predicted in their secondary structure (Fig. 4.6) Overall, the hydrophilic and hydrophobic regions of NeabNPV *vp39* appeared to be more conserved with the NPVs than the GVs, with the hydrophilic profile of NeabNPV being most similar to SpltMNPV.

The difference in the hydrophilic nature of the VP39 proteins may be due to the unique cytopathology demonstrated by NPVs and the GVs. GV infection causes the disintegration of the cell nucleus and release of viral nucleocapsids into the cytoplasm where the nucleocapsids become enveloped. In contrast, NPV infection leaves the cell nucleus intact. NPV nucleocapsids are assembled within the nucleus and, during maturation, acquire their envelope from the inner nuclear membrane (Federici, 1997). The differences in the hydrophilic nature of both NPV and GV VP39 proteins may cause a slight variation in protein's conformation that is required for nucleocapsid maturation (envelopment) in their respective environments (nucleus vs. cytoplasm).

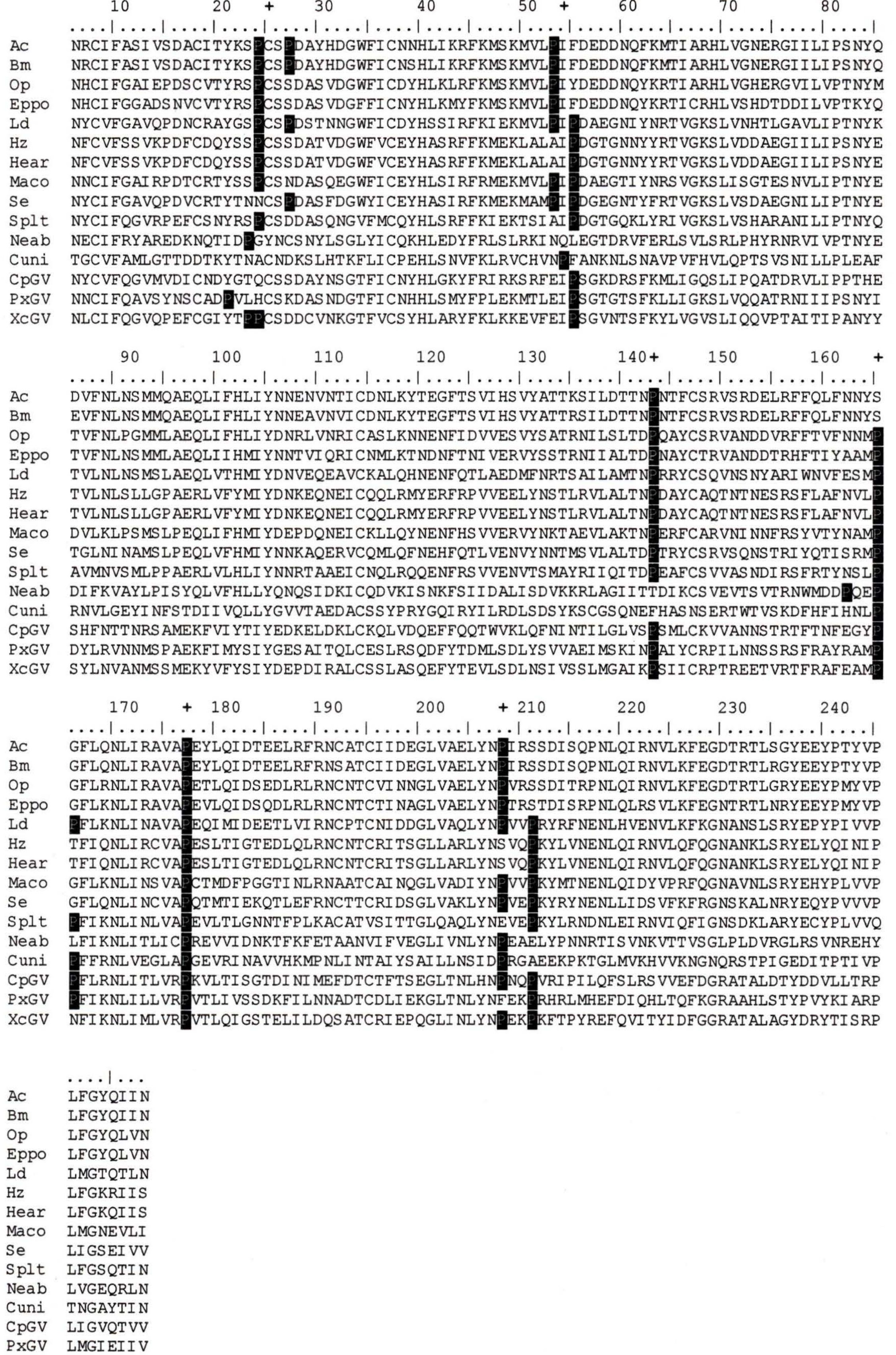

10 20 + 30 40 50 + 60 70 80
Ac NRCIFASIVSDACITYKSPCSPDAYHDGWFICNNHLIKRFKMSKMVLPIFDEDDNQFKMTIARHLVGNERGIILIPSNYQ
Bm NRCIFASIVSDACITYKSPCSPDAYHDGWFICNSHLIKRFKMSKMVLPIFDEDDNQFKMTIARHLVGNERGIILIPSNYQ
Op NHCIFGAIEPDSCVTYRSPCSSDASVDGWFICDYHLKLRFKMSKMVLPIYDEDDNQYKRTIARHLVGHERGVILVPTNYM
Eppo NHCIFGGADSNVCVTYRSPCSSDASVDGFFICNYHLKMYFKMSKMVLPIFDEDDNQYKRTICRHLVSHDTDDILVPTKYQ
Ld NYCVFGAVQPDNCRAYGSPCSPDSTNNGWFICDYHSSIRFKIEKMVLPIPDAEGNIYNRTVGKSLVNHTLGAVLIPTNYK
Hz NFCVFSSVKPDFCDQYSSPCSSDATVDGWFVCEYHASRFFKMEKLALAIPDGTGNNYYRTVGKSLVDDAEGIILIPSNYE
Hear NFCVFSSVKPDFCDQYSSPCSSDATVDGWFVCEYHASRFFKMEKLALAIPDGTGNNYYRTVGKSLVDDAEGIILIPSNYE
Maco NNCIFGAIRPDTCRTYSSPCSNDASQEGWFICEYHLSIRFRMEKMVLPIPDAEGTIYNRSVGKSLISGTESNVLIPTNYE
Se NYCIFGAVQPDVCRTYTNNCSPDASFDGWYICEYHASIRFKMEKMAMPIPDGEGNTYFRTVGKSLVSDAEGNILIPTNYE
Splt NYCIFQGVRPEFCSNYRSPCSDDASQNGVFMCQYHLSRFFKIEKTSIAIPDGTGQKLYRIVGKSLVSHARANILIPTNYQ
Neab NECIFRYAREDKNQTIDPGYNCSNYLSGLYICQKHLEDYFRLSLRKINQLEGTDRVFERLSVLSRLPHYRNRVIVPTNYE
Cuni TGCVFAMLGTTDDTKYTNACNDKSLHTKFLICPEHLSNVFKLRVCHVNPFANKNLSNAVPVFHVLQPTSVSNILLPLEAF
CpGV NYCVFQGVMVDICNDYGTQCSSDAYNSGTFICNYHLGKYFRIRKSRFEIPSGKDRSFKMLIGQSLIPQATDRVLIPPTHE
PxGV NNCIFQAVSYNSCADPVLHCSKDASNDGTFICNHHLSMYFPLEKMTLEIPSGTGTSFKLLIGKSLVQQATRNIIIPSNYI
XcGV NLCIFQGVQPEFCGIYTPPCSDDCVNKGTFVCSYHLARYFKLKKEVFEIPSGVNTSFKYLVGVSLIQQVPTAITIPANYY
90 100 110 120 130 140+ 150 160 +
Ac DVFNLNSMMQAEQLIFHLIYNNENVNTICDNLKYTEGFTSVIHSVYATTKSILDTTNPNTFCSRVSRDELRFFQLFNNYS
Bm EVFNLNSMMQAEQLIFHLIYNNEAVNVICDNLKYTEGFTSVIHSVYATTRSILDTTNPNTFCSRVSRDELRFFQLFNNYS
Op TVFNLPGMMLAEQLIFHLIYDNRLVNRICASLKNNENFIDVVESVYSATRNILSLTDPQAYCSRVANDDVRFFTVFNNMP
Eppo TVFNLNSMMLAEQLIIHMIYNNTVIQRICNMLKTNDNFTNIVERVYSSTRNIIALTDPNAYCTRVANDDTRHFTIYAAMP
Ld TVLNLNSMSLAEQLVTHMIYDNVEQEAVCKALQHNENFQTLAEDMFNRTSAILAMTNPRRYCSQVNSNYARIWNVFESMP
Hz TVLNLSLLGPAERLVFYMIYDNKEQNEICQQLRMYERFRPVVEELYNSTLRVLALTNPDAYCAQTNTNESRSFLAFNVLP
Hear TVLNLSLLGPAERLVFYMIYDNKEQNEICQQLRMYERFRPVVEELYNSTLRVLALTNPDAYCAQTNTNESRSFLAFNVLP
Maco DVLKLPSMSLPEQLIFHMIYDEPDQNEICKLLQYNENFHSVVERVYNKTAEVLAKTNPERFCARVNINNFRSYVTYNAMP
Se TGLNINAMSLPEQLVFHMIYNNKAQERVCQMLQFNEHFQTLVENVYNNTMSVLALTDPTRYCSRVSQNSTRIYQTISRMP
Splt AVMNVSMLPPAERLVLHLIYNNRTAAEICNQLRQQENFRSVVENVTSMAYRIIQITDPEAFCSVVASNDIRSFRTYNSLP
Neab DIFKVAYLPISYQLVFHLLYQNQSIDKICQDVKISNKFSIIDALISDVKKRLAGIITTDIKCSVEVTSVTRNWMDDPQEP
Cuni RNVLGEYINFSTDIIVQLLYGVVTAEDACSSYPRYGQIRYILRDLSDSYKSCGSQNEFHASNSERTWTVSKDFHFIHNLP
CpGV SHFNTTNRSAMEKFVIYTIYEDKELDKLCKQLVDQEFFQQTWVKLQFNINTILGLVSPSMLCKVVANNSTRTFTNFEGYP
PxGV DYLRVNNMSPAEKFIMYSIYGESAITQLCESLRSQDFYTDMLSDLYSVVAEIMSKINPAIYCRPILNNSSRSFRAYRAMP
XcGV SYLNVANMSSMEKYVFYSIYDEPDIRALCSSLASQEFYTEVLSDLNSIVSSLMGAIKPSIICRPTREETVRTFRAFEAMP
170 + 180 190 200 + 210 220 230 240
Ac GFLQNLIRAVAPEYLQIDTEELRFRNCATCIIDEGLVAELYNPIRSSDISQPNLQIRNVLKFEGDTRTLSGYEEYPTYVP
Bm GFLQNLIRAVAPEYLQIDTEELRFRNSATCIIDEGLVAELYNPIRSSDISQPNLQIRNVLKFEGDTRTLRGYEEYPTYVP
Op GFLRNLIRAVAPETLQIDSEDLRLRNCNTCVINNGLVAELYNPVRSSDITRPNLQIRNVLKFEGDTRTLGRYEEYPMYVP
Eppo GFLKNLIRAVAPEVLQIDSQDLRLRNCNTCTINAGLVAELYNPTRSTDISRPNLQLRSVLKFEGNTRTLNRYEEYPMYVP
Ld PFLKNLINAVAPEQIMIDEETLVIRNCPTCNIDDGLVAQLYNPVVPRYRFNENLHVENVLKFKGNANSLSRYEPYPIVVP
Hz TFIQNLIRCVAPESLTIGTEDLQLRNCNTCRITSGLLARLYNSVQPKYLVNENLQIRNVLQFQGNANKLSRYELYQINIP
Hear TFIQNLIRCVAPESLTIGTEDLQLRNCNTCRITSGLLARLYNSVQPKYLVNENLQIRNVLQFQGNANKLSRYELYQINIP
Maco GFLKNLINSVAPCTMDFPGGTINLRNAATCAINQGLVADIYNPVVPKYMTNENLQIDYVPRFQGNAVNLSRYEHYPLVVP
Se GFLQNLINCVAPQTMTIEKQTLEFRNCTTCRIDSGLVAKLYNPVEPKYRYNENLLIDSVFKFRGNSKALNRYEQYPVVVP
Splt PFIKNLINLVAPEVLTLGNNTFPLKACATVSITTGLQAQLYNEVEPKYLRNDNLEIRNVIQFIGNSDKLARYECYPLVVQ
Neab LFIKNLITLICPREVVIDNKTFKFETAANVIFVEGLIVNLYNPEAELYPNNRTISVNKVTTVSGLPLDVRGLRSVNREHY
Cuni PFFRNLVEGLAPGEVRINAVVHKMPNLINTAIYSAILLNSIDPRGAEEKPKTGLMVKHVVKNGNQRSTPIGEDITPTIVP
CpGV PFLRNLITLVRPKVLTISGTDINIMEFDTCTFTSEGLTNLHNPNQPVRIPILQFSLRSVVEFDGRATALDTYDDVLLTRP
PxGV PFIKNLILLVRPVTLIVSSDKFILNNADTCDLIEKGLTNLYNFEKPRHRLMHEFDIQHLTQFKGRAAHLSTYPVYKIARP
XcGV NFIKNLIMLVRPVTLQIGSTELILDQSATCRIEPQGLINLYNPEKPKFTPYREFQVITYIDFGGRATALAGYDRYTISRP
Ac LFGYQIIN
Bm LFGYQIIN
Op LFGYQLVN
Eppo LFGYQLVN
Ld LMGTQTLN
Hz LFGKRIIS
Hear LFGKQIIS
Maco LMGNEVLI
Se LIGSEIVV
Splt LFGSQTIN
Neab LVGEQRLN
Cuni TNGAYTIN
CpGV LIGVQTVV
PxGV LMGIEIIV
XcGV LLGRESIA

Figure 4.7. Partial amino acid sequence alignment of baculovirus VP39 proteins. All gaps from the multiple sequence alignment have been removed. Proline residues that may be important for beta turn secondary structure in the VP39 protein are shown shaded in black. Predicted beta turns (shown in Fig.4.10) are indicated by a plus (+) sign.

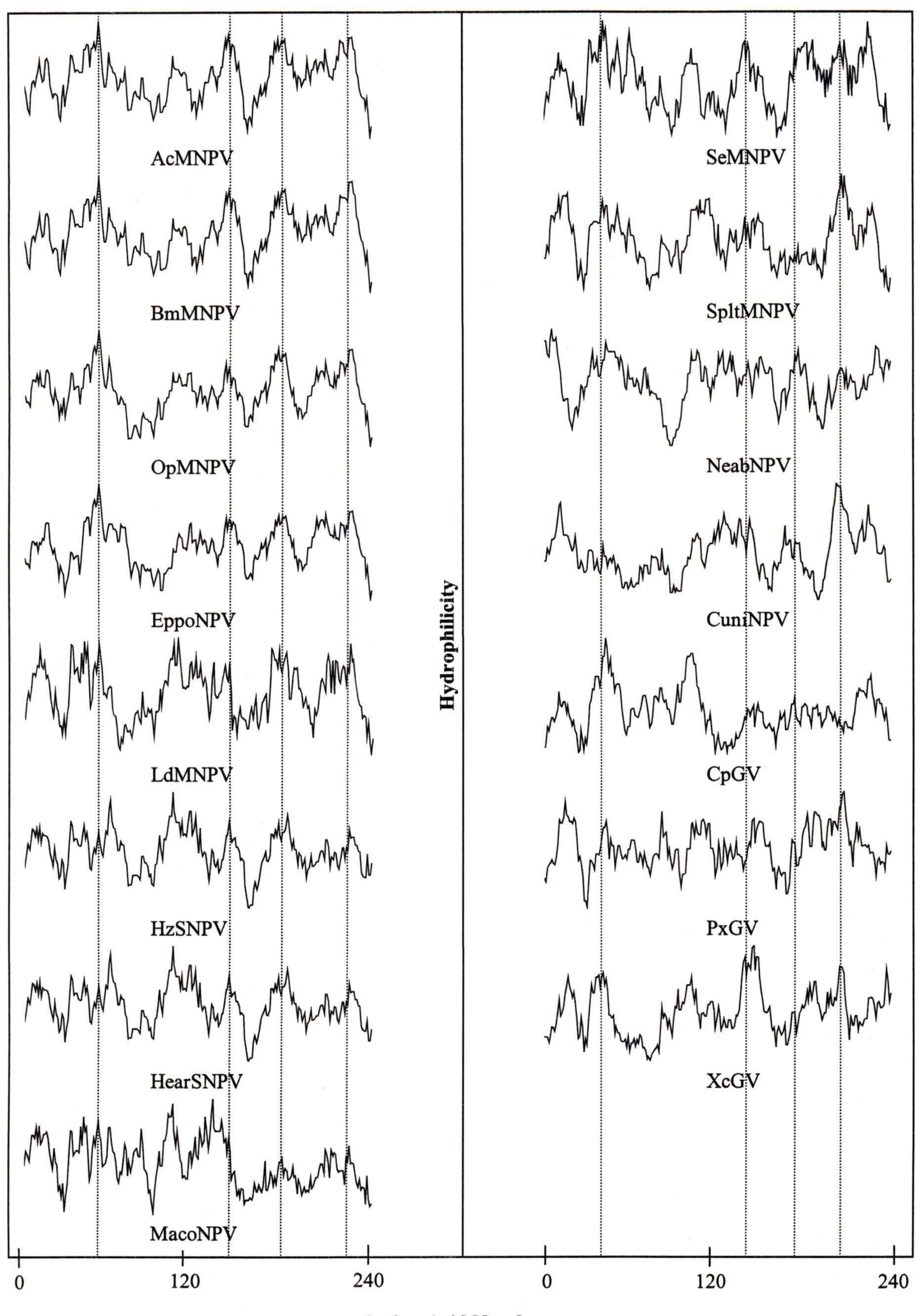

AcMNPV
BmMNPV
OpMNPV
EppoNPV
LdMNPV
HzSNPV
HearSNPV
MacoNPV
SeMNPV
SpltMNPV
NeabNPV
CuniNPV
CpGV
PxGV
XcGV
Hydrophilicity
0
120
240
0
120
240
Amino Acid Number

Figure 4.8. Predicted hydrophilicity of a 240 amino acid region from baculovirus VP39 proteins. Hydrophilicity was predicted using the method of Hopp and Woods (1981) and Bioedit (v5.0.9). Sequences used are the same as those in Fig.4.11, where gaps have been omitted. Vertical lines at amino acids 50, 150, 180, and 210 indicate conserved hydrophilic regions. Amino acid number is given on the abscissa and the level of hydrophilicity is indicated on the ordinate.

4.4. References

Afonso, C. L., Tulman, E. R., Lu, Z., Balinsky, C. A., Moser, B. A., Becnel, J. J., Rock, D. L. & Kutish, G. F. (2001). Genome sequence of a baculovirus pathogenic for Culex nigripalpus. *J Virol.* **75**, 11157-65.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol.* **215**, 403-10.

Braunagel, S. C. & Summers, M. D. (1994). Autographa californica nuclear polyhedrosis virus, PDV, and ECV viral envelopes and nucleocapsids: structural proteins, antigens, lipid and fatty acid profiles. *Virol.* **202**, 315-28.

Chou, P. Y. & Fasman, G. D. (1977). Beta-turns in proteins. *J Mol Biol.* **115**, 135-75.

Chou, P. Y. & Fasman, G. D. (1978). Prediction of the secondary structure of proteins from their amino acid sequence. *Adv Enzymol Relat Areas Mol Biol.* **47**, 45-148.

Federici, B. A. (1997). Baculovirus Pathogenesis. In *The Baculoviruses*, pp. 33-59. Edited by L. K. Miller. New York: Plenum Press.

Funk, J., Braunagel, S. C. & Rohrmann, G. F. (1997). Baculovirus Pathogenesis. In *The Baculoviruses*, pp. 7-32. Edited by L. K. Miller. New York: Plenum Press.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl. Acids. Symp Ser.* **41**, 95-98.

Hames, B. D. & Rickwood, D. (1981). Gel Electrophoresis of Proteins: A Practical Approach. Washington, DC: IRL Press.

Hopp, T. P. & Woods, K. R. (1981). Prediction of protein antigenic determinants from amino acid sequences. *Proc Natl Acad Sci USA.* **78**, 3824-8.

Kneller, D. G., Cohen, F. E. & Langridge, R. (1990). Improvements in protein secondary structure prediction by an enhanced neural network. In *J Mol Biol.* pp. 71-82.

Kyte, J. & Doolittle, R. F. (1982). A simple method for displaying the hydropathic character of a protein. *J Mol Biol.* **157**, 105-32.

Lehninger, L. L., Nelson, N. L. & Cox, M. M. (1993). Principles of Biochemistry: 2nd Ed., pp. 161-197. New York: Worth Publishers.

O'Reilly, D. R., Miller, L. K., and Lucklow, V. A. (1992). Baculovirus Expression Vectors: A Laboratory Manual. W. H. Freeman, New York.

Rohrmann, G. F. (1986). Polyhedrin structure. *J Gen Virol.* **67**, 1499-513.

Volkman, L. E. & Summers, M. D. (1977). Autographa californica nuclear polyhedrosis virus: comparative infectivity of the occluded, alkali-liberated, and nonoccluded forms. *J Invertebr Pathol.* **30**, 102-3.

Chapter 5. Phylogenetic Analysis of Eight Putative NeabNPV Genes

5.1. Introduction

The *Baculoviridae* are a large and diverse family of rod-shaped, enveloped viruses that are pathogenic for invertebrates, particularly insects. The family is divided into two genera: the *Nucleopolyhedroviruses* (NPVs) and the *Granuloviruses* (GVs), which together have been isolated from over 600 insect species (Martignoni and Iwai, 1986). Phylogenetically, the NPVs are the most common and widely distributed baculovirus having been discovered in more than 400 insect species spanning seven different insect Orders (Volkmann *et al.*, 1995). NPVs are found primarily in Lepidopteran hosts but have also been isolated from several other insect Orders including Diptera and Hymenoptera.

Outbreaks of the Hymenopterous insect, the balsam fir sawfly (*N. abietis*), are currently causing severe damage to economically valuable forests in eastern Canada. To control this forest pest it has been proposed that an insect-specific baculovirus be utilized, as environmental regulations prohibit the use of conventional chemical pesticides. A nucleopolyhedrovirus (NeabNPV) has recently been isolated from *N. abietis* and research involving the nucleotide sequence analysis of the NeabNPV genome has begun. This research has led to the discovery of several putative complete baculovirus genes including *polh*, *p74*, *odv-e56,* and *vlf-1* as well as many other partial baculovirus gene sequences including *p143, lef-8, lef-9,* and *vp-91.*

The polyhedrin gene (*polh*) encodes the major occlusion body protein of NPVs and GVs. The polyhedrin protein is approximately 245 amino acids in length and is highly conserved among the baculoviruses (Vlak and Rohrmann, 1985). The *p74* gene encodes a 74 kDa structural protein that is associated with the occlusion-derived virion

and is most likely a nucleocapsid envelope protein (Faulkner *et al.,* 1997). The protein is approximately 650 amino acids long and is essential for the primary infection of the midgut cells of host insect larvae (Faulkner *et al.,* 1997). The o*dv-e56* gene also encodes a nucleocapsid envelope protein found in occlusion-derived virions. The protein is approximately 355 amino acids long and has been shown to be involved in the process of envelopment of the virion in the host cell nucleus (Braunagel *et al.,* 1995). The v*lf*-1 gene encodes a protein that appears to be related to the lambda phage integrase/resolvase family (McLachlin and Miller, 1994), which is a family of genes capable of promoting homologous recombination between two target DNA sites. The protein is approximately 370 amino acids long and had been shown to preferentially stimulate the very late gene expression of several baculovirus genes and assist with the binding of a baculovirus-specific RNA polymerase II to the promoters of these late and very late genes (Yang and Miller, 1999).

Recently the genome sequences of 10 NPVs and 3 GVs pathogenic to Lepidoptera as well as 1 baculovirus pathogenic to Diptera have been described (Chen, 2002, Afonso *et al.,* 2001, Hayakawa, 2000). In contrast, little information has been reported on the DNA sequences from baculoviruses that infect Hymenopteran insects (Moser *et al.,* 2001, Zanotto *et al.,* 1993). An earlier phylogenetic study (Zanotto *et al.,* 1993) suggested that the Hymenopteran NPVs (NeseNPV) separated from the Lepidopteran NPVs before the divergence of the GVs from the NPVs, a hypothesis supported by Rohrmann (1992). In this chapter I describe the phylogenetic relationships between the Hymenopteran NPV, NeabNPV, and 14 previously described baculoviruses through a comparison of the NPV and GV genes *polh*, *p74*, *odv-e56*, *vlf-1,* and the partial gene sequences of *vp-91*, *lef-8*, *lef-9*, and *p143.*

5.2. Materials and Methods

5.2.1. Baculovirus sequences

The baculovirus genomes used for this study are listed in Table 5.1. The complete genome sequence of NeabNPV is unavailable at this time and is currently being sequenced.

Table 5.1. Baculovirus genomes

Baculovirus Genome	Classification	Size (bp)	No. of Genes	Accession no.	References
AcMNPV	NPV Group I	133,894	155	L22858	Ayres *et al.*
BmMNPV	NPV Group I	128,413	143	L33180	Gomi *et al.*
OpMNPV	NPV Group I	131,990	152	U75930	Ahrens *et al.*
EppoNPV	NPV Group I	118,584	136	AY043256	Hyink *et al.*
SeMNPV	NPV Group II	135,611	139	AF169823	Ijkel *et al.*
LdMNPV	NPV Group II	161,046	166	AF081810	Kuzio *et al.*
HearSNPV	NPV Group II	131,403	135	AF271059	Chen *et al.*
HzSNPV	NPV Group II	130,869	139	AF334030	Chen *et al.*
SpltMNPV	NPV Group II	139,342	141	AF325155	Pang *et al.*
MacoNPV	NPV Group II	155,060	169	AF325155	Li *et al.*
CuniNPV	Unclassified	108,252	109	AF403738	Alfonso *et al.*
NeabNPV	Unclassified	~96,000	Unknown	None	None
XcGV	GV	178,733	181	AF162221	Hayakawa *et al.*
PxGV	GV	100,999	120	AF270937	Hashimoto *et al.*

Table 5.2. Parsimoniously-informative characters from each baculovirus amino acid sequence alignment.

Data Set	Total Characters	Informative Characters	% Informative
Whole Genes			
odv-e56	402	284	70.7
p74	783	479	61.2
polh	882	131	14.9
	243	**149**	**61.3**
vlf-1	455	298	65.5
Partial Genes			
p143	1485	1039	70.0
lef-8	1005	561	55.8
lef-9	628	310	49.4
vp-91	964	704	73.0

** Indicate that the CuniNPV occlusion protein has been trimmed so that only those residues that align with baculovirus polyhedrin and granulin proteins were used for sequence alignment.

(A)

	Ac	Bm	Op	Ld	Eppo	Hear	Hz	Splt	Se	Maco	Neab	Cuni	XcGV	CpGV	PxGV
Ac	100	83	68	53	68	46	46	45	45	48	30(46)	22	33	40	39
Bm	---	100	63	50	64	46	46	46	43	47	29(48)	21	33	39	38
Op	---	---	100	52	80	46	46	44	46	47	31(49)	23	34	41	39
Ld	---	---	---	100	52	48	48	46	47	49	36(52)	23	37	41	42
Eppo	---	---	---	---	100	45	45	44	46	48	31(50)	22	33	42	39
Hear	---	---	---	---	---	100	100	43	48	48	33(51)	24	39	42	41
Hz	---	---	---	---	---	---	100	44	48	48	33(51)	24	39	42	41
Splt	---	---	---	---	---	---	---	100	43	44	30(44)	20	33	39	39
Se	---	---	---	---	---	---	---	---	100	56	30(46)	22	35	40	37
Maco	---	---	---	---	---	---	---	---	---	100	32(47)	19	35	40	37
Neab	---	---	---	---	---	---	---	---	---	---	100(100)	23 (40)	30(51)	34(53)	34(51)
Cuni	---	---	---	---	---	---	---	---	---	---	---	100	21	23	24
XcGV	---	---	---	---	---	---	---	---	---	---	---	---	100	56	49
CpGV	---	---	---	---	---	---	---	---	---	---	---	---	---	100	66
PxGV	---	---	---	---	---	---	---	---	---	---	---	---	---	---	100

(B)

	Ac	Bm	Op	Ld	Eppo	Hear	Hz	Se	Splt	Maco	Neab	Cuni	XcGV	PxGV	CpGV
Ac	100	91	79	54	78	47	47	51	50	50	37(58)	31	35	38	37
Bm	---	100	78	54	77	48	48	52	51	51	37(58)	31	35	37	37
Op	---	---	100	55	86	47	47	52	50	49	37(57)	31	32	38	37
Ld	---	---	---	100	53	53	53	54	48	51	34(54)	31	34	34	37
Eppo	---	---	---	---	100	47	47	51	49	48	36(59)	30	31	37	36
Hear	---	---	---	---	---	100	98	51	48	53	33(53)	29	33	35	34
Hz	---	---	---	---	---	---	100	51	47	52	33(53)	29	33	35	35
Se	---	---	---	---	---	---	---	100	49	63	34(56)	29	34	34	37
Splt	---	---	---	---	---	---	---	---	100	48	34(56)	30	35	35	35
Maco	---	---	---	---	---	---	---	---	---	100	34(54)	29	34	33	35
Neab	---	---	---	---	---	---	---	---	---	---	100(100)	31(49)	29(48)	28(45)	31(50)
Cuni	---	---	---	---	---	---	---	---	---	---	---	100	27	26	26
XcGV	---	---	---	---	---	---	---	---	---	---	---	---	100	36	41
PxGV	---	---	---	---	---	---	---	---	---	---	---	---	---	100	45
CpGV	---	---	---	---	---	---	---	---	---	---	---	---	---	---	100

(C)

	Ac	Bm	Op	Eppo	Ld	Se	Splt	Hear	Hz	Maco	Neab	Cuni	CpGV	PxGV	XcGV
Ac	100	86	89	89	80	85	83	85	85	88	47(67)	5	55	52	52
Bm	---	100	91	92	79	81	78	81	81	83	48(69)	5	55	54	53
Op	---	---	100	98	79	84	82	83	83	85	48(68)	5	55	55	52
Eppo	---	---	---	100	80	85	81	82	82	85	49(69)	5	56	56	53
Ld	---	---	---	---	100	81	80	80	80	83	46(66)	5	54	53	53
Se	---	---	---	---	---	100	83	87	87	90	47(68)	5	54	54	53
Splt	---	---	---	---	---	---	100	84	84	83	46(66)	5	55	53	52
Hear	---	---	---	---	---	---	---	100	100	87	47(68)	5	55	54	52
Hz	---	---	---	---	---	---	---	---	100	87	47(68)	5	55	54	52
Maco	---	---	---	---	---	---	---	---	---	100	48(69)	5	55	54	54
Neab	---	---	---	---	---	---	---	---	---	---	100(100)	4(10)	36(57)	37(58)	37(58)
Cuni	---	---	---	---	---	---	---	---	---	---	---	100	4	4	4
CpGV	---	---	---	---	---	---	---	---	---	---	---	---	100	85	87
PxGV	---	---	---	---	---	---	---	---	---	---	---	---	---	100	87
XcGV	---	---	---	---	---	---	---	---	---	---	---	---	---	---	100

(D)

	Ac	Bm	Op	Ld	Eppo	Hear	Hz	Splt	Se	Maco	Neab	Cuni	XcGV	CpGV	PxGV
Ac	100	98	78	63	78	63	62	62	61	59	23(48)	20	30	29	28
Bm	---	100	78	63	78	62	61	61	60	59	23(47)	20	29	29	29
Op	---	---	100	62	89	61	61	58	58	58	24(48)	19	28	28	26
Ld	---	---	---	100	62	60	60	60	61	59	23(47)	20	27	28	27
Eppo	---	---	---	---	100	61	61	57	58	58	24(48)	19	28	27	27
Hear	---	---	---	---	---	100	99	62	56	58	24(44)	17	29	28	26
Hz	---	---	---	---	---	--	100	62	56	58	24(44)	17	29	27	26
Splt	---	---	---	---	---	---	---	100	57	57	26(48)	20	30	32	30
Se	---	---	---	---	---	---	---	---	100	89	26(47)	19	26	27	26
Maco	---	---	---	---	---	---	---	---	---	100	25(46)	19	26	27	26
Neab	---	---	---	---	---	---	---	---	---	---	100(100)	17(37)	25(44)	25(46)	26(46)
Cuni	---	---	---	---	---	---	---	---	---	---	---	100	17	18	17
XcGV	---	---	---	---	---	---	---	---	---	---	---	---	100	53	52
CpGV	---	---	---	---	---	---	---	---	---	---	---	---	---	100	44
PxGV	---	---	---	---	---	---	---	---	---	---	---	---	---	---	100

Table 5.3. Amino acid sequence identities (%) of the baculovirus genes *odv-e56* (A), *p74* (B), *polh* (C), and *vlf-1* (D) from 15 baculoviruses. AcMNPV (Ac), BmMNPV (Bm), OpMNPV (Op), LdMNPV (Ld), EppoNPV (Eppo), HzSNPV (Hz), HearSNPV (Hear), SpltMNPV (Splt), MacoNPV (Maco), NeabNPV (Neab), CuniNPV (Cuni), XcGV, CpGV, and PxGV. Numbers in ( ) indicate the amino acid similarity (%) of NeabNPV to other the baculovirus sequences.

5.2.2. Phylogenetic inference based on baculovirus gene sequences

To date, 14 baculovirus genomes have been completely sequenced (Table 5.1). Partial DNA sequence data has become available for the NeabNPV genome and has been compared to sequences in the NCBI Genbank database using BLAST (v2.0) (Altschul *et al.*, 1990). Pairwise sequence alignments predicted the presence of several putative partial baculovirus genes and four putative complete baculovirus genes from the available NeabNPV genomic sequence data. Four complete baculovirus genes predicted to be contained in the NeabNPV genome were used in this study and include the structural proteins *polh*, *p74*, and *odv-e56* and the very late gene expression factor *vlf-1.* As well, four partial baculovirus gene sequences predicted to be contained in the NeabNPV genome were used in this study and include the structural protein *vp-91*, the late gene expression factors *lef-8* and *lef-9*, and the DNA helicase *p143.* All of the baculoviruses used in this study share these eight genes. For each individual gene, the inferred amino acid sequence was predicted and aligned using ClustalX (v1.81) (Jeannmougin *et al.,* 1998) with the default parameters. The multiple sequence alignments were checked and refined by eye using Bioedit (v5.0.9) (Hall, 1999) prior to being compiled into a single file of 6788 amino acid characters of which 3798 characters were determined to be parsimoniously informative. Each individual gene represents a defined subset of this file. Gaps were treated as missing data. Maximum parsimony analyses were performed using PAUP (v4.0b10) (Swafford, 1998). Phylogenies of the individual genes were estimated using the branch and bound algorithm of PAUP (v4.0b10) and branch support was evaluated by bootstrap analysis (1000 replicates) using a heuristic search algorithm. Phylogenies of the combined gene data set were estimated using the heuristic search algorithm of PAUP (v4.0b10) and branch support was evaluated by bootstrap analysis

(1000 replicates) using a heuristic search algorithm. All trees were rooted using the GVs as a sister outgroup to the NPVs as no unambiguous outgroup for the *Baculoviridae* is currently known. Phylogenetic trees were visualized and edited using Treeview (v1.6.6) (Page, 1996) and Windows Metafile Companion (v1.11).

5.3. Results

The inferred amino acid sequences of four completely sequence (*odv-e56*, *p74*, *polh*, and *vlf-1*) and four partially sequenced (*lef-8*, *lef-9*, *vp-91*, and *p143*) putative NeabNPV genes have each been compared to 14 other similar baculovirus gene sequences. Sequence comparisons indicated a low amino acid identity between NeabNPV and the other baculoviruses (Table 5.3). The NeabNPV *odv-e56* amino acid identity ranged from 23% with CuniNPV to 36% with LdMNPV. For NeabNPV *p74,* the amino acid identity ranged from 28% with PxGV to 37% among AcMNPV, BmMNPV, and OpMNPV. The amino acid identity between NeabNPV *vlf-1* and the other baculoviruses ranged from 17% with CuniNPV to 26% among SeMNPV, SpltMNPV, and PxGV. For NeabNPV *polh*, amino acid identities ranged from 4% with CuniNPV to 49% with EppoNPV.

Comparison of the NeabNPV genes also indicated an overall low amino acid sequence similarity. For NeabNPV *odv-e56*, amino acid similarity ranged from 40% with CuniNPV to 51% - 53% among the GVs (XcGV, PxGV, and CpGV). NeabNPV *p74* demonstrated an amino acid similarity ranging from 45% with PxGV to 57% - 59% among the Group I NPVs (AcMNPV, BmMNPV, OpMNPV, and EppoNPV). NeabNPV *vlf-1* showed an amino acid similarity ranging from 37% with CuniNPV to 44 - 48% among all the other NPVs and GVs used in this study. For *polh*, amino acid similarity to

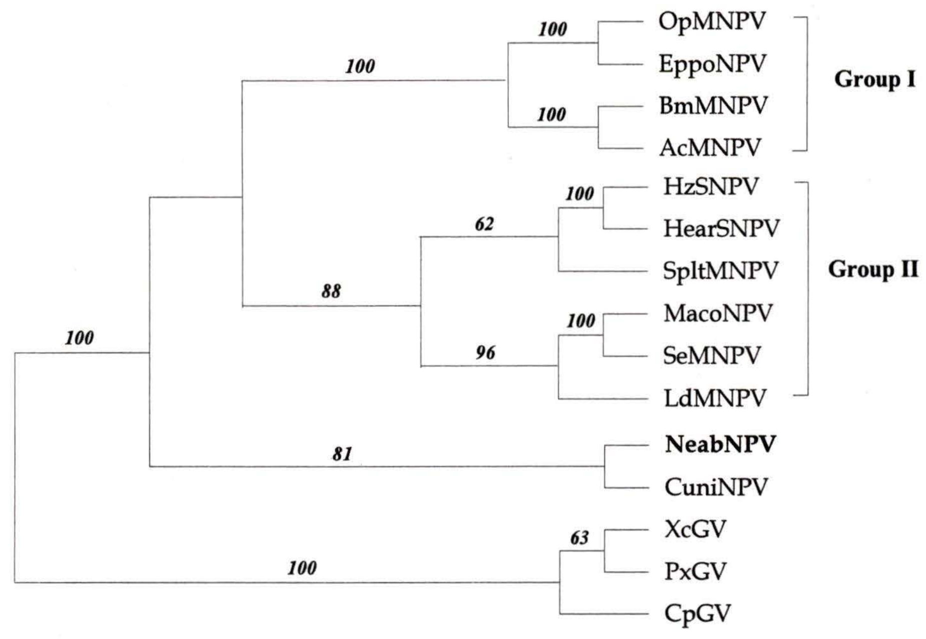
NeabNPV ranged from 10% with CuniNPV to 66% - 69% among the Group I NPVs (AcMNPV, BmMNPV, OpMNPV, and EppoNPV) and Group II NPVs (SeMNPV, SpltMNPV, HearSNPV, HzSNPV, MacoNPV, and LdMNPV).

A comparison of the 15 baculovirus genomes used in this study revealed that *polh*, *p74*, *odv-e56, vlf-1, vp-91*, *lef-8, lef-9*, and *p143* are all common to these viruses. The CuniNPV genome, however, appeared to lack an analogue of the highly conserved baculovirus polyhedrin gene and instead contained an ORF (CUN085) that partially aligns with the baculovirus occlusion protein sequences. CUN085 encodes an 882 amino acid long protein, which is significantly longer than the polyhedrin and granulin proteins (~245 amino acids). It has been suggested that the large CuniNPV occlusion protein is cleaved into smaller peptide fragments, some of which combine to form the viral occlusion body (Afonso *et al.*, 2001). As CUN085 is significantly longer than the other sequences, only a portion of it aligned, resulting in a very low overall sequence identity with the other baculoviruses. As a result, only the portion of CUN085 that aligned with the occlusion protein sequences was considered. This partial sequence showed an amino acid identity ranging from 16% with NeabNPV to 37% among the GVs.

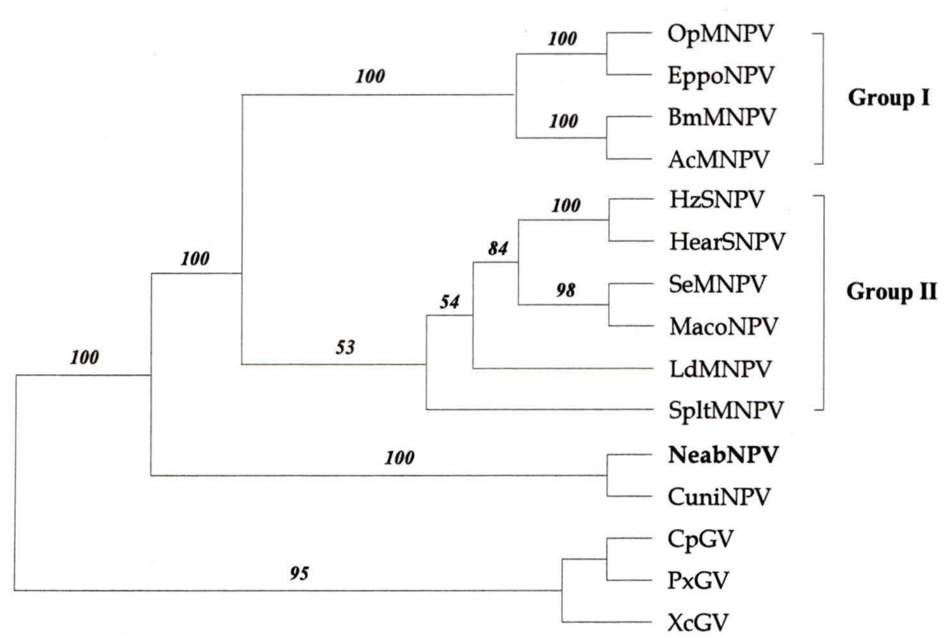
Phylogenetic trees were generated for each of the shared genes resulting in several different tree topologies. The most parsimoniously informative amino acid sequence came from the *p143* gene, which contained 1039 informative residues (Table 5.2), whereas the least parsimoniously informative amino acid sequence came from the *polh* gene, which contained 131 informative residues (Fig. 5.1 and Fig. 5.2, respectively). Most of the topological variation was seen in the arrangement of the Group II NPVs. The GV clade showed much less topological variation than the Group II NPVs and the Group I NPV clade showed the least variation, remaining the same in 7/8 phylogenies. All of the

phylogenetic trees generated supported the NPV-GV division, which was substantiated by bootstrap resampling frequencies greater than 86% in 7/8 phylogenetic trees. None of the trees had an identical topology.

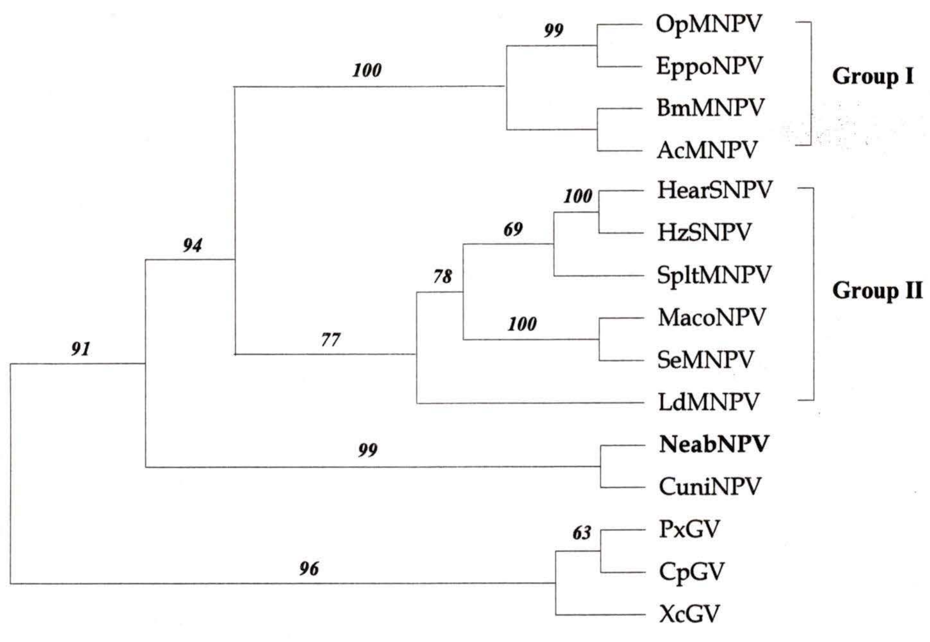
A phylogenetic tree was also generated for the combined gene data set. This analysis generated a single most parsimonious tree with high bootstrap support (Fig 5.3). The tree supported the NPV-GV division as well as the subdivision of the NPVs into the well-recognized Group I NPVs (Zanotto *et al.,* 1993) and Group II NPVs (Bulach *et al.,* 1999). In contrast to previous phylogenetic studies that placed LdMNPV as the basal Group II NPV taxon, the most parsimonious phylogenetic tree obtained here, based on the combined gene data set, places SpltMNPV as the basal taxon to the Group II NPV clade. This branching pattern is supported by a bootstrap resampling frequency of 93%. NeabNPV and CuniNPV formed a unique clade, which was distinct from the rest of the NPVs and was well supported by bootstrap resampling (100%). This branching pattern was also seen in 6 of the 8 individual gene phylogenies and was also well supported by bootstrap analysis, demonstrating an average bootstrap resampling value of greater than 90%.



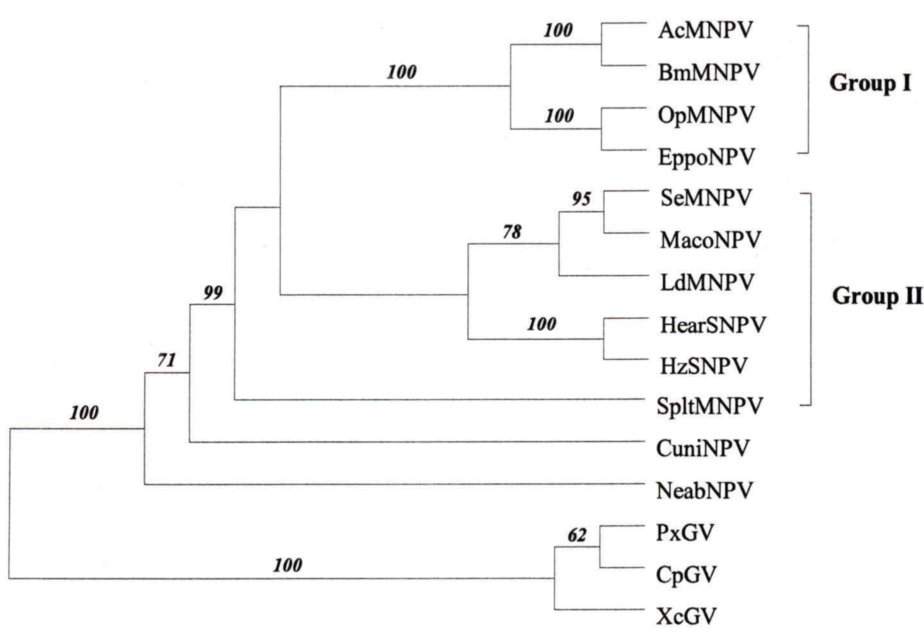
(A)



(B)

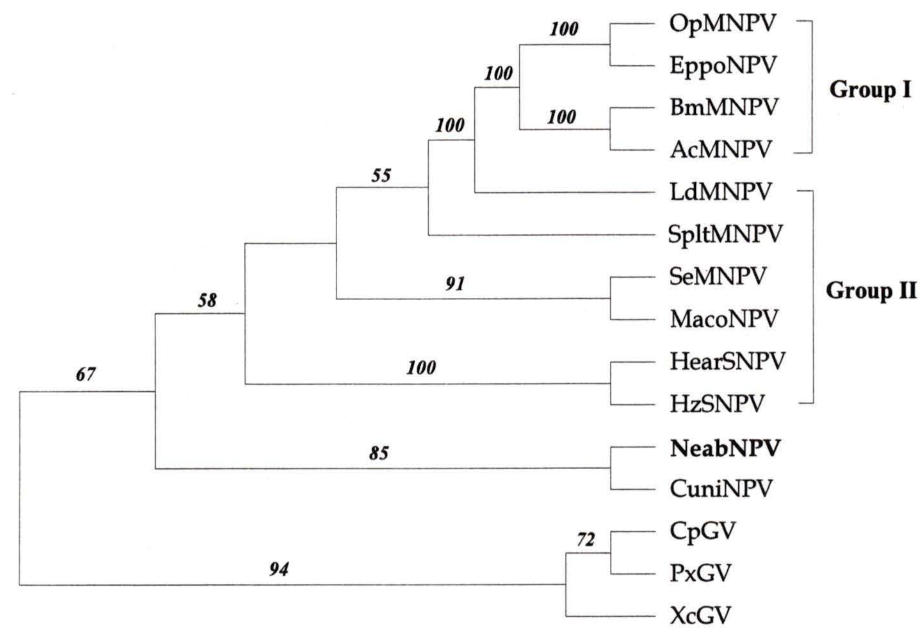


(C)

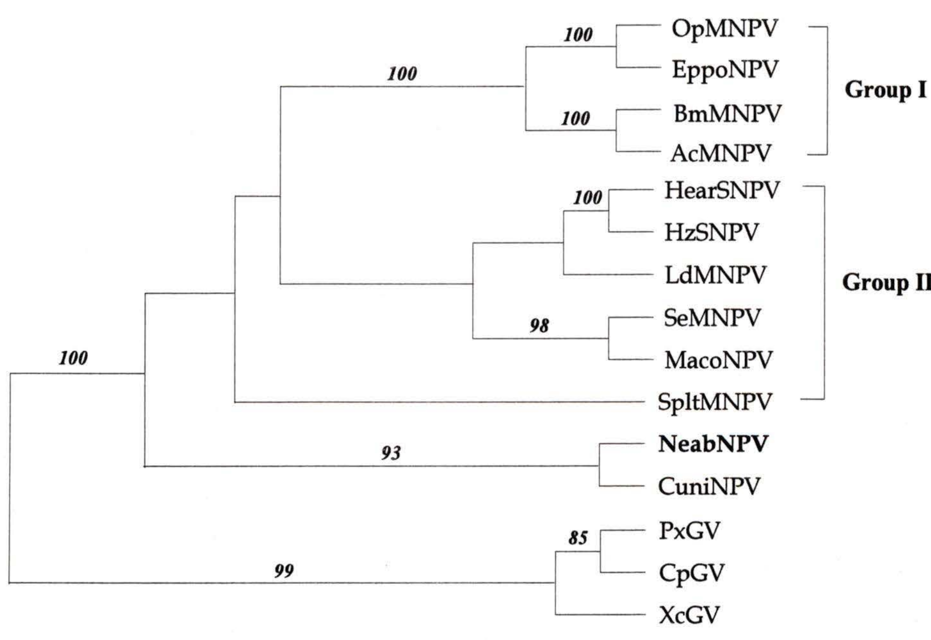


(D)

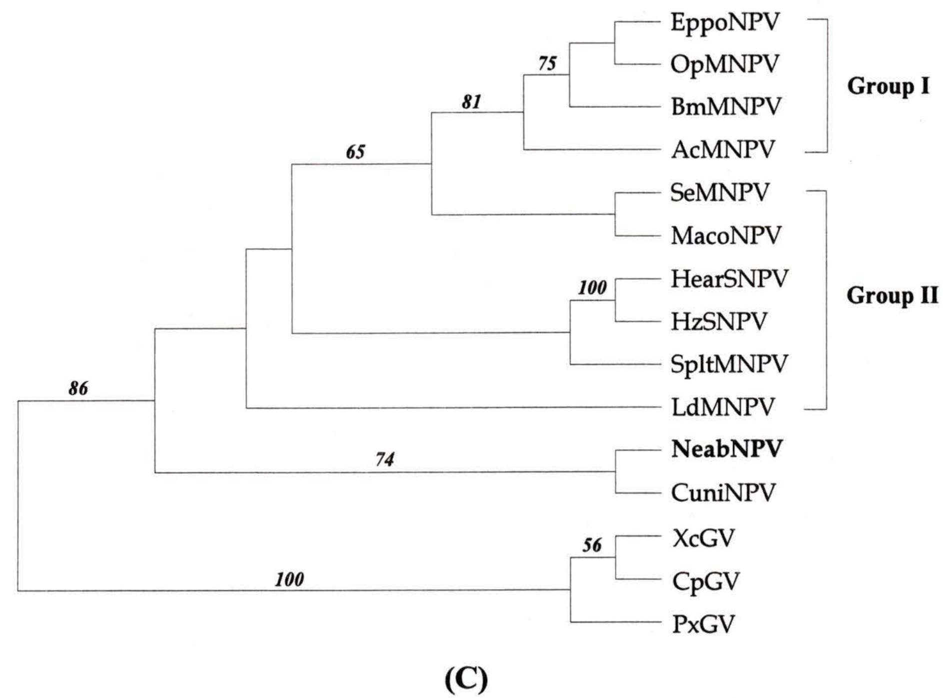
Figure 5.1. The single most parsimonious tree topologies obtained for the individual phylogenetic analyses of 4 baculovirus genes: *p143* (A), *lef-8* (B), *lef-9* (C), and *vp-91* (D). Only a partial, putative gene sequence was used for NeabNPV in each case. The inferred amino acid sequence of each baculovirus genes was used and the most parsimonious tree for each gene was constructed using the 'branch and bound' algorithm of PAUP (v4.10b). Percentages of bootstrap support (1,000 replicates) greater than 50% are shown along the branches. Trees are rooted using the GVs as a sister group to the NPVs.



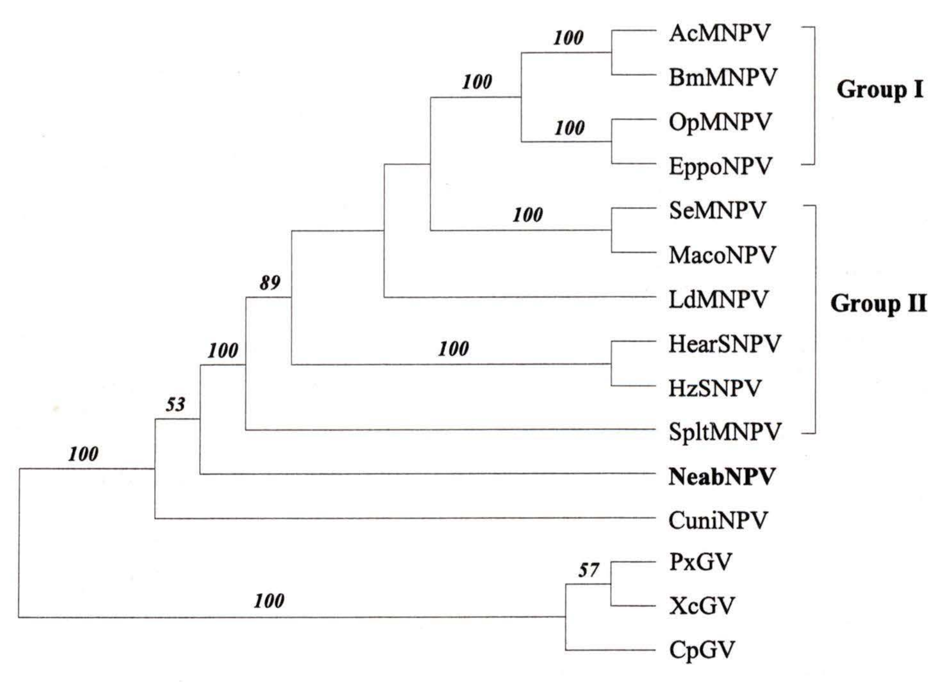
(A)



(B)



(C)



(D)

Figure 5.2. The single most parsimonious tree topologies obtained for the individual phylogenetic analyses of 4 baculovirus genes: *odv-e56* (A), *p74* (B), *polh* (C), and *vlf-1* (D). The inferred amino acid sequence of each baculovirus genes was used and the most parsimonious tree for each gene was constructed using the 'branch and bound' algorithm of PAUP (v4.10b). Percentages of bootstrap support (1,000 replicates) greater than 50% are shown along the branches. Trees are rooted using the GVs as a sister group to the NPVs.

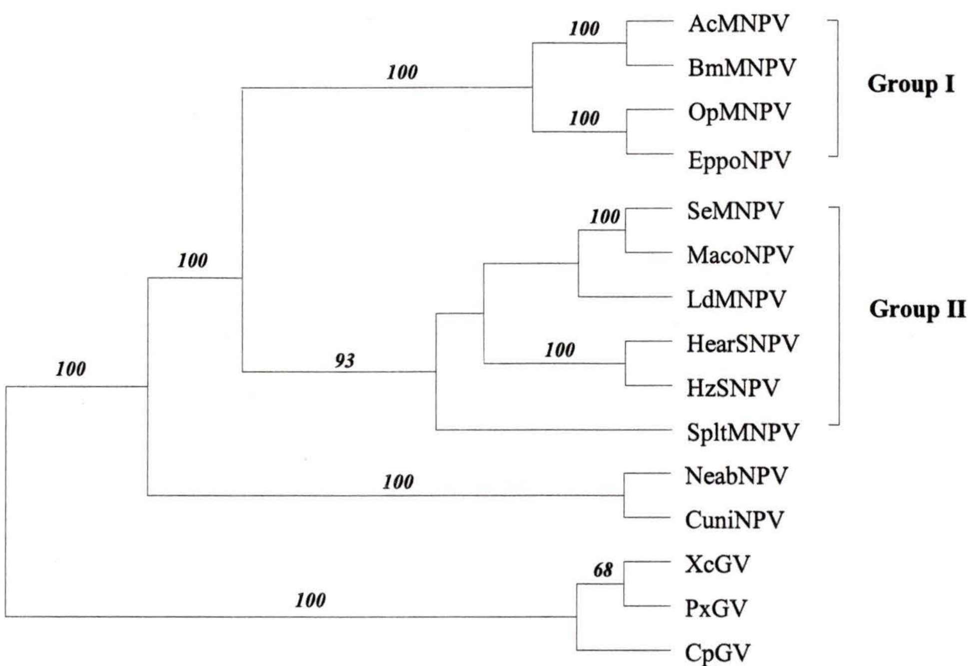


Figure 5.3. Phylogenetic tree of the combined sequence of *odv-e56*, *p74*, *polh*, and *vlf-1*, *p143*, *lef-8*, *lef-9*, and *vp-91*. All 8 genes are shared by the 15 baculovirus taxa used in this study. The inferred amino acid sequence from each gene was used for this phylogeny estimation. The most parsimonious tree was constructed using the 'heuristic' algorithm of PAUP (v4.0b10). Percentages of bootstrap support (1,000 replicates) greater than 50% are shown. Trees are rooted using the GVs as a sister group to the NPVs.

5.4. Discussion

All of the individual phylogenetic gene trees described here support the separation of the NPVs and GVs and the division of the NPVs into two major clades previously described by Zanotto *et al.,* (1993) and Bulach *et al.,* (1999). With the exception of the polyhedrin gene tree, all the individual gene trees were able to resolve the relationships between the Group I NPVs, with EppoNPV and OpMNPV forming a unique clade and AcMNPV and BmMNPV forming another clade. The inability of the polyhedrin gene tree to resolve the Group I NPVs has been previously noted (Herniou *et al.,* 2000). The overall bootstrap support for the polyhedrin gene tree was much lower than the other gene trees, which reflects the weak phylogenetic signal in polyhedrin amino acid sequence alignments.

A comprehensive analysis of baculovirus phylogeny conducted by Herniou *et al.* (2000), which used 63 genes common to 9 baculovirus genomes, showed that the polyhedrin gene tree was the only phylogenetic tree found to disagree with their consensus tree (a majority rule consensus tree representing 63 single most parsimonious gene trees). Their polyhedrin gene tree placed BmMNPV and OpMNPV into a unique Group I NPV clade whereas the other 62 gene trees constructed placed AcMNPV and BmMNPV together.

While the majority of published baculovirus gene sequences are for the major occlusion protein (polyhedrin and granulin) and the majority of phylogenetic studies that have been performed involve the use of the polyhedrin gene, it has been suggested that this protein is not an ideal candidate for phylogenetic studies (Bulach *et al.,* 1999). This is due to the fact that the protein is small (245 to 250 amino acid residues) and is highly conserved among the baculoviruses (more than half the amino acid residues are

invariant). Due to this high degree of invariance, relatively few amino acids are parsimoniously informative (residues present more than once, but not in all). As well, it was noted by Cowan *et al.* (1994) that estimations of more distant phylogenetic relationships from polyhedrin genes may be biased by the degree to which the nucleotide sequence has been saturated by mutation, which can cause conflicting placement of some branches depending on the procedure that is used for tree construction. As a result, caution should be taken when basing phylogenies solely on the polyhedrin gene.

Relationships between the viruses in the Group II NPVs were resolved in a monophyletic clade in 3 of the 8 individual gene trees, including the *lef-8, lef-9* and *p143* genes, but none of these trees had identical topologies. In contrast, the Group I NPVs and the GVs are often resolved as a monophyletic clade. Previous studies conducted by Herniou *et al.* (2000) demonstrated that of 63 individual gene trees, 7 had topologies that agreed with their consensus tree (mentioned above). Two of these 7 genes were the *lef-8 and lef-9* genes. As a result, it appears that the Group II NPVs are difficult to consistently resolve.

The combined gene sequence analysis performed in this study clearly resolved the Group I NPVs, Group II NPVs and GVs (Fig 5.3). This tree is a more likely representation of the relationships between the baculoviruses than any of the individual gene trees for several reasons. First, the size of the data set is much larger than the individual gene trees providing more data to sample from. It has been previously observed that when using a consistent method, combining genes reduces sampling error and causes the phylogenies to converge towards the correct solution with good support (Mitchell *et al.,* 2000, Herniou *et al.,* 2000). Although only 8 genes were used in this study, the combination of these genes versus using them by themselves should help to

produce a more plausible baculovirus phylogeny. Second, the phylogenetic tree constructed for the multiple gene sequence alignment is well supported by bootstrap analysis (>93% at 10 of 13 nodes and ~88% support overall). While none of the individual gene trees shared an identical topology to the combined gene tree, the *p143* and *lef-9* gene trees had topologies that were very similar (Fig 5.1). Third, there is evidence of gene exchange between co-infecting baculoviruses (Kondo and Maeda, 1991), baculoviruses and their hosts (Crozier and Ribeiro, 1992), and other infectious agents (Rohrmann and Karplus, 2001). As a result, phylogenies constructed from individual, non-orthologous genes can provide misleading phylogenetic information. In the combined gene tree, SpltMNPV is placed as the basal taxa to the Group II NPVs. Previous phylogenetic studies (that included SpltMNPV) using *p74* (Afonso *et al.,* 2001), DNA polymerase (Bulach *et al.,* 1999), and polyhedrin (Hyink *et al.,* 1998) did not place SpltMNPV as the basal Group II NPV taxa. More multi-gene phylogenetic studies including SpltMNPV will have to be performed to further verify this branch placement.

An interesting pattern that emerged from this data was the placement of both NeabNPV and CuniNPV into a clade, separate from the Group I and Group II NPVs. This is the first time that gene sequences from both a Hymenopteran and Dipteran baculovirus have been included in a molecular phylogenetic study together and as a result this is the first time that this unique clade has been seen. Very little sequence data is currently available for non-Lepidopteran baculoviruses and as such, few phylogenetic studies have been performed which include Hymenopteran NPVs or GVs. In a previous phylogenetic study of 18 baculovirus polyhedrin amino acid sequences, NeseNPV was shown to branch before the separation of the Lepidopteran GVs and NPVs. It was hypothesized that the Lepidopteran baculoviruses separated from the Hymenopteran (NeseNPV) lineage

before the divergence of the NPVs from the GVs, preceding Lepidopteran baculovirus radiation (Zanotto *et al.,* 1993). In another phylogenetic study, using baculovirus p74 and DNA polymerase amino acid sequences, (Afonso *et al.,* 2001) a Dipteran baculovirus (CuniNPV) was included in the datasets. This study found that CuniNPV formed a clade that was separate from the Lepidopteran NPVs and GVs and suggested that CuniNPV may be a member of a new genus within the family *Baculoviridae*. The pattern of branching found in the trees constructed by Afonso *et al.,* (2001) showed CuniNPV branching from the tree prior to the separation of the Lepidopteran GVs and NPVs.

In our phylogeny, based on 8 conserved baculovirus genes, NeabNPV and CuniNPV branch off from the GVs before the separation of the Lepidopteran NPVs. It remains to be seen which topology accurately reflects the evolutionary history of *Baculoviridae*. A more accurate interpretation of the relationship of the baculoviruses will be obtained when an unambiguous outgroup to the *Baculoviridae* is determined and when more Hymenopteran and Dipteran NPV sequences become available.

5.5. References

Afonso, C. L., Tulman, E. R., Lu, Z., Balinsky, C. A., Moser, B. A., Becnel, J. J., Rock, D. L. & Kutish, G. F. (2001). Genome sequence of a baculovirus pathogenic for Culex nigripalpus. *J Virol.* **75**, 11157-65.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990). Basic local alignment search tool. *J Mol Biol.* **215**, 403-10.

Braunagel, S. C., Elton, D. M., Ma, H. & Summers, M. D. (1996). Identification and analysis of an Autographa californica nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. *Virol.* **217**, 97-110.

Bulach, D. M., Kumar, C. A., Zaia, A., Liang, B. & Tribe, D. E. (1999). Group II nucleopolyhedrovirus subgroups revealed by phylogenetic analysis of polyhedrin and DNA polymerase gene sequences. *J Invertebr Pathol.* **73**, 59-73.

Chen, X. (2001). Genomics and Genetic Engineering of *Helicoverpa armigera* Nucleopolyhedrovirus. In *Virology, Plant Sciences, Research School 'Production Ecology and Resource Conservation*. Netherlands: Wageningen University.

Chen, X., Ijkel, W. F., Dominy, C., de Andrade Zanotto, P. M., Hashimoto, Y., Faktor, O., Hayakawa, T., Wang, C. H., Prekumar, A., Mathavan, S., Krell, P. J., Hu, Z. & Vlak, J. M. (1999). Identification, sequence analysis and phylogeny of the lef-2 gene of Helicoverpa armigera single-nucleocapsid baculovirus. *Virus Res.* **65**, 21-32.

Cowan, P., Bulach, D., Goodge, K., Robertson, A. & Tribe, D. E. (1994). Nucleotide sequence of the polyhedrin gene region of Helicoverpa zea single nucleocapsid nuclear polyhedrosis virus: placement of the virus in lepidopteran nuclear polyhedrosis virus group II. *J Gen Virol.* **75**, 3211-8.

Crozier, G., and H. C. T. Ribeiro. (1992). Recombination as a possible major cause of genetic heterogeneity in *Anitcarsia gemmatalis* nuclear polyhedrosis virus populations. *Virus Res.* **26**, 183-196.

Doolittle, W. F. (1999). Phylogenetic classification and the universal tree. *Science.* **284**, 2124-9.

Hall, T. A. (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp* Ser. **41**, 95-98.

Hayakawa, T., Rohrmann, G. F. & Hashimoto, Y. (2000). Patterns of genome organization and content in lepidopteran baculoviruses. *Virol.* **278**, 1-12.

Herniou, E., Luque, T., Chen, X., Vlak, J. M., Winstanley, D., Cory, J. S. & O' Reilly, D. R. (2001). Use of Whole Genome Sequence Data to Infer Baculovirus Phylogeny. *J Virol.* **75**, 8117-8126.

Jeanmougin, F., Thompson, J. D., Gouy, M., Higgins, D. G. & Gibson, T. J. (1998). Multiple sequence alignment with Clustal X. *Trends Biochem Sci.* **23**, 403-5.

Kang, W., Tristem, M., Maeda, S., Crook, N. E. & O'Reilly, D. R. (1998). Identification and characterization of the Cydia pomonella granulovirus cathepsin and chitinase genes. *J Gen Virol.* **79**, 2283-92.

Kondo, A., and S. Maeda. (1991). Host range expansion by recombination of the baculoviruses *Bombyx mori* nuclear polyhedrosis virus and *Autographa californica* nuclear polyhedrosis virus. *J Gen Virol.* **65**, 3625-3632

Martignoni, M. E., and Iwai, P. J. (1986). *A catalog of viral diseases of insects, mites, and ticks*, 4th edn. Portland Oregon: USDA Forest Services PNW-195.

McLachlin, J. R. & Miller, L. K. (1994). Identification and characterization of vlf-1, a baculovirus gene involved in very late gene expression. *J Virol.* **68**, 7746-56.

Mitchell, A., Mitter, C. & Regier, J. C. (2000). More taxa or more characters revisited: combining data from nuclear protein-encoding genes for phylogenetic analyses of Noctuoidea (Insecta: Lepidoptera). *Systematic Biology.* **49**, 202-224.

Moscardi, F. (1999). Assessment of the Application of Baculoviruses for the Control of Lepidoptera. *Annual Review of Entomology* **44**, 257-289. Nei, M. (1996). Phylogenetic analysis in molecular evolutionary genetics. *Annu Rev Genet.* **30**, 371-403.

O'Reilly, D. R., Miller, L. K. & Luckow, V. A. (1992). Baculovirus Expression Vectors: A Laboratory Manual, pp. 139-142. New York: W. H. Freeman and Company.

Page, R. D. M. (1996). TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer Applications in the Biosciences*. **12**, 357-358.

Rohrmann, G. F. (1992). Baculovirus structural proteins. *J Gen Virol*. **73**, 749-61.

Rohrmann, G. F. and P. A. Karplus. (2001). Relatedness of baculovirus and gypsy retrotransposon envelope proteins. *BMC Evol Biol*. **1**, 1.

Swofford, D. L. (1998). PAUP*. Phylogenetic Analysis Using Parsimony (* and Other Methods), 4 edn. Sunderland, Massachusetts: Sinauer Associates.

Vlak, J. M. & Rohrmann, G. F. (1985). The Nature of Polyhedrin. In *Viral Insecticides for Biological Control*, pp. 489-544. Edited by K. Maramorosh & K. E. Sherman. Orlando: Academic Press.

Yang, S. & Miller, L. K. (1998). Expression and mutational analysis of the baculovirus very late factor 1 (vlf-1) gene. *Virol*. **245**, 99-109.

Yang, S. & Miller, L. K. (1999). Activation of baculovirus very late promoters by interaction with very late factor 1. *J Virol*. **73**, 3404-9.

Zanotto, P. M., Kessing, B. D. & Maruniak, J. E. (1993). Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. *J Invertebr Pathol*. **62**, 147-64.

Conclusions

An initial molecular characterization of the Hymenopteran baculovirus NeabNPV has been performed through the analysis of DNA sequence data derived from NeabNPV genomic libraries. Preliminary findings have determined that the NeabNPV genome is approximately 94.5 kilobases in size and contains many putative baculovirus specific sequences including homologues of structural (*polh, p74*), replication (*p143, dnapol*), transcription (*lef-8, lef-9*), and auxiliary (*iap-3*) genes common to the *Baculoviridae.* Further research will be necessary in order to characterize NeabNPV in more detail. This should involve sequencing the entire NeabNPV genome, development of a complete physical and genetic map, and in-depth analyses of a variety of conserved and unique open reading frames and genes.

The complete DNA sequence of the NeabNPV polyhedrin gene was determined and analysed. Sequence analysis indicated that NeabNPV is highly diverged from Lepidopteran baculoviruses and is closely related to the Hymenopteran baculovirus NeseNPV. The average amino acid similarity of NeabNPV polyhedrin with that of Lepidopteran baculovirus polyhedrins was approximately 63%, whereas the amino acid similarity with NeseNPV was 86%. The NeabNPV polyhedrin gene sequence was determined to contain putative baculovirus specific motifs including two late promoter motifs (ATAAG and TTAAG), a single nuclear localization signal (KKRK), and a single polyadenylation signal (AATTATTTTT). The 5' UTR of the NeabNPV polyhedrin gene was also analysed and determined to be more similar (length and nucleotide identity) to the Lepidopteran GVs than NPVs.

Phylogenetic analysis of 43 baculovirus polyhedrin amino acid sequences placed NeabNPV into a unique clade with NeseNPV, separate from the Lepidopteran NPVs and GVs. Another baculovirus phylogeny was also estimated using the combined sequence of four complete, and four partial baculovirus specific open reading frames. This analysis demonstrated that NeabNPV forms a unique clade within the NPVs, separate from the traditional Group I and Group II NPVs, when GVs are used as a sister outgroup. This clade, which also included the Dipteran baculovirus CuniNPV, may represent the non-Lepidopteran baculoviruses. Future phylogenetic studies should include more non-Lepidopteran baculoviruses as well as whole genome phylogenetic data (gene content, gene order), which should help to confirm the placement of NeabNPV into this unique NPV clade.

The identification and characterization of a putative partial NeabNPV VP39 protein revealed a pattern of secondary structure and hydrophilicity that appears to be conserved among baculovirus VP39s. The average amino acid sequence identity between Lepidopteran baculoviruses and NeabNPV VP39s was low (~23%). However, the placement of hydrophilic amino acids along the length of the protein and the position of predicted regions of secondary structure suggests that the function of the protein has been conserved in the putative NeabNPV VP39 structural protein.

Structural protein profiles (derived by separating proteins using SDS-PAGE) of putative NeabNPV BVs and ODVs were compared with AcMNPV (AcMNPV protein profiles obtained from Braunagel and Summers, 1994). A comparison of these structural protein profiles demonstrated that NeabNPV has several proteins with molecular weights similar to AcMNPV (e.g. BV proteins at 67, 45, and 35 kDa). However, the overall

protein profiles of the putative NeabNPV BVs and ODVs were significantly different from AcMNPV. These differences were not attributed to specific viral structural proteins as the purification procedures used for NeabNPV may have co-purified non-viral proteins. Protein comparisons made in this thesis were only preliminary in nature and were not performed with the intent of drawing any major conclusions. Future structural protein analyses of NeabNPV should include more stringent purification procedures (e.g. sucrose gradient centrifugation of putative OBs, ODVs and BVs) as well as confirmation steps to verify the presence of NeabNPV virions in sample preparations (e.g. electron microscopy to confirm presence of nucleocapsids after removal of viral envelope). As these steps were not taken during the purification of NeabNPV virions for protein analysis, no definitive conclusions could be made about the similarities or differences between NeabNPV and AcMNPV structural proteins.

Appendix. Translation BLAST results for NeabNPV genomic DNA sequences. Nucleotide data was obtained, translated in all six frames, and compared to the NCBI Genbank sequence database. (*) indicates a entomopoxvirus homologue.

Sequence Name	Aligned With	Putative ORF	E- Value
2	CpGV	IAP-3	1.00E-33
	OpSNPV		2.00E-33
	EppoNPV		2.00E-32
5	LdMNPV	ODV-E56	1.00E-22
	AcMNPV		5.00E-21
	HzSNPV		6.00E-21
7	PxGV	ODV-E56	7.00E-46
	HearSNPV		1.00E-44
	CfGV		2.00E-44
121	HzSNPV	AcORF115	3.00E-27
	HearSNPV		3.00E-27
	BmMNPV		2.00E-23
148	SeMNPV	P143	3.00E-15
	LdMNPV		4.00E-12
	SpltMNPV		5.00E-12
214	LdMNPV	P49	8.00E-18
	HearSNPV		2.00E-17
	EppoNPV		4.00E-17
215	AcMNPV	Alkaline Nuclease	4.00E-15
	PxGV		5.00E-15
	BmMNPV		9.00E-15
233	XcGV	AcORF119	4.00E-25
	PxGV		1.00E-21
	AcMNPV		2.00E-21
235	OpSNPV	IAP-3	9.00E-26
	EppoNPV		9.00E-26
	CpGV		3.00E-24

Sequence Name	Aligned With	Putative ORF	E- Value
247	PxGV	LEF-8	9.00E-34
	SpltMNPV		3.00E-33
	XcGV		6.00E-33
261	NeseNPV	AcORF22	1.00E-100
	CuniNPV		3.00E-55
	PxGV		4.00E-55
263	NeseNPV	HOAR	5.00E-44
264	SeMNPV	LEF-9	5.00E-34
	BmMNPV		1.00E-33
	AcMNPV		2.00E-33
269	EppoNPV	AcORF96	5.00E-17
	CuniNPV		1.00E-16
	SeMNPV		3.00E-16
274	SpltMNPV	VLF-1	1.00E-30
301	EppoNPV	P74	4.00E-43
	OpMNPV		5.00E-42
	SpliNPV		1.00E-41
302	PxGV	ODV-E56	2.00E-42
	HearSNPV		1.00E-41
	LdMNPV		1.00E-41
321	LdMNPV	AcORF119	2.00E-48
	PxGV		9.00E-46
	AcMNPV		2.00E-45
322	LdMNPV	AcORF119	6.00E-41
	AcMNPV		3.00E-39
	SeMNPV		3.00E-39
335	NeseNPV	HOAR	2.00E-13
350	LdMNPV	AcORF91	6.00E-29
	PxGV		4.00E-28
	AcMNPV		1.00E-27

Sequence Name	Aligned With	Putative ORF	E- Value
394	HearSNPV	VP-91	2.00E-16
	EppoNPV		7.00E-16
	OpMNPV		2.00E-15
395	SeMNPV	AcORF119	7.00E-20
	OpMNPV		3.00E-19
	LdMNPV		1.00E-16
434	CpGV	IAP-3	1.00E-16
	SeMNPV		7.00E-16
	OpMNPV		7.00E-14
509	SpltMNPV	P143	8.00E-18
	HearSNPV		8.00E-18
552	PxGV	ORF73	8.00E-22
	OpMNPV	ORF92	4.00E-21
	AcMNPV	ORF91	9.00E-18
588	CpGV	ODV-E56	2.00E-24
	CfGV		2.00E-23
	HearSNPV		4.00E-22
631	LdMNPV	ORF155	1.00E-31
	AcMNPV	ORF119	1.00E-30
	PxGV	ORF7	3.00E-30
734	XcGV	ORF84	5.00E-23
	PxGV	ORF7	2.00E-19
	HearSNPV		
926	SeMNPV	P74	1.00E-33
	CfMNPV		8.00E-33
	AcMNPV		2.00E-32
930	HycuNPV	Polh	1.00E-17
	SeMNPV		2.00E-17
	SfNPV		2.00E-17
953	SeMNPV	LEF-9	1.00E-27
	LdMNPV		1.00E-25
	BmMNPV		3.00E-25

Sequence Name	Aligned With	Putative ORF	E- Value
955	SpltMNPV	P143	2.00E-17
	HzSNPV		3.00E-17
	HearSNPV		3.00E-17
959	HearSNPV	DNA Pol.	1.00E-12
	HzSNPV		1.00E-12
	OranNPV		3.00E-12
964	CpGV	ODV-E56	3.00E-17
	PxGV		3.00E-17
	HearSNPV		3.00E-16
1020	AcMNPV	P74	1.00E-22
	LdMNPV		2.00E-22
	CfMNPV		4.00E-22
1026	NeseNPV	ORF22	2.00E-16
1028	HearSNPV	VP91	2.00E-26
	HzSNPV		5.00E-26
	MacoNPV		2.00E-24
1036	LdMNPV	ODV-E56	2.00E-22
	CfMNPV		2.00E-22
	EppoNPV		3.00E-22
1039	CuniNPV	P74	2.00E-29
	LdMNPV		2.00E-26
	CfMNPV		1.00E-25
1065	SpltMNPV	VP39	7.00E-12
	MacoNPV		6.00E-11
	OpMNPV		2.00E-09
1084	EppoNPV	P74	3.00E-20
	CfMNPV		3.00E-18
	AcMNPV		6.00E-18
1101	LdMNPV	ORF155	2.00E-38
	XcGV	ORF84	6.00E-38
	HearSNPV		8.00E-38

Sequence Name	Aligned With	Putative ORF	E- Value
1112	HzSNPV	ORF101	1.00E-29
	HearSNPV		4.00E-29
	SeMNPV	ORF50	7.00E-26
1114	SpltMNPV	AcORF81	8.00E-15
	EppoNPV		6.00E-14
	HearSNPV		1.00E-13
1117	CpGV	IAP-3	2.00E-14
	BmMNPV		2.00E-14
	SeMNPV		7.00E-14
1121	SpltMNPV	VLF-1	2.00E-11
	HearSNPV		5.00E-09
	HzSNPV		5.00E-09
1135	HearSNPV	LEF-1	1.00E-25
	HzSNPV		1.00E-25
	SpltMNPV		5.00E-25
1136	PxGV	ORF87	1.00E-14
1145	CpGV	LEF-8	4.00E-24
	SpltMNPV		1.00E-23
	XcGV		3.00E-23
1171	SpltMNPV	LEF-1	1.00E-34
	XcGV		6.00E-29
	HearSNPV		4.00E-27
1172	NeseNPV	HOAR-like	2.00E-11
1181	NeseNPV	HOAR-like	2.00E-11
1205	BusuNPV	DNA Pol.	7.00E-13
	SeMNPV		8.00E-12
	MabrNPV		8.00E-12
1210	AcMNPV	AcORF119	3.00E-15
	EppoNPV		4.00E-15
	LdMNPV		4.00E-15
1213	NeseNPV	HOAR-like	2.00E-14

Sequence Name	Aligned With	Putative ORF	E- Value
1217	HearSNPV	GP-41	8.00E-13
	HzSNPV		8.00E-13
	MacoNPV		4.00E-11
1224	NeseNPV	ORF22	3.00E-96
	HearSNPV		2.00E-59
	AcMNPV		5.00E-59
1236	NeseNPV	ORF22	7.00E-42
	PxGV	ORF37	3.00E-12
	CuniNPV	ORF38	3.00E-12
1252	MacoNPV	P143	2.00E-11
	SeMNPV		2.00E-10
	LdMNPV		1.00E-09
1275	ArseNPV	Polh	9.00E-19
	BusuNPV		5.00E-18
	AnfaNPV		5.00E-18
1325	EppoNPV	P74	8.00E-28
	CfMNPV		9.00E-26
	OpMNPV		9.00E-26
1332	AgseNPV	Polh	1.00E-37
	WisiNPV		1.00E-37
	EcobNPV		3.00E-36
	MacNPV		4.00E-11
1457	NeseNPV	ORF22	7.00E-53
	AcMNPV		2.00E-35
	MacoNPV		2.00E-35
1482	NeseNPV	HOAR-like	3.00E-28
1491	OpMNPV	AcORF119	2.00E-23
	LdMNPV		3.00E-24
	SeMNPV		3.00E-24
1494	BusuNPV	P74	2.00E-13
	SpltMNPV		2.00E-12
	SeMNPV		3.00E-10

Sequence Name	Aligned With	Putative ORF	E- Value
1502	NeseNPV	ORF22	1.00E-80
	SpltMNPV		2.00E-57
	AcMNPV		2.00E-55
1504	NeseNPV	ORF22	4.00E-95
	AcMNPV		5.00E-65
	BmMNPV		2.00E-63
1516	SpltMNPV	LEF-8	1.00E-11
	SpliNPV		2.00E-10
	MacoNPV		9.00E-10
1518	SpltMNPV	VLF-1	9.00E-26
	HzSNPV		7.00E-20
	HearSNPV		7.00E-20
1520	EppoNPV	P74	1.00E-20
	CfMNPV		2.00E-18
	AcMNPV		3.00E-18
1522	XcGV	LEF-9	1.00E-49
	CpGV		6.00E-49
	SpltMNPV		1.00E-46
1530	BusuNPV	DNA Pol.	8.00E-12
	OranNPV		9.00E-10
	HzSNPV		2.00E-09
1538	SpltMNPV	VP39	4.00E-12
	MacoNPV		3.00E-11
	OpMNPV		1.00E-09
1607	AcMNPV	P74	7.00E-12
	SeMNPV		2.00E-11
	BusuNPV		2.00E-11
1626	CpGV	IAP-3	6.00E-33
	BmMNPV		8.00E-32
	TniGV		4.00E-31
1634	PxGV	LEF-1	1.00E-15
	CpGV		9.00E-15
	HzSNPV		3.00E-13

Sequence Name	Aligned With	Putative ORF	E- Value
1675	AcMNPV	P74	3.00E-22
	CfMNPV		8.00E-22
	LdMNPV		2.00E-21
1678	AnfaNPV	Polh	3.00E-55
	HycuNPV		8.00E-55
	EppoNPV		1.00E-54
1692	BmMNPV	VP91	1.00E-10
	MacoNPV		2.00E-10
	OpMNPV		4.00E-10
1694	LdMNPV	P74	1.00E-15
	AcMNPV		7.00E-15
	BmMNPV		7.00E-15
1787	PxGV	LEF-9	2.00E-11
	XcGV		3.00E-11
	CpGV		5.00E-10
1824	NeseNPV	HOAR-like	1.00E-41
1828	PxGV	ODV-E56	1.00E-28
	HearSNPV		2.00E-28
	HzSNPV		2.00E-28
1839	EppoNPV	IAP-3	9.00E-18
	TniGV		9.00E-17
	AmmoEPV		1.00E-16
1840	SpltMNPV	VP39	3.00E-11
	MacoNPV		2.00E-10
	OpMNPV		6.00E-09
1850	LdMNPV	VP91	2.00E-14
	HearSNPV		1.00E-13
	SpltMNPV		2.00E-13

Sequence Name	Aligned With	Putative ORF	E- Value
1876	PxGV	ODV-E56	1.00E-28
	HearSNPV		2.00E-28
	HzSNPV		2.00E-28
1887	XcGV	LEF-8	4.00E-15
	MacoNPV		5.00E-15
	SpltMNPV		1.00E-14
1888	SpltMNPV	VP39	8.00E-12
	MacoNPV		5.00E-11
	OpMNPV		2.00E-09
1913	AnfaNPV	Polh	2.00E-50
	AngeNPV		2.00E-50
	HearSNPV		4.00E-50
1918	HearSNPV	ODV-E56	4.00E-44
	HzSNPV		7.00E-44
	LdMNPV		8.00E-43
1940	SpltMNPV	LEF-8	1.00E-10
	SeMNPV		2.00E-10
	CpGV		3.00E-10
1946	BusuNPV	P47	6.00E-10
	HearSNPV		8.00E-10
	HzSNPV		8.00E-10
1952	SpltMNPV	VP91	6.00E-31
	AcMNPV		3.00E-28
	SeMNPV		6.00E-28
1956	HearSNPV	Polh	4.00E-43
	BmMNPV		9.00E-43
	AnfaNPV		2.00E-42

VITA

Surname: Young Given Names: Aaron Michael

Place of Birth: Revelstoke, British Columbia, Canada

Educational Institutions Attended:

University of Victoria	2000 to 2002
University of Victoria	1996 to 1999
Okanagan University College	1994 to 1996

Degrees Awarded:

B.Sc.	University of Victoria	2000

Honours and Awards:

Canadian Forestry Service Graduate Research Supplement	2001
NSERC Postgraduate Scholarship (PGS-A)	2001 to 2002
University of Victoria President's Research Scholarship	2001 to 2002
University of Victoria Graduate Teaching and Research Fellowship	2000
The Ethel N. Lohbrunner Award	2000
The Gerald G. Few Graduate Award	2000
The Steven Peterson Memorial Award	1998
The James Ellison Campbell Taylor Award	1997
The RE Postill & Sons Ltd. Award	1996
The Kalamalka Centre Award	1995

Publications:

None

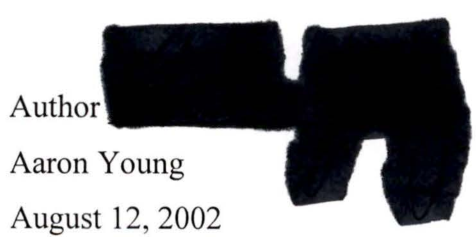
PARTIAL COPYRIGHT LICENSE

UNIVERSITY OF VICTORIA PARTIAL COPYRIGHT LICENSE

I hereby grant the right to lend my thesis (or dissertation) to users of the University of Victoria Library, and to make single copies only for such users or in response to a request from the Library of any other university, or similar institution, on its behalf or for one of its users. I further agree that permission for extensive copying of this thesis for scholarly purposes may be granted by me or a member of the University designated by me. It is understood that copying or publication of this thesis for financial gain by the University of Victoria shall not be allowed without my written permission.

Title of Thesis:

Molecular Characterization of the *Neodiprion abietis* Nucleopolyhedrovirus



Author

Aaron Young

August 12, 2002