

**THE MOLECULAR CHARACTERIZATION OF *TRICHOPLUSIA NI*
SINGLE NUCLEOCAPSID NUCLEOPOLYHEDROVIRUS:
A STUDY ON EARLY REGULATORY FEATURES**

BY

WEIZHOU WANG

**A thesis submitted in partial fulfillment of the requirements for the
Degree of Doctor of Philosophy in the Department of Biotechnology
University of the Western Cape**

Supervisor: Professor S. Davison

July 2003

**THE MOLECULAR CHARACTERIZATION OF *TRICHOPLUSIA NI*
SINGLE NUCLEOCAPSID NUCLEOPOLYHEDROVIRUS:
A STUDY ON EARLY REGULATORY FEATURES**

BY

WEIZHOU WANG

**A thesis submitted in partial fulfillment of the requirements for the
degree of Doctor of Philosophy in the Department of Biotechnology
University of the Western Cape**

Supervisor: Prof. Sean Davison

July 2003

DECLARATION

I declare that the research work titled "The molecular characterization of *Trichoplusia ni* single nucleocapsid nucleopolyhedrovirus: a study on early regulatory features" is my own work and all the resources I have used or quoted have been indicated and acknowledged by means of complete references.

Full name Weizhou Wang

Signed _____

Date July 24, 2003

Acknowledgements

I would like to express my sincere gratitude and appreciation to the following:

Prof. Sean Davison, my supervisor, whose inspiration, encouragement, guidance and great help, not only through this project but also in my personal life, will always be remembered.

Prof. Chris Gehring for providing me access to his laboratory and for his encouragement and help.

Prof. Peter Krell for allowing me to work in his laboratory for a period at the Department of Microbiology, University of Guelph, Canada

Colleagues in the virology laboratory for the valuable discussions, advice and assistance towards part of the project.

Steve Lord and J. D. de Jong, Department of Microbiology, University of Guelph, Canada, for technical assistance towards part of the work.

Pastor Arthur Volker and Mrs. Joy Volker, Mowbray Baptist Church, for their love, spiritual guidance and invaluable help to myself and my family which have strengthened and encouraged me to face difficult challenges during my studies.

Mrs. Gill Lanham and Mr. John Lanham for their kind caring for my family, especially for their love towards my son.

Mr. Don Waterson and Mrs. Sheila Waterson, Mowbray Baptist Church, for their special care and friendliness towards my family and myself.

Friends at the Mowbray Baptist Church, for their encouragement, fellowship and sharing God's Words with me.

My wife Yanqing for her love, hard work and devotion to the family.

My son Zijng for his understanding and excellent performance at school for which we are always proud.

List of Figures

Figure 1.1	Structural diagram of the two baculovirus pheontypes.....	4
Figure 1.2	Schematic diagram of the baculovirus life cycle.....	8
Figure 2.1	The location of the <i>ie-1</i> gene within an 11 kb <i>EcoRI</i> fragment and an adjoining 2.3 kb <i>EcoRI</i> fragment.....	33
Figure 2.2	Nucleotide and deduced amino acid sequences of the TnSNPV <i>ie-1</i> gene.....	36
Figure 2.3	Multiple sequence alignment of <i>ie-1</i> promoter regions.....	38
Figure 2.4	Alignment of amino acid sequences of eleven baculovirus <i>ie-1</i> genes.....	42
Figure 2.5	Acid-base profiles of the eleven baculovirus IE-1 proteins.....	44
Figure 2.6	Phylogenetic relationships of the predicted full length (a) and the C-terminal two thirds (b) of the IE-1 proteins.....	48
Figure 3.1	Schematic representation and amino acids sequences of the wild type IE-1s and their truncations.....	64
Figure 3.2	Structural alignments of the putative C-terminal binding regions and the b-HLH-like domains of eight IE-1s.....	70
Figure 3.3	Expression, purification and detection of the AcNPV and TnSNPV IE-1 truncations.....	71
Figure 3.4	Efficacy of purification of the 6xHis tagged truncation ($\Delta 1-452$).....	72
Figure 3.5	DNA-binding activities of the IE-1 truncation and controls.....	74
Figure 4.1	The location of the me-53 gene within an 11 kb and an adjoining 2.3 kb <i>EcoRI</i> fragment and the AcMNPV recombinant.....	86

Figure 4.2	Nucleotide and deduced amino acid sequences of the TnSNPV <i>me-53</i> - containing fragment.....	92
Figure 4.3	Alignment of the deduced amino acid sequences of twelve baculovirus <i>me-53</i> genes.....	97
Figure 4.4	Pattern of the zinc finger-like motifs in the baculovirus ME-53s.....	99
Figure 4.5	Uninfected and AcBacTnme53GFP-infected Sf 9 and High5 cells.....	100
Figure 4.6	Temporal expression of the TnSNPV <i>me-53</i> transcripts by RT-PCR analysis.....	103
Figure 4.7	Comparison of regulatory elements within the upstream regions of ten NPV <i>me-53s</i>	105
Figure 4.8	Rooted neighbor-joining trees of deduced baculovirus ME-53 (a) and IE-1 (b) proteins.....	108
Figure 4.9	Point mutation of the TATA motifs and the RNA start site of the TnSNPV <i>me-53</i> promoter region.....	116

List of Tables

Table 2.1	The N-terminal net negative charge regions of eleven baculovirus IE-1s.....	45
Table 2.2	Identity values of the amino acid sequences of eleven baculovirus IE-1s.....	46
Table 2.3	Comparison of full length (a) and the C-terminal two third portions (b) of ten deduced NPV IE-1s with the TnSNPV homologue.....	47
Table 4.1	Identity values of the amino acid sequences of twelve baculovirus ME-53s.....	106
Table 4.2	Amino acid sequences of eleven NPV IE-1s compared with the TnSNPV IE-1.....	106

CONTENTS

Declaration	I
Acknowledgements	II
List of Figures	IV
List of Tables	VI
Chapter 1 General Introduction	1
1.1 Introduction to Baculovirus.....	2
1.2 Taxonomy and Structure.....	3
1.3 Infection Cycle.....	6
1.4 Genome Organization.....	9
1.5 Gene Regulation, Expression and DNA Replication.....	12
1.6 Genetic Engineering and Expression Vectors of Baculoviruses.....	16
1.7 <i>Trichoplusia ni</i> and TnSNPV.....	18
1.8 Scope of the Thesis.....	19
Chapter 2 Identification, Sequence Analysis and Phylogeny of the Immediate Early Gene 1 of the <i>Trichoplusia ni</i> Single Nucleocapsid Polyhedrosis Virus	25
2.1 Abstract.....	26
2.2 Introduction.....	27
2.3 Materials and Methods.....	31
2.3.1 Insects, virus isolation and purification.....	31
2.3.2 DNA extraction, manipulation and sequencing.....	32

2.3.3	Computer analysis.....	33
2.4	Results.....	34
2.4.1	Nucleotide and protein sequence analyses.....	34
2.4.2	Comparison and phylogenetic analysis of IE-1 proteins.....	46
2.5	Discussion.....	49
Chapter 3	Experimental Prediction of a DNA-Protein Binding Region of the <i>Trichoplusia ni</i> Single Nucleocapsid Nucleopolyhedrovirus IE-1.....	58
3.1	Abstract.....	59
3.2	Introduction.....	60
3.3	Materials and Methods.....	62
3.3.1	Prediction of the C-terminal secondary structures of NPV IE-1s.....	62
3.3.2	Plasmid constructs.....	63
3.3.3	Protein expression and purification.....	65
3.3.4	Western blot assay.....	66
3.3.5	Radio-labeling of the probe and electrophoretic mobility shift assay.....	67
3.4	Results.....	68
3.4.1	Prediction of the secondary structures and the experimental design....	68
3.4.2	Expression, purification and detection of the AcNPV and TnSNPV IE-1 truncations.....	71
3.4.3	Gel shift assay.....	73
3.5	Discussion.....	75

Chapter 4	Identification and Characterization of a Major Early-Transcribed Gene of <i>Trichoplusia ni</i> Single Capsid Nucleopolyhedrovirus Using the Baculovirus Expression System.....	80
4.1	Abstract.....	81
4.2	Introduction.....	82
4.3	Materials and Methods.....	85
4.3.1	DNA manipulation of the TnSNPV <i>me-53</i> , sequencing and Computer analysis.....	85
4.3.2	Cells and virus.....	87
4.3.3	Generation of the recombinant baculoviruses.....	88
4.3.4	Total RNA isolation, 5' and 3' RACE and RT-PCR analyses of the TnSNPV <i>me-53</i> transcripts.....	89
4.3.5	Phase contrast and fluorescence microscopy.....	94
4.4	Results.....	94
4.4.1	Sequence analysis of the TnSNPV <i>me-53</i>	94
4.4.2	Infection of the recombinant viruses and GFP expression.....	99
4.4.3	Transcription analyses of the TnSNPV and authentic <i>me-53s</i> of AcBacTnme53GFP.....	101
4.4.4	Analysis of the 5' non-coding regions of the NPV <i>me-53s</i>	103
4.4.5	Comparative phylogenetic analysis of IE-1 and ME-53 proteins.....	105
4.5	Discussion.....	108
4.5.1	Amino acid sequence analysis of the TnSNPV ME-53.....	108

4.5.2	Accurate transcription initiation of the TnSNPV <i>me-53</i> Promoter.....	110
4.5.3	Early transcriptional analysis of the TnSNPV <i>me-53</i>	111
4.5.4	Early regulatory elements of the baculovirus <i>me-53s</i> and strategy for determination of the functional role of the TATA motifs in the TnSNPV <i>me-53</i> promoter region.....	113
4.5.5	Late transcription analysis of the TnSNPV <i>me-53</i>	118
4.5.6	Late transcription initiation element of the NPV <i>me-53s</i>	119
4.5.7	Host factor binding sites of the NPV <i>me-53s</i>	120
4.5.8	Phylogenetic analysis of the TnSNPV ME-53.....	121
4.5.9	A potential model system for improving bio-insecticidal agents...	123
Chapter 5	Summary and Concluding Remarks.....	125
	References.....	132
	List of Publications.....	171

Chapter 1

General Introduction

1.1 Introduction to Baculoviruses

Baculoviridae are a diverse family of insect virus pathogens. They are infectious for arthropods but the majority of the infecting hosts are insects of the order Lepidoptera (Blissard, 1996). Some baculoviruses have also been isolated from the other insect orders Hymenoptera, Diptera and Coleoptera, as well as from the crustacean order Decapoda (Rohrman, 1992). Although over 800 species of insects are infected by baculoviruses, individual isolates normally show a limited host range and infect only closely related insects within a single order (Ijkel, 2001). Due to the characteristic of causing epizootic in nature (the viruses infect insect hosts as an epidemic does to humans), which may play a role in controlling insect population, baculoviruses have attracted more and more attention in scientific research and become more popular as biological control agents. Unlike wide-spectrum chemical insecticides with disadvantages such as non-specific insect targets, causing chemical resistance in insects and environmental damage, baculoviruses have provided an ecological advantage as an alternative to chemical insecticides with little direct effect on human health, hazard to other vertebrates and non-resistant to insects (Persley, 1996).

Although baculoviruses have potential as biological insecticides, they also have limitation in insect control. Many attempts have been made to improve molecular properties of baculoviruses as biological pesticides for pest control. The technologies for manipulation and genetic engineering of the viral genome have been successfully used to improve molecular properties of baculoviruses as biological pesticides and also allow to

develop baculovirus expression vectors which have been widely used for production of recombinant protein of pharmaceutical interest as well as for basic and applied biomedical and veterinary research purposes (Possee, 1997; Kost and Condreay, 1999).

To date, only some of baculoviruses have been identified. The most intensively studied baculovirus is *Autographa californica* nucleocapsid polyhedrosis virus (AcMNPV). A wealth of information is available for this virus and other identified baculoviruses. In recent years a South African *Trichoplusia ni* single capsid nuclear polyhedrosis virus (TnSNPV) isolate has been partially characterized, but not much is known about the virus, especially for its gene regulation. This chapter will provide some background information concerning the biology and molecular genetics of baculoviruses in general and TnSNPV in particular.

1.2 Taxonomy and Structure

The family Baculoviridae is divided into two taxonomic genera: Nucleopolyhedrovirus (NPV) and Granulovirus (GV) (Murphy *et al.*, 1995), which is based on the number of virions occluded in a protein matrix called the occlusion body, polyhedron or granulum. In NPVs several or as many as 200 virions are found in one polyhedron. In GV a single or at most a few virions are occluded within a granulum. Polyhedron and granulum mainly consist of the two proteins, respectively, named polyhedrin and granulin. Polyhedrin and granulin are structurally and functionally related proteins. NPVs have two morphotypes, one referred to as single enveloped nucleocapsid NPV (SNPV) and other multiple enveloped nucleocapsid NPV (MNPV). In the latter, up to nine nucleocapsids

are assembled in a single virion envelope before several of these packages are occluded into a polyhedron (van Strien, 1997). In contrast, GVs contain a single nucleocapsid per envelope. The other feature of NPVs is that they have two virion phenotypes. The virions released from the occlusion bodies are termed occlusion-derived virus (ODV), while others produced in epithelial cells of the midgut and found in the tracheal system and hemocoel of infected insect host is termed budded virus (BV). The structural difference between ODV and BV is shown in Fig. 1.1.

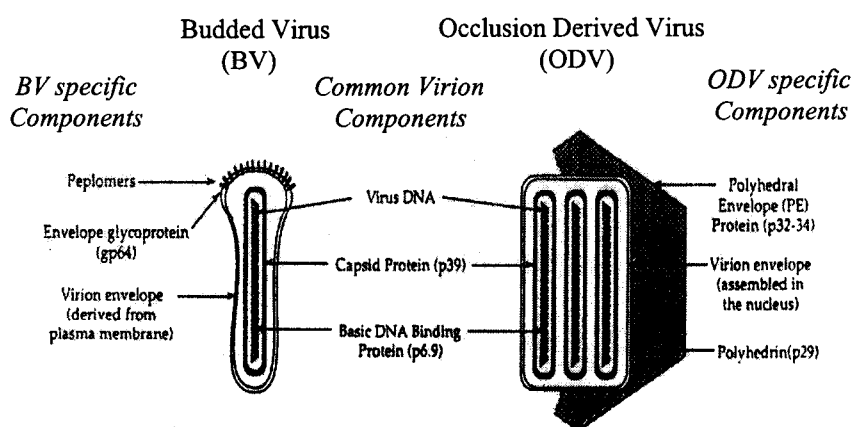


Figure 1.1 Structural diagram of the two baculovirus phenotypes (Blissard and Rohrmann, 1990). The locations of relevant virion components are indicated by arrows and solid lines.

To date, twelve different baculovirus species have been classified into NPV genus and five to GV genus (Blissard *et al.*, 2000). All the baculoviruses are named after the host from which they were first isolated. To establish the identity of baculovirus species, individual genes such as polyhedrin/granulin and DNA polymerase (Zanotto *et al.*, 1993; Cowan *et al.*, 1994; Bluach *et al.*, 1999; Bideshi *et al.*, 2000) have been used to study genetic relatedness of baculoviruses in the recent years. The phylogenetic analyses of

several other genes also contributed to infer baculovirus phylogenies. These analyses proposed that baculoviruses were separated into two groups, NPVs and GVs, while NPVs were further subdivided into group I and II. However, phylogenies based on different genes are often in conflicts with one another (Herniou *et al.*, 2001). With availability of increased complete genome sequence data, the different approaches such as genome sequence comparison, gene content and gene order have been used to infer baculovirus and other virus phylogenies (Hannenhalli *et al.*, 1995; Hu, Z. H., 1998; Montague and Hutchison, 2000; Herniou *et al.*, 2001, 2003). Herniou *et al.* (2001) combined three analyses based on trees constructed from data sets of gene order, gene content and a large number of individual genes shared by nine baculoviruses, to study baculovirus phylogenies. Although the latter has been shown to be the most robust approach for inference of baculovirus phylogenies, as suggested by Herniou *et al.* (2001), two other different data sets that are phylogenetically informative should be used to support the combined gene analysis. More recently, reconstruction of baculovirus phylogenies was conducted by Herniou *et al.* (2003). The data set generated from a core set of 30 genes conserved among thirteen baculovirus genomes, including the dipteran NPV (*Culex nigripalpus* NPV: CuniNPV), was used to yield a majority rule consensus tree with four major groups, the GV, group I and II lepidopteran baculovirus NPV and the dipteran NPV (CuniNPV). This phylogeny was supported by analyses of gene order and gene content. The comparative genomic approaches have provided correct resolution to baculovirus phylogenies and will give insights into the understanding of baculovirus biology, host-interaction and evolution (Herniou *et al.* 2003).

1.3 Infection Cycle

In the baculovirus life cycle, two structurally distinct phenotypes play specific roles in the infection process (Fig. 1.2). Transmission of baculovirus in an insect population occurs via either ingestion of food or other sources (soil, water, or crevices of plants) contaminated with occlusion bodies (OBs) in the environment (Blissard *et al.*, 2000). During the ingestion the OBs rapidly dissolve in the alkaline environment of the larval midgut and then released enveloped virions termed 'occlusion-derived virus' (ODV) enter the host cells by fusion of the virion envelope with microvilli of midgut columnar epithelial cells (Granados and Lawler, 1981; Horton and Burand, 1993). After fusion process, the nucleocapsids (NCs) are released into the cytoplasm, transported into, and uncoated in the nucleus where viral gene expression, DNA replication and assembly of progeny NCs occur. The newly assembled progeny nucleocapsids then bud through the nuclear membrane and subsequently through plasma membrane on the basal side of the epithelial cells into hemocoel or tracheoblast, becoming budded virions (BVs). During the budding process, nucleocapsids become enveloped by the plasma membrane where a glycoprotein, *gp64*, is incorporated (Volkman *et al.*, 1984; Blissard and Rohrmann, 1989; Oomens *et al.*, 1995). This protein is required for efficient budding (Monsma *et al.*, 1996; Oomens and Blissard, 1999). When BVs spread through tracheal systems and hemolymph to other tissues (fat body, muscle, salivary gland, gonads and cuticular epithelium, etc.) and enter other cells by adsorptive endocytosis, a secondary infection and viral DNA replication occur. The larvae insect tracheal system was found to be mainly responsible for virus systemic infection (Engelhard *et al.*, 1994). Its lymph system and

epidermal feet (cytoplasmic extensions, which is interdigitated to form the contact among tracheal epidermis) are likely to facilitate the rapid systemic spread of virus because the lymph channels and the epidermal feet could facilitate tissue-specific cell communication over long distances (Locke, 1985), while tracheoblasts (newly produced daughter cells of the tracheal epidermis on tracheoles) facilitate immediate secondary infection by crossing the basal lamina barriers and serve as principle carriers to deliver virus to the insect tracheal epidermal cells and to hemocoel, because tracheoblasts directly contact the tracheal epidermis and are also closely associated with other tissues. BVs are essential for systemic infection of insect hosts (tissue to tissue and cell to cell) (Keddie *et al.*, 1989; Granados and Lawler, 1981), while the larvae insect tracheal system is mainly involved in the process. It provides not only a productive target tissue but also a conduit for facilitating movement of the virus from midgut to other tissues (Engelhard *et al.*, 1994).

The ODVs, the virions formed later in the infected cells, are not released by budding. They are enveloped *de novo* in the nucleus and later occluded with polyhedrin (polyhedra-derived virus, PDV). Typically, occlusion bodies are first observed at 24 h p.i. (Blissard and Rohrmann, 1990). Except for being responsible for the initiation of an infection in the insect, the ODV provides protection for the virus against harsh environment and can allow the virus to remain viable outside of the hosts for years. Although the two virion phenotypes (BV and ODV) appear to be genetically identical (Smith and Summer, 1978; Blissard, 1996), they differ in morphology, protein composition, source of virion envelopes (Fig. 1.1), tissue specificity and mode of viral entry into host cells.

It is not until the final stage of infection that infected larvae become sluggish and cease feeding. The disintegration of the larvae occurs and is enhanced by two virus-encoded enzymes, chitinase and cathepsin. Both of them are associated with liquefaction of the host at the end of the infection process. The insect cuticle, the structure of which is formed through chitin crosslinked with sclerotized proteins, must be degraded in this process. Cathepsin may trip the protein from cuticular chitin so that chitinase can act on the naked chitin and then cuticle degrades (Ohkawa, et al., 1994; Hawtin, *et al*, 1995; Hawtin, *et al*, 1997; Slack *et al.*, 1995; Hill, *et al.*, 1995). After cells and tissues disintegrate, OBs are released into the environment and the next infection cycle begins as occlusion bodies are ingested by other susceptible insects.

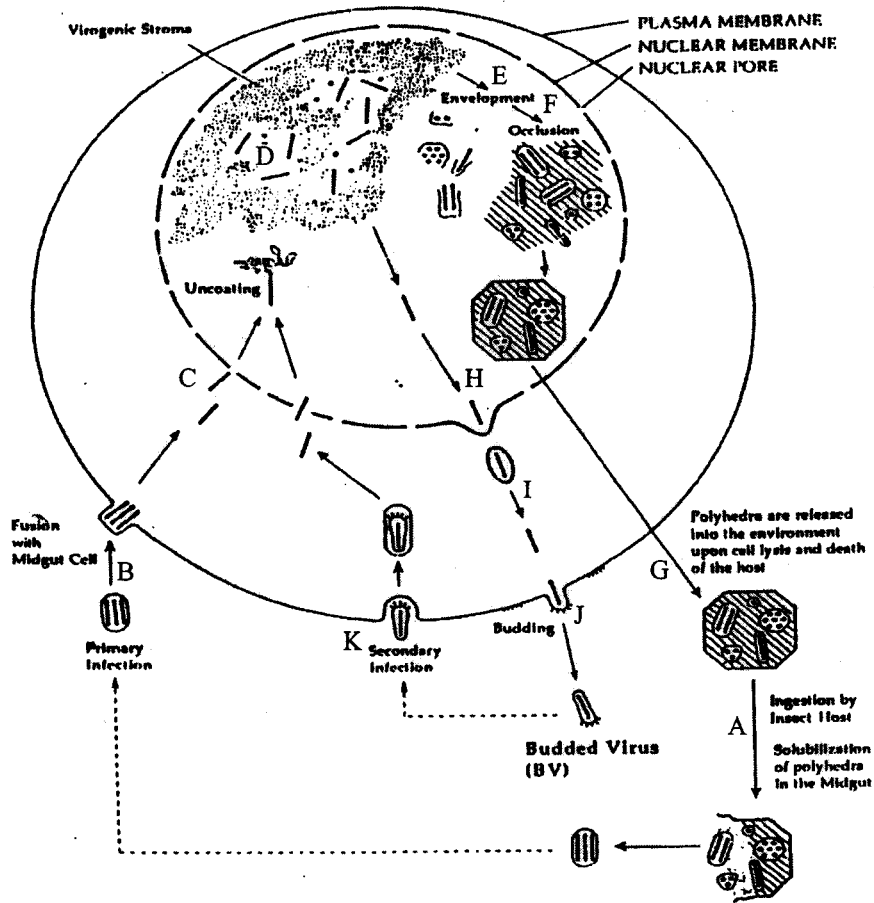


Figure 1.2 Schematic diagram of the baculovirus life cycle (from van der Beek, 1980). Ingested polyhedra (OB) are solubilized in the midgut and virions are released (A). Virions of the occlusion-derived virus (ODV) phenotype are released and enter midgut epithelial cells by fusion with microvilli (B). Nucleocapsids are transported to the nucleus and uncoated, followed by gene expression and viral replication (C). Progeny nucleocapsids are synthesized and assembled in the virogenic stroma (D). Following envelopment in the nucleus (E), progeny nucleocapsids go through the budding process (H, I, J). Budded virions infect other host cells by endocytosis (K). Nucleocapsids produced in the late stages of an infection are occluded within polyhedrin protein (F) and become matured form. Upon insect death and cell lysis, occlusion bodies are released into the environment (G).

1.4 Genome Organization

Baculoviruses have circular, supercoiled, double-stranded DNA genomes that vary in size from 90 to 180 kb depending on the species. Until now, genome sequences of ten lepidopteran NPVs, a dipteran NPV and four GV's are available. They are: *Autographa californica* NPV (NC001623) (Ayres *et al.*, 1994); *Bombyx mori* NPV (NC001962) (Gomi *et al.*, 2001); *Orgyia pseudotsugata* NPV (NC001875) (Ahrens *et al.*, 1997); *Helicoverpa armigera* NPV (NC003094) (Chen *et al.*, 2001); *Helicoverpa zea* NPV (AF334030) (Chen *et al.*, 2002); *Spodoptera litura* NPV (NC003102) (Pang *et al.* 2001); *Spodoptera exigua* NPV (NC002169) (Ijkel *et al.*, 1999); *Lymantria dispar* NPV (NC001973) (Kuzio

et al., 1999); *Mamestra configurata* NPV A (NC003529) (Erlandson *et al.*, 2002); *Epiphyas postvittana* NPV (NC003083) (Hyink *et al.*, 2001); *Culex nigripalpus* NPV (AF403738) (Afonso *et al.*, 2001); *Xestia c-nigrum* GV (NC002331) (Hayakawa *et al.*, 1999); *Plutella xylostella* GV (NC002593) (Hashimoto *et al.*, 2000); *Phthorimaea operculella* GV (NC004062) (Croizier *et al.*, unpublished); *Cydia pomonella* GV (NC001623) (Luque *et al.*, 2001); The sizes of these genome range from 101(PxGV) to 179 (XcGV) kb. The number of potential open reading frames (ORF) differs from 109 (CuniNPV) to 181(XcGV) in the different genomes.

Although genome arrangements among either group I NPVs (AcMNPV, BmNPV and OpMNPV) or GVs are similar (Blissard and Rohrmann, 1990; Hu *et al.*, 1998; Herniou *et al.*, 2003), gene order, gene content and gene conservation among different baculovirus groups have shown a high degree of variability. The comparison of genomic sequences of either 9 or 13 baculoviruses (Herniou *et al.* 2001; 2003) revealed that gene order is quite different among different baculovirus groups, and even among group II NPVs, while gene content showed that 17 genes are only found in group I NPVs, 14 genes specific to lepidopteran NPVs and 27 to GVs. These studies indicated that 63 genes are conserved among GVs, group I and II NPVs, while the dipteran NPV (CuniNPV) only shares 30 genes with three other group baculoviruses. The differences of genomes in different data sets reflect differences of both biological features and evolutionary relatedness among baculovirus groups (Ijkel, 2001; Herniou *et al.*, 2003). During the course of baculovirus evolution, some genes have been lost and some acquired, and extensive genome rearrangements have occurred. Although these changes occurred during the baculovirus

evolution, the cluster of four genes were preserved in the same relative position in all the sequenced genomes. Two of them (*helicase* and *lef-5*) are essential for baculovirus replication and transcription, and the functions of two others are unknown. The conservation of the gene order suggested that some essential functional requirements in the virus infection cycle may be required to preserve the gene arrangement (Herniou et al., 2003). With the growth of knowledge in baculovirus gene function, differences of gene conservation and genome rearrangements among baculoviruses will be better understood.

Among baculoviruses the AcMNPV is most intensively studied. It has a genome size of 133,894 bp, encoding 154 proteins (Ayres et al., 1994). Several homologous regions (*hrs*) with repeat sequences are interspersed at specific sites along the length of the genome. Hrs are believed to be origins of DNA replication and serve as enhancers of gene expression. They are well conserved among all the baculoviruses, including CuniNPV (Afonso, et al., 2001), but excluding CpGV which does not have the homologues of baculovirus typical *hrs*. However, several different classes of repeats were found in the CpGV genome, the functions of which have not been identified (Luque et al., 2001). Although the number of *hrs* varies among different baculovirus species, ranging from 4 (PxGV) to 13 (LdMNPV) (Hayakawa et al., 2000), with the exception of CpGV, they are generally presented in baculovirus genomes, especially in NPVs, and therefore, can be considered as a characteristic feature of baculoviruses (Chen, 2001; Ijkel, 2001).

1.5 Gene Regulation, Expression and DNA Replication

Baculovirus gene expression is temporally regulated in a cascaded manner in which each successive phase is dependant on the previous one (Blissard and Rohrmann, 1990). Three main classes of genes are recognized: early, late and very late (Blissard and Rohrmann, 1990; 1991). Early genes are transcribed by host RNA polymerase II, while late genes by a viral encoded RNA polymerase (Guarino *et al.*, 1998). The typical eukaryotic consensus transcription motifs, such as a TATA box and a CATG motif, usually function as core elements of baculovirus early promoters. TATA motif plays a critical role in regulating the rate of transcription and selecting the RNA start site for early gene transcription, especially in the absence of CAGT motif (Dickson and Friesen, 1991; Guarino and Smith, 1992; Blissard *et al.*, 1992), while CAGT motif not only serves as early transcription start site for many early genes but also as an initiator in the case of the AcMNPV *ie-1* (Pullen and Friesen, 1995). Unlike TATA and CAGT motifs responsive to host factors in uninfected cells, an unusual promoter motif, CGTGC, positioned at the RNA start site of *dnapol* or *p143*, is inactive in uninfected cells but highly responsive to transregulator IE-1 for *dnapol* (Ohresser *et al.*, 1995) and to IE-1, IE-2, PE38 and *cis*-liked *hr5* enhancer for *p143* (Lu and Carstens, 1993). In addition to the early promoter motifs, two *cis*-acting elements, CGT and GC, presented in upstream activating regions (UAR) of *p35*, *39k* and *gp64*, stimulate basal promoter activities from these genes (Dickson and Friesen, 1991; Guarino and Smith, 1992; Kogan and Blissard, 1994). Another *cis*-acting motif in UAR is GATA that was identified as a nuclear factor binding site in host insect cells. Mutation analysis of the motif resulted in reduction of the

promoter activity in the case of *gp64*, but not in *pe38* (Krappa et al., 1992; Kogan and Blissard, 1994). As observed in GATA, the mutation of the *cis*-acting CACGTG motif in UAR, a host factor binding sequence, eliminated host factor binding and reduced *gp64* promoter activity in the mutagenesis assay (Kogan and Blissard, 1994). This reflects a correlation between host factor interaction and transcriptional stimulation. Until now, four very early genes have been identified as transregulators: *ie-1*, *ie-0*, *ie-2* and *pe38* (Friesen, 1997). The product of *ie-1* (IE-1) is the principle transactivator of baculovirus gene expression. It stimulates transcription of promoters for the *39k*, *p35*, *gp64*, *p143*, *dnapol*, *pe38*, *lef-1*, *lef-2* and *lef-3* genes and also its own promoter (Pullen and Friesen, 1995; Ribeiro *et al.*, 1994; Lu and Carstens, 1993; Passarelli and Miller, 1993; Guarino and Smith, 1992; Blissard and Rohrmann, 1991; Kovacs *et al.*, 1991; Theilmann and Stewart, 1991; Nissen and Friesen, 1989). Moreover, IE-1 negatively regulates expression of *ie-0*, *ie-2* and *p38* genes (Kovacs *et al.*, 1991; Leisy *et al.*, 1997). In addition, immediate early gene 1 (*ie-1*) is indirectly and directly involved in DNA replication by stimulating late expression factors (*lefs*) and binding to *hr* elements which are thought to serve as origins of viral DNA replication (Kool *et al.*, 1993; Lu and Miller, 1995; Lu *et al.*, 1997). IE-0 is thought to be the only baculovirus gene that can be spliced. The AcMNPV IE-0 transactivates early *39K* only in the presence of the *cis*-linked enhancer (Kovacs *et al.*, 1991), while the OpMNPV and LdMNPV IE-0s can activate expression of the gene (*39K*) in the enhancer-dependent or -independent manner, although in the latter case the transactivation level from a *39K*-GUS reporter cassette, mediated by the IE-0, is comparatively low in the absence of the *hr* enhancer (Pearson and Rohrmann, 1997; Theilmann *et al.*, 2001). The AcMNPV IE-0 stimulated expression

from *ie-1* gene but was not autoregulatory (Kovacs et al., 1991), while the OpMNPV IE-0 can autoregulate its own promoter (Theilmann et al., 2001). IE-2 transactivates *39k*, *ie-0* and *ie-1* genes, although IE-1 negatively regulates *ie-2* (Leisy et al., 1997) and autoregulates its own promoter (Carson et al., 1991; Theilmann and Stewart, 1993). PE38 stimulates transcription of *p143* gene (Lu and Carstens 1993) and augments apoptosis induced by IE-1 (Prokhod'ko and Miller, 1999).

Unlike the early gene expression prior to the initiation of viral DNA replication, the late gene expression occurs during and after DNA replication. A virus-encoded RNA polymerase is responsible for transcription of late and very late genes. The late and very late promoter classes contain a consensus A/G/T(TAAG) sequence that serves as the late transcription initiation site. Only limited sequence beyond a few nucleotides surrounding the TAAG motif is required for late promoter activity (Morris and Miller, 1994; Ooi et al., 1989; Rankin et al., 1988). Eighteen late expression factors (*lefs*), *ie-1* *ie-2*, *lefs 1-11*, *dnapol*, *p143*, *p43*, *p35* and *39K*, were previously identified (Lu and Miller, 1995). They are required for optimal transactivation of expression from late (*vp39*, major capsid protein, and *p6.9*, the basic core protein) and very late (*polh* and *p10*) promoters (Todd et al., 1995, 1996; Lu and Miller, 1994; Morris et al., 1994; Passarelli et al., 1994; Li et al., 1993; Passarelli and Miller, 1993a,b, 1994). Deletion of any of the *lef* clones from the *lef* library in transient expression assay dramatically reduced expression level from the *vp39* promoter (Lu and Miller, 1995). Recently, a 19th late expression factor (*lef-12*) has been reported to be necessary for expression of a reporter gene under control of late *vp39* and very late *polh* promoters (Rapp, et al., 1998). In addition, each of four *lefs*, *p47*, *lef-4*, *lef-*

8 and *lef9* were identified to be one of four subunits of the virus-encoded DNA-direct RNA polymerase (Guarino et al., 1998). Apart from 19 *lefs*, very late expression factor-1 gene (*vlf-1*) was observed in transient assay to stimulate expression from very late *polh* and *p10* promoters, but has little or no effect on the late *vp39* and *p6.9* promoters (Todd et al., 1996). This suggested that *vlf-1* may be the major transactivator of very late gene expression. The two very late virus proteins, polyhedrin and *p10*, are expressed during the process of OB formation and ultimately ends at cell lysis. Their strong promoters are extensively utilized in baculovirus insect cell expression system (Smith et al., 1983; Martens et al., 1995).

The six *lefs*, *ie-1*, *lef 1-3*, *p143* and *dnapol* are essential for transient DNA replication (Kool et al., 1994) and the *lef 11* was also reported to be one of the essential genes for viral DNA replication (Lin and Blissard, 2002). Five other *lefs*, *ie-2*, *lef-7*, *p35*, *dnapol* and *pe38* stimulated transient DNA replication (Todd et al., 1995; Lu and Miller, 1995; Ahrens et al., 1996; Kool et al., 1995; Lu et al., 1997). Some replication-associated *lefs*, such as *dnapol*, *lef-3* and *p143*, are directly involved in viral DNA replication, and some are probably indirectly involved in the replication process, such as *ie-2* and *pe38*, by stimulating expression of other *lefs* (Lu et al., 1997), while IE-1 is indirectly and directly involved in viral DNA replication by transactivating *lef-3*, *lef-7*, *p147*, *dnapol*, *p35* and *ie-2* promoters and binding to origins (*hrs*) of viral DNA replication (Lu and Miller, 1997).

Homologous regions (*hrs*) are present in all the baculovirus genomes. Evidence suggests that *hrs* act as enhancers to stimulate transcription of early genes in *trans* and *cis* (Guarino and Summers, 1986; Guarino et al., 1986; Nissen and Friesen, 1989; Carson et al., 1988; Lu and Carstens, 1993), and serve as origins (*oris*) of DNA replication (Kool et al., 1993a,b; Pearson et al., 1992). They also indirectly contribute to late gene expression (Lu and Miller, 1995). IE-1 binds to the *hr* through interaction with the 28-bp imperfect palindrome. This is the minimal sequence required for IE-1-mediated *hr* enhancer activity (Rodems and Friesen, 1995) and plasmid DNA replication (Pearson et al., 1992; Leisy et al., 1995). Current evidence indicates that IE-1 interact with the 28-mer as dimer (Rodems and Friesen, 1995; Rodems et al., 1997). The IE-1 oligamerization and DNA binding are required for its transactivation function (Olson et al., 2001) and the binding of IE-1 to the *hr* may directly contribute to DNA replication (Lu et al., 1997).

1.6 Genetic Engineering and Expression Vectors of Baculoviruses

In nature the wild types of baculoviruses have some disadvantages of being commercialized. One of the major limitations is relatively slow speed of action, typically far less virulent in later star insects which cause the majority of damage to crops (Evans, 1981). A relatively narrow host range is also a drawback. Genetic engineering of baculoviruses may improve baculovirus properties as insecticidal agents for pest control. Many studies have been conducted to enhance recombinant baculovirus insecticidal efficacy. Two main approaches are applied to explore the possibilities. Deletion of the viral *egt* gene (an ecdysteroid UDP-glucosyltransferase), the function of which is to either delay or block insect molting and pupation, and the introduction of a scorpion toxin

gene and a mite toxin gene can increase viral killing speed to some extent. (O'Reilly and Miller, 1991; Slavicek *et al.*, 1999; Stewart *et al.*, 1991; Maeda *et al.*, 1991; McCutchen *et al.*, 1991; Gershburg *et al.*, 1998; Tomarski and Miller, 1991; 1992). However, over-expression of insect genes in recombinant baculoviruses, such as a modified juvenile hormone esterase gene (JHE), did neither significantly increase the baculovirus speed of action nor reduce the food consumption of the infected insects (Hammock *et al.*, 1990). The different promoters, either from a powerful late polyhedrin and a p10 promoter or an early promoter, have been studied for improvement of the insecticidal activity (King and Possee, 1990; O'Reilly *et al.*, 1992; Luckow, 1991; Jarvis *et al.* 1996a; 1996b). The effectiveness of the genetically modified baculoviruses has also been tested in the field and their potential to enhance crop protection has been reported (Cory *et al.*, 1994; Gard, 1997; Hernandez-Crespo *et al.*, 1999). The AcMNPV and HaSNPV recombinants with an *egt*-deletion and an insertion of the insect-specific neurotoxin (*AalT*) had shown a faster killing speed over their wild types and the *egt*-deletion forms respectively. (Black *et al.*, 1997; Chen *et al.*, 2000;). The recombinants thus exhibit a possible commercial potential as bio-insecticides.

Other application of baculoviruses is to produce recombinant proteins for biotechnological and pharmaceutical interests by baculovirus expression vectors (Possee, 1997; Krost and Condreay 1999). The baculovirus-derived vectors have also been used to transfer genes into mammalian cells (Hofmann *et al.*, 1995; Krost and Condreay, 1999; Sarkis *et al.*, 2000) and it thus could be a possible tool for gene therapy. The polyhedrin or p10 locus are usually adopted to introduce the foreign genes as both genes have a very

strong promoter and are not essential for infectivity of NPVs in cultured insect cells (King and Possee, 1992; O'Reilly *et al.*, 1992; Luckow, 1991).

1.7 *Trichoplusia ni* and TnSNPV

Trichoplusia ni (*T.ni*) (Lepidoptera: Noctuidae) is commonly referred to as cabbage looper. The average generation time of *T.ni* is between 26-40 days. Following mating, females lay between 200-300 eggs on the upper or lower surfaces of upper canopy leaves. Once eggs hatch, the pale green larvae feed actively on veins on the underside of lower leaves for 14-21 days, after which they pupate. Nine days later, after initial pupae develop, an adult moth emerges. Mature moths feed on plant nectar and can live for up to 24 days (Adlerz, 1971; Kishaba *et al.*, 1976; Marsden, 1979; Hofmann and Fordsham, 1993).

T.ni is a polyphagous crop pest distributed throughout Africa, North America, Hawaii, Europe and Asia. The larvae cause damage on cabbage, spinach, sugarbeets, peas, celery, potatoes, alfalfa, beans, tomato, certain ornamental plants and mint, etc. Although *Trichoplusia ni* has been successfully controlled with synthetic chemical pesticides, increased resistance to these insecticides such as *Bacillus thuringiensis* (Bt) has been reported (Dornan *et al.*, 1995). The combined use of traditional pheromone releasers and black light traps has proven inadequate in crop protection (Debolt *et al.*, 1979). Thus, an effective long-term pest control agent is urgently needed.

Single nucleocapsid polyhedrosis virus is one of morphotypes of baculoviruses. In this morphotype viruses are packaged as a single nucleocapsid within each viral envelope. To date, several single nucleocapsid polyhedrosis viruses were either fully sequenced or under investigation. Since TnSNPV was cultured *in vitro* by Granados *et al.* (1986), a TnSNPV isolate has been characterized by Davis and Wood (1996). In recent years a South African TnSNPV isolate was isolated from a field population of *Trichoplusia ni* and partially characterized (Fielding, Ph.D. thesis, 2001). Although a previous attempt to construct a physical map of the entire TnSNPV genome was unsuccessful, an approximate genome size of 158 kb was determined by restriction endonuclease analyses (Fielding, Ph.D. thesis, 2001). The size is about 40 kb greater than those of the previous TnSNPV isolates (Bilimoria, 1983; David and Wood, 1996). Five genes within the TnSNPV had been fully sequenced. A few attempts made to phylogenetically define the relation of the virus to other baculoviruses have suggested that the TnSNPV belong to group II NPVs (Wang *et al.*, 2001; Fielding and Davison, 1999; Fielding *et al.*, 2002).

1.8 Scope of the Thesis

Along with the development of biological insecticides, many research efforts have been made in baculoviruses to investigate fundamental molecular aspects of these viruses, such as the function and regulation of genes, genome organization, mode of entry, DNA replication and virus factors that determine the host range and virulence. Previously, a South African TnSNPV isolate was partially characterized as a novel baculovirus. During the process of the characterization, a few late genes of the virus were identified. This thesis describes a molecular characterization of the TnSNPV early genes to gain insight

into the functional roles of these genes, their unique features and further determination of the placement of TnSNPV in baculovirus phylogeny.

In the cascade of baculovirus regulatory events, successive stages of virus replication are dependent on proper expression of genes within the preceding stages. Thus, critical to baculovirus replicative success is the appropriate expression and regulation of early genes (Friesen, 1997). Immediate early gene 1 (*ie-1*) is thought to be a principal transregulator, and conserved among baculoviruses. It potently stimulates transcription of promoters for some early and late genes as well as its own promoter. *Ie-1* gene is also directly or indirectly involved in DNA replication. The N-terminal half of IE-1 has been identified as a transactivating domain which has an acidic feature and is poorly conserved. The C-terminal half of the gene has a DNA-binding function and is comparatively better conserved. The sequence analyses of the TnSNPV *ie-1* gene will provide useful information for further understanding of the essential regulatory gene, as well as for phylogenetic classification of the TnSNPV due to its advantages over polyhedrin and other conserved late genes with small sizes which were previously used as models for phylogenetic study (**Chapter 2, 4**).

Current evidence suggests that IE-1 transactivation involves DNA-independent and *hr* enhancer-dependent mechanisms (Friesen, 1997). In a DNA-independent model, IE-1 interacts directly or indirectly with a transcriptional initiation complex without involvement of DNA binding in a sequence-specific manner. In a *hr* enhancer-dependent model the binding of IE-1 to the *hr* occurs through interaction with the 28-mer

palindromes within the *hr*, and IE-1 binds to the *hr* enhancer element as a dimer which is a requirement for IE-1 transregulatory function. The binding stimulates *hr* enhancer activity and may lead to transcriptional stimulation through direct and indirect protein contact with TFIID-associated factors of the preinitiation complex (Friesen, 1997). Although deletion and insertion studies on IE-1 binding have suggested that deletion of more than 25 amino acids at the C-terminus demolished the IE-1 DNA-binding capacity (Kovacs et al, 1992), and a few C-terminal residues are responsible for oligomerization of IE-1 (Rodems et al., 1997; Olson et al., 2001), which regions or domains of IE-1 directly bind to the *hr* enhancer element remains to be determined. Based on conservation of *hrs* in all the lepidopteran baculovirus genomes, it was thought that TnSNPV may have *hr* sequence homologs as presented in other baculoviruses. However, due to lack of entire TnSNPV genome information, the TnSNPV *hr* sequences or other possible substitutes of enhancer elements of the TnSNPV IE-1 could not be determined, and therefore it was decided to use AcMNPV IE-1 as a model to experimentally predict the existence of the possible direct binding region in the TnSNPV IE-1, based on conserved primary and secondary structures between both protein analogs (**Chapter 3**). This study will contribute to the further understanding of the functional role of IE-1 in DNA binding.

Another early gene was previously identified as a major early-transcribed gene. Its homolog was found in all the available baculovirus genomes. Although it was suggested that ME-53 may play a role in gene regulation, based on the presence of a zinc finger-like motif in its amino acid sequence, the regulatory role of the gene is unclear. Thus, identification and characterization of the TnSNPV *me-53* early gene will enrich

knowledge we have obtained only from its AcMNPV counterpart and give a better understanding of the common features around all the *me-53s* as well as the unique features of the TnSNPV *me-53*. In addition, the comparative phylogenetic analysis of ME-53s and IE-1s provides information for further defining phylogenetic relationships of TnSNPV among baculoviruses.

The second goal of this study is to test multiple foreign gene expressions under the control of heterologous early promoters in the context of AcMNPV genome. Although an insect-specific toxin gene (AaIT) under the control of a baculovirus early promoter was tried to improve the properties of baculovirus recombinant as a bio-control agent (Jarvis et al., 1996), multiple early expressions of insect-specific toxin genes have not been reported. The introduction of multiple toxins into baculoviruses, expressed early in infection may reduce insect feeding and, finally, result in feeding arrest or death quicker than those induced in the very late phase of infection, because different toxins may act on the different physiological pathway of larvae development and/or the same multiple toxins may be rapidly accumulated to a killing threshold dosage. However, baculovirus native early promoters are comparatively weaker than very late ones and therefore overexpression level of a toxin gene may be much lower than those which are under control of late strong promoters, such as *polh* and *p10* promoters. Selection of comparatively strong early promoters, such as *ie-1* or *gp64* promoters, can be taken into account. Associated with the promoter selection, enhancer elements can be applied to facilitate overexpression of selected early promoters. With respect to enhancement of early promoters, several enhancer-dependent early genes can be chosen for this purpose.

In addition, appropriate heterologous early promoters, such as enhanced CMV promoter (Clontech) or even hybrid early and late promoters, to drive expression of insect-specific toxins, can also be candidates in this regard. However, these possibilities still need to be tested in both laboratory and field.

It was reported that very late gene expression does not occur until all the virus-encoded factors become available during a very late phase of infection (Li, et al., 1993; Passarelli and Miller., 1993a,b,c; McLachlin and Miller, 1994; Passarelli, et al., 1994; Morris et al., 1994; Todd et al., 1995; Xu et al., 1995), and it is also undesirable for very late promoters to produce secreted proteins and glycoproteins because it appeared that cellular processing pathways may be compromised during the late stage of infection (Jarvis and Summers, 1989; Jarvis et al., 1990; Murphy et al., 1990). Although the expression of toxin genes from very late promoters appears to have such disadvantages for pest control purposes, the overexpression of insect-specific toxin genes controlled by strong very late promoters, *polh* or *p10*, showed a dramatic increase of killing speed of infected insect hosts which was faster than that of the wild type virus (McCutchen et al., 1991; Chen, et al., 2000). However, the recombinant virus carrying a single toxin gene under control of an immediate early promoter did not show promising results, although it showed the reduction of insect growth rate (Jarvis et al., 1996). The main reasons for this may be summarized as 1) a threshold dose for killing was not reached due to the accumulation of the toxin expressed by the early promoter which was slower and less than that of the very late promoters, specifically at the very late postinfection; 2) the duplication of the early promoter may slow down the viral replication cycle because the duplicate of the promoter

competes limited cellular transcription factors with its native counterpart; 3) the expression of the toxin (AaIT) in the midgut epithelial cells could impede efficient virus spread to other tissues in the infected insect host (Jarvis et al., 1996). Although expression of a single toxin gene under the control of an early promoter could not produce as much active impact on enhancing baculovirus recombinant killing speed of infected insects, as demonstrated in the case of a very late promoter, the strategy for early expression of multiple toxin genes may bring encouraging views into the challenging topic.

Due to insertion of two heterologous genes into the baculovirus expression vector system, the expression of these genes in the recombinant virus will provide evidence for developing a model system, possibly used for introduction of multiple insect-specific toxin genes under the control of early promoters without interfering with virus replication (**Chapter 4**). Thus, virulence and killing speed of the virus as a bio-insecticide may be dramatically boosted and improved.

Chapter 2

**Identification, Sequence Analysis and Phylogeny of the Immediate Early
Gene 1 of the *Trichoplusia ni* Single Nucleocapsid Polyhedrosis Virus**

2.1 Abstract

The immediate early gene 1 (*ie-1*) as a principle transactivator protein (IE-1) plays an essential role in gene regulation and DNA replication in the case of of *Autographa californica* multicapsid nuclear polyhedrosis virus (AcMNPV) and the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus (OpMNPV). In this study an *ie-1* homolog from *Trichoplusia ni* single nucleocapsid polyhedrosis virus (TnSNPV) was identified. It has a comparatively larger open reading frame (ORF), encoding a protein of 739 amino acids with a molecular mass of 84.5 kDa. Each promoter region of compared NPV *ie-1*s contains the TATA and CAGT/T motifs. Protein analysis showed that the N-termini of IE-1s are poorly conserved, but rich in acidic amino acids. Two negative charge dominant regions were identified in each N-terminus. The SwissProt database search showed that N-terminal region of the TnSNPV IE-1 significantly matched the SET-domain. On the other hand, the C-terminal regions of the deduced IE-1 proteins were much more conserved than the N-termini. Several conserved regions were found in C-termini. The TnSNPV *ie-1* shared the highest sequence homology with the *Spodoptera exigua* MNPV (SeMNPV) homologue in comparison with others. The phylogenetic analysis showed that TnSNPV is a member of the group II NPVs and it is most closely related to SeMNPV.

2.2 Introduction

Baculoviruses are a diverse group of insect viruses with circular double stranded DNA genomes ranging in size from 90kb to 180 kb (Herniou *et al.*, 2003). The most well-characterized baculovirus, AcMNPV encodes approximately 150 genes (Ayres, 1994). Gene expression of baculoviruses are transcriptionally regulated in a cascade manner, which includes *three* temporal phases: early, late and very late (Friesen, 1997). Early gene expression precedes viral DNA replication, dependent on host RNA polymerase II and does not require the presence of any viral gene products, while late and very late genes are transcribed by a virus-encoded RNA polymerase during or after the initiation of viral DNA synthesis (Blissard and Rohrmann, 1990; Guarino *et al.*, 1998). Some early genes such as *ie-0*, *ie-1*, *ie-2* and *pe38* play critical roles in the regulation of viral gene expression (Frisen, 1997).

The *ie-1* gene encodes a principal early transregulator protein (IE-1) and expresses at a very early stage of infection and remains detectable *throughout* infection (Frisen, 1997; Theilmann and Stewart, 1990). The early transcription of the AcMNPV, OpMNPV and LdMNPV *ie-1*s was initiated from a baculovirus consensus early promoter motif CAGT (Theilmann and Stewart, 1990; Pullen and Friesen, 1995; Van Strien *et al.*, 2001), but in the case of the AcMNPV *ie-1* the motif can also function as a late promoter (Pullen and Friesen, 1995). IE-1 stimulates the transcription of some early, late and its own promoters (Pullen and Friesen, 1995a; Ribeiro *et al.*, 1994; Lu and Carstens, 1993; Passarelli and Miller, 1993; Guarino and Smith, 1992; Blissard and Rohrmann, 1991; Kovacs *et al.*,

1991; Nissen and Friesen, 1989), and also negatively regulates *ie-0*, *ie-2* and *pe38* promoters (Kovacs *et al.*, 1991; Leisy *et al.*, 1997). IE-1 mediates the transcriptional activation of early promoters by enhancer-dependent and -independent mechanisms (Guarino and Summers, 1986b,c; 1987; Guarino *et al.*, 1986a; Nissen and Friesen, 1989; Blissard and Rohrmann, 1991; Blissard *et al.*, 1992). In the former when homologous region (*hr*) enhancer elements were cis-linked to the upstream of early promoters (*39K*, *p35* and *p143*), expression of reporter genes was dramatically stimulated by IE-1, while in the latter case the early promoters (*39K* and *gp64*) were transactivated by IE-1 in the absence of *hr* enhancer elements. *Ie-1* is one of five genes that are essential for transient DNA replication (McDougal and Guarino, 2000). It is indirectly and directly involved in viral DNA replication by transactivating some *lefs* and binding to *hr* regions (Lu and Miller, 1997; Lu *et al.*, 1997). In addition, IE-1 alone is sufficient to induce apoptosis in Sf21 cells, the function of which can be augmented by the involvement of *pe38* (Prihod'ko and Miller, 1996; 1999).

It is known that *ie-1* is the only baculovirus gene that can be spliced. Its spliced form is IE-0 which differs from IE-1 by addition of different numbers of amino acids to the N-termini of different baculovirus IE-1s (Kovacs *et al.*, 1991; Pearson and Rohrmann, 1997). Unlike AcMNPV and OpMNPV IE-1s essential for both transactivation and DNA replication, the LdMNPV IE-0 showed the dual functions in the Ld652Y cell line but the LdMNPV IE-1 was inactive. This suggested that splicing in the case of the LdMNPV IE-1 is important for producing an active transactivator that has the same functions as those of other unspliced IE-1s (Pearson and Rohrmann, 1997). It is possible that IE-1 and IE-0

are translated from the LdMNPV *ie-0* transcripts as in the case of the OpMNPV *ie-0* (Theilmann *et al.*, 2001). The function of the LdMNPV *Ie-1* has not been identified.

The previous studies on IE-1s showed that the amino acids at the N-terminal half of IE-1s are poorly conserved and rich in acidic features, while the C-terminal residues of IE-1s are more conserved (Theilmann and Stewart, 1990; van Strien *et al.*, 2000). Several domains have been identified as being involved in transactivation and DNA replication at the N-termini of IE-1s. One or two acidic transactivation domains were mapped in the different regions of N-termini of the AcMNPV or OpMNPV IE-1s, respectively (Slack and Blissard, 1997; Rodems *et al.*, 1997; Forsythe *et al.*, 1998), whereas in the case of the AcMNPV IE-1s it was suggested that two N-terminal regions with basic charges may be possibly involved in negatively regulating transcription from *ie-0*, *ie-2* and *pe38* (Slack and Blissard, 1997). More recently, amino acids 1 to 65 within the N-terminal acidic activation domain of OpMNPV IE-1 were identified to be essential for DNA replication but functionally separated from the transcriptional activation domain (Pathakamuri and Theilmann, 2002). The C-terminal half of IE-1 contains a few residues and a helix-loop-helix-like domain which are critical for IE-1 oligomerization and DNA binding. The binding of IE-1 to *hr* enhancer elements as a dimer is required for transactivation of early promoters (Rodems *et al.*, 1997; Olson *et al.*, 2001).

Previously, based on the morphology of the occlusion bodies, baculoviruses were classified into two genera: nucleopolyhedrosis virus (NPV) and granulosis virus (GV) (Murphy *et al.*, 1995; Blissard *et al.*, 2000). The phylogenetic relatedness of lepidopteran

NPVs have been further divided into group I and II based on a single gene (*polh*) phylogeny (Zanotto *et al.*, 1993). The analyses from a number of individual gene phylogenies also contributed to the classification (Chen *et al.*, 1999; Clarke *et al.*, 1996; Jin *et al.*, 1999; Kang *et al.*, 1998; Li *et al.*, 2001; Liu *et al.*, 1999). In the recent years with increase of available genome sequences, the classification has been clarified and analyzed in details by several combined analyses based on baculovirus genome information. Most recently, a dipteran NPV (*Culex nigripalpus* NPV, CuniNPV) has been identified and fully sequenced (Moser *et al.*, 2001; Afonso *et al.*, 2001). Herniou *et al.* (2003) suggested that baculovirus phylogeny can be reconstructed to four major groups based on a number of major differences of CuniNPV from other lepidopteran baculoviruses (Afonso *et al.*, 2001). In spite of the major change of baculovirus classification, the phylogenies of group II NPVs are not well resolved because of a lack of sufficient sequence data available in the divergent group. The sequence data of South Africa TnSNPV isolate will provide useful information to contribute to baculovirus phylogeny. Previous studies suggested that the TnSNPV is more closely related to group II NPVs based on phylogenies of a few late genes (Fielding and Davison, 1999; Fielding *et al.*, 2002). However, phylogenetic relationship in deep branches among TnSNPV and other baculoviruses has not yet been established. Identification of the TnSNPV *ie-1* homolog, which is conserved among all the lepidopteran baculoviruses, will contribute to the TnSNPV phylogeny.

In this study the TnSNPV *ie-1* nucleotide and deduced protein sequences were analyzed and discussed in detail. The eleven baculovirus IE-1 sequences were compared to identify

common features of the early regulatory gene. The phylogenetic analysis of TnSNPV has also been conducted for further establishment of the virus phylogeny among baculoviruses.

2.3 Materials and Methods

2.3.1 Insects, virus isolation and purification

The original virus isolate was prepared from diseased *Trichplusia ni* (Noctuidae: Lepidopteran) larvae collected from the Eastern Cape, South Africa (Fielding and Davison, 1999). Third star Larvae were fed on small plugs of an artificial lepidopteran diet, containing 1 µl purified polyhedral inclusion bodies (PIBs) per piece. After the entire diet plug was consumed, each of the infected larvae were transferred to fresh diet in glass containers and reared at 26°C and 65% humidity on a 12 hour day/night cycle until death. Polyhedral inclusion bodies (PIBs) were purified from virus-killed larvae following a modified procedure based on the previously described method (Miller and Dawes, 1978). Briefly, ten collected insect cadavers were homogenized in 10 ml H₂O within a pestle using a mortar, and the mixture was filtered through two layers of muslin cloth into a beaker. Debris was removed by centrifugation at 2000 RPM for 5 min. The supernatant was centrifuged at 15000 RPM for 10 min. The pellet was resuspended in 10 ml 0.1% SDS and centrifuged at the same speed for 10 min, and the step repeated with the same volume of 0.5M NaCl. The following pellet was resuspended with gentle shaking in 5 ml H₂O overnight. The 1.2 ml preparation was loaded onto a discontinuous sucrose gradient (40, 45, 50, 55 and 60 %) and centrifuged at 26500 RPM for 3 hrs.

Visible band at 55% sucrose layer was collected and diluted up to 25 ml with H₂O. PIBs were centrifuged at 15000 RPM for 10 min. The pellet was resuspended in 1 ml H₂O and store at 4°C.

2.3.2 DNA extraction, manipulation and sequencing

TnSNPV DNA was isolated from the occluded form of the virus (Fielding and Davison, 1999). PIBs (2×10^6 PIB/ml) were placed in 0.5M Na₂CO₃ alkaline solution at 37°C for 30 min. SDS was added to the solution to a final concentration of 0.1%, and PIBs were incubated at 60°C for 30 min. After cooling down, PIBs was digested by proteinase K with a final concentration of 500 µl/ml at 37°C for 1 hour (or until clear), followed by phenol-chloroform-isoamyl alcohol (25:24:1) and chloroform-isoamyl alcohol (24:1) extractions. Viral DNA was recovered by ethanol precipitation and the concentration of DNA was determined by measuring the OD₂₆₀ and OD₂₈₀ of the solution.

The virus genomic DNA was used to construct genomic libraries with *Pst*I and *Eco*RI. Ten of the 29 *Pst*I and 22 of 27 *Eco*RI fragments were cloned into pSK-Bluescript (Stratagene) (Fielding, 2001). As part of a preliminary attempt to map the genome, the ends of the *Eco*RI library were sequenced. Subsequent genome analysis led to the identification of the TnSNPV homolog. The *Ie-1* gene was found to be truncated with its 5' and 3' ends on a 2.3 kb and an 11 kb *Eco*RI fragment, respectively. The 11 kb and 2.3 kb DNA fragments were digested by exonuclease III (Henikoff, 1984) and then cloned into pSK-Bluescript vector. Appropriate sizes of fragments were selected and DNA templates were prepared for nucleotide sequencing. Sequencing was conducted using the

Sequitherm kit (Epicentre Technologies) using CY-5 labeled primers. Nucleotide sequence was resolved on an Alfexpress automated DNA sequencer (Pharmacia). Sequence was obtained in both the sense and anti-sense directions before the final sequence was confirmed (Fig. 2.1).

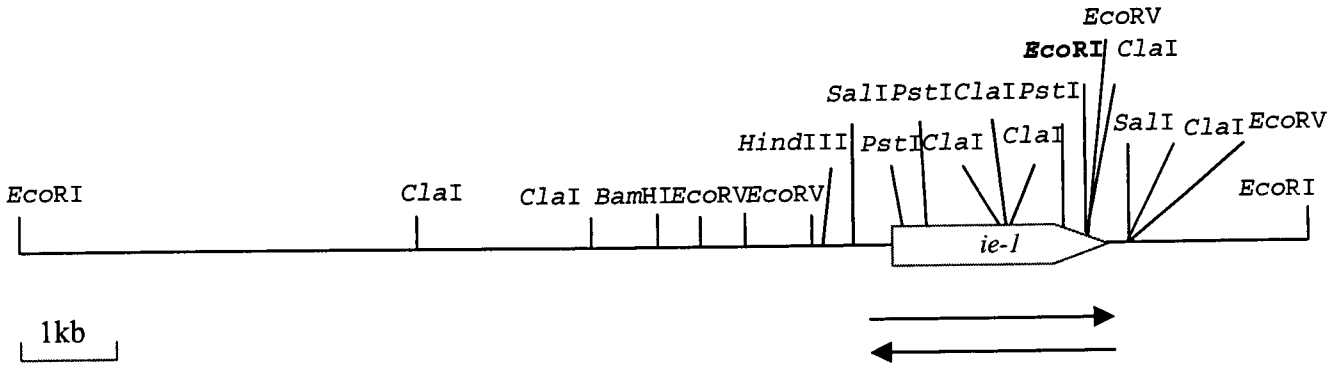


Figure 2.1 The location of the *ie-1* gene within an 11 kb *EcoRI* fragment and an adjoining 2.3 kb *EcoRI* fragment. An *EcoRI* restriction site between the 11 kb and 2.3 kb fragments is in bold. The orientation of the gene and sequencing directions are indicated by arrows.

2.3.3 Computer analysis

Nucleotide and amino acid sequence manipulation was carried out using the University of Wisconsin, Genetics Computer Group (GCG) sequence analysis package. The BLAST algorithm of Altschul *et al.* (1990) was used to compare sequences generated in this study with entries in non-redundant nucleotide and protein sequences databases accessed by the National Center for Biotechnology Information (NCBI). IE-1s were analyzed using the ExpASY server (Appel *et al.*, 1994) for charge analysis and domain prediction. Multiple

sequence alignments were conducted using the ClustalW program of Thompson *et al.* (1994). The alignment was used as the input to construct the phylogenetic trees by using neighbor-joining method implemented in the Clustal W program and the robustness of the data sets was estimated using bootstrap resampling procedure. GenDoc software was used for similarity shading and scoring among the aligned sequences.

2.4 Results

2.4.1 Nucleotide and protein sequence analyses

The TnSNPV *ie-1* gene has an open reading frame (ORF) of 2217 nucleotides, encoding a protein of 739 amino acids, with a molecular mass of 84.5 kDa (Fig. 2.2). The ORF is slightly larger than other compared *ie-1*s in this study which range from 1578 to 2142 nucleotides for McNPV and SeMNPV species respectively.

```

1      TTA-55ACTCGGTCAGCGCGCCCGCGTTCGTAGCCGCTTACGAGTCTTGTGGAAGCGAAATTG
61     CGGTTATTACCATCGCCGTCAGTGTGCATCGCTTTCGTTTATTTGAAACGGTATGATATTT
121    ATCGAGTTGATCGCATTATACAGCTAAATGTAATTAGGCTATCTTTATCTTGAGGATAG
181    TATAAATTGACAATCATT-24TTTTATTTTTACTTCCAGTTCCTCACTTACAAGCGTTGTGATGTC
                                     M S 2
241    GCATCCAAACATCATCAGTGCCATGGACAACAACGACGCGTCTATCAAATATAAAAATTA
        H P N I I S A M D N N D A S I K Y K N Y 22
301    TATCGACAACGCCATCAACACTCCTACGCATACGATTCTGCAGAACGTCAGTATGGATTT
        I D N A I N T P T H T I L Q N V S M D F 44
361    CGACGACAGCAATATTCTGGATTTCGGCAACGAAAATGATATGAATGTCTACGACAGACG
        D D S N I L D F G N E N D M N V Y D R R 64
421    AGACAGTAACAGCAGTAAAATTGTGAATGATGCTTGCATGAAAACCTCTCAATTTTCTGA
        D S N S S K I V N D A C D E N S Q F S D 84

```

481 TGTCACGTCAACAATAATGCGGACAACAATGATTATATAAAAAATCATGAAAACCTGCTAC
V N V N N N A D N N D Y I K I M K T A T 104
541 CGATGTCGTCGAAAAAAGAATGAATATACGAATAAACATAAAAACCTGCAGTCGTTTCGAC
D V V E N K N E Y T N K H K T A V V S T 124
601 TAAACCATTCAAGAAAAATCCTAAAAAAGGCCATCGTCATCGTTGACGACTACGACGAC
K P F K K N P K K R P S S S L T T T T T 144
661 GACGACGACAGAAAAGAAGAACAAGTCAAGACCCCAACAGACCTCCCAATTCCTACTGTAAT
T T T E K K N K S R P N R P P N S T V I 164
721 CGCTGATGGTAGTATTCCACCACAACCTGTGATAAAGCCATCGAAGAAGCAGACAGTTTT
A D G S I P P Q P V I K P S K K Q T V F 184
781 TGTTCGCTTTAATCAATAGAGGAGAAAAAACTTGAATGTGTGCGCAACGACAATAA
V S P L I N R G G K N L N V L R N D N N 204
841 TAATAATTTCAATAATGACAGTGACGATAGCAACGGAAGCGACAGTGAAGATAGCGATT
N N F N N D S D D S N G S D S E D S D S 224
901 TACGCATCCGCCGCCTTCGAAAAAGACAAAAATGACATCAAATCATCTAAAATGTCCGT
T H P P P S K K T K M T S K S S K M S V 244
961 GACGCCGCAACAACAAATGCCCGAGATTTTGAAAATTAATGCTGCCGACAAAAATAAAGT
T P Q Q Q M P E I L K I N A A D K N K V 264
1021 CAATGACGAGAAACAACAGTGAAATATAACAAAAAAGCAACAATCTCAAGACGCTGG
N D E K Q T V K Y N K K K Q Q S Q D A G 284
1081 TGCCGTCGTGGTCGTGAAACAACAAAACTTGATAACGAATCAACAAGTCAAACCTCCGT
A V V V V K Q Q K L D N E S T S Q T S V 304
1141 TAATGATGATCAACAACGGTCGAAAGATTGCGATTCTCCAACAATGACTTGTTTGAAA
N D D Q Q R S K D C D S P T N D L F E N 324
1201 TAAAATAATCCCAACATGATGACCATGGAAAGAGACAATAACCGCAAGTTTGTGCAATA
K I I P N M M T M E R D N N R K F V Q Y 344
1261 TATTCTCAACGCTCACAATTATCTGTTTATAGTATACGAAAACAAGTATAATGCCAAGAC
I L N A H N Y L F I V Y E N K Y N A K T 364
1321 TTTTAACAAAACTCCAACGCATCGATTATAAAAATAGAGTATGTGAATTGCGTCCAGTC
F N K N S N A S I Y K I E Y V N C V Q S 384
1381 CATATACAAGTATTATAACGCCAATTACTCGCATATCGATAGAACATGCAAAGTCGTGTC
I Y K Y Y N A N Y S H I D R T C K V V S 404
1441 TTTCAATCGATTGAGATTGCGCATATCTGTGAACCTTTTAAATAAAATGCAGATTGAATT
F N R F R F A I S V N L L N K M Q I E L T 424
1501 GCCTCCTACGGAACAATTTAAAAAGGAAGACCTCAAGAAGATTTCTCCGAAGAACAATTT
P P T E Q F K K E D L K K I S P K N T F 444
1561 TTGCCATTAAATGAAGTCAAAGATCCGGATTTTCATTTCCAAGCTCACTAACACATTCCG
C L L N E V K D P D F I S K L T N T F G 464
1621 CTTGGACAATATTTATATTCAGGGTCAACTCACTATGCTGCTCTCGTCGATTGGTGAGAA
L D N I Y I Q G Q L T M L L S S I G E N 484
1681 TCGGGCAAAGATTTGAATCAGCATATCAGTGCAATGATTGAAGATAAAAAGCCTATTAC

R A K I L N Q H I S A M I E D K S L F T 504

1741 TATACCTTTGCATTTGTCTCGATCCAAGGAATTGGAAGAAATTGTCGATGACGATCTGAA
I P L H L S R S K E L E E I V D D D L N 524

1801 CCCCACAACAGTAACGTCTCGTCGGCCTACATTCGAGACATAATAGAACTCTCGAACAA
P N N S N V S S A Y I R D I I E L S N K 544

1861 ACTCAAGTTTAAGGCTCCTATTATTCCGTCATATGTCATAAAACCAAGGAACAAAACAT
L K F K A P I I P S Y V H K T K E Q N I 564

1921 TGAGAATGTTCTTAGTTTTTGGATCAACACTCAGAAGAACAACAACGAGCGGATAAAAC
E N V L S F W I N T Q K N N N E R D K T 584

1981 TTTGGCAAAATCTCTGCAGTTTACATAACAAGTTTACCAGTGTGCTCGAGTGTCTCTCGA
L A K S L Q F T Y K F T S V A R V L F D 604

2041 CGAAACCGATGGCGACGTCAATAAACTGTTTAAAGTGAAGAAAGAGCCTGGATCTGTGGC
E T D G D V N K L F K V K K E P G S V A 624

2101 AATGATTGAAGATTATCTACAGGCTTGTGAAAAAATACCCAATGGCAACAACCTTTATTAT
M I E D Y L Q A C E K I P N G N N F I M 644

2161 GATCAACACACTCAACGATGAACGCGTGACAATCATCAAGGCCAAAAATGAATTCTTTTG
I N T L N D E R V T I I K A K N E F F W 664

2221 GATTTCGTACTAATAATCCTAATAATTTAATTCACTGTATCGATATCATTATGGCCTTTAA
I R T N N P N N L I H C I D I I M A F K 684

2281 AAACCTTAATCATCATTGGCTCTCTTTGATTCCCAGCAATCGTAAGGATTGAACAATCG
N F N H H L L S L I P S N R K D L N N R 704

2341 TCACAGTGGATTAATAAAGCTAGTGGCCTATCATTTAGGTGGTGATGTTGACATTAATTT
H S G L I K L V A Y H L G G D V D I N F 724

2401 TGTACGTGCCATGGCTGAGAAGTTTAAATGTAATTATCTATATAAAAAATTTTAAATGTA
V R A M A E K F K C N Y L Y K K F * 739

2461 TAATTTTTCTGTCTACTGTAATTTGAATAAAATTTTAAAGGATCGTATTTTGTATAG

Figure 2.2 Nucleotide and deduced amino acid sequences of the TnSNPV *ie-1* gene. A putative TATA box and a CAGT motif are underlined. The positions of these elements in the *ie-1* promoter region are indicated by numbers. An *EcoRI* restriction site located at the truncated ends of two fragments (11 kb and 2.3 kb) and a single polyadenylation signal in the 3' non-coding region are double underlined. The SET-domain-like region of 47-120 amino acids at the N-terminus of the deduced IE-1 protein are in bold.

A consensus baculovirus early promoter motif CAGT and a TATA box were found at -24 nt and -55 nt, relative to the predicted translational start codon of the TnSNPV *ie-1*, respectively. (Fig. 2.2). A comparison of the promoter regions of the eleven *ie-1* genes indicated that a TATA box is completely conserved among these regions, and a CAGT element is highly conserved with the exception of the HzSNPV and HaSNPV *ie-1* promoters, both of which have a CATT motif presented at the same position of -47, relative to its translation start codon (Fig. 2.3). Two CAGT motifs were found downstream of a conserved TATA box in the case of the CfMNPV and SpltNPV *ie-1*s, respectively. A single polyadenylation signal incorporating the TAA termination codon was found at the 3' non-coding region of the TnSNPV *ie-1* (Fig. 2.2).

```

AcMNPV : -GGCAAAATAAAGAAAATTGTTATCGTGTTCGCCATTAGCGCAG : 45
BmNPV : GGGCAAAATAAAAAAATTATTATCGTGTTCGCCATTAGCGCAG : 46
CfMNPV : TGGCAAGTAGAACCTAAATATCGTTATCTGTTCACCATCGGCTTGG : 46
OpMNPV : AGGCAAGTAGAAGGTTACATCAATTATCTGTTCGCCATCGGCTTGG : 46
LdMNPV : -GGAACGCTGATGGTCAATATTTTATCTAACTGTTGCCGCGATAA : 45
TnSNPV : TATCAACGCTAAATGAAATAGGCTATCTT-TATCTTGAG--GATAG : 43
SeMNPV : TAACAACATAACGTTACAATAATTAGTAT-TATCTTGCGC--GTAG : 43
McNPV : CAACAACATAACGTTATAATAATTGTCTG-TGTCTTTGGC--GCAG : 43
HzSNPV : CACTATCGTAACACAATAAAATTATCGATATGCGATAACGGTTGA : 46
HaSNPV : CACTATCGTAACACAATAAAATTATCGATATGCGATAACGGTTGA : 46
SpltNPV : ---CAACCAGCAACTGTTCACGAACAGCAGCAACCGGTTCCCGAC : 43

AcMNPV : TATAAATGACCTTCATGTTGGATAATGTTTCAGTTGCAAGTTGAC : 91
BmNPV : TATAAATGACCTTCATGTTGAATAATGTTTCAGTTGCAAGTTGAC : 92
CfMNPV : TATAAATAGACCTTCATGTTGGTTTTTGTTCAGTTGCAAGTTGAC : 92
OpMNPV : TATAAATAGACCTTCATGTTGGTTTTTGTTCAGTTGCAAGTTGGC : 92
LdMNPV : TATAAATGATATTCATTTTTTC---TTGTTTCAGTTGATCGTCCGC : 88
TnSNPV : TATAAATGACAATCATTTTTATTTTTACCTCCAGTTC-TCACITAC : 88
SeMNPV : TATAAATGACCTTCATGTTGTAAATGTTTCAGTTT-GCTGAGAC : 88
McNPV : TATAAATGACCTTCATTTGACGAAATGTTTCAGTCT-GCCCGGTC : 88
HzSNPV : TATAAATGATCTTCATTTTGTTTCAGAAITTAATG-ANAATTGAA : 91
HaSNPV : TATAAATGATCTTCATTTTGTTTCAGAAITTAATG-ANAATTGAA : 91
SpltNPV : TATAACAGTTTACAACAATTACAATAC-----CAGTACCAACTCC : 84
TATA TATA CAGT CAGT

AcMNPV : ACTGGCGGCACAAAGATCGTGAACAACCAAGTGACTATG : 130
BmNPV : ATTGGCGG-CACACGATCGTGAACAACCAACGACTATG : 130
CfMNPV : CGCGACGC-GCCCAGTACCTC-GCAATG----- : 118
OpMNPV : TGCGGCGC-GCCCAGTACCTTTGCGATG----- : 119
LdMNPV : GGGAA--ACGCCGAGTGGCAGGTGGACTCGCTTATG--- : 122
TnSNPV : AA--GCGTTCTCAGT----- : 101

```

```

SeMNPV : AA--TCATG----- : 95
McNPV : AC--ACGTCGTCATG----- : 104
HzSNPV : CTTTACCGCAACTATG----- : 107
HaSNPV : CTTTACCGCAACTATG----- : 107
SpltNPV: AAGTGTCAAGACGTCGAAGAAATTATG----- : 111

```

Figure 2.3 Sequence alignment of *ie-1* promoter regions. Putative TATA boxes and consensus baculovirus early promoter elements (CAGT or CATT only for the HzSNPV and HaSNPV *ie-1s*) are indicated by either letters or in bold italic. Gaps are introduced with dashes for optimization of the alignment. Putative translation start codons (ATG) are in bold.

```

AcMNPV : -----MTQINFNAS-YTSASTPSRASFDNSYSEFCDKQ- : 32
BmNPV : -----MTQINFNAS-YTSAPTPSRASFDNGYSEFCDKQQ : 33
CfMNPV : -----MPKNMAALQQSLYTGPSTPSHTQFSR-STEPENL- : 34
OpMNPV : -----MPKNMETLQRS-YMGPSTPNHNLNN-ATELPDDL- : 33
McNPV : -----MHQAYKNCEGRSTPARDKLGSFLTFAPSTPEVIDQVF : 37
SeMNPV : -MHTPSHHNQISPATYKTTG-STPLRDSLGPFLKFQHSIPOPPLDMF : 45
HzSNPV : --MANRITTPLRDQVGNQVTINYPFQSQESCNYNNDSDSYMNRNNDV : 45
HaSNPV : --MANRITTPLRDQVGNQVTINYPFQSQESCNYNNDSDSYMNRNNDV : 45
TnSNPV : MSHPNIISAMDNNDASIKYKNYIDNAINTPTHILQNVSMDFDDSNI : 47
LdMNPV : ----- : -
SpltNPV: -----MDKLNDLNDHINIEKAFDESQYENFNSLFLEDQQEIM : 37

AcMNPV : PND----YLSYYNHPTPDGADTVSDSETAAASNFLASV----- : 67
BmNPV : PND----YLNYYNNPTPDGADTVSDSRLQAASNFLASV----- : 68
CfMNPV : -----NFDVLNDSYETFSSVSLTT-----AE----- : 55
OpMNPV : -----NFSTMDVPYDGSMPMMS----- : 51
McNPV : G-----TDEPYTDNNYINLIENVGINNTYD----- : 63
SeMNPV : NNNNNMNTNQFVDIDYTDTNYNLLINNAEMINQNYDAAAAAAAAASHH : 92
HzSNPV : DVKKLIKTVENASNKTVENASAFFASYIPPTSSNKPSR----- : 84
HaSNPV : DVKKLIKTVENASNKTVENASAFFASYIPPTSSNKPSR----- : 84
TnSNPV : LDFG--NENDMNVYDRRDSNSSKIVNDACDENSQFSDVNVNNNADNN : 92
LdMNPV : -----MEPFLMYNEEMKTPAR----- : 16
SpltNPV: PTPP--PSVREQPAPSVRKTPACDRPVLYGKGKRMKMSG----- : 74

AcMNPV : ----NSLT-DNDVECEL-----KTTDN----LEEAVSSA : 93
BmNPV : ----NSLTDDNDMECEL-----KTTDN----LGEAVSSA : 95
CfMNPV : -----QDNQIDKL-----QESAA----MNRDVNSE : 77
OpMNPV : -----SDSEMNLE-----DRSKK----LACAVDIE : 73
McNPV : ----DASNLQINLNFFN-----ALTSEN----FAPSMECNK : 92
SeMNPV : HDPFNASSPQEDLDLFPP--EQ--VPFVETKSNVDAVPTIKIDK : 135
HzSNPV : --PNHLRFGDELVMSPLAMS--PQRITPRSERSENVIESLPESLSL : 127

```


Ac213 I

AcMNPV : D-----DNDYNSNRFS~~SDHMSETGYMFVVVKKSEV~~KP----- : 235
 BmNPV : D-----DNDYNSNRFS~~SDHMSETGYMFVVVKKSEV~~KP----- : 237
 CfMNPV : KS----DCQMSANRFS~~SDYISETGYVFVVVKKSEH~~KP----- : 217
 OpMNPV : KS----ECQMSANRFS~~SDYISETGYVFVVVKKGDR~~KP----- : 218
 McNPV : PS---DMHLQNNR~~LFA~~SH~~L~~LD~~T~~GY~~M~~FL~~V~~VY~~D~~Q~~N~~SS~~D~~DS----- : 241
 SeMNPV : LQQNDDVVVDNSGI~~TTSH~~L~~D~~TSY~~M~~FI~~L~~SK~~S~~SN~~V~~DEV----- : 344
 HzSNPV : VQNMPTFRVDKSR~~R~~FV~~D~~TI~~Q~~Q~~K~~NY~~H~~MF~~I~~V~~K~~EQENINSS~~S~~IE----- : 298
 HaSNPV : VQNMPTFRVDKSR~~R~~FV~~D~~TI~~Q~~Q~~K~~NY~~H~~MF~~I~~V~~K~~EQENINSS~~S~~IE----- : 305
 TnSNPV : IPNMMTMERDNNR~~K~~FV~~Q~~IL~~N~~NA~~N~~L~~F~~I~~V~~YENKYN~~A~~KT~~F~~N~~K~~NS~~N~~ASI : 371
 LdMNPV : AEDARVSGGVRD~~K~~R~~F~~ST~~V~~Y~~N~~AN~~Y~~M~~F~~I~~V~~SEDAET~~P~~ARP----- : 193
 SpltNPV : DFS---AETDQSR~~R~~FV~~D~~E~~F~~Y~~T~~SK~~L~~Y~~H~~M~~F~~I~~L~~SPK~~T~~SD~~V~~D~~N~~DP-----AA : 261

Tn338 II

AcMNPV : FEIIFAKYVSNV~~V~~Y~~E~~Y~~N~~NY~~M~~V~~D~~NR~~V~~F~~V~~V~~T~~F~~K~~IR~~F~~M~~I~~S~~N~~L~~V~~K~~E~~T : 282
 BmNPV : FEIIFAKYVSNV~~V~~Y~~E~~Y~~N~~NY~~M~~V~~D~~NR~~V~~F~~V~~V~~T~~F~~K~~IR~~F~~M~~I~~S~~N~~L~~V~~K~~E~~T : 284
 CfMNPV : FEVVF~~A~~K~~F~~V~~N~~N~~V~~T~~N~~E~~Y~~N~~N~~Y~~M~~V~~D~~NR~~V~~F~~V~~V~~S~~L~~N~~V~~K~~F~~M~~V~~S~~K~~L~~V~~R~~E~~Q~~ : 264
 OpMNPV : FEVVF~~A~~K~~F~~V~~N~~N~~A~~T~~N~~E~~Y~~N~~N~~Y~~M~~V~~D~~NR~~V~~F~~V~~V~~S~~L~~N~~V~~K~~F~~M~~V~~S~~K~~L~~V~~R~~E~~Q~~ : 265
 McNPV : FVIRYV~~N~~C~~V~~H~~S~~V~~N~~E~~Y~~V~~A~~R~~H~~M~~H~~D~~R~~F~~V~~V~~T~~Y~~E~~R~~Y~~R~~F~~M~~V~~S~~K~~L~~L~~L~~H~~L : 288
 SeMNPV : YSLRY~~N~~C~~V~~H~~S~~V~~H~~N~~E~~Y~~T~~A~~H~~M~~H~~D~~R~~F~~V~~L~~V~~V~~T~~I~~E~~R~~Y~~R~~F~~M~~I~~S~~N~~L~~L~~L~~G~~M : 391
 HzSNPV : HVILY~~A~~N~~T~~V~~A~~S~~I~~N~~E~~Y~~E~~S~~S~~Y~~N~~V~~D~~K~~L~~V~~H~~V~~V~~T~~F~~N~~R~~Y~~R~~F~~M~~I~~S~~H~~R~~L~~L~~T~~K~~L : 345
 HaSNPV : HVILY~~A~~N~~T~~V~~A~~S~~I~~N~~E~~Y~~E~~S~~S~~Y~~N~~V~~D~~K~~L~~V~~H~~V~~V~~T~~F~~N~~R~~Y~~R~~F~~M~~I~~S~~H~~R~~L~~L~~T~~K~~L : 352
 TnSNPV : YKIEYV~~N~~C~~V~~Q~~S~~I~~Y~~K~~Y~~N~~A~~N~~Y~~S~~H~~I~~D~~R~~T~~C~~K~~V~~V~~S~~F~~N~~R~~F~~R~~F~~A~~I~~S~~V~~N~~L~~L~~N~~K~~M : 418
 LdMNPV : FKVTY~~A~~N~~C~~V~~W~~C~~I~~N~~Q~~D~~Y~~R~~R~~N~~Y~~R~~H~~V~~D~~N~~K~~V~~M~~V~~L~~S~~I~~E~~K~~C~~R~~F~~M~~I~~S~~K~~L~~L~~K~~K~~M~~ : 240
 SpltNPV : YELRYV~~N~~T~~V~~H~~S~~V~~L~~S~~E~~Y~~R~~K~~Y~~E~~S~~K~~L~~S~~N~~K~~V~~L~~V~~V~~T~~M~~A~~R~~Y~~R~~F~~M~~I~~V~~E~~R~~V~~L~~S~~A~~M~~ : 308

AcMNPV : GIEI~~P~~H~~S~~O~~D~~V~~C~~N~~G~~E~~T~~A~~A~~Q~~N~~C-----KK~~C~~H~~F~~V~~D~~V~~H~~H-~~T~~E~~K~~A~~A~~L~~T~~S~~Y~~E~~N~~ : 323
 BmNPV : GIEI~~P~~H~~S~~O~~D~~V~~C~~N~~D~~E~~T~~A~~A~~Q~~N~~C-----KK~~C~~H~~F~~V~~D~~V~~H~~H-~~T~~E~~K~~A~~A~~L~~T~~S~~Y~~E~~N~~ : 325
 CfMNPV : GIDIP~~P~~H~~V~~N~~L~~C~~D~~D~~A~~Q~~A~~E~~R~~N~~P~~-----Y~~D~~C~~Y~~E~~F~~P~~V~~K~~N~~-V~~F~~Q~~T~~L~~I~~N~~H~~F~~H~~ : 305
 OpMNPV : GIDIP~~P~~H~~V~~N~~L~~C~~N~~D~~A~~Q~~A~~E~~R~~T~~P~~-----L~~N~~C~~Y~~E~~F~~P~~V~~K~~N~~-A~~F~~Q~~A~~T~~L~~I~~N~~H~~F~~H : 306
 McNPV : EIDIP~~Q~~Q~~D~~Q~~F~~S~~E~~T~~Q~~L~~K~~N~~T~~N~~P~~-----K~~E~~C~~Y~~E~~F~~E~~V~~K~~N~~F~~E~~L~~T~~F~~L~~T~~N~~Y~~F~~H : 330
 SeMNPV : NIEIP~~T~~Q~~D~~Q~~F~~S~~E~~K~~Q~~L~~S~~D~~T~~N~~K~~-----N~~M~~C~~I~~F~~E~~E~~V~~K~~D~~F~~K~~L~~S~~L~~I~~N~~T~~E~~R~~ : 433
 HzSNPV : NVHIP~~E~~S~~E~~Q~~F~~P~~M~~R~~V~~H~~Q~~D~~R~~S-----T~~K~~C~~H~~E~~N~~E~~I~~K~~D~~Y~~V~~E~~M~~N~~E~~L~~N~~H~~M~~E~~N~~ : 386
 HaSNPV : NVHIP~~E~~S~~E~~Q~~F~~P~~M~~R~~V~~H~~Q~~D~~A~~S-----T~~K~~C~~H~~E~~N~~E~~I~~K~~D~~Y~~V~~E~~M~~N~~E~~L~~N~~H~~M~~E~~N~~ : 393
 TnSNPV : QIEI~~P~~P~~T~~Q~~F~~K~~K~~E~~D~~L~~K~~K~~I~~S~~P~~K~~N~~---T~~F~~C~~L~~L~~N~~E~~V~~K~~D~~P~~D~~E~~I~~S~~K~~L~~T~~N~~T~~E~~G~~ : 462
 LdMNPV : NIPI~~P~~S~~E~~D~~I~~E~~R~~Q~~A~~A~~A~~E~~A~~A~~E~~A~~A~~E~~A~~A~~R~~E~~D~~K~~C~~Y~~E~~N~~E~~I~~K~~D~~F~~E~~L~~L~~I~~N~~T~~E~~N~~ : 287
 SpltNPV : NITIP~~L~~L~~R~~I~~E~~D~~P~~K~~D~~N~~E~~I~~S~~-----E~~N~~E~~V~~K~~D~~S~~N~~E~~F~~N~~L~~L~~V~~H~~T~~E~~N~~ : 345

III

AcMNPV : LD~~Y~~Y~~A~~Q~~T~~F~~V~~T~~L~~L~~Q~~IG~~E~~R~~K~~C~~G~~F~~L~~L~~S~~K~~L~~Y~~E~~M~~Y~~Q~~D~~N~~L~~F~~A~~L~~P~~I~~M~~L~~S~~R : 370
 BmNPV : LD~~Y~~Y~~A~~Q~~T~~F~~V~~T~~L~~L~~Q~~IG~~E~~R~~K~~C~~G~~F~~L~~L~~S~~K~~L~~Y~~E~~M~~Y~~Q~~D~~N~~L~~F~~A~~L~~P~~I~~M~~L~~S~~R : 372
 CfMNPV : LD~~Y~~Y~~S~~Q~~T~~F~~V~~T~~L~~M~~Q~~MG~~E~~S~~K~~SG~~M~~L~~L~~N~~K~~L~~Y~~MF~~Q~~D~~R~~S~~L~~F~~A~~L~~P~~I~~M~~L~~S~~R : 352
 OpMNPV : LD~~F~~Y~~A~~Q~~T~~F~~V~~T~~L~~M~~Q~~AV~~G~~E~~N~~K~~T~~N~~L~~L~~N~~K~~L~~Y~~M~~Y~~Q~~D~~R~~S~~L~~F~~A~~L~~P~~I~~M~~L~~S~~R : 353
 McNPV : LD~~K~~V~~A~~Q~~G~~K~~S~~L~~L~~L~~A~~IG~~E~~Y~~K~~A~~R~~I~~Y~~N~~T~~L~~T~~E~~M~~I~~N~~D~~K~~S~~L~~F~~R~~L~~P~~F~~H~~M~~C~~K : 377
 SeMNPV : LD~~Q~~V~~I~~Q~~G~~K~~S~~L~~L~~L~~A~~V~~G~~E~~S~~K~~S~~R~~V~~I~~F~~D~~Q~~L~~T~~M~~I~~D~~T~~G~~M~~M~~F~~A~~L~~P~~M~~S~~V~~T~~K~~ : 480
 HzSNPV : LD~~V~~M~~V~~Q~~E~~L~~Y~~F~~L~~M~~S~~A~~I~~G~~P~~D~~K~~G~~K~~L~~I~~K~~S~~V~~M~~E~~H~~I~~N~~D~~D~~H~~L~~F~~V~~L~~P~~I~~N~~L~~S~~R : 433
 HaSNPV : LD~~V~~M~~V~~Q~~E~~L~~Y~~F~~L~~M~~S~~A~~I~~G~~P~~D~~K~~G~~K~~L~~I~~K~~S~~V~~M~~E~~H~~I~~N~~D~~D~~H~~L~~F~~V~~L~~P~~I~~N~~L~~S~~R : 440
 TnSNPV : LD~~N~~I~~I~~Q~~Q~~T~~M~~L~~L~~SIG~~E~~N~~R~~A~~K~~L~~N~~Q~~H~~I~~S~~A~~M~~E~~D~~S~~L~~F~~A~~I~~P~~L~~H~~L~~S~~R : 509
 LdMNPV : LD~~M~~C~~T~~R~~V~~K~~F~~M~~L~~L~~S~~MG~~D~~S~~K~~S~~K~~L~~W~~N~~W~~V~~Y~~R~~V~~K~~D~~E~~T~~L~~F~~H~~I~~P~~V~~N~~Y~~G~~H~~ : 334
 SpltNPV : L~~N~~T~~T~~I~~V~~Q~~I~~D~~A~~F~~M~~Y~~S~~A~~L~~S~~O~~S~~K~~A~~M~~Y~~V~~H~~N~~K~~M~~N~~K~~L~~A~~D~~K~~T~~L~~F~~A~~L~~P~~I~~N~~V~~S~~R : 392

AcMNPV : ~~K~~E~~S~~N~~E~~I~~E~~-----T~~A~~S~~N~~N~~F~~F~~V~~S~~P~~V~~S~~Q~~I~~L~~K~~Y~~S~~E~~S~~---V~~Q~~----- : 400
 BmNPV : ~~K~~E~~S~~N~~E~~I~~E~~-----T~~A~~S~~N~~N~~F~~F~~V~~S~~P~~V~~S~~Q~~I~~L~~K~~Y~~S~~E~~S~~I~~R~~K~~V~~K----- : 405

CfMNPV : EPTIEN-----TPLSRNYTSSYVAQLIKYSKN---VR----- : 382
 OpMNPV : EPVNEIN-----APQKNHAFSYVAQIMKYSKN---LR----- : 383
 McNPV : EITD-----DEMSRAYSSAYVSDIILKTEN---IK----- : 405
 SeMNPV : EAPNQ-----DELKKYDMSMYVEDIMKYTTG---LH----- : 509
 HzSNPV : QESKLED-----IORTVASVSLYVONIVSLSKD---VQ----- : 463
 HaSNPV : QESKLED-----IORTVASVSLYVONIVSLSKD---VQ----- : 470
 TnSNPV : SKELEIVDDDL---NPNNSNVSSAYIRDIIELSNK----- : 542
 LdMNPV : QPIVLEED-----FLAAPAAPECGASASSADSEHVKSIV----- : 368
 SpltNPV : RDVVLEASEIAANVATPTSRAIAIATIESATNMSTKTPMIASDEOPL : 439

AcMNPV : -----FPDNPPNKYVVDNLNLIIVNKKSTLIVKYSS-- : 430
 BmNPV : -----FPDNPPNKYVVDNLNLIIVNKKSTLIVKYSS-- : 435
 CfMNPV : -----PENNPNGVTSRLEETVTVQKSSLIVKYSS-- : 412
 OpMNPV : -----PQGDPTQQVTRLEETVTVQKSSLIVKYSS-- : 413
 McNPV : -----KVLQYKKKHTRPRVTKNVLNALSFWLRSKD : 437
 SeMNPV : -----INKFEE-DRKTSRAQIVDSVSKLSFWYENKQ : 540
 HzSNPV : -----KQTAE--NFMNRDDVINYVTVALKFWLRSKN : 493
 HaSNPV : -----KQTAE--NFMNREDVINYVTVALKFWLRSKN : 500
 TnSNPV : -----LKFKAPIIPSYVHKTKEQNIENVLSFWINTQK : 574
 LdMNPV : -----SAGEGLSFRVADAKLTAEQALDSVRFWLRITKS : 400
 SpltNPV : YISDIVKACHCASFVVNDHGAKSSPNKLESYIKKLTEELKFWLPNAS : 486

AcMNPV : -----VANLLENNYKYHDNIASINNAENLKKVKKE-DG : 462
 BmNPV : -----VANLLENNYKYHDNIASINNAENLKKVKKE-DG : 467
 CfMNPV : -----VANLLENSRYGHQ---RDNADSLKVKKE-DG : 440
 OpMNPV : -----VANLLENNRYGR---RDNADALKVKKE-DG : 440
 McNPV : TKSSSEFK---EKNNFTYKYGSVRVLYDATDKNVSKLFIKKE-NG : 479
 SeMNPV : TIKNRNKQQQEKSNFTYKYGCIARQFYDPHKGVKKLEKVKKE-NG : 586
 HzSNPV : EKNVVKE----QSDFFTYYKYGSVVRLLEK-ESIHTNALLKIKRE-TG : 534
 HaSNPV : EKNVVKE----QSDFFTYYKYGSVVRLLEK-ESIHTNALLKIKRE-TG : 541
 TnSNPV : NNNERDK-TLAKSLQFTYKYGSAVLADETDGTVNKLKVKKE-PG : 619
 LdMNPV : NDAVQKT----KDCYINRYACTVRLLEDEQDKRIANLLKIKKPGAG : 443
 SpltNPV : GEDVKKIR----NSFTYKYGSSVALFLK--DRDLYLEFKVVRKT-KG : 525

IV

AcMNPV : SMHIVEQYLTQNVQNV-KGHNFIVLSF--NEERLTIKKNKEFYWI : 506
 BmNPV : SMHIVEQYLTQNVQNV-KGHNFIVLSF--NEERLTIKKNEEFYWI : 511
 CfMNPV : NRLLVEQYMSQENDE-TSHNFIVLQFGGVNDERLTIKKGIEFFWI : 486
 OpMNPV : NRLLVEQYMSYNENDD-TSHNFIVLQFGGVNDERLTIKKGKEFYWI : 486
 McNPV : SVKLIBEYLSACKQFP-ESHNFILITR--SDERITAK-----WA : 517
 SeMNPV : STKLIENYLNACKERF-ENYSFILITR--SDERITIKNGMEFLWI : 630
 HzSNPV : HAGLIDNYLEANQNDT-TSNSFILINR--MDERITIKKGPVFLWI : 578
 HaSNPV : HAGLIDNYLEANQNDT-TSNSFILINR--MDERITIKKGPVFLWI : 585
 TnSNPV : SVAMIEDYLQACEKIP-NGNMFIMINL--NDERVTIKKAKNEFFWI : 663
 LdMNPV : TAEIIVEHYLVCAKLPRDSQNFLLVTRK--NEERLTIKNGPRLVWI : 488
 SpltNPV : NKSLVEAYLNAMNHS--VHSFILVDRK--NEERLTIKKNVEFVWI : 568

AcMNPV : SGEKDDV--SQVIQKRNRFKHMFVIGKVNRRRESTLHNNLLKLL : 551
 BmNPV : SGEKDDA--SQVIQKRNRFKHMFVISKVNRRRESTLHNNLLKLL : 556
 CfMNPV : AAEDNV--DLVKQTRNVHVERIINVNRRESTTWHNNLLKLL : 531
 OpMNPV : AGEKDSV--DLIKQARNVHVERIINVNRRESTTWHNNLLKLL : 531
 McNPV : LSLSGSPS----- : 526
 SeMNPV : TSVKDIIV--TEIKKQKMYNHVYVNLNNGNRKEINIRHNGMIKLL : 675

```

HzSNPV : TSIKDLIA--MLLIEYKKHTHVFENLSNTNRKEMNNKHNGMIKLL : 623
HaSNPV : TSIKDLIA--MLLIEYKKHTHVFENLSNTNRKEMNNKHNGMIKLL : 630
TnSNPV : RTNNPNNLTHCIILIMAKNFNHLLSLIPSNRKDLNNRHSGLIKLV : 710
LdMNPV : SGVADLCYGG-DIINKFDGQFEHVFKLNKVSRKELNNRHNGLLKLV : 534
SpltnPV : NCISSDIP-ELIHHKDGTHYVEAMKRTNRKEVHARLNGMMKLI : 614
                                     521
AcMNPV : ALILQGLVPLSDAITFAEQKLN--CKYKQFEFN----- : 582
BmNPV : ALILQGLVPLSDAITFAEQKLN--CKYKQFEFN----- : 587
CfMNPV : QLLLQNLIRIDDVQQYSNKGDSK-FIYKQL----- : 560
OpMNPV : QLLLQNLIRLEDVQRYSDKSDTK-FVYKQV----- : 560
McNPV : ----- : -
SeMNPV : SNYTGGRLTINEATGIAVESFN--CNFEKVIYDKKNAKSIN----- : 714
HzSNPV : SFYTSNLLMLDEKFAVNNFN--CSYDCKHYA----- : 654
HaSNPV : SFYTSNLLMLDEKFAVNNFN--CSYDCKHYA----- : 661
TnSNPV : AYHLGGVDINFRAMAEKFKCN-YLYKQF----- : 739
LdMNPV : SLYTSAAVDLSVVEIQTOFE--CDYRCSQTSM----- : 566
SpltnPV : STYVDESLEKMKHVIKIADI FAANCEI IYDDGRLVDRRDSA AVAAA : 661
                                     570
AcMNPV : ----- : -
BmNPV : ----- : -
CfMNPV : ----- : -
OpMNPV : ----- : -
McNPV : ----- : -
SeMNPV : ----- : -
HzSNPV : ----- : -
HaSNPV : ----- : -
TnSNPV : ----- : -
LdMNPV : ----- : -
SpltnPV : VQQVKPAAIKRSVDDSSSIKSKKQKV : 688

```

Figure 2.4 Alignment of amino acid sequences deduced from nucleotide sequences of eleven NPV IE-1s. The comparatively conserved regions of IE-1s start from the amino acids 213 (AcMNPV) and 338 (TnSNPV). The four conserved regions at C-termini are numbered in Roman. The position of the helix-loop-helix-like domain was indicated by numbers and arrows. The order of alignment sequences with GenBank accession number is: *Autographa californica* (Ac) MNPV (P11138); *Bombyx mori* (Bm) NPV (P33245); *Choristoneura fumiferana* (Cf) MNPV (L04945); *Orgyia pseudotsugata* (Op) MNPV (P22114); *Mamestra configurata* (Mc) NPV(U59461); *Spodoptera exigua* MNPV (AF169823); *Helicoverpa zea* (Hz) SNPV (U67264); *Helicoverpa armigera* (Ha) SNPV

(NC003094); *Trichoplusia ni* (Tn) SNPV (AF317709); *Lymantria dispar* (Ld) MNPV (AF006656); *Spodoptera litura* (Splt) NPV (NC003102).

A comparison of eleven deduced amino acid sequences of NPV IE-1s (Fig. 2.4) showed little similarity among the N-terminal regions. The acid-base maps of IE-1s (Fig. 2.5) showed that each of N-terminal regions contains a high proportion of acidic residues. The number of acidic residues in each of these regions was different from one another. A cluster of basic amino acids appeared to form the boundary of each acidic region, but the pattern and position of the clusters were different in each case. The comparison of charge distribution of N-termini of IE-s showed that the net negative charge remained in each case (Table 2.1). Two regions with net negative charge, designated as region I and II, were identified at the N-terminus of each NPV IE-1 (Table 2.1). The number of amino acid residues in each of the region Is was more than that of the corresponding region II, but the number of residues in the region II of the SpltNPV IE-1 was more than that of its region I. The net negative charge among the region Is ranged from -3 to -19, while among the region IIs from -3 to -15. SwissProt database search for all the full length IE-1 proteins revealed that the amino acids 47 to 120 aa of the deduced TnSNPV IE-1 protein significantly matched the SET-domain of transcriptional regulators, while no matches to the SET-domain were found in each of the other IE-1s.

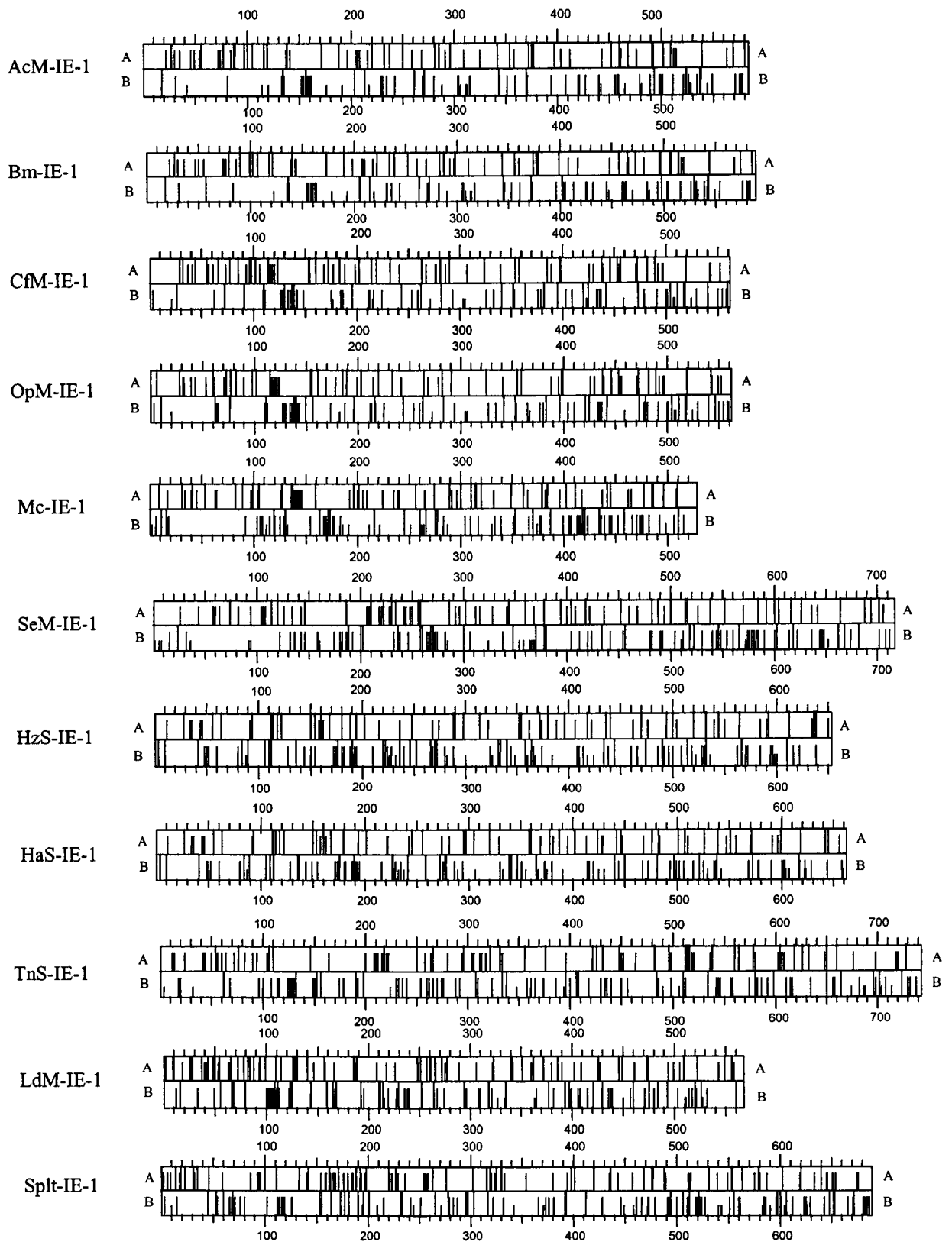


Figure 2.5 Acid-base profiles of the eleven IE-1 proteins. The acidic and basic amino acids were indicated in lane A and lane B of each diagram, respectively. Medium and full bars in lane A represent aspartic and glutamic acids. Small, medium and full bars in lane B depict histidine, lysine and arginine.

	AcMNPV 582aa	BmNPV 587aa	CfMNPV 560aa	OpMNPV 560aa	McNPV 526aa	SeMNPV 714aa	HsSNPV 654aa	HaSNPV 661aa	TnSNPV 739aa	LdMNPV 566aa	SpltNPV 688aa
Region I	1-151	1-153	1-125	1-126	1-146	1-120	1-47	1-47	1-113	1-99	1-46
Segment size (151aa)	(1-153)	(125aa)	(126aa)	(146aa)	(120aa)	(47aa)	(47aa)	(113aa)	(99aa)	(46aa)	
Net charge	-16	-15	-14	-13	-15	-10	-3	-3	-13	-16	-8
Region II	161-228	163-230	148-183	143-211	190-230	199-262	149-170	149-170	192-228	122-166	139-230
Segment size (67aa)	(67aa)	(35aa)	(68aa)	(40aa)	(63aa)	(21aa)	(21aa)	(36aa)	(44aa)	(91aa)	
Net charge	-5	-5	-3	-9	-11	-4	-5	-6	-4	-15	

Table 2.1 The N-terminal net negative charge regions of eleven baculovirus IE-1s. The total amino acids of each deduced protein are shown under the corresponding IE-1. The position of each segment is indicated with numbers. The sizes of two amino acid segments of each IE-1 are indicated with numbers in brackets. Negative charge (-) in each case refers to net charge of the segment.

In contrast, the C-terminal regions were comparatively conserved, which contained four well conserved regions (Fig. 2.4). Apart from region I with 16 amino acids, the rest of three regions consisted of over 50 amino acids and most residues in each region were well conserved. More than half of the previously suggested helix-loop-helix-like domain was also conserved among the compared IE-1s. However, SwissProt database search showed that either these conserved regions, or the entire C-termini of the IE-1s, matched no homologies with relevant functional domains.

2.4.2 Comparison and phylogenetic analysis of IE-1 proteins

Previously, AcMNPV, BmNPV, CfMNPV and OpMNPV were identified as group I NPVs and HzSNPV, HaSNPV, McNPV, SeMNPV, TnSNPV, LdMNPV and SpltNPV defined as group II NPVs (Cowan *et al.*, 1994; Bulach *et al.*, 1999). The alignment of deduced IE-1 amino acid sequences showed high degree of homologies among the AcMNPV, BmNPV, CfMNPV and OpMNPV IE-1s, whereas the rest of the NPV IE-1s (McNPV, SeMNPV, TnSNPV, LdMNPV and SpltNPV) shared much lower homologies, except HzSNPV and HaSNPV which displayed a higher degree of amino acid homology (Table 2.2). Among the group I and II IE-1s exist lower homologies (Table 2.2). This is in good agreement with the previously defined baculovirus phylogeny in which group I and II NPVs and GVs are distinct from one another (Zanotto *et al.*, 1993). The proposed IE-1s of subgroup II-A, SeMNPV and McNPV, shared 35 % identity, while subgroup II-B, between SpltNPV and either HzSNPV or HaSNPV, shared 21% identity (Table 2.2).

	BmNPV (%)	CfMNPV (%)	OpMNPV (%)	McNPV (%)	SeMNPV (%)	HzSNPV (%)	HaSNPV (%)	TnSNPV (%)	LdMNPV (%)	SpltNPV (%)
AcMNPV	96	43	44	20	18	21	21	16	20	16
BmNPV		43	43	20	18	21	21	16	19	15
CfMNPV			73	20	19	20	20	16	21	14
OpMNPV				19	19	20	20	16	21	14
McNPV					35	19	19	19	17	15
SeMNPV						24	25	22	19	17
HzSNPV							96	20	21	21
HaSNPV								20	21	21
TnSNPV									17	14
LdMNPV										17

Table 2.2 Identity values of the amino acid sequences of eleven baculovirus IE-1s. The percentage of identity amino acids shared by relevant sequences are indicated by

numbers. The scores shared between TnSNPV and SeMNPV, and between SeMNPV and McNPV are in bold.

The TnSNPV IE-1 shared the highest homology with the SeMNPV homologue (22% identity and 40% similarity) and lower with the rest of IE-1s (Table 2.3a). The comparison between the deduced full length TnSNPV IE-1 and other IE-1 homologues is proportionally less conserved than that of the C-terminal two third portions of the TnSNPV IE-1 with the other corresponding counterparts (Table 2.3a,b). In the latter case the TnSNPV IE-1 shared the highest homologies with SeMNPV (29% identity and 52% similarity) homologue (Table 2.3b).

(a)

	AcMNPV (%)	BmNPV (%)	CfMNPV (%)	OpMNPV (%)	McNPV (%)	SeMNPV (%)	HzSNPV (%)	HaSNPV (%)	LdMNPV (%)	SpltNPV (%)
Identity	16	16	16	16	19	22	20	20	17	14
Similarity	31	31	32	31	33	40	38	39	31	32

(b)

	AcMNPV (%)	BmNPV (%)	CfMNPV (%)	OpMNPV (%)	McNPV (%)	SeMNPV (%)	HzSNPV (%)	HaSNPV (%)	LdMNPV (%)	SpltNPV (%)
Identity	22	22	22	22	27	29	26	27	23	18
Similarit	39	39	39	39	44	52	46	47	41	35

Table 2.3 The full length (a) and the C-terminal two third portions (b) of ten deduced NPV IE-1s are compared with the TnSNPV homologue, respectively. The greatest homologies shared by TnSNPV and SeMNPV IE-1s in table a and b are in bold.

Based on alignment data sets of the predicted full length IE-1 proteins and the C-terminal two third portions of the amino acid sequences, two rooted phylogenetic trees, A and B, were constructed, respectively (Fig. 2.6a,b).

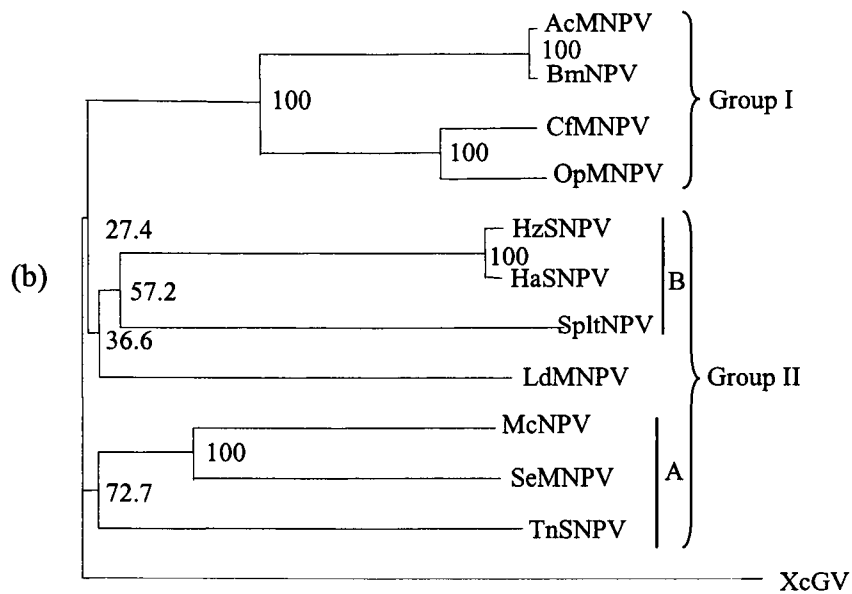
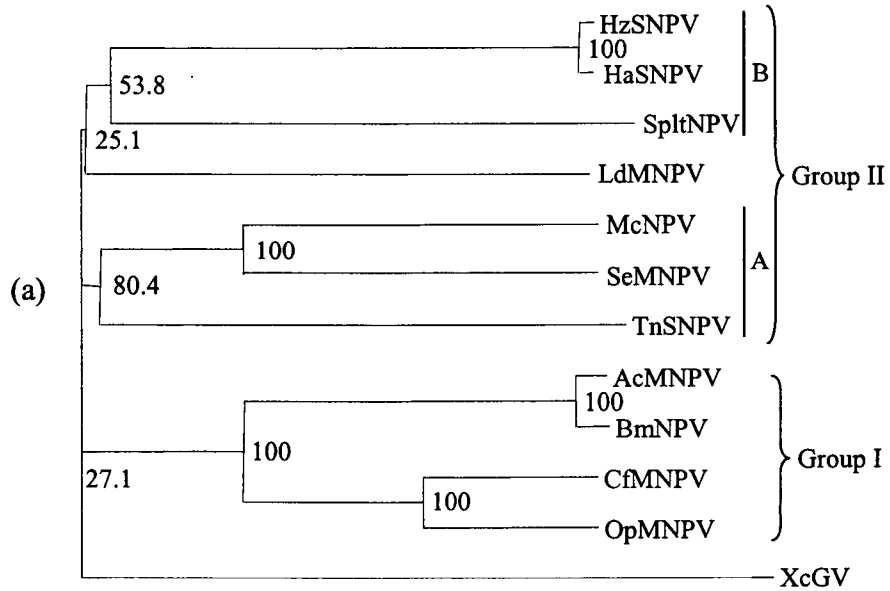


Figure 2.6 Phylogenetic relationships of the predicted full length (a) and the C-terminal two thirds (b) of the IE-1 proteins. The bootstrap scores are present at each node as percentages of 1000 trials. Branch lengths are proportional to relatedness. The members of group I and II and proposed subgroup II-A and II-B are indicated in the trees. The XcGV taxa was used as an out-group.

The high bootstrap supports derived from trees were obtained for inferring phylogenetic relationships between the IE-1s of group I, subgroup II-A (SeMNPV and McNPV) and II-B (HzSNPV and HaSNPV) NPVs, but the separation of group I and II was supported by low bootstrap values in each tree, with a slightly higher score in tree B than in tree A. TnSNPV IE-1 obtained comparatively high scores (80.4% and 72.4 % for trees A and B) for being included in subgroup II-A, while the comparatively low bootstrap scores (53.8% and 57.25 % for trees A and B) were observed for inclusion of the SpltNPV in Subgroup II-B. Although the LdMNPV IE1 appeared to be clustered in subgroup II-B with the support of low bootstrap values (below 50%) in both trees, the bootstrap score of this branching in tree B was slightly higher than that in tree A. Similarly, the phylogenetic relationship between subgroup II-A and II-B NPV IE-1s was supported by low bootstrap values. In this case tree B showed a slightly higher bootstrap value than tree A in support of the separation.

2.5 Discussion

Many baculovirus early promoter regions contain a conserved early consensus motif

CAGT and a TATA box (Friesen, 1997). To examine the functions and interactions of these early regulatory elements in directing transcription of the AcMNPV *ie-1*, a previous study (Pullen and Friesen, 1995b) demonstrated that site mutations of the promoter CAGT reduced the rate of transcription, but the RNA start site remained at or near the initiation site of the wild type *ie-1* for each mutation. In contrast, deletion of the TATA element reduced the rate of transcription much more than that of the former experiment but failed to abolish transcription from the promoter CAGT. However, deletion of the CAGT motif abolished early and late transcription (Pullen and Friesen, 1995a). This suggested that CAGT can function as a faithful initiator in a TATA-less promoter, while a TATA element in the *ie-1* promoter plays a role in stimulating transcription from the CAGT. Although Pullen and Friesen (1995b) demonstrated that the *ie-1* promoter CAGT motif and its flanking sequences were sufficient for proper transcription in a TATA-independent manner, the TATA plays an important role in cooperation with the CAGT initiator to properly regulate *ie-1* promoter activity. Since a TATA box is present in each of eleven NPV *ie-1* promoter regions, and an early promoter sequence CAGT highly conserved in the same region of most *ie-1*s (Fig. 2.3), both elements cooperating with each other may be a common feature for properly directing transcription of *ie-1*s. Although only a CATT motif was found in both HzSNPV and HaSNPV *ie-1* promoters, instead of a usual CAGT element, it may act as a transcription initiator since it conforms to the consensus early promoter CAG/TT (Friesen, 1997).

By observation of eleven NPV *ie-1* promoter regions, two CAGT motifs are found downstream of a putative TATA box in each case of CfMNPV and SpltNPV *ie-1*s,

respectively. It has not yet been determined which CAGT motif could be utilized as RNA start site, but there is no available evidence that both CAGT motifs could act as initiators in the case of early transcription of *ie-1*. In addition, the study of Pullen and Friesen (1995b) suggested that the early promoter CAGT can function as a late promoter element in the case of the AcMNPV *ie-1* late transcription. The sequence gazing of these compared *ie-1* promoter regions did not find any late consensus promoter TAAG. Perhaps, as suggested earlier, a CAGT motif may play a role in late *ie-1* transcription.

Many transcriptional activator proteins contain an acidic domain as a common characteristic (Brent and Ptashne, 1985; Hope and Struhl, 1986; Trezenberg *et al.*, 1988). The role of the N-terminal acidic region of the AcMNPV IE-1 as a transcriptional activation domain has been extensively studied in recent years. It was demonstrated that the N-termini of AcMNPV and OpMNPV IE-1s contain independent acidic transactivation domains, although sizes of the domains were mapped differently in the first 222 N-terminal amino acids (Kovacs *et al*1992; Rodems *et al.*, 1997; Slack and Blissard, 1997; Forsythe *et al.*, 1998;). Slack and Blisard (1997) identified two N-terminal acidic activation domains of the AcMNPV IE-1. One was mapped at amino acids 1-125, while the other at 168-222. Both domains contain net negative charge. These data are similar to those identified in this study. A comparison of eleven IE-1 amino acid sequences, including the AcMNPV IE-1 showed that the N-terminal residues are not conserved but rich in acidic features and dominant in negative charge. Although in most cases sizes of the acidic region Is are bigger than those of the region IIs, and the sizes between each of the region Is and each of the region IIs may be different from one

another, the patterns of the net negative charge in the two identified acidic regions of each IE-1 are similar to those of AcMNPV IE-1 (Table 2.1). This appeared to suggest that IE-1 contains two acidic activation domains at its N-terminus as identified in the case of the AcMNPV and OpMNPV IE-1s (Slack and Blissard, 1997; Forsythe *et al.*, 1998). SwissPort database search revealed that the N-terminal 47-120 amino acids of the deduced TnSNPV IE-1 protein significantly matched the SET-domain of transcriptional regulators, although it was not the case for other IE-1s. The SET-domain has been found in more than 40 transcriptional regulation proteins from yeast to mammals. It is involved in modulating transcriptional activities as well as protein-protein interactions (Jenuwein *et al.*, 1998; Cui *et al.*, 1998). The position of the SET-like domain in the TnSNPV IE-1 is similar to that of one of the two acidic activation domains mapped previously in the case of the AcMNPV IE-1 (Slack and Blisard, 1997; Rodems *et al.*, 1997). This suggests that the SET-like domain of the TnSNPV IE-1 may serve as an acidic activation domain as did one of the AcMNPV IE-1 domains. In addition, it was reported that several basic amino acids were clustered together consisting of a boundary of the first N-terminal acidic region in the case of the OpMNPV and AcMNPV IE-1 (Theilmann and Stewart, 1990; Slack and Blisard, 1997). The clustered basic amino acid region was suggested to be a transcriptional inhibitory domain that may be negatively involved in ie-1 transcriptional regulation (Slack and Blisard, 1997). It appeared that all the IE-1s contain a cluster of basic amino acids at the border of its acidic region (Fig. 2.5). The proposed functional role of the clustered basic amino acids may be in common among IE-1s.

On the other hand, the most recent study (Pathakamuri and Theilmann, 2002) demonstrated that the N-terminal acidic activation domain contains an essential replication domain, but the functional roles of both domains are independent and separable from one another. This study suggested that since IE-1 binds to *hrs*, which are known as baculovirus replication origins, as a dimer, IE-1 may be involved in viral replication as an origin binding protein to interact with the other replication proteins to form the replication complex. However, it appeared that the acidic domain for replication could not be exchangeable for either AcMNPV or other heterologous counterparts. This conflicts with results from other studies which showed that native acidic activation domains for replication activation can be functionally replaced by heterologous counterparts (Cheng *et al.*, 1992; He *et al.*, 1993; Hu *et al.*, 1999; Li *et al.*, 1998). Thus, whether or not the independent acidic replication domain of the OpMNPV IE-1 is conserved in other IE-1s, or only specific to the OpMNPV IE-1, remains to be investigated.

Previously, the C-terminal half of the AcMNPV IE-1 has been identified as being involved in oligomerization and DNA binding (Rodems *et al.*, 1997 and Olsen *et al.*, 2001). Recently, van Strien *et al* (2000) compared the seven IE-1 amino acid sequences and found that the deduced proteins were comparatively conserved at their C-terminal two third regions. A comparison of eleven deduced baculovirus IE-1 protein sequences revealed that the sequence conservation at the C-terminal two thirds of the proteins mainly centered in four separate regions (Fig. 2.3). The region IV contains a helix-loop-helix-like domain that is involved in IE-1 oligomerization and DNA binding in the case

of AcMNPV (Rodems *et al.*, 1997 and Olsen *et al.*, 2001). However, no significant match to any class of identified DNA binding domains were found within either four conserved regions, or entire C-termini, compared with entries in SwissProt database. Thus, apart from the helix-loop-helix-like domain, the functions of these conserved regions need to be further determined.

Previously, most analyses of baculovirus phylogenies were based on different genes such as *polyhedrin*, *dnapol*, *egt*, *gp41*, *chitinase*, *cathepsin*, *lef2*, *gp37*, *ie-1* and *vp39* (Bulach, *et al.*, 1999; Kang, *et al.*, 1998; Chen, *et al.*, 1999; Jin, *et al.*, 1999; Clarke, *et al.*, 1996; Li, *et al.*, 2001; Liu and Maruniak, 1999; van Strien *et al.*, 2000; Hayakawa *et al.*, 2000). However, these analyses have produced some conflicts between phylogenies based on different genes (Herniou *et al.*, 2003). In particular, *polh* phylogeny often disagrees with other gene phylogenies (Clarke, *et al.*, 1996; Kang, *et al.*, 1998). To resolve the conflicts among phylogenies of different genes, entire genomes have been used to reconstruct baculovirus phylogeny. Based on complete genome sequences several different approaches have been explored to infer baculovirus phylogenies, such as gene order, gene content, combined analysis of all shared genes in baculovirus genomes and comparison of complete genome sequences. With regard to the first two, each has some disadvantages for phylogenetic inference, such as lack of quantitative estimate and long branch attraction artifacts, while last two represent the most advanced and robust approaches to infer baculovirus phylogenies (Herniou *et al.*, 2001; 2003). However, the phylogenetic analysis of individual genes can still be used to bring some insight into the

baculovirus phylogenies, as did the previous analyses using different genes for inference of baculovirus phylogenies, especially when there is a lack of entire genome information.

In previous attempts to study the TnSNPV phylogeny the virus polyhedrin protein and a 2.97kb region, containing *p26*, *p10*, AcMNPV ORF-29, *lef 6* and the C-terminal portion of *p74*, were employed separately to infer the phylogenetic relationship of TnSNPV among baculoviruses (Fielding and Davison, 1999; Fielding *et al.*, 2001). The following data sets were used in these studies: 1) sequence alignment of a single gene (*polh*) which provided limited phylogenetic information because of the small size of its amino acid sequence and high sequence conservation (Bulach *et al.*, 1999); 2) alignment comparisons of five deduced protein sequences (making up a total amino acids of 849 in size only) from TnSNPV; 3) the gene order of the five genes in a 2.97 kb fragment compared with the corresponding regions of four other NPVs. The studies suggested that TnSNPV is a member of group II NPV. The previous phylogenetic estimates were mainly based on comparing several deduced protein sequences, and were not supported by bootstrap analyses and, therefore, TnSNPV phylogenies could not be well resolved by such analyses, especially in the divergent group II NPVs. In contrast, *ie-1* gene, as one of six essential genes involved in viral DNA replication in AcMNPV and OpMNPV (McDougal and Guarino, 2000), is present in all the lepidopteran baculovirus genomes. The conservation of these essential genes in baculovirus genomes suggested that a common mechanism of the viral DNA replication cannot be substituted by host genes (Hayakawa *et al.*, 2000) and, therefore, the genes can be preserved during evolution. By contrast, some genes, such as *p26*, early *gp64* gene and *p10*, are not conserved in all the

baculovirus genomes (Herniou *et al.*, 2001; Luque, *et al.*, 2001) and may not be suitable for phylogenetic studies because genes, such as the above, may not reflect whole baculovirus evolutionary history. On the other hand, the TnSNPV IE-1 has much longer amino acid sequence (739 aa) than that of *polh* (245-250 aa). Unlike *polh* with invariance of sequence in more than half among its homologues, the IE-1 contains both poorly conserved and comparatively better conserved amino acid sequences. In addition, the XcGV *ie-1* gene provides an appropriate outgroup taxa since GVs are distinct from NPVs.

Recently, van Strien *et al.* (2000) conducted a phylogenetic analysis based on alignment of the C-terminal part of IE-1 proteins, but the bootstrap analysis of the full length proteins was not included in the study. To define the TnSNPV phylogeny in deep branch of group II baculoviruses, two data sets, based on alignments of the deduced full length IE-1s and the C-terminal two third conserved portions of the proteins, were combined to further explore the IE-1 phylogenies. The topologies of two trees constructed with the above two data sets were consistent with one another. Both trees each separated NPVs into groups I and II, and resolved group II NPVs into two subgroups (Fig. 2.6a,b). This is in agreement with previous baculovirus phylogenies based on two individual gene analyses (Bulach *et al.*, 1999). The phylogenetic relatedness between the TnSNPV IE-1 and its homologues (SeMNPV and McNPV IE-1s) of subgroup II-A was well supported by both trees with high bootstrap values and higher homologies compared between these IE-1s (Table 2.3a,b). Interestingly, it appeared that the weak bootstrap signals of branching in some areas of tree A were slightly reinforced in tree B (Fig. 2.6). This

implied that the poorly conserved amino acid sequences in the N-terminal first third portions of the full length proteins may produce a slightly negative impact on the confidence level of branching in some areas of tree A. However, the two phylogenetic estimates based on the two trees consistently supported the inference of TnSNPV phylogeny among baculoviruses. The analyses further confirm that TnSNPV is a member of group II NPVs and suggested that the virus is most closely related to the members of subgroup II-A SeMNPV and McNPV. In addition, although branching in the same areas of the trees for separation of group I and II and of subgroup II-A and II-B is weakly supported by the low bootstrap figures, the TnSNPV IE-1 protein appeared to share the most recent common ancestor with the subgroup II-A NPV homologues, and so did group I and II. Since relationships within group II NPVs are not well resolved (Herniou *et al.*, 2003), analysis to further resolve phylogenetic relatedness of baculovirus species in this group should be done when more *ie-1* gene sequences become available. In addition, owing to a lack of the IE-1 sequence data from subgroup II-C defined earlier (Bulach *et al.*, 1999), what the relationship exists between TnSNPV and the members of group II-C, remains to be established.

It is worth noting that, although the phylogenetic study of TnSNPV based on its IE-1 phylogeny suggested that TnSNPV may be a member of subgroup II-A, to accurately infer the virus phylogenies, analyses based on all combined genes shared between available baculovirus genomes, or comparison of complete genome sequences, need to be done.

Chapter 3

Experimental Prediction of a DNA-Protein Binding Region of the *Trichoplusia ni*

Single Nucleocapsid Nucleopolyhedrovirus IE-1

3.1 Abstract

The interaction of IE-1 (immediate early gene 1) with homologous region (*hr*) enhancer sequences is required for *hr* enhancer-dependent transactivation of IE-1 in baculovirus early gene regulation. To define which IE-1 region or domain directly binds to a *hr* enhancer, the *Autographa californica* (Ac) and *Trichoplusia ni* (Tn) nucleopolyhedrovirus (NPV) IE-1 deletion truncations fused with 6xHis-Tag of pET expression vector were expressed in the *E. coli* and purified. The IE-1 truncations contain a basic helix-loop-helix (b-HLH)-like domain. An electrophoretic mobility shift assay revealed that the AcMNPV IE-1 deletion truncation containing the C-terminal 130 amino acids can bind to a 60 bp repeat within *hr5*, directly and specifically. Due to the presence of highly conserved primary and secondary structures in the truncation regions of eight NPV IE-1s, the binding result suggested that the same protein-DNA binding activity may remain in the corresponding regions of the TnSNPV and other compared IE-1s. In addition, it was also suggested that it is likely that the b-HLH-like domain may be mainly responsible for the direct DNA binding. This study also provided protein sources for further Nuclear Magnetic Resonance (NMR) and crystallography assays.

3.2 Introduction

The immediate early gene 1 (IE-1) is the principle transregulator for gene expression of *Autographa californica* nucleopolyhedrovirus (AcMNPV). Its transcriptional activation may occur in both sequence-independent (Blissard *et al.*, 1992) and dependent manners (Guarino and Summers, 1986b). The former requires only a basal promoter motif. In this case the specific DNA sequence appears not to be involved in recognition by IE-1. The latter is believed to be associated with enhancer elements for activation of promoters. IE-1 also negatively regulates transcription in the case of *pe38* and *ie-2* (Leisy *et al.*, 1997). On the other hand, it was suggested that IE-1 may also contribute directly or indirectly to viral DNA replication (Kool *et al.*, 1993; Lu and Miller, 1995).

Homologous region (*hr*) enhancer sequences ranging in size from 30 to 800 bp are dispersed throughout the AcMNPV genome (Ayres *et al.*, 1994; Cochran and Faulkner, 1983). They may function not only as origin of viral DNA replication (Kool *et al.*, 1993; Leisy and Rohrmann, 1993) but also as enhancer elements for early gene expression (Rodems and Friesen, 1993 and 1995). IE-1-mediated transcriptional activation can be significantly amplified when the promoter is *cis* linked to homologous region (*hr*) enhancer elements from AcMNPV (Guarino and Summers 1986b; Guarino *et al.*, 1986a; Pullen and Friesen 1995b), whereas interaction of IE-1 with these sequences is essential for *hrs* to function as transcriptional enhancer (Leisy *et al.*, 1995). Two previous studies were involved in identifying minimal sequences for DNA binding and enhancement of early gene expression. One suggested that a 24 bp palindrome within a 60 direct repeat

(DR60) was the smallest DNA binding unit, although not functioning sufficiently as an enhancer element (Guarino and Dong, 1994). The other study indicated a 28-mer palindromic repeat was a minimal sequence for orientation- and position-independent enhancer activity (Rodems and Friesen 1995). Both palindromes are located in the left half of *hr5*. Current evidence shows that IE-1 binds to the 28 bp repeats as a dimer and both half-sites of the repeats are required for the optimal interaction of IE-1, but a single half of the repeats is not sufficient for the proper interaction of IE-1 with the 28-mer enhancer repeats (Rodems and Friesen 1995).

Many transcriptional activator proteins have two domains: one involved in regulation of other genes, and the other in the DNA-binding function (Ptashne, 1988; Mitchell and Tjian 1989; Brent and Ptashne 1985). The functional mapping studies of the AcMNPV IE-1 revealed that its N-terminus contains two transactivation domains, residue 8 to 118 and 168 to 222 (Fig. 3.1A), that are dispensable for DNA binding (Kovacs *et al.*, 1992; Rodems *et al.*, 1997; Slack and Blissard 1997), while the C-terminal half participates in DNA binding but not in transactivation (Fig. 3.1A) (Kovacs *et al.*, 1992; Rodems *et al.*, 1997). It was also reported that the basic-helix-loop-helix (b-HLH)-like motif at the C-terminus of IE-1 plays a critical role in the protein-DNA interaction and oligomerization that is required for IE-1 transactivation and contributes to the DNA-binding (Olson *et al.*, 2001; Rodems *et al.*, 1997). The basic-helix-loop-helix (b-HLH) domain was also identified in other transcriptional activators (Murre *et al.*, 1989). It plays the role in creating dimer combination by interaction between the amphipathic helices. This is an efficient mechanism for regulation of gene expression (Robinson and Lopes, 2000).

In an attempt to define the role of the highly conserved C-terminal half region of the *Trichoplusia ni* single capsid nuclear polyhedrosis virus (TnSNPV) IE-1 in DNA-protein binding, and to provide information for studying the structure of the possible binding region by NMR or crystallography assay, N- and one half of C-terminal deletion truncations of the AcMNPV and TnSNPV IE-1s were constructed and expressed in the pET expression system (Novagen). The advantage of this approach was not only to eliminate the possible influences from the other potential binding residues or regions of the AcMNPV IE-1, but also to reduce the impact of possible cellular binding factors on DNA binding (Guarino and Dong, 1991) by protein purification. Unlike previous binding studies, the focus of this study was to identify an independent direct binding region other than determining which residues and regions contribute to DNA binding. However, due to lack of information of the TnSNPV *hrs*, which were thought to be a common feature in baculovirus genomes (Chen, *et al.*, 2000, 2001, 2002; Chapter 1), the AcMNPV IE-1 truncation was used as a model to experimentally predict the role of the corresponding region of the TnSNPV IE-1 in DNA-protein binding. In addition, the relationship between two independent enhancer elements (a 28-mer palindrome and a 60 bp direct repeat) was also discussed.

3.3 Materials and Methods

3.3.1 Prediction of the C-terminal secondary structures of NPV IE-1s

PSIPRED server was used to generate the predicted secondary structures of NPV IE-1s. The secondary structures of the relevant C-terminal regions were manually aligned

according to the corresponding amino acid sequence alignment of NPV IE-1s (Wang *et al.*, 2001) generated from ClustalW (Thompson *et al.*, 1994).

3.3.2 Plasmid constructs

The N- and C-terminal deletion mutants used in this study were generated by PCR using a AcMNPV-IE-1 plasmid kindly obtained from Dr. Knebel-Morsdorf (University of Cologne, Germany) and a TnSNPV IE-1 fragment (Wang *et al.*, 2001) as templates. The deletion-specific primers used in PCR were as follows: AcIE453/421F, 5'-cacacacaggatccAATTTAAAAAAGGTTAAGAAG GAG-3'; AcIE582R, 5'-tctctctcaagcttTTAATTAAATTCGAATTTTTTATATTTAC-3'; TnIE647/310F, 5'-cacacacaggatccAACGATGAACGCGTGAC-3'; TnIE739R, 5'-ctctctctaagcttTTAAAATTTTTTATATAGATAATTAC-3'. A *Bam*HI and a *Hind*III sites (underlined in the deletion-specific primers above) were introduced into the 5' and 3' ends of the mutants by PCR using *pfu* DNA polymerase (Promega), respectively. The PCR products were purified using a PCR purification kit (Qiagen) and digested with *Bam*HI and *Hind*III endonucleases according the manufacturers' instructions (Promega). The standard cloning techniques (Sambrook *et al.*, 1989) were used to insert the digested PCR products into the corresponding sites of the pET28a expression vector (Novagen). The resulting plasmids were confirmed to be in frame by sequencing (Wang *et al.*, 2001) using both T7 promoter and terminator primers. Predicted sizes of the IE-1 deletion truncations, AcΔ1-452 and TnΔ1-646, and their amino acid sequences, are shown in Fig. 1A and B.

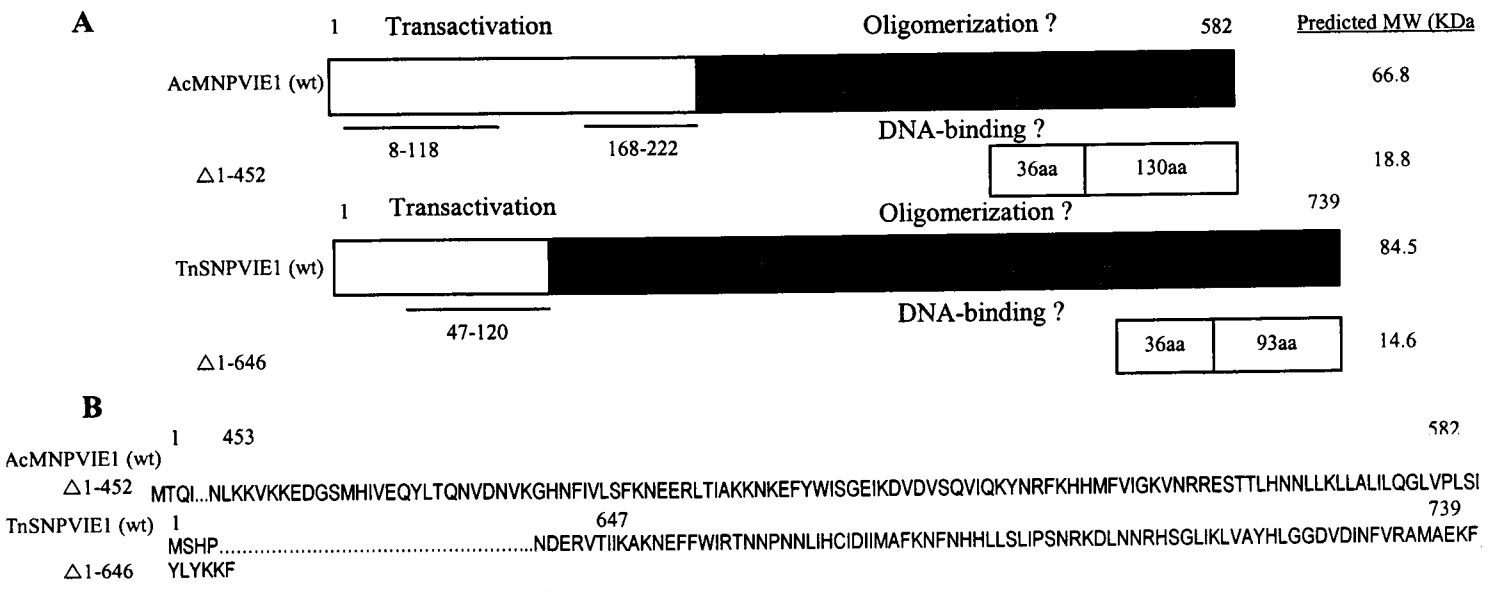


Figure 3.1 Schematic representation and amino acids sequences of the wild type IE-1s and their truncations. (A) Functional structures of the IE-1s and their 6xHis-tagged truncation fusions. Transactivation domains of AcMNPV (residue 8 to 118 and 168 to 222) and TnSNPV IE1s (residues 47 to 120) (see the text) are within their N-terminal half and unmapped DNA-binding and oligomerization domains within their C-terminal half. A N-terminal His-tag was fused with IE1 truncations. The sizes of the full length proteins and their truncation fusions in MW are indicated. The entire His-tag configuration and the two truncations have 36, 93 and 130 amino acids, respectively. The symbol Δ and following numbers denote the deleted amino acids from the IE-1s. (B) The amino acid sequences of the wild type IE-1s and their truncations. The different dashed lines represent the amino acids omitted and removed from the wild types and truncations, respectively. The positions of the amino acids are indicated in numbers.

3.3.3 Protein expression and purification

All the procedures for the protein expression were followed according to the manufacturer's instruction (Invitrogen). Briefly, all the plasmids were transformed into *E.coli* BL21(DE3) pLysS strain. The cells were propagated at 37°C until OD₆₀₀ reached 0.6. The following protein induction of cells was carried out at 18°C. A final concentration of 1mM IPTG was used for induction of the proteins. The samples were collected before and after 1h, 2h and 3h induction and analyzed by a 12% polyacrylamide gel according to a standard protocol (Sambrook *et al.*, 1989). The buffer components used for the protein purification under the denaturing condition were as follows: 1) Binding buffer: 8M urea, 20 mM sodium phosphate and 500 mM sodium chloride (pH to 7.8); 2) Wash buffer 1: 8M urea, 20 mM sodium phosphate and 500 mM sodium chloride (pH 6.0); 3) Wash buffer 2: 8M urea, 20 mM sodium phosphate and 500 mM sodium chloride (pH 5.3); 4) Elution buffer: 8M urea, 20 mM sodium phosphate and 500 mM sodium chloride (pH 4.0). Protein purification was carried out following the manufacturer's instruction (Invitrogen). Briefly, 4 ml of the resin was prepared by washing with H₂O three times. The 100 ml cell pellets for each sample collected after 3h induction were lysed in a 10 ml urea denaturing buffer and sonicated using six 10-second bursts at a medium intensity setting with a 10-second cooling period between each burst. The lysates were centrifuged at 3000 x g for 15 min at 4°C. The 5 ml supernatant aliquot of each sample was mixed with the 2ml prepared resin. The mixture was gently rocked for 30 min and then the resin was settled with low speed centrifugation (800 x g). The supernatant was carefully aspirated. The procedure was repeated with a second 5 ml aliquot. The lysate-resin mixture (supernatant) was loaded into an empty column,

followed by three washes. The His-tagged fusion proteins were eluted by applying 5 ml of denaturing elution buffer with pH 4.0. The purified protein fractions were confirmed by SDS-PAGE. The fractions were dialyzed and refolded using 3500 MW cut off dialysis tubing (Pierce) against 10 mM Tris, pH 8.0 with 0.1% Triton X-100 overnight at 4°C. The dialyzed protein fractions were stored at 4°C.

3.3.4 Western blot assay

The supernatants and precipitants of the purified protein fractions were separated by centrifugation (10000 x g) for 10 min and resolved onto a 15 % polyacrylamide gel. The cell lysates obtained from the *E. coli* strain, which contained pET vector only, were used as negative controls. The proteins on the gel were transferred onto a nitrocellulose membrane, blocked with blocking buffer, washed and detected by Ni-NTA AP conjugate according to the manufacturer's protocol (Qiagen). Ni-NTA conjugates consisted of Ni-NTA coupled to calf intestinal alkaline phosphatase. They can be used for direct detection of any fusion proteins with 6xHis tag. The detection was based on the affinity of nickel-nitrilotriacetic acid (Ni-NTA) metal-affinity chromatography matrices to 6 consecutive histidine residues which were tagged with the N-termini of the target proteins in this study. The reliability of the technology for 6xHis tagged protein detection has been recommended by the manufacturer's handbook (Qiagen). Although the positive control was not available, the results from the detection were positive and reliable, based on the sizes of the truncations, strong positive signals and the indications of the negative controls (the cell lysate obtained from the propagated *E. coli strain* containing a pET

vector only) which were similar to the test samples but lacking the target proteins (Ac Δ 1-452 and Tn Δ 1-646).

3.3.5 Radio-labeling of the probe and electrophoretic mobility shift assay (EMSA)

The complementary sequences of a 60 bp repeat within the left half of the AcMNPV *hr5* were synthesized and annealed together. The SP1 21 bp oligo used for a positive control was provided from a gel shift kit (Promega). The two oligos (the 60 bp repeat and SP1 21 bp oligo) were end labeled with ^{32}P using T4 polynucleotide kinase and purified by G-25 spin columns according to the manufacturer's protocol (Roche). The gel shift assay was performed using the gel shift kit, following the protocol recommended by the manufacturer (Promega). The HeLa nuclear extract provided from the kit and the SP1 21 bp probe were used as a positive control in the experiment. A purified fish viral protein, cloned into the same sites of the pET28a vector as those of the truncation Δ 1-452, was used as a negative control. 1 μl ^{32}P labeled 60 bp probe (2 pmol/ μl) was incubated with 3 μl of the fish viral protein. The purified AcMNPV IE-1 truncation (Δ 1-452) was centrifuged at 10000 x g for 10 min at 4°C and the supernatant (soluble fraction) was used in EMSA. Due to the low concentration of the soluble fraction of the purified truncation Δ 1-452, 6 μl of the fraction and 1 μl ^{32}P labeled 60 bp probe (2 pmol/ μl) were mixed in the reaction. The amounts of the cold specific oligos used in the competition assay were 2 μl SP1 21 bp oligo (2 pmol/ μl) for the positive control and 1.5 μl and 2 μl 60 bp oligo (2 pmol/ μl) for Δ 1-452, respectively. The 2 μl non-specific AP2 oligo (2 pmol/ μl) was incubated with HeLa nuclear extract and the same amount of the cold non-specific SP1 oligo with Δ 1-452. The control and Δ 1-452 groups (Fig. 3.5A and B) were

resolved in a 4 % and 6% polyacrylamide gel in 0.5x TBE buffer, respectively, and ran for 20 min at 350V. The gels were covered with plastic wrap and exposed to X-ray film overnight at -70°C.

3.4 Results

3.4.1 Prediction of the secondary structures and the experimental design

To identify the DNA-protein binding region of the TnSNPV IE-1, the Δ 1-646 and Δ 1-452 deletion mutants of the TnSNPV and AcMNPV IE-1s were constructed and expressed. Initially, one of the applications for the AcMNPV IE-1 truncation Δ 1-452 was designed to be a positive control. However, due to lack of information of the TnSNPV *hr* repeats, which may bind to the Δ 1-646 of the TnSNPV IE-1, prediction of the secondary structures of eight NPV IE-1s was conducted. It was found that the structures of half of the C-termini of IE-1s are highly conserved (Fig. 3.2A), especially in the previously suggested basic-helix-loop-helix (HLH)-like domain (Rodems *et al.*, 1997) in which the predicted secondary structures are almost completely conserved (Fig. 3.2A). Therefore, it was decided to employ the AcMNPV IE-1 truncation Δ 1-452, instead of the Δ 1-646 of the TnSNPV IE-1, to carry out the DNA-protein binding assay. The purpose of this was not only to identify whether the AcMNPV IE-1 truncation may directly bind to the AcMNPV enhancer element, but also to experimentally predict the corresponding binding region of the TnSNPV IE-1.

A

AcMNPV	-LIFNNYKYHDNI---- <td>472</td>	472
BmNPV	-LIFNNYKYHDNI---- <td>477</td>	477

CfMNPV -LIFSRVGHQ-----RDNNADSLKVKVKE-DGNRLLVEQYMS 450
 OpMNPV -LIFNRYGR-----RDNNADALKVKVKE-DGNRLLVEQYMS 450
 HzSNPV -SDEFTYKYGSVVRLLFK-ESIHTNALLKIKRE-TGHAGLIDNYLE 544
 SeMNPV EKSNTYKYGCIARQFYDPTHKGVKLFKVKKE-NGSTKLIENYLN 596
 LdMNPV -DCYINVKYACIVRLLYDEQDKRIANLLKIKKPGACTAELVEHYLN 453
 TnSNPV KSLQFTYKFTSVARVLFDETDGIVNKLKFKVKKE-PGSVAMIEDYLO 629

AcMNPV C----C-CCCCHHHHHHHCCC-CCCHHHHHHHHH
 BmNPV C----C-CCCCHHHHHHHCCC-CCCHHHHHHHHH
 CfMNPV -----CCCCHHHHHHHCCC-CCCHHHHHHHHH
 OpMNPV -----CCCCHHHHHHHCCCC-CCCHHHHHHHHH
 HzSNPV EEEEEC-CCCCHHHHHHHCCC-CCCHHHHHHHHH
 SeMNPV HHHHHCCCCCHHHHHHHHCCC-CCCHHHHHHHHH
 LdMNPV EEEEECCCCCHHHHHHHHCCCCCHHHHHHHHH
 TnSNPV HHHHHCCCCCHHHHHHHHCCC-CCCHHHHHHHHH

AcMNPV QNVDN-VKGHNFIIVLSFK--NBERLTIANKNEFYWISGEIKD--V 513
 BmNPV QNVDN-VKGHNFIIVLSFK--NBERLTIANKNEFYWISGEIKD--V 518
 CfMNPV QNEND-ETSHNFIIVLQFGVNEERLTIANKNEFYWIAAEIKD--I 493
 OpMNPV QNEND-DTSHNFIIVLQFGVNEERLTIANKNEFYWIAAGEIKD--I 493
 HzSNPV ANOND-TTSNSFILINTK--MERITLIKNEPIELWISIIKD--I 585
 SeMNPV ACKER-FENYSFILITTK--SERITLIKNEEFLWISVIKD--I 637
 LdMNPV VCAKLPRDSQNFLLVTTK--NEERLTLVKNPRLVWISGVARD--I 495
 TnSNPV ACEKI-PNGNMFIMINTL--NERVTLIKAKNEFWIRTNPNPNI 672

AcMNPV HCCCC-CCCCEEEEEEEC--CCEEEEEECCCEEEEEHHHHH--C
 BmNPV HHHCC-CCCCEEEEEEC--CCHHHHHHCCCEEEEECCCC--C
 CfMNPV HHCC-CCCCEEEEECCCCCHHHHHHHHCCCEEEEECCCC--C
 OpMNPV HHCC-CCCCEEEEECCCCCHHHHHHHHCCCEEEEECCCC--C
 HzSNPV HHCC-CCCCEEEEEEC--CCHHHHHHCCCEEEEECCCC--C
 SeMNPV HCCCC-CCCCEEEEEEEC--CCEEEEEECCCEEEEEEECC--C
 LdMNPV HHHCCCCCCCCEEEEEC--CCHHHHHHCCCEEEEECCCC--C
 TnSNPV HHCC-CCCCEEEEEEC--CCEEEEEECCCEEEEEEECCCEEC

AcMNPV DVSQVIQKYN-RKHHMFVIGKNRRESTTLHNNLLKLLAIIQGL 558
 BmNPV DASQVIQKYN-RKHHMFVIGKNRRESTTLHNNLLKLLAIIQGL 563
 CfMNPV NYDDLKKYT-INVHVFRIINRRESTTWHNNLLKLLQILLQNL 538
 OpMNPV SYDDLKKYA-INVHVFRIINRRESTTWHNNLLKLLQILLQNL 538
 HzSNPV IAMDLIEKYK-IHTHVFNLNTRKEMNKHNGMIKLLSFYTSNL 630
 SeMNPV IVTDIINKYK-MYNHYVYNIINGNRKEINIRHNGMIKLLSNYTGGR 682
 LdMNPV CVGDIINKPDGQ-EHHVFKLNKYSRKELNRRHNGLLKLVSLYTSAA 541
 TnSNPV HCIDIIMAFK-NENHLLSLIPSNRDKLNNRHSGLIKLVAYHLGGD 717

521

543

AcMNPV CHHHHHHHH-HCCEEEEEEECCCHHHHHHHHCHHHHHHHHCCC
 BmNPV CHHHHHHHH-CCCCEEEEECCCCCHHHHHHHHHHHHHHHHHCCC
 CfMNPV CHHHHHHHH-CCCCEEEEECCCCCHHHHHHHHHHHHHHHHHCCC
 OpMNPV CHHHHHHHH-CCCCEEEEECCCCCHHHHHHHHHHHHHHHHHCCC
 HzSNPV CHHHHHHHH-CCCCEEEEECCCCCHHHHHHHHHHHHHHHHHCCC
 SeMNPV CHHHHHHHH-HCCEEEEEEECCCHHHHHHHHCHHHHHHHHCCC
 LdMNPV CHHHHHHHHCCCCEEEEEECCCCCHHHHHHHHHHHHHHHHHCCC
 TnSNPV CHHHHHHHH-CCCCEEEEEEECCCHHHHHHHHHHHHHHHHHCCC

AcMNPV VPLSDAITEAQKLN-**KYK**KFEFN----- 582
 BmNPV VPLSDAITEAQKLN-**KYK**KFEFN----- 587
 CfMNPV IRIDDYQQSNKGD**S**KFI**K**YKRL----- 560

3.4.2 Expression, purification and detection of the AcMNPV and TnSNPV IE1 truncations

The uninduced and induced samples (3h induction) of the truncations $\Delta 1-452$ and $\Delta 1-646$ were resolved on a 12% polyacrylamide gel (Fig. 3.3A). The evidence for expression was clearly exhibited on the gel. As expected, the sizes of the truncations were the same as predicted ones (Fig. 3.3A). The expression levels from the samples after 3h induction appeared not to be high. However, the purified fractions 3 and 4 of the $\Delta 1-452$ and 3,4, and 5 of $\Delta 1-646$ showed much higher intensity in the gel than those of induced samples (Fig. 3.3A, B), although only 6 μl of each fraction was loaded in the gel. The sizes of the purified truncations are the same as those of unpurified samples.

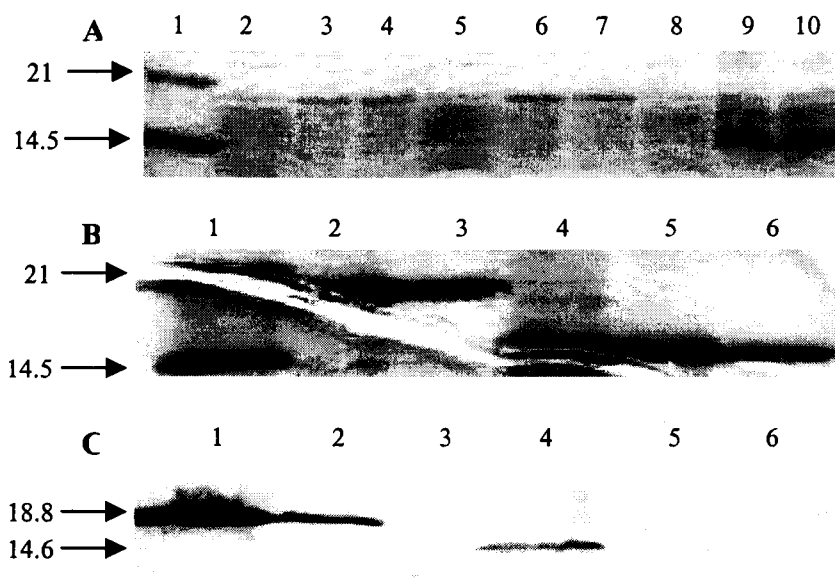


Figure 3.3 Expression, purification and detection of the AcMNPV and TnSNPV IE-1 truncations. (A) Expression of IE-1 $\Delta 1-452$ and $\Delta 1-646$. The uninduced sample of $\Delta 1-452$ is shown in lane 2 and 5, while the uninduced sample of $\Delta 1-646$ in lane 8. The

sample of Δ 1-452 collected from 3h induction was loaded in lane 3, 4, 6 and 7 (duplicates of the same sample), while the samples of Δ 1-646 from 3h induction in lane 9 and 10 (duplicates). (B) Purified fractions of the IE-1 truncations. The fraction 3 and 4 of Δ 1-452 are in lane 2 and 3. The fraction 3, 4 and 5 of Δ 1-646 are indicated in lane 4, 5 and 6. (C) Detection of the two purified truncations. The precipitants of both truncations are detected in lane 1 (Δ 1-452) and 4 (Δ 1-646) and supernatants are in lane 2 (Δ 1-452) and 5 (Δ 1-646), respectively. Nothing is observed in the negative controls (lane 3 and 6). The sizes of the truncations (KDa) are shown on the left side of the Figure C, while the molecular sizes of the protein standard in KDa on the left side of Figure A and B.

Most of the proteins produced from *E.coli* were washed away through the purification column and a strong signal of the 6xHis tagged truncation (Δ 1-452) was detected by Coomassie blue staining (Fig. 3.4).

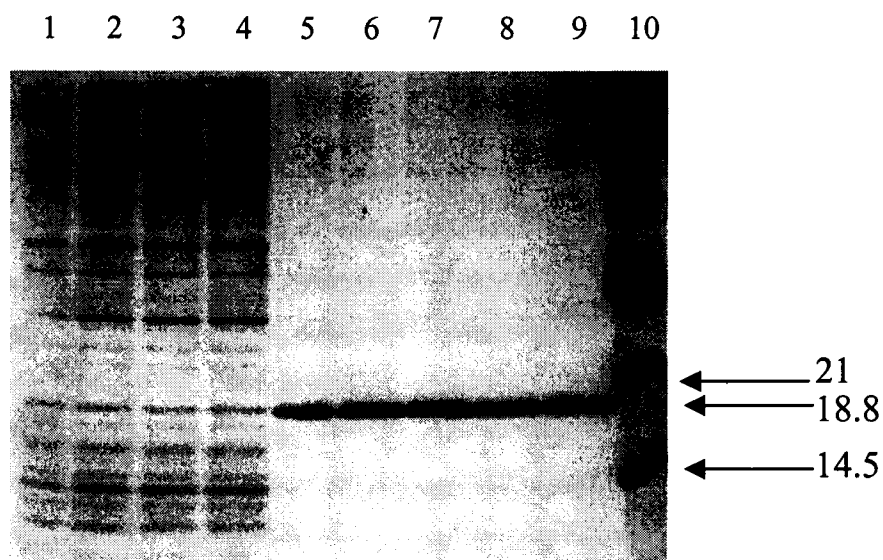


Figure 3.4 Efficacy of purification of the 6xHis tagged truncation (Δ 1-452). The flow-through (non-6xHis tagged proteins) of the wash was loaded on lane 1-4. The purified truncation (Δ 1-452) was eluted from the column in lane 5-9. The numbers represent the sizes of the protein ladder and the truncation in KDa.

Upon dialysis the supernatants and precipitants (aggregated protein fractions) of both purified samples were detected by western blot. Although soluble fractions of the truncations in the gel were hardly to be seen, as was observed by a low intensity of the Coomassie blue staining (data not shown), they were still detected in the western blot (Fig. 3.3C). In this experiment the concentrations of the proteins were not quantified but, according to the manufacturer's instruction (Promega), the two different amounts (1.5 μ l and 2 μ l) of the DNA probe with a low concentration (2 pmol/ μ l) were used for the DNA-protein reaction. A comparison of these two results could give some idea whether an excess of protein (if any) was in the binding reaction.

3.4.3 Gel shift assay

An EMSA was first performed on the control groups to verify whether the controls and the system can work as expected. A fish viral protein carried by pET28a vector was incubated with the 32 P labeled 60 bp repeat which was previously identified within the left half of AcMNPV *hr5* (Guarino and Dong, 1991). In this case there was no indication showing that a DNA-protein complex was formed on the gel (Fig. 3.5A). This indicated that the 60 bp repeat does not respond to the vector-encoded N-terminal fusion and the fish viral protein. In the positive control a transcription factor protein in the HeLa nuclear extract interacted with the 32 P labeled SP1 probe but not with AP2 oligo. The excess of

the cold SP1 oligo blocked the formation of the protein-DNA complex (Fig. 3.5A). To identify whether the truncation Δ 1-452 of the AcMNPV IE-1 binds to the AcMNPV *hr5*, the 60-bp direct repeat and the soluble fraction of the purified truncation Δ 1-452 were employed to perform the gel retardation assay (Fig. 3.5B). When the truncation Δ 1-452 was incubated with the 32 P labeled 60 bp probe, a single DNA-protein complex was formed and there was no non-specific complex detected. The addition of excess amount of the unlabeled 60 bp oligo with two different dosages inhibited the binding of the Δ 1-452 to the 32 P labeled 60 bp probe, whereas the addition of an excess dose of the cold non-specific SP1 oligo did not block the formation of the protein-DNA complex (Fig. 3.5B).

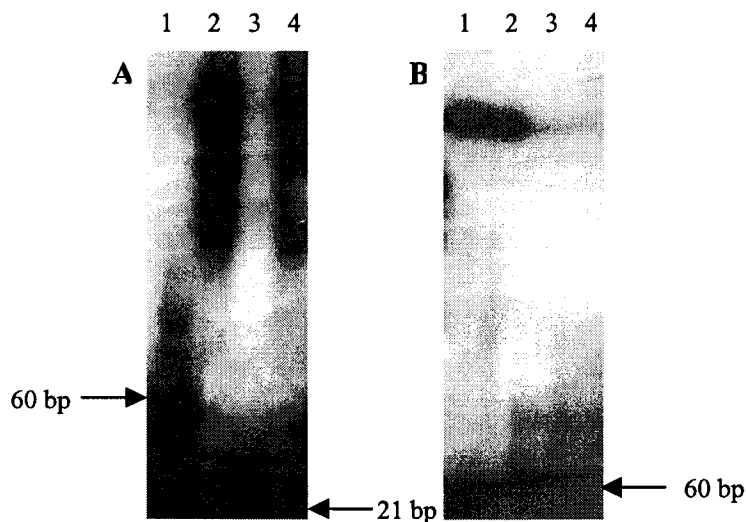


Figure 3.5 DNA-binding activities of the IE-1 truncation and controls. (A) EMSA of positive and negative controls. The 60 bp DNA probe did not interact with the fish viral His-tag fusion protein (lane 1). The SP1 oligo (21 bp) bound to a transcription factor in HeLa extract (lane 2). Prior to addition of the SP1 21 bp probe, excess amount of the

competitor (cold SP1 21 bp oligo) and a non-specific AP2 oligo (26 bp) to the transcription factor were added to the reaction mixtures (lane 3 and 4). (B) Gel-retardation assay of the IE1 truncation. The IE1 Δ 1-452 interacted with the 60 bp probe in the absence and presence of the non-specific 21 bp SP1 oligo (lane 1 and 2). The non-radio labeled competitor (60 bp) with the different excess doses was incubated with the IE1 Δ 1-452 before addition of the 60 bp probe (lane 3 and 4).

3.5 Discussion

Oligomerization is believed to be required for many HLH-containing transcriptional activators to bind to palindromic DNA recognition sites (Murre *et al.*, 1989; Massari and Murre, 2000; Patikoglou and Burley, 1997). This also occurs in baculovirus IE1 mediated transactivation (Olson *et al.*, 2001). Current evidence indicates that IE1 oligomerization occurs prior to *hr* binding (Olson *et al.*, 2001) and IE1 binds to *hr* as a dimer (Rodems and Friesen, 1995). Furthermore, the transactivation is directly correlated with the capacity of IE-1 to bind to the palindromic 28-mer, a *hr* enhancer element (Guarino and Dong, 1994; Kremer and Knebel, 1998; Rodems and Friesen, 1995). Therefore, the oligomerization of IE1 contributes to the DNA binding and transactivation function.

The previously functional mapping of the AcMNPV IE-1 demonstrated that deletion of 25 or more amino acid residues from the C-terminus abolished its DNA binding and enhancer-dependent transactivation function (Kovacs *et al.*, 1992). According to the

inspection in this study, the 25 amino acid (aa) deletion actually removed almost half of the IE-1 basic helix-loop-helix (b-HLH)-like domain. On the other hand, the insertions at the different position of the b-HLH-like domain greatly reduced or eliminated the capacity of each mutated IE1 to bind DNA as a dimer (Olson *et al.*, 2001). For example, a series of insertions constructed at the positions of the eight hydrophobic residues (Fig. in a failure of 3.2B) within the HLH-like domain, caused the loss of both homo- and heterodimers, thus resulting transactivation, while the replacement of basic residues 537 and 538 in the domain appeared to exhibit a similar impact on oligomerization and transactivation *in vitro* to that which occurred in the cases of the insertion mutations of the hydrophobic residues. This event in which multiple residues were involved in oligomerization in such a small region was not observed in other N- and C-terminal regions. The findings revealed that the b-HLH-like domain of the IE1 indeed is critical for oligomerization and DNA binding. The comparative analysis of the basic HLH-like domains in the present study indicated that the domains are highly conserved in the putative primary and predicted secondary structures of the C-terminal half of those IE-1s (Fig. 3.2A). This suggested the event occurring in the b-HLH-like domain of the AcMNPV IE1 may possibly happen in the case of other NPV IE1s.

As suggested previously, the multiple regions may be involved directly or indirectly in DNA-binding or oligomerization (Rodems and Friesen ,1995; Rodems *et al.*, 1997). To investigate whether a b-HLH-like domain-containing region could bind to DNA independently, a deletion mutant Δ 1-452 was used in the binding assay, (where only the b-HLH-like domain was identified to contribute to IE-1 oligomerization). Although it

was reported that an insertion at the conserved residue 425 exhibited a disruption of DNA binding (Rodems *et al.*, 1997), due to exclusion of the residue from Δ 1-452, the binding function of the b-HLH-like domain could be distinguished from that of the residue or other regions of IE-1 which are possibly involved in DNA-binding and oligomerization. On the other hand, several hydrophobic residues which are required for oligomerization and contribute to IE-1 stability (Olson *et al.*, 2001), were centered in the HLH-like domain which has about 50 aa in size. Such a structure was not found in other binding- and oligomerization-related regions of the AcMNPV IE-1 where only certain specific residues were identified as being directly or indirectly involved in oligomerization and DNA binding (Olson *et al.*, 2001; Rodems, *et al.*, 1997). The binding activity of Δ 1-452 showed that the HLH-like domain-containing truncation could bind to the 60 bp repeat directly and specifically. Therefore, it is more likely that the b-HLH-like domain could be mainly responsible for the direct binding. However, except for Δ 1-452, whether or not other regions of the IE1 could also bind to the repeat independently, was not excluded. The analyses of both the binding result and the structural comparisons of the C-terminal regions of the IE1s suggested that the binding activity of the AcMNPV IE1 deletion mutant (Δ 1-452) may remain in the corresponding region of other NPV IE1s, including TnSNPV IE-1. Furthermore, expression and detection of the TnSNPV IE-1 Δ 1-646 (93 aa) (Fig. 3.3) suggests that it is possible to further reduce the binding region down to a smaller size such as 93 aa. This study also provided the purified protein sources for identifying the HLH-like domain by NMR or X-ray crystallography studies.

The previous binding experiment of IE1 (Guarino and Dong, 1991) indicated that two protein-DNA binding complexes were formed when a ³²P-labeled 252 bp fragment consisting of the left half of *hr5* was used. However, when a single 60 bp repeat within the left half fragment was applied, only one complex was formed in the experiment. It was suggested that this was because IE1 protein bound simultaneously to one or two copies of a 60 bp direct repeat (DR60) contained in the *hr5* fragment. In this study to eliminate the impact of possible cellular binding factors on the binding reaction by protein purification, and using a 60 bp probe only, the result was consistent with the previous one.

Current evidence showed that a 28 bp palindrome was a minimal sequence for DNA binding (Rodems and Friesen, 1995). It represented the smallest functional unit of *hr5* capable of position- and orientation-independent transcription activation (Rodems and Friesen, 1993). However, a functional dissection of the left half (252 bp) of the *hr5* revealed that a conserved 24-bp palindrome within the 60 bp direct repeat (DR60) was mapped as a minimum sequence for DNA-protein interaction, but a complete copy of the DR60 was required for enhancer function (Guarino and Dong, 1994). The conclusion was different from that in which a 28 bp repeat was essential for DNA-protein interaction and enhancer-dependent transactivation of IE1 (Rodems and Friesen, 1995). In accordance with the inspection of the entire *hr5* sequence (Margot *et al.*, 1992) in this study, only half of the 28 bp palindrome, referred to as *hr5*-A, was found in the 252 bp fragment (Guarino and Dong, 1991) and its sequence differed from that of the 60 bp repeat (*hr5*-B) in which the 24-bp palindrome is present. According to the differences of

the locations and the sequences of 28 bp (*hr5-A*) and DR60 (*hr5-B*) in *hr5*, it appears that there are two different essential repeats in the left half of the *hr5* which are capable of DNA-protein binding and enhancement of transactivation. It will be interesting to investigate whether there is any functional difference between the two independent enhancer elements.

In summary, an independent DNA-protein binding region of the AcMNPV IE-1 which contains a b-HLH-like domain was identified. Although this result did not define the smallest protein binding unit of the IE-1, it indicated that there is a direct binding domain in the 130 amino acid region. This study also suggested that the HLH-like domain in the binding region could be mainly responsible for the DNA binding as several residues involved in oligomerization were centered in the domain. It will be interesting to investigate the functional role of the domain further by analyzing the domain fusion protein. Due to the structural similarity of the C-terminal binding region in AcMNPV and other NPV IE-1s, the protein-DNA binding reaction may occur, not only in the case of AcMNPV IE-1, but in the corresponding regions of the TnSNPV and other NPV IE-1s. The study also provided the purified protein source for studying the structure of the b-HLH-like domain of IE-1 by NMR or crystallography.

Chapter 4

Identification and Characterization of a Major Early-Transcribed Gene of *Trichoplusia ni* Single Capsid Nucleopolyhedrovirus Using the Baculovirus Expression System

4.1 Abstract

An early transcribed gene (*me-53*) of *Trichoplusia ni* single capsid nucleopolyhedrovirus (TnSNPV) was sequenced and identified. It has an open reading frame of 1146 nucleotides that encode a protein of 382 amino acids with a molecular mass of 45.2 KDa. Upon introduction of the gene and a green fluorescent protein (GFP) reporter gene into the baculovirus expression vector system (BEVS), the transcriptional analysis of the gene in two cell lines infected with the *Autographa californica* nucleopolyhedrovirus (AcMNPV) recombinant revealed that an early transcript can be detected at 1h postinfection (p.i.) until 12 h p.i., and a late one from 18 h p.i. up to 48 p.i. A unique early transcription start site was identified at two bases upstream of a conserved late transcription start site TAAG. The putative regulatory elements within the upstream regulatory regions of ten NPV *me-53s* were also identified and discussed. The protein analysis showed that zinc finger-like motifs at the C-termini of twelve baculovirus ME-53s are highly conserved with similar patterns. The phylogenetic analysis of these ME-53s indicated that the TnSNPV ME-53 is more closely related to the members of NPV subgroup II-A, *Spodoptera exigua* multiple (SeM) and *Mamestra configurata* (Mc) nucleopolyhedrovirus ME-53s. Due to two heterologous genes (TnSNPV *me-53* and GFP) introduced into the AcMNPV polyhedrin locus, the recombinant was applied to test a new model system for improvement of bio-insecticidal agents for insect pest control.

4.2 Introduction

Baculovirus gene expression is believed to be regulated in a cascade fashion in which each successive phase precedes the previous phase (Blissard and Rohrmann, 1990). An early phase occurs prior to viral DNA replication and late phase during or after the viral DNA synthesis (Blissard and Rohrmann, 1990). The early transcription is mediated by host RNA polymerase II (α -amanitin-sensitive) (Fuchs *et al.*, 1983; Hoopes and Rohrmann, 1991) and independent of viral replication and late gene expression (Friesen, 1997). Many baculovirus early promoters contain a conserved sequence CAGT at or near the transcription start site (Theilmann and Stewart, 1991 and 1992; Krappa and Knebel, 1991; Blissard and Rohrmann, 1989) and a functional TATA box (Blissard *et al.*, 1992; Theilmann and Stewart, 1991; Dickson and Friesen, 1991; Guarino and Smith, 1992). The two *cis*-regulatory elements, TATA and CAGT are essential for start site selection and initiation efficiency of early genes respectively, and may functionally work in a cooperative manner (Blissard *et al.*, 1992). On the other hand, some sequence elements in the baculovirus *gp64* may be involved in transcription initiation by a TATA-independent mechanism (Kogan *et al.*, 1995). Except for the early transcription, some of the early genes, with a conserved late transcription motif (A/G/T)TAAG (Friesen, 1997), such as the *me-53* and *p35* (Knebel *et al.*, 1993; Nissen and Friesen, 1989), can be recognized and transcribed by a α -amanitin-resistant virus-encoded RNA polymerase (Guarino *et al.*, 1998) in the late phase of viral infection, and others, such as the *ie-2*, *pe38* and *he65*, have the same conserved motif which has no function display during the late phase of infection (Carson *et al.*, 1991; Krappa and Knebel, 1991; Becker and Knebel, 1993). The