INTRODUCTION

1 The Baculovirus Expression Vector System

2 Baculovirus Structure and Molecular Biology

3 The Polyhedrin Gene Promoter of AcNPV

4 Aims and Objectives
1 THE BACULOVIRUS EXPRESSION VECTOR SYSTEM

The expression of heterologous genes in insect cells infected with recombinant baculoviruses has become a popular and powerful tool for the synthesis and analysis of a wide variety of eukaryotic and prokaryotic proteins (for reviews see Luckow, 1991; Jarvis and Summers, 1992; O'Reilly et al., 1992; Sridhar et al., 1994).

The baculovirus expression vector system (BEVS) uses a group of DNA viruses, the baculoviruses, that infect certain insects and insect cell lines cultured in-vitro. Late in the virus-infection process, the infected cell devotes much of its energy to the synthesis of a viral protein called polyhedrin that comprises 50% of the total infected cell protein. Polyhedrin is not essential for the infection process, so it is possible to replace the polyhedrin (polh) gene with a heterologous gene while retaining the polh regulatory signals. Infection of insect cells with such a recombinant baculovirus leads to the synthesis of the foreign protein to high levels.

Baculovirus-infected insect cells perform many of the post-translational and proteolytic processing events seen in higher eukaryotes (O'Reilly et al., 1992) including glycosylation, fatty acid acylation, amino-terminal acetylation, carboxy-terminal α-amidation, myristylation, and phosphorylation. In most cases, the recombinant proteins are targeted to their natural locations within the cell. Proteins containing signal peptides are recognized and properly cleaved before insertion into cellular membranes or secretion from the cell. Also, hetero- and homo-oligomeric assemblies have been demonstrated for a wide variety of proteins in baculovirus-infected cells. This, along with the high expression levels that can
be achieved, accounts for the immense popularity of BEVS. This system has been used in over 1,000 different laboratories throughout the world with more than 500 genes from viruses, bacteria, fungi, plants, and animals having been already expressed.

2 BACULOVIRUS STRUCTURE AND MOLECULAR BIOLOGY

BACULOVIRUS STRUCTURE

Baculoviruses are a diverse group of viruses found mostly in insects. The baculo portion of the name refers to the rod-shaped capsids of the virus particle. The capsids are usually 40-50 nm in diameter and 200-400 nm in length (Harrap, 1972b). Within the capsid, the DNA is condensed into a nucleoprotein structure known as the core. The DNA genome of a baculovirus is double-stranded, covalently closed, and circular (Summers and Anderson, 1972). The DNAs of the two baculoviruses commonly used as expression vectors - Autographa californica nuclear polyhedrosis virus (AcNPV) and Bombyx mori nuclear polyhedrosis virus (BmNPV) - are both approximately 130 kbp.

Nucleocapsids are synthesized in the nucleus of infected cells and are enveloped by one of two processes:

i) Nucleocapsids can bud through the plasma membrane into the extracellular fluid. This budded virus (BV) acquires a loosely fitting membrane envelope.

ii) Nucleocapsids may also acquire a "de-novo" envelope within the nucleus (Stoltz et al., 1973). These virions are then occluded within a crystalline protein matrix within the
nucleus, and such virus particles are called occluded virus (OV). The protein making up the crystalline protein matrix is polyhedrin and the polyhedral occlusion bodies are known as polyhedra (Harrap, 1972a).

VIRUS INFECTION CYCLE

In nature, insect larvae are infected with the baculovirus upon ingestion of polyhedra along with their food. The polyhedrin matrix dissolves in the alkaline midgut of the insects (Harrap and Longworth, 1974), releasing the embedded virions that enter the midgut cells by membrane fusion (Granados and Williams, 1986). The BV released by the infected midgut cells spread the infection to other tissues of the host via the hemolymph (Keddie et al., 1989). The infection process takes 5 to 7 days and culminates in the disintegration of the infected larva and release of polyhedra into the environment. The polyhedrin matrix protects the virions from environmental damage until they are ingested by another insect host.

In cell culture, the infection cycle has been characterized using AcNPV and the Sf21 cell line [derived from Spodoptera frugiperda pupal ovarian tissue (Vaughn et al., 1977)] as the model baculovirus-insect cell system. The infection cycle can be divided into three phases - early, late, and very late.

Early phase: The first 6 hours of infection constitutes the early phase and it is during this period that the cell is reprogrammed for virus replication. Infection in cell culture is mediated by BV that enter the cells by adsorptive endocytosis. Nucleocapsids migrate to the nucleus where the nucleoprotein core is released by the uncoating of the capsid within
Introduction

1 hour post infection (h pi), (Granados and Williams, 1986). Infected cells undergo cytoplasmic and nuclear changes during this period - cytoskeletal rearrangements occur, the host chromatin disperses, and the nucleus becomes enlarged.

Late phase: This phase extends from 6 h pi to approximately 20 h pi. During this period viral DNA replication, late gene expression, and BV production take place (Knudson and Harrap, 1976).

Very late phase: This phase, which begins around 20 h pi and corresponds to the production of OV, is microscopically characterized by the accumulation of polyhedra within the nucleus (Fraser, 1986). Also, large arrays of fibrous material, composed of a 10-kDa protein - p10 - accumulate in the nucleus. This fibrous material may play a role in the lysis of host cells (Williams et al., 1989). Lysis usually begins around 60 h pi and is complete by about 72 h pi.

BACULOVIRUS GENE REGULATION

Baculovirus genes are transcribed in three phases - early, late, and very late - in a regulated cascade, with gene products of one temporal class transactivating (directly or indirectly) transcription of the genes of the next temporal class (for a review see Blissard and Rohrmann, 1990).

Early gene transcription: Early genes have been defined as those that are transcribed in the absence of any viral gene expression. Expression of several early genes is dependent on the product of the ie-1 gene, IE-1, that is believed to enter the cell as a component of the virion (Guarino and Summers, 1986a). Induction of early promoters by IE-1 can be
enhanced by the presence of the \textit{ie-n} gene product and homologous region (hr) sequences, the latter being repeats of palindromic sequence present at five locations in the AcNPV genome (Guarino and Summers, 1986b). Early gene transcription is mediated by host RNA polymerase II (sensitive to \textit{\alpha}-amanitin) and no viral gene expression is required.

**Late and very late gene transcription:** Late and very late gene transcription is insensitive to \textit{\alpha}-amanitin (Huh and Weaver, 1990a) and involves a novel or modified RNA polymerase and virus specific factor(s). Late and very late gene transcription is dependent on early gene expression and on DNA replication - both cycloheximide and aphidicolin block transcription of late and very late genes (Huh and Weaver, 1990b); uninfected \textit{Spodoptera frugiperda} cells do not support late or very late gene expression when transfected with a reporter plasmid (Hoopes and Rohrmann, 1991). However, the basis for this dependence of late gene transcription on DNA replication is not known.

Late and very late promoters can be distinguished by their relative activities during the late and very late phases. Very late promoters display low activity during the late phase (6 h pi-20 h pi) but become highly active during the very late phase (20 h pi-72 h pi). On the other hand, late promoters are more active during the late phase.

The primary determinant of late and very late promoter activity is the pentanucleotide (A/G)TAAG, that is located at the transcription start point of all known late and very late genes (Possee and Howard, 1987; O’Reilly \textit{et al.}, 1992).

Five genes, designated \textit{lef-1}, \textit{lef-2}, \textit{lef-3}, \textit{lef-4}, and \textit{lef-5} (for late expression factor), are essential but not sufficient for late and very late gene expression (Li \textit{et al.}, 1993; Passarelli and Miller, 1993a, 1993b, 1993c). However, it is not known whether the \textit{lef} gene products
act at the level of replication, transcription, or translation.

3 THE POLYHEDRIN GENE PROMOTER OF AcNPV

The polh gene promoter has been the "workhorse" promoter of BEVS. It is a very strong promoter that is expressed very late in infection and is responsible for high levels of synthesis of polyhedrin.

DETERMINANTS OF PROMOTER ACTIVITY

The polh promoter of AcNPV (Fig. 1) has been extensively characterized by deletion and linker-scan mutation analyses. The transcription start point is at -50 relative to the translation start site (designated as +1) and lies within a highly conserved octanucleotide sequence TAAGTATT. The essential polh promoter has been defined (by experiments described below) as a 69 bp stretch (-1 to -69) with the TAAGTATT motif being absolutely essential for transcription initiation.

It was shown by Matsuura et al., (1987) that the full 5' leader sequence of the polh gene (-1 to -50) was required for maximum expression of foreign genes in the baculovirus system. In order to determine the effect of the integrity of the leader sequence on the expression of a reporter gene (the hemagglutinin gene of influenza A virus in this case), Matsuura et al. made a series of recombinant viruses with variable lengths of the leader sequence. Sequential deletions into the leader sequence upstream from the polh ATG
FIG. 1: Schematic representation of the AcNPV polyhedrin gene promoter. The essential polh promoter (→) extends from -1 to -69 (the translation start point is designated as +1). The transcription start point (℅) is at -50 and lies within a conserved, transcriptionally important TAAGTATT motif (●).
adversely affected reporter gene expression.

Possee and Howard (1987) constructed recombinant viruses with deletions in the promoter that progressively removed sequences upstream to the transcription start point. The effect of these deletions on polh promoter activity was judged by expression of the beta-galactosidase gene inserted in lieu of the polh coding sequences. A sequence 69 nucleotides upstream to the polh translational initiation codon was sufficient for maximum promoter activity. Likewise, insertions of varying sizes - 8, 95, and 785 nucleotides - into the EcoRV site upstream to the polh promoter (at -92) did not affect reporter gene expression. This data, together with that of Matsuura et al., (1987), suggests that the polh promoter resides within a 69 bp region (-1 to -69). It is pertinent to mention here that in all of these experiments the ORFs upstream to the polh promoter were intact. Therefore, the effect of such upstream sequence elements on polh promoter-driven transcription cannot be excluded.

Ooi et al., (1989) constructed a series of recombinant baculoviruses containing linker-substituted polh promoters attached to a reporter gene encoding chloramphenicol acetyl transferase (CAT) and tested for expression of the gene. The major determinant for promoter activity was narrowed down to the octanucleotide motif, TAAGTATT, at the transcription start point. Mutations within TAAGTATT blocked initiation of transcription from this site and resulted in a 2000-fold decrease in CAT activity. Mutations downstream from TAAGTATT and within the region specifying the untranslated RNA leader diminished transcription initiation and decreased CAT activity 2- to 20-fold. The half-lives of CAT RNAs were not affected by mutations in the untranslated RNA leader region; nuclear run-on analysis showed that these mutations decreased the rate of transcription initiation. Thus,
the low steady-state RNA levels in such linker-scan mutants were due to less efficient transcriptional initiation, and not due to an alteration in RNA turnover rates. Transcription initiation thus appeared to be the major means of polh gene regulation.

Ooi et al. proposed a model wherein the (A/G)TAAG sequence serves as an essential component for the recognition of both late and very late promoters by the virus-induced RNA polymerase activity. Additional factors (binding to the cis-acting sequences downstream to the (A/G)TAAG sequence) then dictate the differences in late and very late transcription.

Therefore, the major determinant of polh gene transcription is located at the transcription start point, and additional determinants are found between the RNA start point and the translation initiation codon.

SEQUENCE CONSERVATION OF THE POLYHEDRIN PROMOTER

The polh promoter sequence is relatively well conserved between otherwise distantly related baculoviruses (Rohrmann, 1986). Comparison of a number of polh genes from different baculoviruses has highlighted the presence of a 12 nucleotide consensus sequence spanning the mRNA start site - (A/T)(A/G)TAAGNA(T/A/C)T(T/A)T (the transcriptionally important TAAGTATT motif of the AcNPV polh promoter is part of this consensus sequence; the (A/G)TAAG motif is present at the transcription start point of all late and very late promoters). This 12 nucleotide consensus sequence is also present in the very late p10 genes that have been analyzed. Another sub-group of baculoviruses (granulosis viruses) also have the same consensus sequence upstream of the granulin gene (that encodes an
occlusion body protein of similar function to polyhedrin).

The promoters of the AcNPV polh gene and the other very late gene, p10, though sharing a similar organization, display only limited sequence homology (Roelvink et al., 1992). The polh promoter is composed of a 50 nucleotide 5' non-coding leader sequence and a short sequence of ~20 nucleotides upstream from the transcription start point. The AcNPV p10 promoter is composed of a 70 nucleotide 5' non-coding leader sequence and an additional 30 nucleotides upstream from the transcription start point. The 5' leaders of both promoters are extremely AT-rich (80%). A consensus nucleotide sequence (AATAAAGTA) is present that includes the transcriptionally important (A/G)TAAG motif present in all late and very late promoters. Interestingly, the p10 gene has a transcriptional pattern distinct from the other very late polh gene - the p10 gene is active at least 4 h earlier than the polh gene (Reoelvink et al., 1992). The nucleotide sequences responsible for this difference in transcriptional activity remain to be determined.

UNUSUAL STRUCTURE OF THE POLYHEDRIN PROMOTER

Transcription of the polh promoter may result from the virus-induced RNA polymerase recognizing the transcriptionally important TAAGTATT motif. The polh promoter, with a short motif at the transcription initiation site, is unusual in structure as compared to typical promoters recognized by nuclear eukaryotic RNA polymerases and bacterial RNA polymerases. Both bacterial and nuclear eukaryotic promoters usually consist of multiple non-contiguous blocks of sequence information present upstream or downstream to the transcription start point. In contrast, the structure of the polh promoter appears to be
similar to that of promoters recognized by RNA polymerases specific to yeast mitochondrial DNA and certain bacteriophages (T7 and T3), (Masters et al., 1987). These promoters also consist of short sequences located at the transcription start point. Yeast mitochondrial RNA polymerase, for example, recognizes and initiates at the sequence TATAAGTATT that is remarkably, although perhaps coincidentally, similar to the polh initiation sequence. On the basis of the promoters they recognize, Yang et al (1991) proposed that the virus-induced RNA polymerase and the mitochondrial RNA polymerase may be related.

Also, all the additional determinants of polh promoter activity are present in the region specifying the untranslated mRNA leader (Ooi et al., 1989) - another unusual structural feature of this promoter.

**THE POLYHEDRIN PROMOTER IS RECOGNIZED BY A NOVEL RNA POLYMERASE**

The polh promoter, like all late and very late promoters, is recognized by a novel, α-amanitin-resistant, baculovirus-induced RNA polymerase. Fuchs et al., (1983) monitored AcNPV-specific RNA synthesis in isolated nuclei of Spodoptera frugiperda cells in culture at different times post infection. They showed that early viral RNA synthesis was sensitive to 5 μg of α-amanitin per ml. During the course of infection this sensitivity decreased, and by 24 h pi RNA synthesis was completely resistant to α-amanitin. Their studies revealed that the α-amanitin-resistant transcription begins just after 6 h pi, simultaneous with the beginning of late phase of transcription. This novel polymerase was isolated at 24 h pi and shown to be a new, chromatographically and immunologically distinct form whose kinetics and response to divalent cations differed from those of the host RNA polymerases (Fuchs
et al., 1983; Yang et al., 1991).

Fuchs et al. proposed that early in infection, the viral genes are transcribed by host RNA polymerase II. One or more early genes code for a virus-specific α-amanitin-resistant RNA polymerase or factors that modify one of the host polymerases. The virus-induced or virus-modified polymerase then plays a major role in late and very late transcription including polh gene transcription.

**THE LEF GENES ARE ESSENTIAL FOR POLYHEDRIN PROMOTER ACTIVATION**

Passarelli and Miller (1993a) developed a method to identify baculovirus genes required for late and very late gene expression. Their method was based on the subtraction of clones from an AcNPV genomic library that was able to transactivate promoters of reporter plasmids in transient expression assays. Using this method they were able to identify and sequence five early genes, designated lef-1, lef-2, lef-3, lef-4, and lef-5 (for late expression factor) that are essential but not sufficient for late and very late gene expression (Li et al., 1993; Passarelli and Miller, 1993a, 1993b, 1993c). These lef-1, lef-2, lef-3, lef-4 and lef-5 genes are dispersed throughout the AcNPV genome and are located on 7.35-8.65, 2.2-2.6, 43.4-45.2, 57.6-58.8, 64.0-65.4 map units, respectively. However, the authors have not been able to determine whether the lef gene products act at the level of replication, transcription, or translation. In fact, experiments carried out by Kool et al. (1994) suggest that lef-1 and lef-2 may be directly involved in viral DNA replication. Since late and very late gene expression is dependent on viral DNA replication (Rice and Miller, 1986), lef-1 and lef-2 may, as a consequence, be indirectly involved in late and very late gene expression. It is
also possible that the lef genes affect the expression of other early genes that, in turn, are required for late gene expression.

A PUTATIVE NEGATIVE REGULATOR BINDS TO THE POLYHEDRIN PROMOTER

Recently, Etkin et al. (1994) have reported the identification of a 200-kDa host factor from Spodoptera frugiperda cells that binds specifically to the -72 to -86 region of the polh promoter of AcNPV. This binding activity was found in uninfected cells and also in cells during early stages of viral infection, but decreased by 18-24 h pi (concomitant with hyperexpression of the polh gene). The authors proposed that the host-encoded DNA-binding protein may act as a negative regulator and the virus may utilize this factor to control the differential expression of late versus very late genes. Binding of this factor to the polh promoter could inhibit transcription by the virus-induced RNA polymerase. Degradation of this factor by 18-24 h pi, therefore, would have a stimulatory effect on the polh promoter. These results, however, do not imply that this is the only mechanism governing very late expression of the polh gene.

4 AIMS AND OBJECTIVES

Progression through the AcNPV infection cycle is governed by a cascade of early, late, and very late gene transcription. Although one of the most intriguing aspects of baculovirus replication concerns the control of this viral transcription cascade, few of the host and viral
Introduction

Factors participating in the regulation of baculovirus gene expression have been identified.

The fundamental nature of the RNA polymerase(s) and other factors involved in transcription from late and very late viral promoters remains largely unknown. It is generally thought that an RNA polymerase switching mechanism is involved in the transition from early to late phase of transcription. This view is supported by biochemical evidence of a novel late phase-specific RNA polymerase and molecular evidence that the primary determinant of late and very late promoters is a (A/G)TAAG sequence present at the transcription start point. However, the nature of the new polymerase and the nature of the factors that recognize or distinguish between these unusual, late and very late promoters remains to be determined.

Although a lot of research has gone into the development of BEVS, very little is known about the mechanism governing transcription from the very important polh promoter. Regulation of the polh promoter is of considerable interest as this is a very strong promoter with an unusual structure, and is recognized by a novel RNA polymerase. Also, an understanding of the polh promoter could contribute to the development of a virus-free expression system that will naturally overcome the major problem of "secretory load" encountered in this system (Nakhai et al., 1991; Jarvis and Summers, 1992; Sridhar et al., 1993; Sridhar and Hasnain, 1993).

An understanding of polh promoter regulation requires knowledge of the protein factors involved and the sequence elements upon which they act. However, to date, no factor directly interacting with transcriptionally important regions of the polh promoter has been identified.
In an attempt to develop some understanding of how the polh promoter is regulated, we carried out studies aimed at the identification and characterization of factor(s) binding specifically to the promoter. The objectives of the study described in this thesis are as follows:

1. Identification of factor(s) interacting with the polh promoter,

2. Identification of sequence motif(s) important for binding,

3. Analyses of binding characteristics of factor(s) identified,

4. Determination of molecular mass of the factor(s),

5. Study of the effect of protein phosphorylation on DNA-binding activity, and finally,

6. Purification of the factor(s) to homogeneity.