CHAPTER I

INTRODUCTION

1. The Baculovirus Expression Vector System
2. Virus Structure
3. The Infection Process
4. Baculoviruses as Expression Vectors
5. Baculoviruses as Engineered Insecticides
6. Baculovirus Gene Expression and Replication
7. Structure of AcNPV Very Late Gene Promoters
8. Aims and Objectives
1. THE BACULOVIRUS EXPRESSION VECTOR SYSTEM

The baculovirus, *Autographa californica* nuclear polyhedrosis virus (AcNPV) which normally infects the larvae of the alfalfa leaf hopper, apart from causing disease in other hosts such as the fall army worm, cabbage looper and salt marsh caterpillar, has been harnessed extensively as a vector system to produce recombinant proteins of economic and academic interest. Over the last 10 years, these have become a very popular and effective system to produce foreign proteins in large quantities (Luckow and Summers, 1988a; Kidd and Emery, 1993). Several mammalian viruses including simian virus 40, vaccinia virus, adenovirus, parvovirus, herpes simplex virus, semliki forest virus, cytomegalovirus, bovine papillomavirus, mouse mammary tumor virus have also been used for these purposes but most of these systems are limited in terms of safety and the quantity of recombinant protein produced. They often do not yield correct posttranslational processing of the gene product. In many respects, the baculovirus expression vector system (BEVS) circumvents most of these deficiencies. The expressed proteins are correctly folded and glycosylated in a manner closely approximating that achieved in their native forms. Post-translational modifications of the gene products of these insect viruses closely parallel glycosylation, fatty acid acylation, and phosphorylation in mammalian cells (Luckow, 1991). Scale up of insect cells in culture has also been largely perfected, making purification of large quantities of recombinant proteins a reality (Van Lier et al., 1992). AcNPV possesses a wider host range than most other baculoviruses, but it fails to infect mammalian cells, making its safety in large-scale recombinant protein production very appealing. The very strong polyhedrin and p10 promoters transcribing their respective genes of this virus have been used to drive the transcription of the foreign gene leading to large quantities of recombinant proteins (upto 400 mg/l) in insect cells. Baculoviruses also offer an ecologically acceptable and effective alternative to chemicals for the control of forest and agricultural insect pests (Huber, 1986; Wood and Granados, 1991).
The development of the baculovirus expression system was facilitated by the establishment of insect cell lines that support the replication of one subgroup, the nuclear polyhedrosis viruses (NPVs). AcNPV multiplies readily in cell lines derived from both the fall army worm (*Spodoptera frugiperda*) and the cabbage looper (*Trichoplusia ni*). Most expression vectors are based on AcNPV infection of *Spodoptera frugiperda* cells. However, the production of heterologous proteins in silkworm (*Bombyx mori, Bm*) larvae relies on infection with recombinant *BmNPV* (Maeda, 1989). Normally *Bm* cells are not permissive to AcNPV but recombinant baculoviruses have been recently generated (Croizier et al., 1994) which possess an extended host range to *BmNPV* larvae. This has been possible by replacing three AcNPV specific amino acids with corresponding *BmNPV* amino acids within the p143 helicase gene of the AcNPV genome. The baculovirus expression system is based on the introduction of the foreign gene into nonessential regions of the viral genome through allelic replacement utilizing the host insect cell in vivo recombination machinery. Production of the recombinant protein is achieved following infection of insect cells or larvae with the newly engineered recombinant virus having a replacement of the native polyhedrin gene with the desired foreign gene.

2. VIRUS STRUCTURE

The AcNPV nucleocapsid is bacilliform in shape, measures 35-40 x 200-400 nm and contains a circular, double-stranded DNA genome of ~133.89 Kb, which has been recently sequenced in its entirety (Ayres et al., 1994). Baculovirus DNA is tightly associated with a protein known as p6.9 (Tweeten et al., 1980; Wilson et al., 1987). The resulting complex forms the core of the nucleocapsid. Several other genes associated with the nucleocapsid have been identified. The most abundant proteins are the major capsid protein p39 (Thiem and Miller, 1989) and the minor nucleocapsid protein, p24 (Wolgamot et al., 1993). A baculovirus encoded phosphoprotein, pp31, binds DNA nonspecifically, co-localizes with the virogenic stroma, is tightly associated with the nuclear matrix and may play a role in
packaging or, alternatively, in viral transcription and/or replication (Guarino and Smith, 1990; Guarino et al., 1992). At the cell surface, the nucleocapsid acquires a loosely fitting envelope that contains the budded virion (BV) envelope glycoprotein, gp67 (Whitford et al., 1989). This protein, which may be present in peplomer-like structures at one end of the virion, is required for BV infectivity by pH-dependent fusions (Blissard and Wenz, 1992). At least three distinct proteins are associated only with occluded or polyhedra-derived virion (PDV), but not with BV virions. Two of these, p25 and gp41, appear to be associated with the PDV envelope (Russell and Rohrmann, 1993; Whitford and Faulkner, 1992). The other protein, p74, is not essential for viral replication in cell culture but is required for larval infection following ingestion of occlusion bodies (OBs, Kuzio et al., 1989).

The major component of the OB is polyhedrin (polh), a 29-kDa protein that is highly conserved among the NPVs (Adams and McClintock, 1991). The polyhedral envelope (PE) or calyx which surrounds the matrix of the OB and increases its stability is rich in carbohydrates (Minion et al., 1979) and also contains a proteinaceous component called pp34 or PE protein (Whitt and Manning, 1988; Gombart et al., 1989). Insertion of cellular DNA that interrupts the p25 gene results in a few polyhedra (FP) phenotype (Beames and Summers, 1989). A second hyper expressed protein, p10, forms fibrous networks in the nucleus and cytoplasm of infected cells (van der Wilk et al., 1987; Williams et al., 1989). An association between p10 and microtubules has also been reported (Volkman and Zaal, 1990). Disruption of the p10 gene results in mutants with varying phenotypes. Certain studies suggest that it is involved in cell lysis late in infection (van Oers et al., 1993). Deletion of the p10 protein prevents release of polyhedra from infected cells, presumably because of impaired nuclear disintegration.

In addition to the structural proteins described above, the baculoviruses encode a number of regulatory proteins, for example, a ubiquitin-like factor (Guarino et al., 1995), protein kinases (Reilly and Guarino, 1994; Li and Miller, 1995a), PTPase (Li and Miller, 1995b), DNA polymerase (Chaeychomsri et al., 1995),
cysteine protease (Slack et al., 1995) and protein that blocks apoptosis (Cartier et al., 1994).

3. THE INFECTION PROCESS

The baculoviridae are a family of double-stranded DNA viruses that infect a variety of arthropods. They can be divided into two sub families (Francki et al., 1991): the Eubaculovirinae (occluded baculoviruses) which infect the larvae of lepidoptera, coleoptera, diptera etc. and the Nudibaculovirinae (nonoccluded baculoviruses). Most baculoviruses isolated thus far are very host-specific, and the majority of Eubaculovirinae have been isolated from larvae of the lepidoptera family.

Baculovirus infection is characterized by the production of two structurally and functionally distinct types of virions, the occluded or polyhedra-derived virion (PDV) and the extracellular or budded virion (BV). The PDV type is responsible for primary infection and is embedded within the matrix of newly formed proteinaceous structures called occlusion bodies (OBs). In a natural infection, the larvae ingest PDV-containing OBs that contaminate their food. The alkaline environment of the insect midgut causes the polyhedra to dissolve releasing the embedded virions. The liberated PDV infect midgut columnar epithelial cells by a process of receptor-mediated membrane fusion (Horton and Burand, 1993). These infected cells produce the BV type which is required for secondary infection and systemic spread within insects. This type also infects cells in culture. Earlier it was thought that the spread of infection within the insect occurs via hemocytes in the haemocoel (Granados and Lawler, 1981; Keddie et al., 1989) but this role has recently been ascribed to cells of the tracheal system (Engelhard et al., 1994). Cellular entry of the BV occurs through receptor-mediated adsorptive endocytosis (Volkman and Goldsmith, 1985; Charlton and Volkman, 1993). Following penetration of the plasma membrane, the nucleocapsids move towards the nucleus by a process that appears to require the formation of actin microfilaments (Charlton and Volkman, 1993). At the nucleus, the nucleocapsids are
uncoated, DNA is released, nucleus becomes enlarged and a distinct electron dense granular structure, the virogenic stroma (Fraser, 1986) is formed. This is thought to be the site of nucleocapsid assembly and is also thought to be the site for viral transcription and replication. By 12 h post infection, progeny BVs are produced and are released into the extracellular compartment. Polyhedra begin to be formed soon thereafter, and mature PDVs (surrounded by an envelope) become occluded. The polyhedrin protein is generally essential for in vivo infections of larvae, but is expendable for infections in cultured cells. Most baculovirus expression vectors exploit this phenomenon by substituting the coding sequence of polyhedrin with a foreign gene.

4. BACULOVIRUSES AS EXPRESSION VECTORS

Two important features of baculoviruses account for the success of this virus as an expression vector. First, the virus contains at least two non-essential genes that can be replaced by an exogenous gene. Second, many of these genes particularly the very late ones, are under the control of powerful promoters that allow abundant expression of the recombinant gene. Most of the expression vectors or transfer vectors in baculoviruses make use of the polyhedrin or p10 promoters together with their associated flanking sequences. Both polyhedrin and p10 are nonessential, since deletion of these genes does not affect the replication of the virus in cell culture (Smith et al., 1983a; Weyer et al., 1990).

The baculovirus genome is too large to permit direct and easy manipulation for insertion of foreign genes. However, several reports have been published (Davies, 1994) describing the direct insertion of foreign genes into the genome via enzymatic ligation (Peakman et al., 1992) through use of large bacterial plasmids and a transposable element (Luckow et al., 1993), or inserting yeast replication elements (Patel et al., 1992). These methods are cumbersome and the usual way to construct a recombinant baculovirus is by introducing the desired foreign coding region into a transfer or transplacement vector. This vector contains a bacterial plasmid, a portion of the baculovirus genome spanning the gene promoter and a
transcription terminator. Two regions of the AcNPV genome have been used to construct expression vectors. These are the 7.3 Kb EcoRI-I fragment in which the polyhedrin gene is located and the 2.0 Kb EcoRI-P fragment containing the p10 gene. Earlier transfer vectors had very long regions flanking the foreign gene insertion site but now the size of the vectors have been considerably reduced to accommodate larger insertions of foreign coding sequences. This is particularly important in the construction of multiple gene expression vectors. A number of improved methods have now been devised to simplify the tedious classical selection process of identifying polyhedrin-negative plaques. These include methods wherein linearized baculovirus DNA (Kitts et al., 1990; Kitts and Possee, 1993) is used to enhance the proportion of recombinants in the progeny virus after co-transfection. A recombinant baculovirus was constructed such that two sites for the restriction enzyme Bsu361 were introduced in the flanking sequences upstream of promoter and within the downstream ORF 1629 encoding an essential gene, the viral replicase. The modified linearized viral genome lacking the sequence between the artificially introduced Bsu361 sites, named BacPAK6, is co-transfected with a conventional polyhedrin based expression vector carrying the deleted portion of the viral genome. Bsu361 digested BacPAK6 viral DNA which lacks the small fragments carrying the essential downstream gene, even after in vivo repair and recircularization is unable to produce viable viruses. Only when this recombines with a transfer vector carrying the missing sequences along with the foreign gene of interest, then the resulting recombinant viral DNA will contain the gene necessary for viral replication and can produce viable progeny viruses. This approach results in an 'assured' recombination event and generates recombinants at frequencies >90%. Clontech (Palo Alto) sells this system under the trade name "BacPAK" and Pharmingen (San Diego) markets it as "BaculoGold". Invitrogen (San Diego) markets a somewhat similar system as "BlueBac" system which has another form of linearized viral DNA containing an Ssel site.

It is now known that the codon usage profile, the secondary structure of the foreign gene and the translation initiation codon context of the Autographa
Introduction

californica nuclear polyhedrosis virus is very different from that utilized by the mammalian systems (Hasnain et al., 1994; Ranjan and Hasnain, 1994; 1995) and this might also be another variable responsible for the varying levels of expression of the foreign proteins. High levels of human basic fibroblast growth factor (Hills and Crane-Robinson, 1995) and human rheumatoid arthritic synovial fluid phospholipase A2 (Kawauchi et al., 1994) have thus been expressed in the baculovirus system by constructing synthetic genes with a favourable codon usage profile. Sridhar and Hasnain (1993) demonstrated differential secretion and glycosylation pattern of βhCG synthesized using the late and very late promoters of the BEVS. They further demonstrated that it is the time of activation and not the overall strength which determines the expression of an extensively processed protein (Sridhar et al., 1993).

Various refinements within the vectors have been introduced for further increasing the level of foreign gene expression. Earlier transfer vectors (Smith et al., 1985) did not contain the complete polyhedrin promoter. A second generation of transfer vectors with the complete promoter showed that the deleted sequences had significant effects on expression levels (Luckow and Summers, 1988b; Matsuura et al., 1987). A series of vectors with modified synthetic polyhedrin gene promoter have also been developed (Thiem and Miller, 1990; Wang et al., 1991). A further refinement has been the integration of the M13 phage or F1 phage origins of replication into the transfer vectors (Livingstone and Jones, 1989; Vialard et al., 1990), thus enabling the production of single-stranded DNA in the bacterial host, after superinfection with M13KO7. These plasmids are very useful when a number of mutagenesis steps have to be carried out on the foreign gene. Foreign gene expression prior to the very late phase i.e. in the late phase may be advantageous for efficient post-translational modifications but most of the late gene products serve as structural virus proteins and they cannot be deleted from the viral genome for the insertion of the foreign gene. To circumvent this problem, the polyhedrin gene locus has been employed as a site for adding a copy of the preferred late
gene promoter to the virus genome. Vectors on these lines have been constructed (Hill-Perkins and Possee, 1990; Thiem and Miller, 1990).

Fusion vectors are normally useful for the expression of a coding region that lacks a translation initiation codon but these are now becoming redundant with the advent of polymerase chain reaction wherein a foreign gene can be tailored to introduce an ATG codon. Nonetheless, these vectors can still be utilized for some foreign proteins which when synthesized as a fusion product are more stable in baculovirus-infected cells. Some vectors that produce fusion proteins with the S-glutathione transferase (Davies and Jones, 1991; Peng et al., 1993) with a protease cleavage site at the junction of the fused protein have been developed so that the proteins can be easily purified using glutathione affinity chromatography. In a similar fashion, histidine residues have been fused to either the carboxy or amino terminal ends of a number of recombinant proteins to facilitate purification on a nickel affinity column (Chen et al., 1993). Further, fusion vectors with amino terminal secretory sequences have also been developed. These include the human placental alkaline phosphatase (Richardson, 1995), honey bee melittin (Chai et al., 1993) for human TNF-β production, myelin-associated glycoprotein, ecdysteroid UDP glycosyltransferase (Richardson, 1995), envelope glycoprotein gp67 and HIV-1 gp120 amino terminus (Murphy et al., 1993). A strategy involving 'artificial chimaeras' (Congote and Li, 1994) has also been adopted to aid in the efficient secretion of the recombinant protein. A synthetic gene with a signal peptide, 10 amino acids of the N-terminus of bombyxin (an insect insulin-like peptide) and 58 amino acids of the C-terminal of human insulin-like growth factor II was constructed and found to be five times more potent than human recombinant IGF II. Garnier et al. (1995) have developed a new system based on the intracellular domain of the rabbit prolactin receptor for secretion of proteins which do not require the post-translational modifications of the classical secretory pathway to be fully active.

The ability to produce two or more foreign proteins simultaneously has been developed for multiple gene expression. This is an attractive approach since it
allows one to investigate protein structure-function, to produce subviral particles for vaccine development and is also cost effective. The first vector of this type was constructed by Emery and Bishop (1987), where a duplicated polyhedrin promoter and transcription termination region was inserted upstream of the native polyhedrin gene. This vector was used to produce the lymphocytic choriomeningitis virus nucleoprotein in addition to the polyhedrin protein. More recently multiple gene expression vectors have been developed which expressed three or four bluetongue virus proteins, which in turn yielded bluetongue virus-like particles in insect cells (Belyaev and Roy, 1993). The latest in the series by the same group is the generation of a quintuple recombinant baculovirus synthesizing up to five foreign proteins (Belyaev et al., 1995) with a fixed ratio comparable to the ratio during the synthesis of native proteins. A novel multiple expression vector which exploits an 'assured' recombination strategy has been developed (Chatterji et al., communicated) which can synthesize proteins at varying levels of stoichiometry.

Attempts to engineer the assembly pathway of the baculovirus insect cell expression system (Hsu et al., 1994a) to lower aggregation and increase production of functionally active proteins and oligomers have also been made. One approach to this is the cloning of potential molecular chaperones *viz.*, the immunoglobulin heavy chain binding protein (BiP) and protein disulfide isomerase (PDI) genes in a baculovirus vector and their subsequent utilization for active murine antibody (IgG) oligomers' production (Hsu et al., 1994b). Baculoviruses can now also be used to display multimeric eukaryotic proteins with fastidious folding requirements in a manner that is analogous to the established bacterial "phage display" systems. As a model system, the marker gene encoding the 26-kDa protein glutathione-S-transferase was used to construct several fusions with the major baculovirus glycoprotein gp64 gene followed by an assessment of the yield and cellular distribution of GST-gp64 protein (Boublik et al., 1995). A general expression vector, pAcSurf-2, was then constructed in which multiple cloning sites were positioned
in-phase between the gp64 signal sequence and the sequence encoding the mature protein under the control of the polyhedrin promoter.

To circumvent large-scale tissue culture, infectivity of a number of recombinant baculoviruses and the stability of their respective gene products in insect larvae have been investigated. Although both the AcNPV and BmNPV systems provide high level expression of foreign genes using larval hosts (Jha et al., 1990; 1992; Sridhar et al., 1994; Price et al., 1989; Medin et al., 1990; Palhan et al., 1995), the silkworm larva has been used more extensively for foreign gene production. The polyhedrin promoter has also been shown to be very active in *Trichoplusia ni* larvae and human terminal deoxynucleotidyltransferase (Tdt) gene has thus been overexpressed (Medin and Coleman, 1992).

5. BACULOVIRUSES AS ENGINEERED INSECTICIDES

The same basic principles as transfer vectors apply to the utilization of these viruses in pest management strategies where the wild-type virus is ineffective in controlling insect pests. In contrast to the use of expression vectors in cell culture where synthesis of polyhedrin is not necessary, the formation of OBs is important for the viral insecticide to survive in nature long enough for the insect to ingest it. Although wild-type baculoviruses have been used as insecticides, the lethal dose and time can be improved by genetic engineering (Bonning and Hammock, 1992). A number of candidate genes with potential insecticidal properties have been tested against the target insects (Maeda et al., 1991; Hammock et al., 1990; Hammock et al., 1993; Tomalski and Miller, 1991). An insect specific toxin that appears to be effective in enhancing AcNPV as an insecticide is derived from the venom of the North African scorpion, *Androctonus australis* Hector (Stewart et al., 1991; McCutchen et al., 1991). The gene product elicited the desired neurotic effects and reduced the median survival time of the infected insect. This modified baculovirus was used recently in a field trial and shown to be very effective in reducing crop damage (Cory et al., 1994).
6. BACULOVIRUS GENE EXPRESSION AND REPLICATION

Given the versatility and safety of the baculoviruses as expression vectors, it is extremely worthwhile and interesting to understand the mechanism of high level expression from this system. Baculovirus gene expression is regulated in a cascade-like fashion where activation of each set of genes relies on the synthesis of proteins from previous classes (Friesen and Miller, 1986). This temporal regulation allows the grouping of baculovirus genes into three phases during infection: early, late and very late. The early genes are transcribed prior to DNA replication, whereas late and very late genes are activated during or after replication. The reason for this dependence on viral replication is not yet known. The early genes generally encode proteins with regulatory functions, such as transcription, replication, and modification of host processes. Late genes include BV and PDV structural proteins, whereas very late proteins are those involved in the processes of occlusion and cell lysis.

Baculovirus early genes are transcribed by the host RNA polymerase II. Consequently, transcription from the early promoters is abolished in the presence of α-amanitin, an inhibitor of RNA polymerase II (Huh and Weaver, 1990). The early promoters resemble typical eukaryotic RNA polymerase II responsive promoters that contain DNA elements that are recognized by host transcription factors (Krappa et al., 1992; Kogan and Blissard, 1994). The late and very late genes are under the control of an α-amanitin-resistant RNA polymerase that is induced during infection (Huh and Weaver, 1990; Grula et al., 1981; Fuchs et al., 1983). This polymerase activity is also resistant to tagetitoxin, an inhibitor of insect RNA polymerase III (Glocker et al., 1993). Partial purification of this activity suggests that its protein composition is different from the three known host RNA polymerases (Yang et al., 1991) and it is not known whether its components are virus encoded, host encoded, or both.

The early genes do not require viral proteins for their activation and are transcribed in uninfected cells (Guarino and Summers, 1986a). This suggests that factors required for early promoter activation are provided during the initial phase
of infection by proteins associated with the virion. For example, IE-1, a transactivator of early genes, has recently been shown to be a component of the BV (Theilmann and Stewart, 1993).

The factors that interact with the late and very late promoters to regulate transcription are not known. The development of an *in vitro* transcription system that utilizes nuclear extracts from virus infected cells at late times post infection may help in the identification of such factors (Xu et al., 1995). Several AcNPV genes thought to have gene regulatory functions have been identified on the basis of their trans-regulatory activity in transient expression assays. Of these, the genes *ie-1*, *ie-n*, *lef-1*, and *lef-2* (Passarelli and Miller, 1993a), *lef-3* (Li et al., 1993), *lef-4*, *lef-5* and *p143* (Passarelli and Miller, 1993b), *lef-6* (Passarelli and Miller, 1994), *lef-7* (Morris et al., 1994), *lef-8* (Passarelli et al., 1994), *lef-9*, *lef-10*, *dnapol* (Lu and Miller, 1994) and *lef-11*, *p47*, *39K*, *p35* (Todd et al., 1995) have been identified as activators of late and very late viral gene expression. The characterization of a very late expression factor gene (*vlf-1*) which has motifs similar to that of RNA helicases has been reported (McLachlin and Miller, 1994). Nine of the identified *lefs* (*ie-1*, *ie-n*, *lef-1*, *lef-2*, *dnapol*, *p143*, *lef-7* and *p35*) are involved in DNA replication (Lu and Miller, 1995b). The product of *ie-1* gene, IE-1, is also a powerful transactivator of early gene expression in transient expression assays. Transactivation of early gene expression by IE-1 can be enhanced by the cis-acting homologous region *hr5* and an *hr5*-binding protein is expressed in cells transfected with the *ie-1* gene. Recently, a virus-encoded host specific factor required for polyhedrin transcription has been identified from *TN360* cells (Lu and Miller, 1995a). Another host factor with unusual characteristics with respect to affinity, stability and specificity has been reported (Burma et al., 1994; Mukherjee et al., 1995a; 1995b; Hasnain et al., 1995). This factor, termed the polyhedrin promoter binding protein (PPBP) binds to transcriptionally important TAAGTATT and the immediately upstream AATAAA motif. No other host or viral protein factor(s) that may be directly involved in transcription of baculovirus very late genes have been identified.
A number of advances have been made in the characterization of replication of baculovirus DNA. The first report of an attempt to identify an origin of baculovirus DNA replication described a region of the genome of the *Galleria melonella* NPV that underwent autonomous replication in insect cells (Blinov et al., 1983). However, the sequence from *AcNPV* failed to replicate in an infection-dependent replication assay in *Spodoptera frugiperda* cells (Kool et al., 1995). Thus far, two strategies have been used to identify potential baculovirus replication origins. One employed the generation of defective interfering virus particles which contain major genomic deletions (Kool et al., 1991; 1993a; Lee and Krell, 1992; 1994) but retain essential cis-acting sequences required for DNA replication. The other approach utilized cloned baculovirus DNA sequences to undergo DNA replication when transfected into infected insect cells. Cochran and Faulkner (1983) originally suggested that homologous regions (*hrs*) may function as baculovirus replication origins and the observation that *hr* sequences were retained in defective genomes (Kool et al., 1993a) lent support to this theory. *AcNPV* contains eight *hrs* called *hr1*, *hr1a*, *hr2*, *hr3*, *hr4a*, *hr4b*, *hr4c*, and *hr5*. *Hrs* are composed of varying numbers of highly conserved repeated sequences of about 70 bp. They have also been demonstrated to be cis-acting enhancers of transcription of baculovirus early genes (Guarino and Summers, 1986b; Lu and Carstens, 1993). Sequences similar to *AcNPV* *hrs* have also been found in *BmNPV*, a close relative of *AcNPV* (Maeda and Majima, 1990; Majima et al., 1993). In addition to *hrs*, sequences that may function as origins of DNA replication, two other non-*hr* containing putative origins have been characterized; the *AcNPV HindIII-K* fragment (Kool et al., 1993b, Kool et al., 1994, Leisy and Rohrmann, 1993; Lee and Krell, 1994) and the *OpMNPV HindIII-N* fragment (Pearson et al., 1993).

7. STRUCTURE OF *AcNPV* VERY LATE GENE PROMOTERS

The baculovirus late and very late gene promoters are very unique and differ from most RNA polymerase II promoters in that they are very compact and do not contain DNA elements, such as the TATA box, present in most eukaryotic
Fig. 1.1 Schematic representation of the polyhedrin gene promoter. The essential polh promoter extends from -1 to -69 (the translation initiation site is depicted as +1). The mRNA start point, space marked with a bent arrow, is at -50 and lies within the PPBP-cognate motifs (Burma et al., 1994): AATAAA and TAAGTATT. The 18 bp minimal promoter as described by Morris and Miller (1994) is marked. The initiator promoter consensus sequence (Javahery et al., 1994) and its comparison with similar regions within the baculovirus very late promoters is shown.
promoters. However, the very late polyhedrin gene promoter, seems to belong to the class of initiator promoters since the initiator sequence, which shares a lot of sequence homology with the consensus initiator sequence around its transcription initiation site (Fig. 1.1), is enough to drive basal transcription. The only element that seems to be present in all late and very late promoters is a consensus core sequence, TAAG, which contains the transcriptional start site and is essential for activity (Wilson et al., 1987; Possee and Howard, 1987; Weyer and Possee, 1989). This element is a part of a very well-conserved sequence, TAATAAGT/AATT. Sequences in the 5' leader region may also influence levels of transcription (Matsuura et al., 1987; Weyer and Possee, 1988). The essential polyhedrin promoter elements have been defined to lie between nucleotides -1 to -69 with the TAAGTATT motif, harboring the transcription start site, being absolutely essential for transcription (Possee and Howard, 1987; Ooi et al., 1989). The minimal polyhedrin promoter has now been demonstrated (Morris and Miller, 1994) to reside only within an 18 bp sequence surrounding this motif. The p10 promoter has not been so extensively characterized but it has been demonstrated that nucleotides -1 to -101 are important for driving transcription from the p10 promoter (Weyer and Possee, 1989).

A pre-requisite to understand the mechanism of transcription from the late or the very late promoters is the identification of factor(s), host or viral, which specifically bind the transcriptionally important sequences and consequently affect the transcription machinery. Etkin et al. (1994) identified a 200-kDa host factor, which was also present in infected cells upto 18 h p.i., binding to regions upstream to the polyhedrin start site. However, direct involvement of this factor in transcription has not been demonstrated.

8. AIMS AND OBJECTIVES

Burma et al. (1994) identified and affinity-purified a 30-kDa cellular factor present in Sf cell lines which specifically bound to transcriptionally important sequence motifs, AATAAA and TAAGTATT, spanning the mRNA initiation site of
the polyhedrin promoter. Polyhedrin promoter binding protein (PPBP) was further shown to be present in other insect cell lines (Mukherjee et al., 1995a) and interestingly its levels varied in different cell lines expressing different levels of reporter gene. It required phosphorylation for binding, a quality shared by few other transcription factors, and showed ssDNA binding activity apart from its ds binding activity (Mukherjee et al., 1995b). The same sequence motifs were utilized for the ssDNA binding activity which was limited to only the coding strand of the promoter. These observations suggested that PPBP was a transcription factor but all studies were limited to in vitro binding analyses. The importance of this factor accompanied with its cognate binding sequences at the in vivo level remains an enigma. Further the influence of context sequences on promoter function has not been investigated for the baculovirus promoters.

PPBP being a host factor and its possible involvement in transcription from one of the very late promoters of AcNPV suggests a more global involvement of this factor, possibly in very late transcription. Thus, it is logical to investigate the binding of this host factor with other baculovirus very late promoters. The closest neighbor to the polyhedrin promoter, also widely used for high-levels of foreign gene expression whose gene product is non-essential for viral replication, is the very late pl0 gene promoter. The pl0 promoter is also extremely A-T rich, contains the same 12 nucleotide consensus sequence identified in polyhedrin genes within which the p10 mRNA CAP site resides and has a 5' leader sequence similar to the polyhedrin promoter.

In an attempt to develop some understanding of how the polyhedrin promoter is regulated at the in vivo level with respect to its PPBP binding capability, the aims and objectives described in this thesis are as follows:

1. Comparative analyses and correlation of the very late promoters, namely, polyhedrin and p10.

2. Evaluation of the role of the identified cis-element(s) and trans-acting factor(s) by in vitro binding assays.

3. In vivo studies using a transient expression system.