1. The baculovirus expression vector system
2. *Autographa californica* Nuclear Polyhedrosis Virus
3. The baculovirus infection process
4. Cellular events during the infection process
5. The late promoters of AcNPV
6. Host gene expression declines during viral infection
7. Origins of replication of AcNPV and its enhancer role
8. *polh* gene promoter: The workhorse of BEVS
9. Evolution of recombinant virus screening techniques
10. Expression in BEVS offers many advantages
11. Variation in heterologous gene expression in BEVS
12. Promoter choice can be important
13. Cell line effects
14. Is there any role of gene sequence itself in influencing its own expression level
15. Translation initiation codon context pattern as a possible factor
16. Codon usage pattern as another possible intrinsic factor
17. Aims and Objectives
1. The baculovirus expression vector system

In recent years, the baculovirus based expression vector system (BEVS) has become highly popular for cloning and expressing heterologous genes of biomedical importance (Luckow and Summers, 1988; Miller, 1988; Luckow, 1991; O'Reilly et al., 1992; Sridhar et al. 1994). More than five hundred genes from diverse sources, ranging from prokaryotes to higher eukaryotes, have been cloned and expressed in this system.

The BEVS employs insect viruses belonging to the baculoviride family. These viruses have very narrow host range and they infect insects belonging to the lepidoptera family only. *Autographa californica* Nuclear Polyhedrosis Virus (AcNPV) and *Bombyx mori* Nuclear polyhedrosis Virus (BmNPV) are the two viruses commonly used for this purpose (Luckow and Summers, 1988; 1989; Maeda, 1989; Maeda and Majima, 1990; Luckow, 1991). Of the two, AcNPV is far more characterized at the genetic level than BmNPV and is considered as a prototype baculovirus.

2. *Autographa californica* Nuclear Polyhedrosis Virus

AcNPV has a double-stranded, covalently closed, circular DNA genome of 134 Kb which has been completely sequenced (Ayres et al. 1994). The viral DNA genome is condensed into a nucleo-protein complex which is enclosed in an enveloped capsid. These enveloped nucleo-capsids are assembled in the nucleus of infected cells. During the infection process two morphologically and biochemically distinct progeny virus forms are produced: a) the Budded Virus (BV) and b) Occluded Virus (OV).
Introduction

The budded viruses (BV) can be morphologically identified as a nucleocapsid with a loosely fitting membrane envelope around it (Adams and McClintock, 1991). These viruses bud out of the infected cells and are released into the extracellular space (Fig 1). This form of progeny viruses is produced early in infection and is primarily responsible for the spread of viral infection to cells belonging to various tissue within the insect larvae.

The occluded viruses (OV) are the enveloped nucleo-capsids embedded in a semi-crystalline matrix (Adams and McClintock, 1991) primarily constituted of a protein called polyhedrin (Harrap, 1972; Rohrman, 1986). In addition to polyhedrin the OV also contain a protein of 25kDa (Russell and Rohrman, 1993) which when absent results in reduced polyhedra formation (Beames and Summers, 1989), p74 protein which is essential for the virulence of the occlusion bodies (Kuzio et al., 1989) and pp31 associated with virigenic stroma (Guarino et al., 1992). As opposed to BV, this form of the virus is produced late in infection and is responsible for dissemination of the virus to other insect larvae. In a natural infection process, the OV has better survival advantage over BV as the polyhedrin matrix embedding the nucleo-capsid protects the virus particles from inactivation by physical factors such as heat, UV, etc. once they are released into the environment.

3. The baculovirus infection process

The infection process begins when insect larvae consume OV, released from disintegrated debris of previously infected insects, as contaminants of their
Fig 1: Typical infection cycle of a baculovirus. The virus is released from occlusion bodies by alkaline pH environment of the insect gut and infects insect cells. The internalized virus then moves to the nucleus where its genome is released and initiates replication of virus. During initial part of infection the progeny viruses are released by budding hence called budded virus (BV). These BVs are very important for spread of infection to neighbouring cells. At late time point, the progeny viruses are embedded in a crystalline matrix called polyhedrin.
food. The polyhedrin matrix of the OV is solubilized under the alkaline pH conditions of the midgut of the insect larvae releasing the virions (Fig 1). The released virions then enter midgut cells by fusing with the microvilli membrane on the lumen side (Granados and Lawler, 1981; Granados and Williams, 1986). The virus replicates inside the infected midgut cells and produces the BV form of progeny virus which are released subsequently from the basement membrane side of the polarized midgut cell (Keddie et al., 1989). BV released from the midgut cells also infect epithelial cells of tracheoles thereby spreading infection along the tracheal network (Engelhard et al., 1994). Finally these BVs gain access to the hemocoel and are transported to other tissues of the infected larvae. Once all the cells of insect larvae get infected, the larvae dies and disintegrates releasing the OVs, produced in late phase of infection. The released OVs spread the infection to other larvae upon ingestion as food and initiate subsequent rounds of infection.

4. **Cellular events during the infection process**

The viroid form of the baculovirus enters the insect cell through membrane fusion assisted by the viral p64 glycoprotein (Blissard and Wenz, 1992). It then migrates through the cytoplasm into the nucleus where nucleoprotein core is released from nucleocapsid (Granados and Williams, 1986). Based on some unique transcriptional features the cellular events of viral infection process can be divided in two distinct phases: a) Early and b) Late.
The **Early Phase** covers the first six hours of the infection process. During this phase the cell is reprogrammed for virus replication. This phase is characterized by transcription of regulatory genes that are required for viral replication and trans-activation of late genes. Transcription of these early genes is not dependent on viral gene product and is carried out by an α-amanitin sensitive host RNA polymerase (Grula et al., 1981; Huh and Weaver, 1990). Although there is no requirement of viral protein synthesis for the expression of many of the early genes (also called immediate early genes) some of the early genes (also known as delayed early genes) are transactivated by the viral ie-1 gene product (Nissen and Friesen, 1989; Guarino and Summers, 1986a).

The **Late Phase** extends from 6 h pi (hours post-infection) to approximately 72 h pi. This is a period characterized by late viral gene transcription, which is carried out by an α-amanitin insensitive RNA polymerase (Grula et al., 1981; Huh and Weaver, 1990). Although the exact nature of this RNA polymerase is unknown it is believed that it could be a virus-modified host RNA polymerase or a new RNA polymerase with virus-encoded subunits (Felgener, et al., 1983; Yang et al., 1991; Beniya et al., 1996). A viral gene having some homology to RNA polymerase domain has been identified (Passarelli et al., 1994). Another important feature of this late phase transcription is its dependence on early viral gene expression (Lu and Carstens, 1994; Passarelli et al., 1994; Passarelli and Miller, 1993a; 1993b; 1994) and on DNA replication as evident from the fact that both
cycloheximide and aphidicolin block transcription of late and very late genes (Rice and Miller, 1986).

The late phase can be further classified into two phases: a) *Late phase* (6 h pi to 20 h pi) and b) *Very late phase* (20h pi to 72 h pi). The former (6 h pi to 20 h pi) is characterized by extensive viral DNA replication, expression of viral structural genes (late genes) and BV production. The latter (20h pi to 72 h pi) is characterized by hypertranscription and expression of the *polh* gene. During this phase *polh* gene product forms the semi-crystalline polyhedrin matrix in the nucleus and entraps the enveloped nucleocapsids. This structure is called occlusion body or polyhedra which is released once the lysis of the infected cells begins (around 72 h pi). As mentioned earlier these structures are responsible for long term survival of the embedded virus and the spread of viral infection to other insect caterpillars.

5. **The late promoters of AcNPV**

The promoters of late genes are unusual in many respect. They are transcribed by an αamanitin resistant RNA polymerase. They possibly belong to the "initiator promoter" class where the sequence around the initiator are capable of acting as promoter. Almost all of them contain a TAAG motif upstream to their 5'UTR, believed to be important for transcription (Rankin et al., 1988; Morris and Miller, 1994). Studies on transcription from two late gene *polh* and *p10* promoters showed that an unusual 30kDa host protein, termed as polyhedrin binding protein or PPBP, interact with sequences at and around TAAG site of late gene promoters (Burma et al., 1994; Mukherjee et
al., 1995a; Mukherjee et al., 1995b; Jain and Hasnain, 1996; Hasnain et al., 1996). This protein is an "initiator promoter" binding protein and is not a TATA binding protein (Hasnain et al., manuscript in preparation). This protein possess both double strand and single strand DNA binding activities (Mukherjee et al., 1995a). As a single stranded DNA binding protein it binds specifically to the coding strand and not to the noncoding strand (Mukherjee et al., 1995a; Jain and Hasnain, 1996). In addition PPBP displays a very strong affinity (dissociation constant in picomolar range), stability (PPBP:DNA complex is stable up to 50°C and in presence of 2M NaCl) and specificity (Burma et al., 1994). How PPBP actually forms the transcription complex with the virus modified RNA polymerase (Huh and Weaver, 1990; Passarelli et al., 1994) and (or) viral encoded trans acting factors, such as lefs (Lu and Miller, 1995a; 1995b 1994; Morris et al., 1994; Todd et al., 1995), leading to hypertranscription in a temporal fashion is yet to be worked out. Attempts are being made to carry out late gene transcription in vitro in order to identify and characterize the transcription apparatus involved (Glocker et al., 1992; 1993; Xu et al., 1995). In addition to PPBP there is also another 120kDa host factor that interacts with upstream region of polh promoter (Etkin et al., 1994). The role of this protein is not known. Recently, another 38 kDa host factor also appear to be involved in enhancing polyhedrin gene promoter driven transcription by binding specifically to the hr1 enhancer element (Habib et al., 1996b).

6. Host gene expression declines during viral infection
AcNPV infection also results in shut down of host gene expression. The decline in steady-state levels of host mRNAs begins around 12 h pi. At 24 h pi, the steady-state levels of cellular mRNAs such as histone and actin are very low (Ooi and Miller, 1988). The decline in host protein synthesis can be observed by 18 h pi leading to a virtual shut-off around 24 h pi (Carstens et al., 1979). By this time, gene expression in the infected insect cells becomes primarily viral-specific. A viral “encoded” RNA polymerase like activity has recently identified at late time point of infection (Beniya et al., 1996). The mechanism by which host RNA and protein levels are down-regulated is not yet known. What is known that the virus encodes a protein, p35, that is responsible for shutting off cellular apoptotic gene function(s) thereby enabling the virus to complete its own life cycle ending in lysis of the host cells (Cartier et al., 1994; Sugimoto et al., 1994).

7. Origins of replication of AcNPV and its enhancer role
The AcNPV genome has eight interspersed homologous regions, designated hr 1, 1a, 2, 3, 4a, 4b, 5 and 5' (Cochran and Faulkner, 1983; Kool et al. 1993a; 1993b). These region are approximately 500-800 bp in length and are believed to function as origins of replication (Kool et al., 1993a; 1993b; 1995; Habib et al., 1996). As do many origins of replication, these regions also act as enhancers for viral promoters present around them (Guarino and Summers, 1986b; Habib et al., 1996) and are even shown to transactivate heterologous promoters both in vivo and in vitro (Venkiah et al. unpublished).

8. polh gene promoter: The workhorse of BEVS
In BEVS, the foreign gene is placed under the transcriptional control of the AcNPV very late polyhedrin gene promoter. Although polh gene is critical for the natural infection cycle of baculovirus, it is not required for replication of the viral DNA and production of the BV form. The BVs are adequate for *in vitro* propagation of the virus in insect cell culture in laboratory conditions. The baculovirus based expression vector system cleverly exploits these observations to clone and express heterologous gene under the hyperactive polh gene promoter control. In this system the heterologous gene is targeted to the polyhedrin locus where the wild type polh gene is replaced with new heterologous gene by *in-vivo* recombination between the viral DNA and a transfer vector carrying the foreign gene flanked by viral sequences. As a result the recombinant progeny baculovirus can no longer express polh gene hence polyhedra or occlusion body formation is not observed in infected cells. Since the occlusion bodies can be visualized microscopically they serve as marker for identifying wild type baculovirus infected cells. The first generation baculoviral vector extensively made use of this feature to distinguish recombinant virus plaques from wild type virus plaque.

9. Evolution of recombinant virus screening techniques

In BEVS, utilizing only the “polyhedra negative” phenotype as selection marker is a very inefficient exercise for selecting recombinant virus. Since the recombinant baculoviruses appear at a very low frequency (0.1-1 %), microscopic screening for the polyhedra negative plaques, which is far less than polyhedra positive plaques, becomes very tedious.
Over the last few years, methods for screening and selection for recombinant viruses have also evolved considerably (Davis, 1994). Some of the new methods further facilitate the screening and production of recombinant virus using a co-expressed reporter genes such as luciferase (Hasnain et al., 1994; Palhan et al., 1995) and $\beta$-gal (Vialard et al., 1990; Weyer et al., 1990) antibody based screening and PCR screening (O'Reilly et al., 1992). Attempt has been made to propagate baculovirus genome in *E.coli* and produce recombinant baculovirus in the same by transposon mediated transfer of heterologous gene (Luckow et al., 1993).

Recently, a major breakthrough has occurred in screening technique. A highly efficient method, which gives extraordinarily high rates of recombination (>95%), was reported (Kitts and Possee, 1993). In this technique, the N terminal end of an essential gene (ORF 1629) immediately downstream to the *polh* locus, is deleted from the wild type viral DNA used for co-transfection experiment. This deletion can only be restored after recombination with a transfer vector that carries the heterologous gene along with the deleted part of the essential viral gene (ORF 1629); thereby making it practically impossible for wild type progeny viruses to appear. This method has greatly simplified the construction of recombinant baculovirus.

10. Expression in BEVS offers many advantages

There are several features of BEVS that are particularly advantageous for expressing heterologous gene in this system (Sridhar et al., 1994; Richardson, 1995):
Baculovirus vector is helper independent and relatively simple to use (Summers and Smith, 1987). It propagates in insect cell which is a far more advanced eukaryote than yeast. The virus infection is primarily restricted to invertebrates thereby making it very popular from biosafety considerations.

It provides a powerful polh gene promoter to drive heterologous gene expression. Since the polh gene of this virus is transcribed to enormous levels during the late phases of baculoviral infection and this gene is also non-essential for the replication of the virus in cell culture (Smith et al., 1983), its protein coding sequence is replaced with the heterologous gene allowing the latter to be transcribed under the control of a powerful polh promoter. Now other AcNPV promoters such as p10 promoter, basic protein promoter are also employed for directing heterologous gene expression (Vlak et al., 1990; Hill-Perkins and Possee, 1990).

It offers an eukaryotic environment for protein production which facilitates synthesis of biologically active proteins. Most post-translational modifications like proteolytic cleavage of signal peptide, N-linked glycosylation, O-linked glycosylation, acylation, amidation, phosphorylation, prenylation, and carboxymethylation can be easily carried out by insect cells (O'Reilly et al., 1992).

The eukaryotic recombinant proteins are targeted to their natural locations in the cell (Hasnain et al., 1994). Proteins containing signal peptides
are recognized and cleaved in the correct fashion (Congote and Li, 1994). The efficiency of secretion can be enhanced much further by fusing heterologous coding sequence with signal peptide coding sequence of the honey bee melittin gene (Tessier et al., 1991). Vectors have also been developed that use a homologous secretory signal from egt and p67 to facilitate efficient secretion of foreign gene product (Murphy et al. 1993).

BEVS allows the production of large quantities of active protein, especially temperature sensitive products as the baculovirus is propagated at 27°C (Reynisdottir et al., 1990). Strategies have been developed for simple one step purification of recombinant protein (Peng et al., 1993) specially by using (his)$_6$ tag (Chen et al., 1993).

It also allows simultaneous multiple gene cloning, homo- and hetero-oligomeric polypeptide assemblies, as has been demonstrated for a wide range of proteins in cells infected with baculoviruses (Emery and Bishop, 1987; Weyer et al., 1990, Wang et al., 1991; Belyaev et al., 1995; Chatterjee et al., 1996). Using a quadruple expression vector, four polypeptide from blue-tongue virus have been co-expressed and shown to assemble into virus-like particles in insect cells (Belyaev and Roy, 1993). Since the nucleocapsids of baculoviruses can accommodate more than 100 kbp of DNA, size of heterologous gene is not a restriction.

Recently BEVS has been successfully used to develop an eukaryotic protein display system. This display system makes use of p67 glycoprotein
that is fused with heterologous gene to display the recombinant protein on the baculoviral membrane surface (Boublik et al., 1995).

As the BEVS operates via the lytic phase, it can be used to produce toxic gene products in large amounts without any concern for host cell toxicity.

Since the baculovirus infection is restricted to insect cells, it has found a new role as viral insecticides (Huber, 1986; Wood and Granados, 1991; Bonning and Hammock, 1992). The potency of this virus based pesticide has been further enhanced by cloning lethal insect specific enzymes such as juvenile hormone esterase (Hammock et al. 1990; 1993) and toxins like mite neurotoxin (Tomalski and Miller, 1991; Stewart et al., 1991). An attempt has also been made to increase its insect host range by introducing mutation in p143 gene (Croizier et al., 1994). These viral insecticide has now gone for field trails (Corey et al., 1994).

The BEVS provides exceptionally high level of expression. The highest level reported represents more than 50% of the total cellular protein corresponding to approximately 1 gram of protein product per $10^9$ cells (per liter culture). Most heterologous proteins, however, are produced at levels ranging from 10-100 mg per $10^9$ cells.

Although more than five hundred genes have already been expressed in BEVS (O’Reilly et al. 1992; Sridhar et al. 1994) the level of expression
varies considerably depending on the gene ranging from 1μg to 1mg/ml/10^6 cells. Inspite of the overwhelming popularity of BEVS very little is known about why some genes are expressed to very high level while others are not. Some of the possible factors could be promoter usage (Sridhar et al., 1993; Sridhar and Hasnain, 1993), host cell line (Mukherjee et al., 1995b) and the heterologous gene sequence per se (Hasnain et al., 1994).

12. Promoter choice can be important

Usually most of the heterologous genes expressed in BEVS are introduced under the transcriptional control of hyper-active and very late polh gene promoter. However, not all genes are expressed efficiently under this promoter especially those genes coding for highly glycosylated proteins such as tissue plasminogen activator, beta subunit of human chorionic gonadotropin hormone, etc.. These proteins require extensive glycosylation that appears to decline late in infection (Sridhar et al., 1993). This is probably due to a decline in host protein synthesis during the late phase of infection which in turn results in poor performance of the endoplasmic reticulum and golgi apparatus. Since polh gene promoter driven heterologous genes are expressed late in infection the time available prior to cell death to process the heterologous protein may not be adequate leading to a "secretory load" (Sridhar et al. 1993; Sridhar et al., 1994) on the host system. This result in intracellular accumulation of polypeptides which are partially processed. The use of promoter activated earlier in the infection cycle, such as core gene promoter, provides more time for the host to process the heterologous proteins before cell lysis and death, and thereby alleviates the secretory load
problem to a large extent (Sridhar and Hasnain, 1993; Sridhar et al., 1993; Hasnain et al., 1994; Sridhar et al., 1994).

13. **Cell line effects**

Choice of cell line also plays an important role in determining the expression level. Although baculovirus has a very narrow host range, as they infect insects mainly belonging to family lepidoptera, AcNPV can infect around 40 different species. There are many cell lines which are available from these species most of which are ovarian in origin. Some of these are Sf21 and Sf9 from *Spodoptera frugiperda* (the fall army worm), TN368 from *Trichoplusia ni* (the cabbage looper). It is known that not all cell lines express the same gene to same level (Hink et al. 1991; Mukherjee et al. 1995b). Expression of heterologous gene in TN368 cell lines is much better than Sf21 and Sf9. Some of the factors that are believed to be responsible for these difference are entry of virus and mRNA stability. TN368 has a large cell surface area as compared to Sf9 and Sf21 facilitating viral adsorption. It has also been shown that reporter gene mRNA has a longer half life in TN368 than in Sf21 and Sf9 cells (Mukherjee et al. 1995b). The differential expression as a function of cell line also appears to be regulated at transcriptional level.

14. **Is there any role of gene sequence itself in influencing its own expression level**

Previous reports on the expression of firefly luciferase and the beta subunit of human chorionic gonadotropin (*βhCG*) genes in insect cells and caterpillars using recombinant baculoviruses vAcpβhCG (Nakhai et al., 1991) and vAcluc
Introduction

(Hasnain and Nakhai 1990, Jha et al. 1990) seem to highlight the importance of gene sequence *per-se* on its expression. Although an exact comparison of expression levels for these heterologous proteins could not be made since these were two independent viruses, nevertheless, what emerged was the striking difference in the expression level of these two genes. This was further supported by the observations that similar differences in expression was obtained with the recombinant baculovirus vAcβhCG-luc, which carried the two heterologous genes βhCG and luc, under the transcriptional control of two viral polyhedrin gene promoters placed back-to-back, integrated within the polyhedrin locus (Hasnain et al., 1994) (Fig 2). When this virus was used to infect Sf9 insect cells and caterpillars (Hasnain et al., 1994; Jha et al., 1992), once again βhCG was expressed to a much lower level compared to luc.

15. Translation initiation codon context pattern as a possible factor

The role of initiation codon context sequence in influencing translation initiation site selection and modulating translation efficiency has been shown in the vertebrate system (Kozak, 1986). The consensus initiation codon context has been identified for various systems such as *E. coli*, yeast and higher eukaryotes like *Drosophila*, vertebrates and plants (Cigan and Donahue, 1987; Kozak 1987; Cavener and Ray, 1991). These sequence do differ from each other suggesting the presence of evolutionary influence on them. Since the context sequences of translation initiation codons have been shown to play an important role in translation initiation site selection, it will be
Fig 2: A. Two heterologous genes expressed in insect cell infected with recombinant baculovirus vAcβhCG-luc. The luc gene is expressed to a very high level as opposed to βhCG.

B. Schematic representation of the polh locus of recombinant baculovirus vAcβhCG-luc showing that the two genes are present under identical transcription control of the viral polh gene promoter.
interesting to know the baculoviral preference for this site. However very little is known in this respect (O'Reilly et al., 1992).

16. Codon usage pattern as another possible intrinsic factor
It is known that in a degenerate genetic code, not all the synonymous codons are utilized with equal frequencies; some are utilized more than others (Wada et al., 1992). The pattern of biased codon usage in a given species reflects the relative tRNA abundance (Bennetzen and Hall, 1982; Ikura, 1982; Shields, 1990) within that species and is often different for different species or taxonomic groups. Theoretically, it has been shown that in selective systems such as unicellular organisms, the biased codon usage and tRNA abundance co-evolves (Bulmer, 1987). Codon usage within the species also varies and this variation can be explained in terms of joint effect of mutation selection and random drift (Sharp and Li, 1986; Sharp and Matassi, 1994). In yeast as well as in E. coli-based expression systems, heterologous genes which utilize-non-optimal codons of the host are expressed poorly (Soerensen, et al, 1989). This could, however, be increased many fold once suboptimal codons were replaced with optimal codons (Kotula and Curtis, 1991). Once again, very little is known about the codon usage of baculovirus and how it influences the gene expression.

17. Aims and Objectives
The level of synthesis of heterologous proteins in BEVS varies considerably. This variation in heterologous gene expression is thought to be due to a number of factors (O'Reilly et al. 1992; Hasnain et al., 1994; Sridhar et al.,
1994; Mukherjee et al., 1995b). In order to increase the expression level of foreign genes in baculovirus systems, most of the efforts seem to have been directed at selecting the promoter (Sridhar et al., 1994), choosing the insect cell lines (Mukherjee et al., 1995b) and optimising poly A signal (Westwood et al., 1993). There has been very little attention directed at the influence of translation initiation codon context and codon usage patterns on heterologous gene expression. These two patterns are known to vary form one species to another (Sharp and Matassi, 1994; Cavener and Ray 1991) and could influence the expression level by influencing the rate of mRNA translation (Sharp and Matassi, 1994; Kozak, 1986). With this in mind the objectives of this study were defined as follows:

1. To study the coding sequence pattern of the prototype baculovirus *Autographa californica* Nuclear Polyhedrosis Virus in terms of
   a) Translation initiation codon context
   b) Codon usage

2. To study the differences at the level of patterns associated with the coding sequence of under- and over- expressed heterologous genes *viz a viz* homologous AcNPV genes.

3. To compare the translation initiation site selection efficiency associated with the consensus Kozak translation initiation context sequence of vertebrates and the computed translation initiation context of AcNPV consensus in BEVS.