1. Patterns associated with nucleotide sequences of genes
2. The preferred translation initiation codon context pattern of the baculovirus
3. The insect virus AcNPV translation initiation codon context is very similar to the pattern observed in a dipteran insect- Drosophila
4. Analysis of translation initiation codon context of individual AcNPV genes
5. Evaluation of the Initiation Codon Context in AcNPV Genes
6. Translation initiation context pattern and heterologous gene expression in BEVS
7. Analysis of two heterologous genes- the over expressed luc and the under expressed βhCG- that are simultaneously expressed in vAcβhCG-luc infected insect cells
8. Vertebrate translation initiation context pattern are poor match to the insect virus (AcNPV) pattern
9. Less efficient translation initiation from the start site when placed under vertebrate translation initiation context pattern
10. The preferred codon usage pattern of the baculovirus
11. Codon usage pattern of AcNPV: Over and Under utilized codons
12. The pattern insect virus AcNPV is similar to the pattern observed for insects like Bombyx mori and Drosophila
13. Analysis of codon usage pattern of individual AcNPV genes
14. Codon usage pattern and heterologous gene expression in BEVS
15. Analysis of two heterologous genes- the over expressed luc and the under expressed βhCG- that are expressed in vAcβhCG-luc infected insect cells
16. Codon usage can also influence the strength of secondary structure
17. Importance of these patterns in protein synthesis at very late time point of infection
18. Poor expression of heterologous genes: A result of nucleotide sequence pattern mismatch
1. Patterns associated with nucleotide sequences of genes
It was previously shown that heterologous genes placed under identical transcriptional control showed different levels of expression in the baculovirus system (Hasnain et al., 1994). Further it was suggested that these expression differences observed for different heterologous genes may arise due to the nucleotide sequence differences (Hasnain et al., 1994) between them as these have been derived from different species separated far apart during evolution. In literature, the two important nucleotide patterns reported to influence gene expression (Caveners and Ray, 1991; Sharp and Matassi, 1994) are the translation initiation context and the codon usage patterns of a gene. These two parameters are also known to vary from one system to another (Caveners and Ray, 1991; Sharp and Matassi, 1994). In this part of work, an attempt was made to characterize the pattern preferred by the prototype baculovirus -AcNPV- genes and address the issue how the patterns of the over and under expressed heterologous genes differ from them.

2. The preferred translation initiation codon context pattern of the baculovirus
It has been shown that for the efficient initiation of translation at the correct start site of the coding gene, there is a requirement for an optimal sequence context around the translation start site, ATG (Kozak, 1987; 1986; 1991). The optimal sequence for efficient translation initiation has been identified for the vertebrate and was found to be a consensus sequence, CC(G/A)CCATGG, derived by aligning the nucleotide sequence around the translation start site of vertebrates gene (Kozak 1987). The present study showed an initiation codon context consensus different from the vertebrates originally proposed
Discussion

by Kozak (1987) and later by Caveners and Ray (1991). The consensus AcNPV context sequence was aAAaATGA as opposed to the Kozak sequence. This AcNPV sequence shows a very high preference for A residue as opposed to C/G observed for the vertebrate. Recently when the complete baculoviral genome sequence became available (Ayres et al., 1994), the analysis was extended to all 154 possible ORFs. The observation made with analyses of 154 ORFs was almost the same as was computed for the 23 protein coding AcNPV genes with a minor difference at position -1, where C residue is utilized slightly more than A residue. Since these ORF's are only predicted genes and the protein products of more than 100 ORF's are yet to be characterized, it is quite possible that some of these predicted ORFs could be noncoding (Ayres et al., 1994).

3. The insect virus AcNPV translation initiation codon context is very similar to the pattern observed in a dipteran insect- Drosophila

Interestingly Drosophila, an other insect system, also utilizes a high A rich translation initiation codon context (Cavener and Ray, 1991). The consensus translation initiation codon context proposed for Drosophila by Cavener and Ray (1991) was also found to be AAAATGA which closely resembled the consensus AcNPV context. These results highlight three very important features: one, it agrees with the evolutionary closeness of the AcNPV hosts, the lepidopteran insect, to the dipteran Drosophila; second it shows that the obligate parasite of insect has tailored its pattern to reflect the host pattern to maximize its expression in them; and third it suggests possible differences in preferences for translation initiation site selection between the invertebrate
insects and vertebrates. It also raises some interesting questions: a) how the context of AcNPV genes match to the consensus pattern observed for the baculovirus? b) Does this preferred pattern has any role in influencing heterologous gene expression level in insect cell. Before one can address these issues, one needs to develop a quantitative method to evaluate the closeness of a context sequence to the pattern employed by AcNPV.

4. Analysis of translation initiation codon context of individual AcNPV genes

In the present study a quantitative method for rating the match quality for translation initiation codon context was developed. The method employed a weight matrix based approach to evaluate a sequences. In this method the base usage percentage at various position were used as weights to evaluate the match quality. For each position a value from weight table was selected which depends on the base used at that position. All these selected values for each position was used to compute an average value that represented the quality of match to the base usage profile on which the consensus is based. In order to restrict the numbers of reported match a physically valid condition was imposed that only those match would be reported that has ATG at a position where they are 100% used.

5. Evaluation of the Initiation Codon Context in AcNPV Genes

The translation initiation codon context pattern of individual genes was analyzed using the percent base usage profile of AcNPV context sequence as weights. When the AcNPV consensus initiation codon context pattern was used to search similar motifs in individual AcNPV gene sequences, it
Discussion

successfully identified the putative translation initiation codons in each of the AcNPV genes. The match quality of initiation codon context of various AcNPV genes varied from 39.95 to 49.95 with the exception of the ie-n gene which showed a relatively poor quality context compared to other AcNPV genes. These value should be compared to a range of match quality values (20.4 to 50.3) that a context sequence can take up. This comparison clearly showed that AcNPV genes had a translation initiation codon context pattern close to the overall AcNPV pattern.

6. Translation initiation context pattern and heterologous gene expression in BEVS

Scatter analysis was carried out to investigate whether the match quality values computed for the context pattern of heterologous genes vis a vis the AcNPV translation initiation codon context pattern have any role in influencing heterologous gene expression levels. The analyses showed an association between the match quality value computed for the context pattern of heterologous genes and its expression level pointing to the role of context pattern in influencing gene expression. A regression line drawn through data clearly showed a positive relationship between the degree of translation initiation codon context match to the AcNPV consensus pattern and the gene expression levels in BEVS. However, it was also observed that not all data point fall on linear regression line suggesting that additional factors may also be responsible in determining heterologous gene expression in BEVS. Since this study was carried out using a sample of heterologous genes cloned under the control of polh gene promoter using an AcNPV based BEVS and were all expressed in Spodoptera frugiperda cells, differences could thus be
attributed to the differences in coding sequence pattern. The codon usage pattern of the gene (Sharp and Matassi, 1994) has been reported to influence the translation elongation rate during the gene expression process (Kurland, 1991; Bulmer, 1987).

7. Analysis of two heterologous genes- the over expressed luc and the under expressed βhCG- that are simultaneously expressed in vAcβhCG-luc infected insect cells

Since the above analysis was based on expression values reported by different groups in different sets of experiment, detail analyses was carried out for the two genes that are placed under identical transcriptional control of the polh gene promoter in the same baculovirus and expressed in the same cell. The over expressed heterologous gene luc showed a higher consensus match quality value which was very close to the average AcNPV value. This similarity was also seen at the level of nucleotide sequences. The consensus match quality value for βhCG gene was found to be much lower than both the over-expressed luc and the average AcNPV values suggesting a poor translation initiation efficiency from its start site. Results of previous work have shown the intracellular retention of around 40% of the βhCG gene products. This retention was unusually high and was attributed to the "secretory load" problem of infected insect cells (Sridhar et al, 1994). But in view of present observations a substantial portion of this retention could be attributed to the inefficient recognition of the first ATG, leading to initiation at a second ATG position placed in frame just after the secretory signal.
8. Vertebrate translation initiation context pattern are poor match to the insect virus (AcNPV) pattern

Using the quantitative parameter consensus match quality value, a comparison of the translation initiation codon context of both insect virus AcNPV's as well as of vertebrates revealed that the vertebrate sequence is a poor match to AcNPV pattern. This observation along with another one made previously, that AcNPV consensus context was similar to the consensus of another insect, Drosophila, raised an interesting possibility that translation selection site in insect may differ in its context requirement from vertebrates. If this is to be true then in insect cell based expression system, like BEVS, the reporter gene with vertebrate context pattern should express less efficiently as compared to the context pattern close to AcNPV genes.

9. Less efficient translation initiation from the start site when placed under vertebrate translation initiation context pattern

The importance of initiation codon context pattern was demonstrated using a luciferase reporter gene based transient expression system. These results confirm the analysis result that the context requirement of insect cell based BEVS is different from vertebrates as the reporter gene using the vertebrate context is expressed with less efficiency. In these set of experiments only 5' sequence (with respect to ATG was changed) not the complete sequence. However sequence 3' to ATG was not changed to avoid changes in the amino acid sequence in the gene which would have made the drawing of inference a complex event.

10. The preferred codon usage pattern of the baculovirus
It has been shown for various organisms that some of the synonymous codons coding for a given amino acid are used more often than others (Sharp and Matassi, 1994; Wada et al., 1992). These non random utilization of synonymous codons results in a biased codon usage pattern for a given organism (Sharp and Matassi, 1994; Sharp and Li, 1986). The observed bias in codon usage pattern vary for different species (Wada et al., 1992; Ikemura, 1982). In yeast as well as in *E.coli* based expression systems, the heterologous genes which utilize non-optimal codons of the host are expressed poorly. It has also been shown that the expression level could be increased by replacing sub-optimal codons with optimal codons (Kotula and Curtis, 1991). The present investigation explored the bias in the codon usage pattern of AcNPV.

11. Codon usage pattern of AcNPV: Over and Under utilized codons

The study identified the over and under utilized synonymous codons coding for different amino acid. Some of the over utilized codons were identified as TTG (Leu), CGC (Arg), AGC (Ser), GGC (Gly), GTG (Val), CCG & CCC (Pro), ATT (Ile), TTT (Phy), AAA (Lys), GAA (Glu), CAA (Gln). TAA was the most favored stop codon. Some of the poorly utilized codons were: CTC & CTT (Leu), CGG & AGG (Arg), TCA & TCC (Ser), etc. Recently, when the complete genome sequence of AcNPV became available with the EMBL Database library, the codon usage analysis was extended to all 154 potential protein coding ORFs (Ayeres et. al., 1994). The pattern computed from these ORFs was almost identical to the overall codon usage profile computed from 25 protein coding AcNPV genes confirming the stability of the pattern.
12. The pattern for insect virus AcNPV is similar to the pattern observed for insects like *Bombyx mori* and *Drosophila*

A comparison of the codon usage pattern of AcNPV with the same of a lepidopteran insect *Bombyx mori* and dipteran insect *Drosophila* revealed that many of the codons highly utilized by AcNPV were also utilized well by these two insects. Some of these codons were CGC (Arg), AGC (Ser), GGC (Gly), CCG & CCC (Pro), GAA (Glu), CAA (Gln), TAC (Tyr). In all the three AcNPV, *B. mori* and *Drosophila* - TAA was the most favored stop codon. However, these insects also showed some differences from AcNPV as they prefer codons, for some amino acid such as Leu, Gly, Val, Ala etc., other than what was highly utilized by AcNPV. Nonetheless, the overall similarity between these three coding pattern highlights two very important features: one, it agrees with the evolutionary closeness of the AcNPV hosts, the lepidopteran insect, to another lepidopteran insect, *Bombyx mori* and the dipteran insect *Drosophila*; second it shows that the obligate parasite of insect has tailored its pattern to reflect the host pattern to maximize its expression in them. It also raised some interesting questions similar to those related to the initiation codon context: a) how the codon usage pattern of AcNPV genes match to the consensus pattern observed for the baculovirus? b) Does this preferred pattern has any role in influencing heterologous gene expression level in insect cells.

13. Analysis of codon usage pattern of individual AcNPV genes

To find out the variation observed in the codon usage profile of individual genes, a correspond analysis procedure of Grantham and colleagues (1982)
was used. The procedure compares the two codon usage profiles and computes a parameter called D square statistic value which measures the difference between the two. Results of analyses showed that although many individual gene pattern do vary from the overall AcNPV pattern, a common interesting pattern still emerges. Many of the early genes use a pattern very close to overall AcNPV pattern as opposed to late genes, which differ relatively more than the early genes. Since most of the early genes were regulatory in nature where as the late genes were mostly structural genes, including the two highly expressed gene polh and p10, it is very tempting to propose that these variations could be a part of baculovirus strategy to maximize its late gene expression in late time point of infection.

It was proposed by Sharp and Li (1986) that the codon usage pattern of the highly expressed genes may differ from the overall codon usage of an organism, as these genes are evolved under a translational selection pressure. These genes utilize mostly those codons that provide them translational advantage. Since the profile of the two highly expressed very late gene differs for the overall AcNPV profile, a combined codon usage pattern of two highly expressed AcNPV gene was computed and compared with AcNPV pattern. The pattern of highly expressed AcNPV genes use GTT (Val), GCT (Ala), TCA (Ser), AAG (Lys), ATC (Ile), ACC (Thr), CTG (Leu), TTC (Phe), CGT (Arg) which differs from the overall AcNPV pattern. However the other highly utilized codons, such as GGC (Gly), GAA (Glu), GAC (Asp), AAC (Asn), TGC (Cys), TAC (Tyr), CAA (Gln), CAC (His), CCG (Pro), ATG(Met) and TGG(Trp), were essentially the same as it were for the overall codon usage profile. Among these two pattern which one of the two show
association with the heterologous gene expression level is an interesting issue.

14. Codon usage pattern and heterologous gene expression in BEVS

Scatter analysis was carried out to investigate whether the D square value, measuring the difference of the codon usage pattern of heterologous genes from that of overall AcNPV codon usage pattern, have any role in influencing heterologous gene expression levels. The results clearly showed an association between the D square value computed for the codon usage pattern of heterologous genes and its expression level confirming the role of context pattern in influencing gene expression. A regression line drawn through data clearly showed a negative relationship between the difference in codon usage from the AcNPV pattern and the gene expression levels in BEVS. As was done previously, for the analysis of translation initiation codon context pattern, these studies were carried out using a sample of heterologous genes that were cloned under identical control of polh gene promoter using an AcNPV based BEVS and were all expressed in Spodoptera frugiperda cells. As observed previously not all data point fall on linear regression line confirming the observations made above that multiple factors including the two pattern mentioned here are responsible for determining heterologous gene expression in BEVS.

A similar analysis was also carried out for the same set of heterologous genes using the D square values computed by taking highly expressed codon usage profile as reference. The analyses showed that both
the patterns were able to explain the variation observed in heterologous gene expression in BEVS to the similar extent.

15. Analysis of two heterologous genes- the over expressed luc and the under expressed $\beta hCG$- that are expressed in vAc $\beta hCG$-luc infected insect cells

Since the above analysis was based on expression values reported by different group in different sets of experiment, detail analyses were carried out for the two genes that are placed under the identical transcriptional control of polh gene promoter in the same baculovirus and expressed in the same cell. The analysis of over expressed heterologous genes, luc, showed a codon usage profile very similar to AcNPV as reflected by a low value. This similarity could also be seen at the level of individual codons. The D square value for $\beta hCG$ gene was found to be much higher than both the over-expressed luc pattern and the overall AcNPV pattern suggesting a poor translation elongation rate from its start site. The comparison of luc and $\beta hCG$ pattern with the highly expressed AcNPV gene pattern also revealed that the codon usage pattern of highly expressed luc gene is close to AcNPV as opposed to the pattern of under expressed $\beta hCG$.

16. Codon usage can also influence the strength of secondary structure

In addition to its influence on translation elongation rate the codon usage can also increase the strength of mRNA secondary structure. Analysis of codon usage profile $\beta hCG$ showed a very high utilization of the C/G rich codons and the codons with C/G at wobble base position, compared to the luc and the
AcNPV genes. This results in stronger secondary structure forming potential of $\beta hCG$ as opposed to $luc$. These strong secondary structures are known to interfere with the heterologous gene expression (Sarkar et al., 1991).

17. Importance of these patterns in protein synthesis at very late time point of infection

In BEVS, most of the heterologous genes are placed under transcriptional control of hyperactive $polh$ and $p10$ gene promoters (O'Reilly et al., 1992; Richardson, 1995). Since these promoters are hyper-activated very late in infection when host genes are shut off, the infected cells face a very critical situation. The infected cells have abundant viral gene transcripts to cope with the finite translational machinery resources. Since the virus passes through this situation each time it infects the insect cell it has optimized it's gene sequence and the pattern associated with the latter. In the baculovirus the optimum translation initiation codon context pattern enhances the efficiency of translation start site selection where as codon usage pattern enhances the rate of translation elongation, thus both together contributing to efficient production of viral proteins essential for propagation and survival of the virus.

18. Poor expression of heterologous genes: A result of nucleotide sequence pattern mismatch

Having assessed the need for an optimal pattern in baculovirus and shown the association between pattern match and the heterologous gene expression efficiency, it is easy to comprehend why the two heterologous genes placed under identical transcriptional control show differential gene expression. The nucleotide sequence of gene has some species specific
pattern associated with it. These are the translation initiation codon context pattern and the codon usage pattern. These patterns are often suitable for its natural location and its normal expression. The problem arise when an attempt is made to over express a gene belonging to one species by cloning it in an expression system belonging to an evolutionary diverged species (Kurland, 1991). Since the surrogate host system, such as the baculovirus differs in its pattern requirement from the origin of heterologous gene, hence, these attempts often result in mismatch of pattern leading to poor expression levels.