Volume 15 Number 24 1987 Nucleic Acids Research

Analysis of the polyhedrin gene promoter of the *Autographa californica* nuclear polyhedrosis virus

R.D.Possee and S.C.Howard

NERC Institute of Virology, Mansfield Road, Oxford OX1 3SR, UK

Received September 22, 1987; Revised and Accepted November 25, 1987

ABSTRACT

The polyhedrin gene promoter of the Autographa californica nuclear polyhedrosis virus was analysed with respect to which sequences are required upstream of the mRNA transcription initiation (CAP) site for efficient promoter activity. Insertions (8, 95 and 785 nucleotides) were made in this region at an EcoR V site between the CAAT- and TATA-like boxes. When these mutations were introduced into the virus they did not affect the activity of the polyhedrin promoter as judged by expression of the beta-galactosidase (lacZ) gene inserted in lieu of the polyhedrin coding sequences. Deletions were made in the promoter which progressively removed sequences upstream from the CAP site. Removal of the TATA motif did not affect lacZ gene expression. A sequence 69 nucleotides upstream to the normal position of the polyhedrin ATG translation initiation codon was sufficient for maximum promoter activity but this was reduced by 90% when only 56 nucleotides upstream remained. The normal CAP site was utilized by each deletion mutant. Promoter activity was undetectable when the CAP site was deleted. The results are discussed in relation to other eukaryotic promoters.

INTRODUCTION

Autographa californica nuclear polyhedrosis virus (AcNPV) is regarded as the prototype insect baculovirus and has become the isolate of choice for most of the molecular studies on this group of viruses. It has a genome of covalently closed circular DNA of about 130 kilobase pairs (kbp) packaged in a rod-shaped nucleocapsid. The nucleocapsids are enclosed within a lipoprotein envelope to form virus particles. These virus particles are further packaged into large proteinaceous inclusion bodies called polyhedra (see ref. 1).

In the infected cell the expression of polyhedrin protein is confined to the late stages of the infection cycle (15-96 h post-infection (p.i.)), after production of the virus particles which spread the infection to other cells. The polyhedrin protein accumulates to very high levels within the infected cell and has been variously estimated to account for 20-50% of the total cell associated protein by the end of the virus replication cycle (2-5). Polyhedrin gene deletion mutants of AcNPV produce infectious virus particles but not polyhedrin inclusion bodies (3). This feature of the virus has allowed the

substitution of foreign genes in place of the polyhedrin and their expression to high levels in cells infected with the recombinant viruses (4, 6, 7).

However, while the use of baculoviruses as expression vectors is well established, our knowledge of the nature of the polyhedrin promoter is confined to an accurate position for transcription initiation (8) and evidence that the full sequence from the mRNA CAP site up to the ATG of the polyhedrin coding sequences is required for the optimal expression of foreign proteins (7).

In the nucleotide sequence data published for the polyhedrin gene and its 5' flanking region, several elements have been identified which may be of importance in understanding its promoter (2). The sequence data reveal a number of features which are characteristic of eukaryotic promoters. A TATA-like box (TAATTAAAAT) and a CAAT-like sequence (TATCAATAT) are located 24-33 nucleotides and 56-64 nucleotides upstream from the CAP site, respectively. Furthermore there are two tandem repeats (CACAAACT) about 80 and 90 bp upstream from the mRNA CAP site . In recent studies the nucleotide sequence data upstream from the polyhedrin gene has been extended (K. Gearing and R.D. Possee, unpublished data) revealing the presence of another open reading frame, 603 nucleotides in length, encoded by the opposite strand of DNA with an ATG codon beginning 156 bp from the polyhedrin gene ATG. Potentially this defines the intergenic region and allows putative promoter elements of the polyhedrin gene to be analysed. In the study reported here we have investigated the effect of making deletions in the promoter region of the polyhedrin gene. The data revealed that it was possible to delete the TATA-like elements without adversely affecting the high level expression of an introduced foreign gene or the point of initiation of mRNA transcription. Furthermore the CAAT-like box and tandem repeats can be separated from the transcriptional elements by up to 785 nucleotides and still not affect the level of expression. The minimal sequence required upstream from the mRNA CAP site for optimal expression is in the region of 20 nucleotides, i.e., 69 nucleotides upstream from the normal polyhedrin ATG codon.

MATERIALS AND METHODS

Viruses and Cells. AcNPV C6 (9) was propagated in Spodoptera frugiperda cells (IPLB-Sf-21) (10) at 28°C using TC100 medium supplemented with 10% foetal calf serum (TC100/10%FCS). Plaque assays were carried out as described by Brown and Faulkner (11). Recombinant viruses containing the beta-galactosidase gene were grown in the same way as the normal virus with the exception that in the plaque assay the plates were incubated with 60µg/ml X-gal at 28°C for 5h at 3 days p.i., the liquid overlay removed and the plates inverted and left at ambient temperature overnight before counting the blue plaques.

Analysis of proteins from infected cells. S. frugiperda cells (10^6/35mm diameter petri dish) were inoculated with virus at a multiplicity of infection (MOI) of 10 pfu per cell, or with medium for a mock infected control. After 1h at room temperature virus was removed and replaced with 2ml of TC100/10% FCS and the plates incubated at 28°C until 24h p.i. The cells were aspirated from the plates, pelleted at low speed, washed with 1ml PBS, repelleted and finally resuspended in 100µl TE (10mM Tris-HCl, pH 7.8, 0.1mM EDTA). A dissociation solution was added to give a final concentration of 30mM Tris-HCl, pH 6.9, 2-mercaptoethanol, 5% SDS, 5% glycerol and a trace amount of bromophenol blue. Samples were heated at 100°C for 5min before analysis in discontinuous polyacrylamide gels (12% resolving gel, 4.5% stacking gel) using Laemmli buffer (12). The gels were stained in 0.2% Coomassie brilliant blue (Bio-Rad) dissolved in 5% glacial acetic acid, 50% methanol and 45% water. Destaining was performed in the same solvent lacking Coomassie blue.

Extraction of RNA from cells. S. frugiperda cells (10^7/75cm^2 flask) were inoculated with AcNPV or recombinant viruses at a MOI of 10 pfu/cell. After 1h at room temperature the virus was removed and TC100/10% FCS added. The cultures were incubated at 28°C until 24h p.i. when the cells were harvested and the RNA extracted using the hot phenol with guanidinium isothiocyanate method (13).

RNA gel electrophoresis and Northern blotting. RNA gel electrophoresis, Northern blotting and hybridization were carried out as described previously (9). Nick translated DNA probes were prepared using standard procedures (14).

Construction of transfer vectors with modified polyhedrin promoters. The transfer vector pAcRP23 was used as the basis for further modifications to the polyhedrin promoter. The derivation of vectors similar to this one have been previously described (7, 9). Its most important feature is that it contains a full length copy of the leader sequence of the polyhedrin gene from the transcription initiation site (8) up to, but not including, the ATG. This construction was found to be critical for optimal expression in the baculovirus system (7).

(a) Additions to the polyhedrin promoter. pAcRP23 was digested with EcoR V, dephosphorylated, ligated to octanucleotide Bgl II linkers and used to transform competent E. coli JM105 cells. Clones were identified with a Bgl II site in place of the original EcoR V and one was designated pAcRP23+8. This plasmid was digested with Bgl II, dephosphorylated and ligated with an 87bp synthetic oligonucleotide or the 777bp chloramphenicol acetyl transferase (CAT) gene cassette (Pharmacia). Both fragments had BamH I compatible ends. Appropriate clones were identified after transforming E. coli JM105 cells and designated pAcRP23+95 or pAcRP23+785 respectively (with due allowance for the Bgl II linker).

(b) Deletion of the polyhedrin promoter. pAcRP23 was digested with EcoR V and then treated with Bal31 exonuclease at 30°C for 1min (13). Samples were taken at intervals of 10 seconds and the reaction terminated with phenol/chloroform. The DNA was precipitated with ethanol, treated with S1 nuclease to trim the ragged ends of the DNA, dephosphorylated and ligated with Bgl II linkers. This was used to transform E. coli JM105 cells, recombinants were isolated and the extent of the deletion from the EcoRV site toward the polyhedrin transcription start site determined by size analysis using polyacrylamide gels. Deletions were confirmed by DNA sequencing (15). Suitable clones were further modified to replace the deleted sequences upstream from the normal position of the EcoR V site by digesting the DNA with Xho I and Bgl II, purifying the large fragment from a low gelling temperature agarose gel and ligating this with the full length Xho I-Bgl II fragment from pAcRP23+8.

Construction of transfer vectors containing the beta galactosidase gene. The plasmid CH110, containing the lacZ gene, was purchased from Pharmacia. It was digested with Hind III, the protruding ends trimmed with S1 nuclease, dephosphorylated and ligated to Bgl II linkers. This construction placed the Bgl II site upstream from the ATG of the lacZ and allowed the entire gene to be excised from the plasmid with Bgl II and BamH I which digests the 3' end of the gene after the SV40 transcription termination sequences. This gene cassette was then inserted into the normal pAcRP23 transfer vector at the BamH I cloning site or the other vectors described above with additions to, or deletions from, the polyhedrin promoter.

Preparation of recombinant viruses containing the beta galactosidase gene. Transfer vectors containing the lacZ gene and various modifications to the polyhedrin promoter were used in co-transfection experiments according to protocols previously described (9). The progeny virus from co-transfections was assayed for recombinants by incubating plaque assays with X-gal as described above. When blue plaques were not detected the plates were stained with 0.01% neutral red for 2h at 28°C then left to clear overnight at room temperature in the dark. Plaques were then screened for the presence or absence of polyhedra using low power microscopy. The putative recombinant plaques identified by both methods were picked into 0.5ml TC100/10% FCS and purified to homogeneity using further rounds of plaque purification (2-3 times). Larger virus stocks were prepared as previously described (9).

Primer extension analysis of the 5' end of mRNA. Primer extension analysis of the RNA produced by lacZ recombinant virus in S. frugiperda cells was carried out as described previously (8).

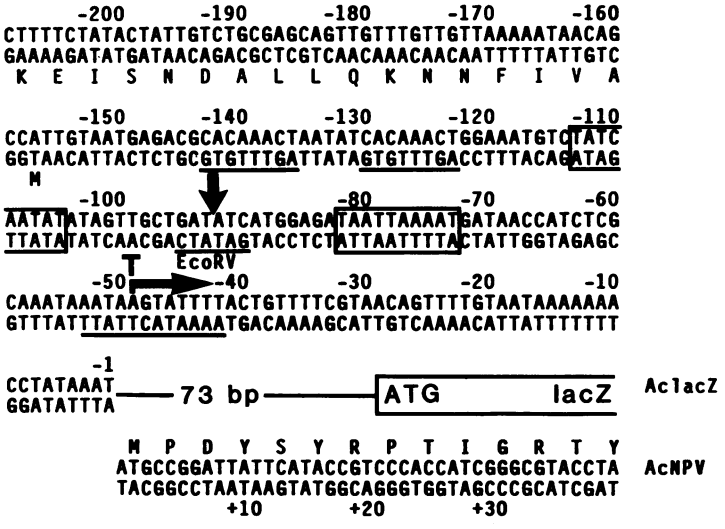

Figure 1. Nucleotide sequence of the intergenic region between polyhedrin and the upstream open reading frame (orf 603). The coding sequences for orf 603 (beginning at -156) are shown together with the amino acid sequence. The two tandem repeats (-143 to -136 and -130 to -123) are underlined. The CAAT-like box (-113 to -105), the TATA-like box (-83 to -73) are completely enclosed and the baculovirus late gene transcription initiation concensus sequence (-53 to -42) underlined (26). The large horizontal arrow and "T" depicts the exact position of transcription initiation (8). The position of the insertions (+8, +95 and +785 bp) made in the EcoR V site are indicated with a large vertical arrow. A schematic representation of the beta-galactosidase (lacZ) gene is shown to illustrate the the 73 bp leader sequence and ATG translation initiation codon (AclacZ). The normal position and sequence of part of the polyhedrin coding region is also shown.

RESULTS

Construction of plasmids with modifications to the polyhedrin promoter

In order to shed some light on the importance of the various elements in the intergenic region between polyhedrin and the upstream gene located on the opposite DNA strand we initially decided to make modifications in the form of insertions to the putative promoter. This approach was adopted because of the uncertainty of whether the two genes were regulated independently. The location of the mRNA CAP site for the upstream gene has not been determined. Fortuitously in the intergenic region there is an EcoR V site about 46 bp upstream from the polyhedrin mRNA CAP site which also effectively divides the CAAT- and TATA-like elements. Figure 1 shows the sequence data for the 5' end of the polyhedrin gene (2) with part of the open reading frame and

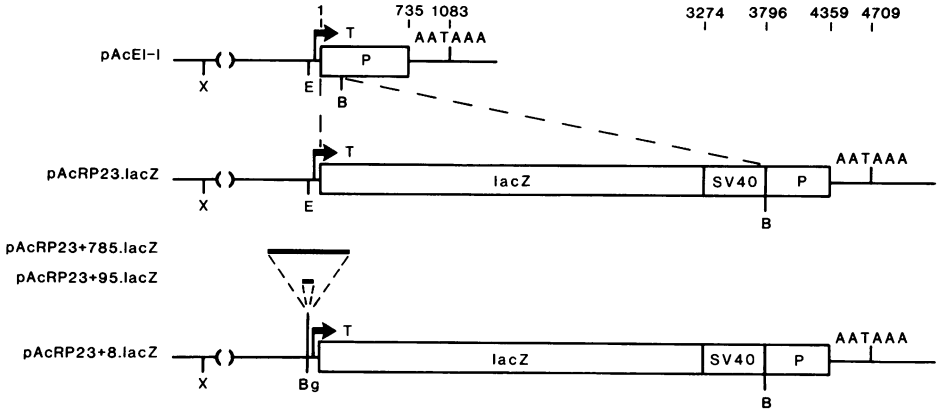

Figure 2. Genetic organisation of the transfer vectors containing promoter insertions. pAcEI-I represents the region of the virus genome containing the polyhedrin gene. pAcRP23.lacZ is a construction with the lacZ gene in place of the polyhedrin gene and pAcRP23+785, +95 or +8.lacZ the plasmids with insertions made in the EcoR V site (E) in the polyhedrin promoter region. The nucleotide positions are shown in the top line. T=transcription initiation, AATAAA=transcription termination, P=polyhedrin, SV40=termination sequences from this virus, lacZ=beta galactosidase, X= Xho I, B=BamH I and Bg=Bgl II.

amino acid sequence for the gene that is upstream of polyhedrin. The positions of the tandem repeats (-143 to -136 and -130 to -123), CAAT-like box (-113 to -105) and TATA-like box (-82 to -73) are indicated relative to the normal position of the polyhedrin ATG. The consensus sequence for the transcription initiation region highlighted by Rohrmann is also shown (16). Insertions were made at this EcoR V site and tested for their effect on expression of a reporter gene in place of the polyhedrin gene. The transfer vector pAcRP23 has a deletion from the polyhedrin ATG to the BamH I site 170 bp downstream. This construction gives high level expression of foreign genes when they are inserted in place of the deleted sequences and then transferred to the virus genome (7). pAcRP23 was digested with EcoR V and a synthetic octanucleotide Bgl II linker inserted using standard procedures. This formed the first insertion mutant and was designated pAcRP23+8. This plasmid was used to construct other insertion mutants by digesting with Bgl II and inserting a synthetic oligonucleotide (87 bp) or the CAT gene cassette (777 bp) into the site to produce the plasmids pAcRP23+95 and pAcRP+785 respectively (with allowance for the Bgl II linker). To test the effect of these modifications on promoter activity a reporter gene was inserted at the cloning site for each of the altered transfer vectors. The gene of choice was the E. coli lacZ which has previously been expressed in insect cells using the baculovirus system (4). Figure 2 shows a summary of the constructions beginning with

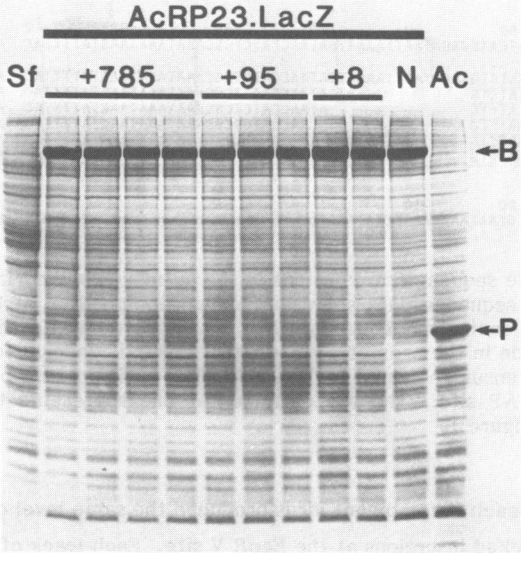

Figure 3. Expression of lacZ by recombinant viruses. Spodoptera frugiperda cells (Sf) were inoculated with 10 pfu/cell of AcNPV (Ac) or lacZ recombinant viruses (AcRP23.lacZ) with a normal promoter (N) or insertions (+785, +95 and +8) at the EcoR V site. After adsorbtion for one hour at room temperature the cells were incubated at 28°C until 24h p.i. and then protein extracts analysed in 12% polyacrylamide gels using Coomassie blue staining. B=beta-galactosidase and P=polyhedrin.

the normal AcNPV EcoRI-I DNA fragment containing the polyhedrin gene and followed by the pAcRP23 transfer vectors containing the lacZ gene with or without additions to the promoter region.

Construction of recombinant viruses and assays of promoter activity

Recombination between infectious virus DNA and the modified plasmids lacking the polyhedrin gene results in the production of polyhedrin deficient virus which can be identified in a standard plaque assay (4, 6, 9). In this study essentially the same approach was adopted but with the advantage that it was possible to screen for recombinants by incubating the cells with X-gal and searching for blue plaques. Performing transfections with the three insertion mutants produced blue plaques in each case. After plaque purification the blue phenotype was stable and accompanied by the expected polyhedrin negative phenotype. Virus DNA from each recombinant was analysed using restriction enzyme mapping and Southern blot hybridization to confirm that the expected insertion (+8, +95 or + 785 bp) was present (data not shown). Expression of lacZ from each virus was assessed by infecting cells at high multiplicity and then analysing protein extracts in polyacrylamide gels stained with Coomassie blue.

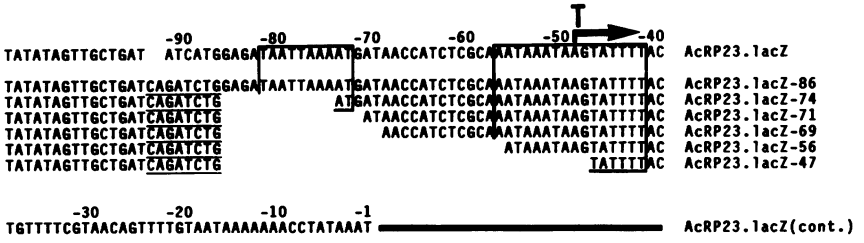

Figure 4. Nucleotide sequence of deletions in the putative polyhedrin promoter. The top and bottom lines of sequence data show the normal composition of the region upstream from the polyhedrin coding sequences (see Fig. 1.). The intervening lines show the extent of each deletion made in the promoter. The Bgl II linker is underlined. The virus TATA-like motif and concensus sequence for transcription initiation are partially boxed, with "T" indicating the CAP site. The thick horizontal line represents the same lacZ gene arrangement as in Figure 1.

Figure 3 shows that each recombinant virus produced the same level of lacZ as the control virus that lacked insertions at the EcoR V site. Each track of the gel represents protein extracts from a different virus isolate and demonstrates the reproducibility of the result. Additionally the level of expression compares favourably with that of the normal polyhedrin protein. Therefore it appears that making insertions in the EcoR V site of the intergenic region does not affect the ability of the polyhedrin promoter to express a foreign gene.

Construction of recombinant viruses with deletions in the polyhedrin promoter

The results from the above experiments indicated that the CAAT-like box does not play a role in expression of the polyhedrin gene, unless this element can have an effect over 785 nucleotides. Attention was turned to the role of the TATA-like box upstream from the polyhedrin. Deletions were made from the EcoR V site towards the polyhedrin gene CAP site. In every construction the sequences removed from the region upstream from the EcoR V were replaced. The series of deletion mutants created are illustrated in Figure 4. Each has the lacZ gene in place of the polyhedrin. The number assigned to each plasmid (-86 to -47) refers to the position of the deletion relative to the first nucleotide of the polyhedrin ATG. These plasmids were used in cotransfections with AcNPV DNA to produce recombinant viruses containing the lacZ gene under the control of the modified polyhedrin promoter. Blue plaques were produced from the progeny of the transfections for all constructions except the deletion where part of the mRNA CAP site had been removed (Fig. 4, pAcRP23.lacZ-47). In this case the recombinant virus was isolated by screening for polyhedrin-negative plaques in the normal way (9). After plaque purification DNA from each recombinant virus was analysed using Southern blot hybridization to confirm that the lacZ gene had

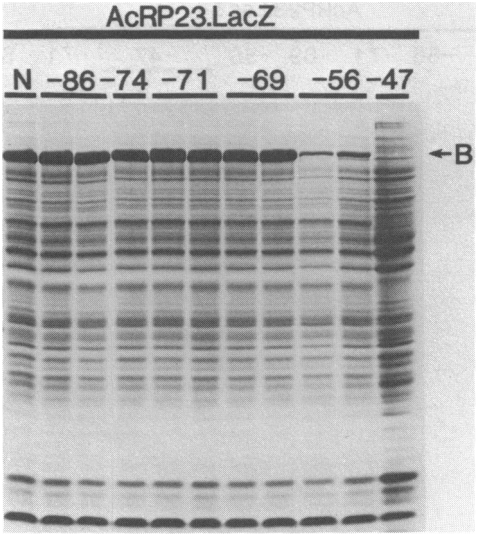

Figure 5. Expression of lacZ by recombinant viruses. S. frugiperda cells were infected with lacZ recombinant viruses (AcRP23.lacZ) with a normal promoter (N) or with promoters containing deletions (-86 through -47) and the proteins analysed as described previously (Fig. 3). Duplicate tracks for some viruses represent different isolates.

been transferred to the virus. DNA from purified virus was also sequenced (15) to confirm the nucleotide composition of each of the promoter deletions (data not shown).

Assaying levels of lacZ expression by recombinant viruses with promoter deletions

The recombinant viruses containing promoter deletions were used to infect S. frugiperda cells at high multiplicities. Protein extracts from these cells were analysed in polyacrylamide gels. Figure 5 shows that the levels of lacZ expression were constant over the deletion series until 56 nucleotides upstream from the normal position of the polyhedrin ATG remained. This point is 7 bp upstream from the mRNA CAP site. Viruses containing this deletion produced about 10-20% of normal levels of the lacZ protein. This yield was consistently observed in other isolates of the same recombinant (data not shown). When the mRNA CAP site was deleted lacZ protein was undetectable in cells infected with the recombinant virus (AcRP23.lacZ-47). Normal levels of lacZ protein production were observed when 69 nucleotides upstream from the normal ATG position were retained. However the complete TATA-like box was absent from this deletion.

Expression was also assayed by analysing the amount of RNA produced by the recombinant virus in infected cells. Figure 6 shows that the levels of lacZ-specific RNA in infected cells were consistent with the data for protein production. Control

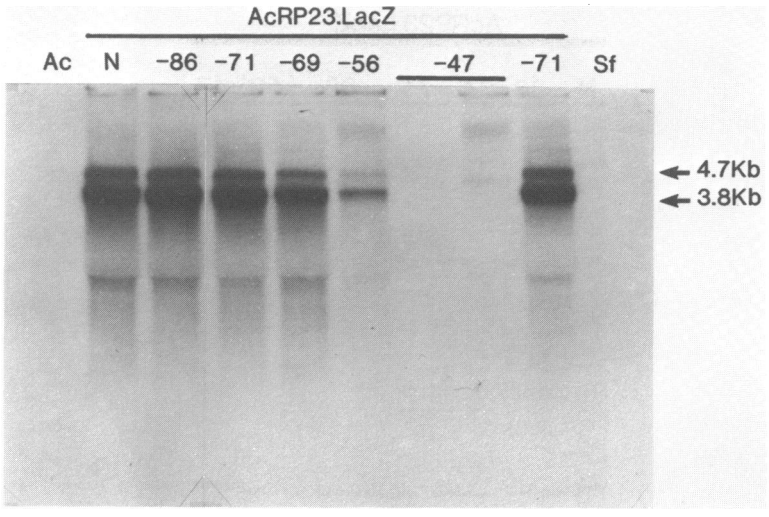

Figure 6. Northern blot analysis of RNA from cells infected with lacZ recombinant viruses. Total RNA was extracted from cells infected with AcNPV (Ac), lacZ recombinant viruses (AcRP23.lacZ) with promoter deletions (-86 through -47) or mock infected S. frugiperda cells (Sf). The RNA was fractionated in a 1% agarose/glyoxal gel, transferred to Genescreen and hybridized with nick translated lacZ-specific DNA. The two sizes of RNA detected are indicated on the right of the autoradiograph.

levels were maintained by mutants lacking all but 69 nucleotides upstream from the polyhedrin ATG. Thereafter the level dropped at a point 56 nucleotides upstream from the same site and was undetectable when the mRNA CAP site was deleted. Figure 6 also shows that two species of RNA (4.7 kb and 3.8 kb) were identified by the lacZ-specific probe in the samples from the recombinant virus-infected cells. Consideration of the genetic organisation of the recombinant viruses containing the lacZ gene (Fig. 2) predicted that a mRNA initiating at the polyhedrin CAP site and terminating by utilization of the polyhedrin polyadenylation signal should yield a species of 4.7 kb. However, if the SV40 transcription termination sequences were used, an RNA of about 3.8 kb would be produced. This hypothesis was tested by Northern blot analysis of RNA produced by lacZ recombinants with probes specific for lacZ or polyhedrin alone (Figure 7). The 4.7 kb RNA was highlighted by both probes but the 3.8 kb RNA only by the lacZ probe. This demonstrated that the 4.7 kb RNA alone had the residual polyhedrin sequences incorporated and that the 3.8 kb RNA terminated at the SV40 sequences. It is also worth noting that RNA from AcNPV-infected cells, included as a control, showed that the amounts of RNA produced by AcNPV and the lacZ recombinants were very similar.

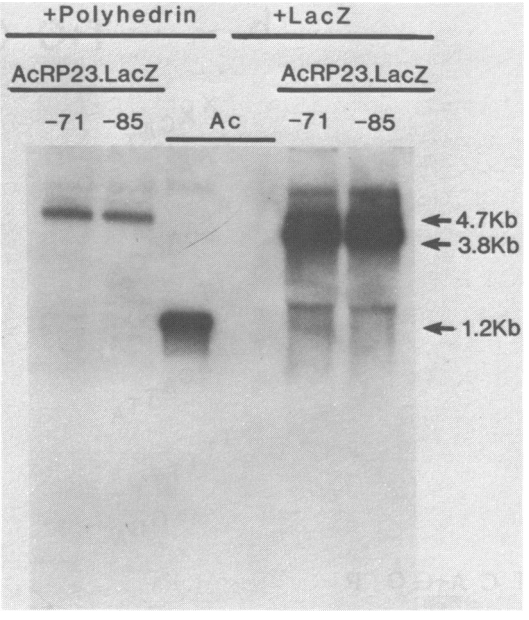

Figure 7. Identification of RNA species produced by lacZ recombinant viruses. Duplicate total RNA samples from cells infected with lacZ recombinant viruses (AcRP23.lacZ-71 or -85) and AcNPV (Ac) were fractionated in a gel and blotted to Genescreen as before (Fig.6). The filter was divided into two and one half hybridized to nick translated polyhedin specific DNA (+polyhedrin) and the other to lacZ specific DNA (+lacZ). The sizes in kilobase pairs are indicated on the right of the figure.

Transcription initiation by lacZ recombinants

The recombinant AcRP23.lacZ-69 completely lacked the TATA-like element but still produced the same amount of RNA and protein as control virus. TATA boxes can affect the point at which transcription is initiated (17), therefore we investigated whether mRNA produced by this recombinant started at the same point as polyhedrin mRNA synthesised by normal AcNPV. Primer extension analyses of the mRNA sequence were performed and the size of the cDNA copy estimated to be 135-136 nucleotides by comparison with a sequence ladder (Fig. 8a). The cDNA was further characterised by sequencing with the chemical method (15). Figure 8b demonstrates that the mRNA initiated at almost the same point (-49) as the normal polyhedrin mRNA (8). The difference observed here (+1 nucleotide) is considered to be within the expected variation of the method and not significant. A similar analysis of the lower amounts of RNA extracted from cells infected with the AcRP23.lacZ-56 recombinant did not produce enough cDNA to permit sequencing, however the product was sized against a

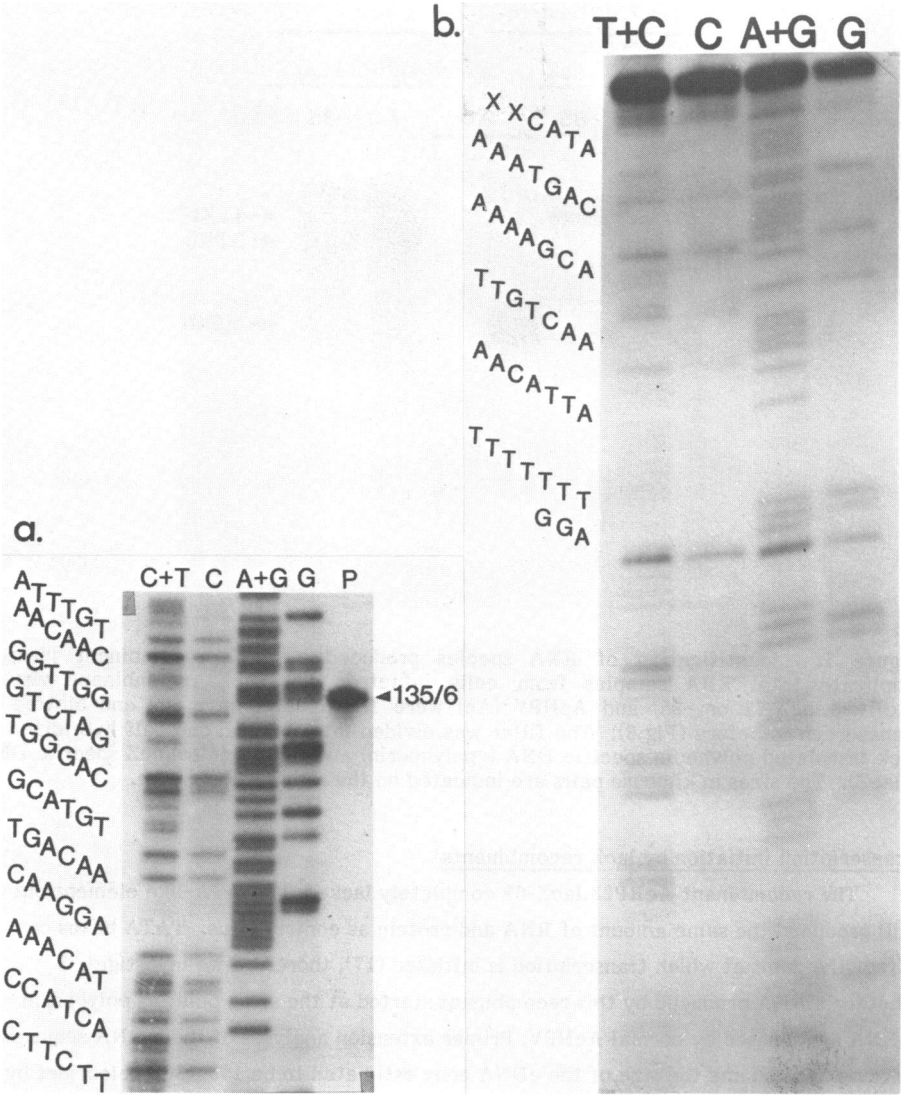

Figure 8. Analysis of cDNA prepared using primer extension of lacZ RNA from AcRP23.lacZ-69 infected S. frugiperda cells. The 5' end of the oligonucleotide lacZ primer was radiolabelled with [^{32}P] ATP and polynucleotide kinase, annealed to the RNA and extended with reverse transcriptase. (a) Estimation of the size (135/6 b) of the primer extension product (P) by comparison with a sequencing ladder. (b) Sequence of the cDNA as determined using the chemical method. The sequence printed to the side of the gel shows the complement of the lacZ RNA beginning 5 nucleotides upstream from the normal position of the polyhedrin AUG.

sequencing ladder to demonstrate that transcription initiation also occurred at the same site (data not shown).

DISCUSSION

The function of the sequences upstream from the baculovirus AcNPV polyhedrin transcription start site have been investigated. The CAAT- and TATA-like elements are separated by an EcoR V restriction enzyme site. When DNA fragments of varying size (8, 95 or 785 bp) were inserted into this site in one of the transfer vectors (pAcRP23) and the modified plasmids used to transfer the lacZ gene into AcNPV, the insertion of these elements into the recombinant virus did not affect the amount of lacZ protein made in infected cells. This suggested that the CAAT structure and the tandem repeats upstream from this could be separated from the polyhedrin transcription unit by at least 785 nucleotides without detriment to its function. It is not known whether these sequences function as enhancer elements. The SV40 enhancer elements can act in either orientation 1400 bp upstream or 3300 bp downstream from the transcription initiation site of the rabbit beta-globin gene (18,19). It would be unwise to completely discount a similar role for the CAAT-like sequences and tandem repeats until appropriate experiments have been performed. The best evidence for enhancer activities of defined sequences in the baculoviruses pertains to the hr5 region of AcNPV which stimulates the transient expression of a delayed-early gene 1000-fold (20, 21). The possible role of hr5 and hr1-4 in the enhancement of polyhedrin gene expression remains to be investigated.

Our conclusions from the experiments where insertions were made into the promoter region left the TATA-like box as a candidate control element in the polyhedrin promoter. However deletion of the entire TATA box did not affect high level expression of lacZ. Retention of a 69 bp region upstream from the normal position of the polyhedrin was sufficient for full promoter activity. Expression was judged by analysis of protein extracts from cells infected with the recombinant viruses and by analysis of RNA from infected cells harvested at 24h p.i. Deletion to a point 7 nucleotides before the mRNA CAP (or 56 bp upstream from the polyhedrin ATG) reduced expression of both RNA and protein to about 10% of the control.

An interesting result was the detection of two lacZ-specific RNA size classes in cells infected with the lacZ recombinants. Differential hybridization analysis indicated that the larger of the two (4.7 kb) was produced as a result of transcription terminating at the normal polyhedrin signal. The smaller RNA (3.8 kb) was consistent with transcription terminating at the SV40 signals included in the recombinant virus. The majority of the RNA was of the smaller size indicating very efficient use of the heterologous termination signals. We also conclude that the larger species was not due

to transcription initiating further upstream from the normal CAP site as the larger RNA was not seen in cells infected with the AcRP23.lacZ-47 recombinant which yielded undetectable lacZ-specific RNA.

The point at which transcription begins can be affected by the TATA box in some genes (17). Primer extension analysis of the RNA from recombinant viruses containing the -69 and -56 end points and consequently lacking the TATA-like box revealed that this was not the case with the polyhedrin promoter. Transcription was determined to begin at the same point (+ 1 nucleotide) as in normal polyhedrin expression (8). The cDNA copy made of the RNA from AcRP23.lacZ-69 was sequenced to confirm this.

The structure and function of eukaryotic promoters in a number of other systems have been analysed. From these data it is apparent that there is not an overall consensus for the constituents of a promoter. For example, the rabbit beta-globin gene has an ATA box (equivalent to the TATA box), a CAAT box and tandemly repeated sequences -30, -74 and -100 nucleotides upstream from transcription initiation (22). All of these regions have a role in globin mRNA transcription. Removal of the ATA box reduced transcription to 1% of control levels however these residual transcripts initiated at the normal CAP site. Deletion of the CAAT box reduced transcription levels by a factor of 5 and again the remaining transcripts initiated at the normal CAP site. Deletions in the tandem repeats reduced transcription by 80-90%. Conversely, for the SV40 early gene promoter removal of the TATA-like box did not affect the level of gene expression but resulted in the generation of multiple CAP sites for the initiation of RNA (17). Removal of the three direct G-C 21 bp repeats (which are adjacent to the TATA box) drastically reduced early promoter function (23, 24). For the herpes simplex virus (HSV) thymidine kinase (tk) gene, promoter functions were identified for a 100 bp region upstream from the CAP site (25). Further analysis using linker scanning methods (26) examined the role of three domains in this sequence. Transcription was abolished when the TATA box located about 25 bp upstream from the CAP site was altered by mutagenesis. However, changes in the CAAT box at -80 bp did not affect transcription, although disruption of G-C rich hexanucleotides on either side of this CAAT box reduced promoter function. Conversely, in the vaccinia virus tk promoter retention of a region from -32 before to +5 after the mRNA start site was sufficient for full promoter activity (27) and a similar requirement was observed with the promoter from the gene encoding the vaccinia 7.5 K protein (28). Analysis of the gene promoter for the 11 K protein in the same virus also showed that a minimal sequence before transcription start (15 bp) was required for efficient expression (29, 30). However, with this example there is some debate as to whether a TATA-like motif at the RNA CAP site should be regarded as a genuine TATA box (30). These examples illustrate that the presence of particular sequences in a promoter region of a gene are not a guarantee of

their playing a functional role in transcription. Also, some vaccinia virus genes can have transcription units comprising very short sequences. The data presented in this paper also points to a minimal sequence requirement for the functioning of the polyhedrin gene promoter. Preservation of only 20 bp upstream from the RNA CAP site maintained a high level of gene expression. It was previously observed that the full 5' leader sequence of the polyhedrin gene was required for maximum expression of foreign genes in the baculovirus system (7). This data, coupled with the results from the present study, suggests that the polyhedrin promoter resides in a fragment with a maximum length of 69 bp (20 bp upstream and 49 bp downstream from the RNA CAP site). This conclusion is supported by other data (31) where it was shown that the EcoR V site upstream from the polyhedrin promoter could be used to insert a duplicate of this promoter to facilitate expression of a second gene in addition to the polyhedrin protein. Presumably the 69 bp region contains the binding site(s) for the protein co-factors and RNA polymerase molecule. These components of the polyhedrin expression system remain to be elucidated.

ACKNOWLEDGEMENTS

We thank C.D. Hatton for photography and D.H.L. Bishop for comments on the manuscript.

REFERENCES

1. Kelly, D.C. (1985) In Maramorosch, K. and Sherman, K.E. (eds), Viral Insecticides for Biological Control, Academic Press, New York, pp. 469-488.
2. Hooft van Iddekinge, B.J.L., Smith, G.E. and Summers, M.D. (1983) Virology **131**, 561-565.
3. Smith, G.E., Fraser, M.J. and Summers, M.D. (1983) J. Virol. **46**, 584-593.
4. Pennock, G.D., Shoemaker,C. and Miller, L.K. (1984) Mol. Cell. Biol. **4**, 399-406.
5. Miyamoto, C., Smith, G.E., Farrell-Towt, J., Chizzonite, R., Summers, M.D. and Ju, G. (1985) Mol. Cell. Biol. **5**, 2860-2865.
6. Smith, G.E., Summers, M.D. and Fraser, M.J. (1983) Mol. Cell. Biol. **3**, 2156-2165.
7. Matsuura, Y., Possee, R.D., Overton, H.A. and Bishop, D.H.L. (1987) J. gen. Virol. **68**, 1233-1250.
8. Howard, S.C., Ayres, M.D. and Possee, R.D. (1986) Virus Res. **5**, 109-119.
9. Possee, R.D. (1986) Virus Res. **5**, 43-59.
10. Vaughn, J.L., Goodwin, R.H., Thompkins, G.J. and McCawley, P. (1977) In Vitro **13**, 213-217.
11. Brown, M. and Faulkner, P. (1977) J. gen. Virol. **36**, 361-364.
12. Laemmli, U.K. (1970) Nature, London **227**, 680-685.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
14. Rigby, P.W.F., Dieckman, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. **113**, 237-251.
15. Maxam, A.M. and Gilbert, W. (1980) Proc. Natl. Acad. Sci. USA **74**, 560- 564.
16. Rohrmann, G.F. (1986) J. gen. Virol. **67**, 1499-1513.
17. Benoist, C. and Chambon, P. (1981) Nature **290**, 304-310.
18. Gruss, P., Dhar, R. and Khoury, G. (1981) Proc. Natl. Acad. Sci. USA **78**, 943-947.

19. Banerji, J., Rusconi, S and Schaffner, W. (1981) Cell **27**, 299-308.
20. Guarino, L. and Summers, M.D. (1986). J. Virol. **60**, 215-223.
21. Guarino, L., Gonzalez, M.A. and Summers, M.D. (1986) J. Virol. **60**, 224-229.
22. Dierks, P., van Ooyen, A., Cochran, M.D., Dookin, C., Reiser, J. and Weissmann, C. (1983) Cell **32**, 695-706.
23. Fromm, M., and Berg, P. (1982) J. Mol. Appl. Genet. **1**, 457-481.
24. Everett, R.D., Baty, D. and Chambon, P. (1983) Nucleic Acids Res. **11**, 2447-2464.
25. McKnight, S.L., Gavis, E.R. and Kingsbury, R. (1981) Cell **25** 385-398.
26. McKnight, S.L. and Kingsbury, R. (1982) Science **217**, 316-324.
27. Weir, J.P. and Moss, B. (1987) Virology **158**, 206-210.
28. Cochran, M.A., Puckett, C. and Moss, B. (1985) J. Virol. **54** 30-37.
29. Hanggi, M., Bannwarth, W. and Stunnenberg, H.G. (1986) EMBO J. **5**, 1071-1076.
30. Bertholet, C., Stocco, P., Van Meir, E. and Wittek, R. (1986) EMBO J. **5**, 1951-1957.
31. Emery, V.C. and Bishop, D.H.L. (1987) Protein Engineering **1**, 359-366.