Characterization of an early gene coding for a highly basic 8.9K protein from the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus

Xiaoning Wu, Sandra Stewart and David A. Theilmann∗

*Agriculture Canada Research Station, 6660 N.W. Marine Drive, Vancouver, British Columbia, Canada V6T 1X2*

A new gene of the baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus (OpMNPV) has been identified that encodes a highly basic 8.9K protein. The gene called p8.9 is expressed as a 0.5 kb transcript by 1 h post-infection and initiates at an early gene motif. The promoter of the 0.5 kb transcript has two upstream elements, repeats I and II, which are similar to motifs previously characterized in the OpMNPV IE-2 gene and the *Autographa californica* nuclear polyhedrosis virus IE-N and PE38 genes. A second p8.9 transcript expressed from 8 to 72 h post-infection was shown to initiate 634 bp upstream from the early gene motif in a region that has no similarity to any previously described baculovirus promoter or initiation site. Transient assays utilizing a reporter gene have shown that the p8.9 early promoter is active in a *Lymantria dispar* (LD652Y) and *Spodoptera frugipeda* (Sf9) cell lines in the absence of other viral factors. In addition, it was also demonstrated that the p8.9 promoter is trans-activated by the OpMNPV IE-2 and p34 genes, but not by IE-1.

**Introduction**

The baculovirus *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus (OpMNPV) is a pathogenic insect virus that has a large DNA genome of approximately 130 kb. The regulation of gene expression of OpMNPV and other baculoviruses is complex and appears to be temporally regulated at the level of transcription (for review, see Blissard & Rohrmann, 1990). Based on their temporal expression, baculovirus genes have been classified into three general groups: early, late, and very late. Early genes are expressed before DNA replication and require host RNA polymerase II whereas late genes are expressed after the onset of DNA replication and utilize an RNA polymerase that is resistant to α-amanitin (Grula et al., 1981; Fuchs et al., 1983; Huh & Weaver, 1990). In this study we describe the identification of a new OpMNPV early gene.

Some baculovirus early genes have high levels of expression in transfected cells and therefore appear to utilize host cell factors to regulate their expression (Guarino & Summers, 1987; Nissen & Friesen, 1989; Blissard & Rohrmann, 1991; Hoopes & Rohrmann, 1991; Theilmann & Stewart, 1991, 1992a). At or near the transcriptional start site of the early gene promoters of the OpMNPV genes IE-1, IE-2, gp64 and p34 (originally called OpPE38) and the *Autographa californica* (Ac) MNPV genes IE-0, IE-1, IE-N, PE38, 39K and gp64, a highly conserved early gene motif that includes a TATA box followed by a CAGT or CACAGT sequence 24 to 28 bp downstream has been identified (Guarino & Smith, 1992). This early gene motif has been shown to be sufficient for detectable or basal level expression in transfected cells (Blissard & Rohrmann, 1991; Theilmann & Stewart, 1991; Krappa et al., 1992).

The AcMNPV early genes IE-N and PE38 are divergently transcribed using symmetrical and identical promoters. The IE-N and PE38 promoters each contain a single copy of three different regulatory sequences 5' to the early gene motifs. These include a GATA motif and two elements called repeat I and repeat II. All three elements have been shown by Krappa et al. (1992) to bind cellular factors. Repeat I has been shown by Carson et al. (1991) to up-regulate the IE-N promoter in transient assays.

We have previously shown that the promoter of the OpMNPV IE-2 gene differs significantly from the promoter of the related AcMNPV gene, IE-N (Theilmann & Stewart, 1992a). In contrast, the sequences upstream from the IE-2 promoter showed significant similarity to the AcMNPV PE38 promoter. This suggested that the OpMNPV sequence upstream of the IE-2 gene was also a promoter of an OpMNPV early gene.

† Present address: Department of Microbiology, University of British Columbia, Vancouver, Canada V6T 1Z3.
Fig. 1. Gene organization, nucleic acid sequence and predicted amino acid sequence of p8.9 gene. (a) Organization of the p8.9 gene region. Large arrows, p8.9 and IE-2 ORFs; CACAGT, early transcriptional start motif. I and II indicate the regulatory repeat elements I and II (Carson et al., 1991; Krappa et al., 1992). GATA, putative GATA transcription factor binding sites (Krappa et al., 1992); ACAGGACGC, IE-2 repeated element (Theilmann & Stewart, 1992a). The double-headed arrow shows the sequenced region reported in (b). Below the map are diagrams representing the p8.9-β-gal constructs used for analysis of the p8.9 promoter in Fig. 4. Ten mm, 119 bp. (b) Nucleic acid and amino acid (single letter code) sequences of p8.9. The orientation of the nucleic acid sequence is presented...
that is transcribed in the opposite direction to IE-2. Since we have previously identified an OpMNPV PE38 homologous gene (p34) downstream from the IE-2 gene (Theilmann & Stewart, 1992b) it suggested that this possible divergent promoter may regulate a previously unidentified early gene. In this report we describe the characterization of this gene and show that it encodes a small highly basic protein that has no homology to any previously identified baculovirus gene. This gene, called p8-9, is transcribed at very early times, initiated at the conserved early-gene motif. In addition, transient assays using a reporter gene were performed to investigate the possible regulatory interactions between the p8-9 early promoter and the OpMNPV transactivators IE-1, IE-2 and p34.

Methods

Cells and virus. Lymantria dispar cells (LD652Y) and Spodoptera frugiperda cells (SF9) were maintained in TC-100 medium as described (Sumners & Smith, 1987). OpMNPV virus was propagated in L. dispar cells as previously described (Quant-Russell et al., 1987). Time course studies of OpMNPV infection were analysed by infecting LD652Y cells at an m.o.i. of 20 and the 0 h post-infection (p.i.) time-point was defined as the time after the virus was allowed to adsorb to the cells for 1 h.

Plasmid constructs and DNA sequencing. Plasmid clones of p8-9 were constructed using restriction fragments of the cosmid 01347 (Leisy et al., 1991). Plasmid constructs were sequenced using double-stranded plasmid DNA templates (Toneguzzo et al., 1988). DNA and predicted protein sequences were analysed and compared to the GenBank/EMBL and SWISSPROT databases using the UWGCG and DNA Strider sequence analysis packages (Devereux et al., 1984; March, 1988).

The p8-9-β-galactosidase fusion (p8-9-β-gal) constructs were made by inserting the lacZ gene from the pMC1871 plasmid (Shapira et al., 1983) in-frame with the p8-9 gene at the EcoRI site. The various promoter lengths were made using convenient restriction endonuclease sites (SalI, NotI and ApaI; Fig. 1a). OpMNPVIE-1 and IE-2 genes and their promoter constructs are described by Theilmann & Stewart (1991, 1992a). The OpMNPV p34 gene and its promoter constructs were made based on the sequences published previously (Theilmann & Stewart, 1992b). The plasmid of the entire p34 gene includes a 164 bp promoter sequence upstream of the p34 early transcription initiation site (p34-NsiI construct).

RNA isolation and Northern blots. Total RNA from OpMNPV-infected LD652Y cells was prepared as previously described (Theilmann & Stewart, 1991). Northern blots (5 μg total RNA per lane in 1× formaldehyde gels) were prepared according to Thomas (1983) and hybridized to 32P-labelled ssRNA probes at 60 °C in 6× SSC (1 × is 0.15 M-NaCl, 0.015 M-sodium citrate, pH 7.0), 5× Denhardt’s solution (1 × is 0.02% polyvinylpyrrolidone, 0.02% BSA, 0.02% Ficoll 400), 0.1% SDS, 100 μg/ml denatured DNA from salmon sperm, 100 μg/ml yeast RNA and 10% polyethylene glycol. ssRNA probes complementary to p8-9 mRNA were synthesized in vitro using T3 or T7 RNA polymerases in the presence of [α-32P]UTP (ICN) (Sambrook et al., 1989). After hybridization the blots were washed twice in 0.1× SSC and 0.1% SDS at 75 °C followed by exposure to Kodak XAR films with an intensifying screen.

Primer extension and S1 nuclease mapping assays. Primer extension assays were performed as described previously (Theilmann & Stewart, 1991). Two 16-base oligonucleotides 5′ TCTTAGCCGTACAG 3′ and 5′ CAACCTGCTCTGGT 3′ were used as primers for mapping the 5′ end of the 0.5 kb and 1.25 kb p8-9 early mRNAs respectively. A 17-base oligonucleotide 5′ ATGTGCTGAAGCGGTTAT 3′ based on the Escherichia coli β-gal gene sequences was used as a primer for mapping the 5′ end of the expressed β-gal mRNA in transfected cells. S1 nuclease mapping assays were also used to map the 5′ and 3′ ends of p8-9 transcripts as previously described (Theilmann & Stewart, 1991).

Transient assays and β-gal assay. Cells were transfected with plasmid DNA as described previously (Theilmann & Stewart, 1991). Transfected cells were harvested 48 h after transfection and the cell extracts were assayed for β-gal activity using o-nitrophenyl-β-D-galactopyranoside according to a standard protocol (Sambrook et al., 1989). The assays were performed in microtiter plates and absorbance was measured at 420 nm. The total RNAs from the transfected cells were prepared as previously described (Xie & Rothblum, 1991).

Results

Isolation and characterization of the p8-9 gene

The p8-9 gene promoter was first identified during the analysis of the OpMNPV IE-2 gene (Theilmann & Stewart, 1992a). A baculovirus early gene transcriptional start site motif CACAGT was found upstream of the IE-2 promoter, but in a divergent orientation (Fig. 1a). To determine whether this was a promoter of a previously unidentified early gene, the region downstream from the early gene motif was sequenced. A small open reading frame (ORF) was identified that coded for a 75 amino acid protein of M, 898 K (Fig. 1b). Analysis of the predicted p8-9 amino acid sequence showed that it has a very basic amino terminus containing 13 basic and no acidic residues in the first 37 amino acids, which is schematically shown in the acid-base profile (Fig. 1c). In

from 5′ to 3′ relative to the p8-9 ORF in (a). Arrows with the dashed lines indicate the start sites of the p8-9 0.5 kb and 1.25 kb mRNAs. Poly(A) signals and the transcription start site of 1.25 mRNA are indicated by a single underline. The TATA box and conserved early transcription start site motif are indicated by a double underline. The termination sites of p8-9 mRNAs are shown (3′) on the nucleic acid sequence. The regulatory elements repeat I and II most proximal to p8-9 are boxed. (+) and (−), basic or acidic amino acids, respectively. The sequences homologous to the oligonucleotides used for primer extension are indicated by the dashed underlines. (c) Acid-base profile of the predicted p8-9 protein. Acidic amino acids, aspartic acid (medium bar) and glutamic acid (full bar); basic amino acids, lysine (medium bar), arginine (full bar), and histidine (small bar).
In addition, the protein is very hydrophilic and contains a high proportion of leucine and glutamine residues. A search for similar proteins in the protein and nucleic acid databases [GenBank, release 74.0 (12/92); EMBL, release 33.0 (12/92); SWISSPROT, release 23.0 (8/92)] using the UWGCG TFASTA and FASTA analysis programs did not identify any polypeptide that was obviously similar to p8-9. Therefore, the predicted early promoter upstream of IE-2 gene appears to regulate a small and previously unidentified baculovirus gene.

**Transcriptional analysis of the p8-9 ORF**

Northern blots of total RNA extracted from OpMNPV-infected LD652Y cells at various times p.i. were used to identify the number and temporal expression of transcripts homologous to the p8-9 ORF. A 120 bp EcoRI–PstI DNA fragment from the p8-9 ORF (Fig. 1a) was cloned and used to generate a ssRNA probe complementary to p8-9 mRNA. The Northern blots detected two major transcripts from 1 to 72 h p.i., a 0.5 kb mRNA that is expressed from 1 to 12 h p.i. and a 1.25 kb mRNA that is expressed from 8 to 72 h p.i. (Fig. 2). At the very late phases of viral infection high levels of longer transcripts were also detected that have homology to the p8-9 ORF. This is consistent with previous observations that baculovirus late gene expression can produce high levels of nested overlapping RNA transcripts with multiple 5’ initiation or 3’ termination sites (Friesen & Miller, 1986; Blissard & Rohrmann, 1990).

To map the 5’ transcriptional initiation site of the very early 0.5 kb p8-9 mRNA precisely, primer extension assays were performed using a 16-base oligonucleotide complementary to the p8-9 ORF (Fig. 3b). At 1 h p.i. a 154 nt primer extension product is observed that places the very early transcriptional start site at the first A of the conserved CACAGT early gene motif. The 5’ and 3’ mapping of the very early mRNA indicates that the transcript that initiates at the early gene motif corresponds in size to the 0.5 kb mRNA detected from 1 to 12 h p.i. on Northern blots (Fig. 2).

To determine the transcriptional start sites of the 1.25 kb p8-9 transcript by S1 nuclease protection assay, three 5’ end-labelled probes were used, the 848 bp EcoRI–SalI fragment, the 694 Apal–SalI and the 218 bp NotI–SalI DNA fragment. The smallest protected fragments of the EcoRI–SalI (Fig. 3c), Apal–SalI (data not shown) and NotI–SalI (Fig. 3d) probes were 697 nt, 543 nt, and 65 or 66 nt, respectively. These results mapped at least one 5’ transcriptional start site to
Identification of an OpMNPV early gene

Fig. 3. Transcriptional mapping of the p8-9 gene. (a) Schematic diagram showing the p8-9 gene region. The p8-9 ORF is indicated by a large arrow and the thin lines with arrows indicate the p8-9 mRNAs. The end-labelled probes used in 5' and 3' S1 nuclease protection assays are indicated by the lines with asterisks. Hatched lines indicate the protected fragments. The oligonucleotide primers used for primer extension analysis of the 0.5 and 1.25 kb early transcripts are represented by the two small arrows. Eight mm represents 100 bp. (b) and (e) Primer extension analysis of the 0.5 and 1.25 kb transcripts respectively. (c) S1 nuclease protection analysis of the 5' initiation site of the p8.9 1.25 kb transcript using a 5' EcoRI-SalI end-labelled probe. (d) S1 nuclease protection analysis of the 5' initiation site of the p8.9 1.25 kb transcript using a NoI-SalI 5' end-labelled probe. (f) S1 nuclease protection analysis of the 3' termination sites of p8.9 transcripts using a 3' end-labelled SalI-EcoRI probe. Numbers on the top of each autoradiograph indicate the time p.i.; M, mock-infected cells. In (b) and (e), the sequencing ladders (GATC) were generated using the same oligonucleotides as in the primer extension assays, whereas those in (c), (d) plasmid and (f) M13 mp18 sequences were used as size markers. The numbers and arrows on the left of each autoradiograph indicate the size of the primer extended or S1 nuclease protected fragments (in nt). The sequences on the right of the primer extension assays (b) and (e) represent the coding strand sequence at the major transcription start sites (*).
apparently -634 relative to the early start site. The -634 initiation site was confirmed using primer extension and the nucleotide at which transcription initiated was precisely mapped (Fig. 3e). A transcript initiating at -634 would produce an mRNA [with no poly(A) tail] of approximately 1048 to 1140 bases depending on which termination site is utilized. The decreasing intensity of the S1 nuclease protected bands at -634 initiation site was confirmed using primer extension analysis of total RNA extracted from transfected cells. As shown in Fig. 5 the results obtained with the SalI 1.25 kb transcript fragment as a probe, none of these sites is used at significant levels at late times p.i. (Fig. 3c).

Transient assays

To examine the promoter requirements for the expression of the p8-9 0.5 kb early transcript the E. coli β-gal gene was fused to the p8-9 start codon as a reporter. Three constructs with various promoter lengths (Fig. 1a) were analyzed for β-gal expression (Fig. 4). The SalI–p8-9–β-gal promoter construct contains 785 bp upstream from the p8-9 early transcription start site, which includes the initiation site of the 1.25 mRNA and all the regulatory motifs of the divergent p8-9 and IE-2 promoters (Fig. 1a). The NotI–p8-9–β-gal promoter construct deletes the 1.25 kb mRNA transcription start site and the Apal–p8-9–β-gal promoter construct deletes all GATA and ACAGGACGC sequences as well as repeat I and only retains a single copy of repeat I (Fig. 1a). All three p8-9–β-gal constructs were transfected into LD652Y cells and, as shown in Fig. 4, the highest β-gal activity was obtained with the SalI construct. Reduced levels were obtained with the NotI construct and only a minimal level of expression was observed from the Apal–p8-9–β-gal construct. These results show that the p8-9 early promoter is actively expressed in transfected cells in the absence of other viral factors. In addition, the varying levels of expression from the three constructs indicate that the β-gal expression detected in transfected cells is not just due to basal level expression from the minimal p8-9 early gene motif CACAGT, but that upstream regulatory elements are also utilized.

Interestingly, between the initiation sites of the 0.5 and 1.25 kb transcripts there are five consensus late transcriptional start sites (G/ATAAG), but, as shown by the S1 nuclease protection assays using the EcoRI–SalI fragment as a probe, none of these sites is used at significant levels at late times p.i. (Fig. 3c).
Fig. 5. Trans-activation analysis of the p8-9 early promoter by IE-1, IE-2 and p34. SalI-p8-9-β-gal-transfected Sf9 cells were assayed for expression levels by primer extension analysis of total RNA using a primer homologous to the β-gal gene. Cells were transfected with SalI-p8-9-β-gal (10 µg) by itself or co-transfected with a plasmid containing a trans-activator (1 µg); mock (lanes 1, 6); SalI-p8-9-β-gal (lanes 2, 7); SalI-p8-9-β-gal and IE-1 (lanes 3, 8); SalI-p8-9-β-gal and IE-2 (lanes 4, 9); SalI-p8-9-β-gal and p34 (lanes 5, 10). Duplicate lanes represent separate transfections. The number and arrow on the left of the autoradiograph indicate the size of the primer extended fragments. The sequence on the right of the autoradiograph was generated by using the SalI-p8-9-β-gal plasmid as a template and the same (unlabelled) primer as in the primer extension assay. The sequence of the coding strand sequence at the SalI-p8-9-β-gal transcription initiation site is shown to the right of the sequencing ladder.

Discussion

Our previous studies on OpMNPV early genes identified an early gene promoter motif that was divergent from the IE-2 promoter (Theilmann & Stewart, 1992a). In this report we show that this promoter regulates a previously unidentified baculovirus gene coding for a protein that is small and highly basic and that we have called p8-9. Transcriptional mapping showed that p8-9 is transcribed as a 0·5 kb mRNA from 1 to 12 h p.i. that initiates at a conserved early gene transcriptional start site motif CACAGT. An additional 1·25 kb transcript is detected at 8 h p.i. that initiates in a region that has no homology to any previously mapped baculovirus initiation sites or promoter elements. The sequence around the 1·25 kb transcript 5′ initiation site is GC-rich with no typical upstream TATA box.

The distinguishing feature of the predicted amino acid sequence of p8-9 is the highly basic domain in the N-terminal half of the protein (Fig. 1b,c). Other highly basic baculovirus proteins have been described, such as the OpMNPV p6-5 and the AcMNPV homologue p6-9 (Wilson et al., 1987; Wilson, 1988; Russell & Rohrmann, 1990). p6-5 is highly basic as it is rich in arginine residues and contains many serines and threonines in addition. This amino acid content is quite similar to that of protamines and it has been hypothesized that p6-5 binds to baculovirus DNA and may aid in the unpackaging of viral DNA in infected cells (Wilson, 1988; Rohrmann, 1992). The predicted p8-9 protein, on the other hand, contains both arginine and lysine residues and only a single threonine and two serine residues. In addition to being arginine/lysine-rich, the p8-9 protein is also rich in leucines and glutamines, and these four amino acids (combined) constitute 41% of the residues. The different amino acid composition of p8-9 in the basic domain indicates that it may not have a similar function as p6-5, although its highly basic character suggests that it may be a nucleic acid-binding protein. Basic domains that are rich in arginine and lysine have been found in many other nucleic acid-binding proteins such as the leucine zipper and the helix-loop-helix DNA-binding protein families (Pabo & Sauer, 1992) as well as in RNA-binding proteins like the human immunodeficiency virus Tat protein (Calnan et al., 1991).

Transient assays using p8-9-β-gal fusion constructs were performed to investigate the requirements for expression from the p8-9 early promoter. The results showed that the p8-9 early promoter was active in the absence of other viral factors in non-infected cells. In addition, we showed that highest levels of p8-9-β-gal expression were obtained with the full-length promoter and that removal of upstream promoter elements caused a reduction in expression from the reporter constructs. This indicates that the p8-9 early promoter is activated by host cell factors and that upstream regulatory sequences are required for its full activity as determined by transient assay.

The p8-9 early promoter region contains sequences homologous to the regulatory motifs repeat I and II that are present in the promoters of the IE-2 gene and the AcMNPV genes IE-N and PE38 (Krappa et al., 1991; Theilmann & Stewart, 1992a, b). The p8-9 early promoter is most similar to the AcMNPV PE38 promoter (67% identity) but does not have sequences homologous to the GATA motif. Interestingly, it has shown that repeat I will up-regulate the IE-N promoter in transfected Sf9 cells but deletion analysis of the promoter of PE38 failed to show any requirement for repeat I or any regulatory motif other than the minimal early gene motif (Carson et al., 1991; Krappa et al., 1992). The ApaI-p8-9-β-gal construct contains a complete copy of the repeat I sequence in the promoter but gives 18 times lower β-gal expression than the NorI-p8-9-β-gal construct, which suggests that the repeat I sequence is not sufficient for high level p8-9 expression. Therefore even though the promoters of p8-9, IE-N and PE38 are all very similar and contain common elements, transient assays appear to indicate that they are regulated differently. Determining the molecular basis for these subtle differences will be the object of future studies.

The OpMNPV trans-activators IE-2 and p34 were shown by transient assays to up-regulate the p8-9 early promoter, whereas IE-1 appeared to have no effect on p8-9 expression (Fig. 5). As with the p8-9 early promoter, the divergent IE-2 promoter has also been demonstrated
to be trans-activated by IE-2 and p34 gene products
(Theilmann & Stewart, 1992a; X. Wu, S. Stewart & D. A. Theilmann, unpublished). This suggests that both IE-2 and p8-9 are transcribed in divergent orientations and simultaneously trans-activated by IE-2 and p34. The early expression of p8-9 combined with its very basic amino acid domain suggests that it is required very early in infection and may be a nucleic acid-binding protein. Future studies will be directed at determining the functional role of this protein in the OpMNPV infection cycle.

We would like to thank D'Ann Rochon and Helene Sanfacon for critical reading of this manuscript. X. Wu was supported by a Canadian Government Laboratory Visiting Fellowship.

References


(Received 3 March 1993; Accepted 6 April 1993)