Analysis of Transcription Initiation in the *Panolis flammea* Nuclear Polyhedrosis Virus Polyhedrin Gene

By REBECCA OAKEY,† IAIN R. CAMERON, BENJAMIN DAVIS, ELIZABETH DAVIS AND ROBERT D. POSSEE*

NERC Institute of Virology, Mansfield Road, Oxford OX1 3SR, U.K.

(Accepted 15 November 1988)

**SUMMARY**

The nucleotide sequence of the polyhedrin gene of *Panolis flammea* multiple nucleocapsid polyhedrosis virus (PfMNPV) has been determined. The coding sequences of this gene shared 82% similarity at the DNA level and 88% similarity at the protein level with the polyhedrin gene from *Autographa californica* (Ac) MNPV. A single nucleotide deviation from the consensus transcription initiation sequence for baculovirus very late genes was identified in the PfMNPV polyhedrin gene. RNA was prepared from *Mamestra brassicae* larvae infected with PfMNPV and compared with RNA harvested at 24 h post-infection from AcMNPV-infected *Spodoptera frugiperda* cells using Northern blotting with an AcMNPV polyhedrin gene-specific probe. The PfMNPV mRNA was estimated to be 1.0 kb compared with a larger size of 1.15 kb for the AcMNPV polyhedrin mRNA. A cDNA copy of the 5' end of the PfMNPV polyhedrin mRNA was made using the technique of primer extension and sequenced to demonstrate that the point of transcription initiation was similar to that of AcMNPV polyhedrin mRNA.

The very late phase of gene expression in baculovirus-infected cells is characterized by a prolonged burst of transcriptional and translational activity which results in the production of large amounts of two proteins, polyhedrin and p10. It has been estimated that up to 90% of the polyadenylated virus RNA present at 48 h post-infection (p.i.) consists of transcripts specific for these two genes (Smith *et al.*, 1983). Polyhedrin is responsible for occluding enveloped virus particles into crystalline structures to form polyhedra. These act as the agents of dispersal for the virus in the natural environment. The p10 is a non-structural protein of uncertain function but it is thought to play a role in polyhedron morphogenesis in the infected cell (van der Wilk *et al.*, 1987; Vlak *et al.*, 1988).

The mechanism by which the infected cell is able to produce such large amounts of these proteins remains to be fully elucidated. For the *Autographa californica* nuclear polyhedrosis virus (AcMNPV) polyhedrin gene it has been ascertained that a sequence extending 69 nucleotides upstream from the ATG is essential for efficient promoter activity (Matsuura *et al.*, 1987; Possee & Howard, 1987). This sequence consists of a 20 nucleotide region upstream from the mRNA transcription initiation (start) site (Possee & Howard, 1987) and a 49 nucleotide 5' non-coding (leader) sequence before the polyhedrin ATG codon (Matsuura *et al.*, 1987). Deletions in either of these two sequences reduced the polyhedrin promoter activity. Analysis of the AcMNPV p10 5' leader sequence has also shown that it is required for high level expression of this gene (Weyer & Possee, 1988).

Comparison of a number of polyhedrin genes from different baculoviruses has highlighted the presence of a 12 nucleotide consensus sequence spanning the mRNA start site [(A/T)-

†Present address: Department of Biochemistry, South Parks Road, University of Oxford, Oxford OX1 3QU, U.K.

0000-8577
ATAAGN(T/A/C)T(T/A)T; Rohrmann, 1986]. This consensus sequence is also present in both of the p10 genes which have been analysed [AcMNPV, Kuzio et al., 1984; Lübbert & Doerfler, 1984: Orgyia pseudotsugata (Op) MNPV, Leisy et al., 1986c]. The subgroup B baculoviruses (granulosis viruses; GVs) also have the same consensus sequence upstream of the granulin gene, which encodes an occlusion body protein of similar function to polyhedrin (Akiyoshi et al., 1985; Chakerian et al., 1985). The polyhedrin mRNA start sites have been accurately determined for AcMNPV (Howard et al., 1986) and OpMNPV (Leisy et al., 1986a). The same two viruses have been used to map the mRNA start site of the p10 gene (AcMNPV, Kuzio et al., 1984; Lübbert & Doerfler, 1984; OpMNPV, Leisy et al., 1986c). The data indicate that transcription of the very late genes always initiates within the 12 nucleotide consensus. In this study we report the cloning and sequencing of the polyhedrin gene from Panolis flammea (Pf) MNPV and the determination of its mRNA start site.

The polyhedrin gene of PfMNPV has been mapped to an 8.08 kb DNA fragment (designated F) produced after digestion with HindIII (Possee & Kelly, 1988). This fragment was isolated from a low gelling temperature agarose gel as described previously (Possee & Kelly, 1988) and inserted into the HindIII site of pAT153, using techniques described by Maniatis et al. (1982), to produce the clone pPFHF. Further subcloning of SalI digest products from pPFHF produced a plasmid with a 1.2 kb insert more convenient for DNA sequencing. The virus DNA in this plasmid was sequenced by performing chemical degradation analysis (Maxam & Gilbert, 1980) on fragments radiolabelled at one end with [32P]dNTP using the Klenow fragment of Escherichia coli DNA polymerase I for 'fill-in' reactions, or [γ-32P]ATP and T4 DNA polynucleotide kinase for the SalI digest. The sequencing strategy is outlined in Fig. 1. Preliminary analysis of the data revealed that a SalI site was only 79 nucleotides upstream from the polyhedrin ATG codon (Fig. 2). Therefore the sequence was further extended by returning to the pPFHF clone and isolating the regions of interest for sequence determination.

In Fig. 2 the nucleotide sequence data for PfMNPV are presented; the polyhedrin gene consists of a 738 nucleotide open reading frame encoding a 246 amino acid polypeptide of predicted Mr 28,935. When the sequence data for the PfMNPV polyhedrin gene-coding region were compared with the sequence of the homologous gene from AcMNPV (Hooft van IJdekinge et al., 1983) they were found to share 82% similarity at the DNA level and 88%
**Short communication**

Fig. 2. Nucleotide sequence of the PfMNPV polyhedrin gene and flanking regions. The coding sequence for the polyhedrin begins at position 581 and terminates at 1318. The predicted amino acid sequence is shown above that of the DNA. The 12 nucleotide consensus sequence spanning the mRNA transcription initiation site (T) is underlined from 533 to 544, with a gap representing the position of the A normally present (Rohrmann, 1986). The sequence complementary to the primer used in the cDNA extension of the mRNA is underlined from positions 605 to 621 and labelled C. The positions of the SalI sites bounding the initial fragment used for DNA sequencing are indicated.

Similarity at the amino acid level. A computer-assisted search of the 3' non-coding region of the PfMNPV polyhedrin gene did not reveal the presence of any potential polyadenylation signals (Birnstiel et al., 1985). When the 344 nucleotide 3' non-coding sequence of the AcMNPV polyhedrin gene (Howard et al., 1986) was compared with a similar length from the PfMNPV polyhedrin gene only 28% homology was calculated. Conversely, the 5' untranslated ends of each virus polyhedrin mRNA were 70% homologous. It was noted that the 12 nucleotide sequence spanning the putative mRNA start site of the PfMNPV polyhedrin gene was not a perfect match with the consensus proposed by Rohrmann (1986). The second nucleotide in this example was a G, whereas in all other very late genes an A nucleotide has been recorded. The other 11 nucleotides agree with the remainder of the consensus.

In order to determine what effect the substitution of a G nucleotide at the second position of the consensus sequence might have on the site of transcription initiation from the PfMNPV polyhedrin gene it was necessary to extract mRNA transcripts specific for this gene from infected cells. In common with many other NPVs this example does not replicate in cell culture,
even when using *P. flammea* cell lines (C. J. Allen & R. D. Possee, unpublished data). An additional problem is that *P. flammea* insects cannot easily be maintained in laboratory culture. Therefore RNA was isolated from *Mamestra brassicae* larvae, which are permissive for PfMNPV, 5 days after infection with PfMNPV using the CsCl-guanidinium isothiocyanate method (Maniatis *et al*., 1982). Northern blot analysis of this RNA and AcMNPV mRNA extracted from *Spodoptera frugiperda* cells at 24 h p.i. was carried out as described previously

---

**Fig. 3.** Northern blot analysis of polyhedrin mRNA. Total RNA extracted from *M. brassicae* larvae 5 days after infection with PfMNPV (lane 1) or from AcMNPV-infected *S. frugiperda* cells (24 h p.i.) (lane 2) was fractionated in a 1% agarose-glyoxal gel and transferred to Genescreen membrane. The filter was hybridized with a nick-translated probe specific for the AcMNPV polyhedrin gene (pAcHV; Possee, 1986). The sizes of the respective mRNA species identified are indicated on each side of the autoradiograph.

**Fig. 4.** Sequence analysis of cDNA prepared using primer extension of PfMNPV polyhedrin mRNA. The oligonucleotide polyhedrin primer was radio labelled at its 5' end with [γ-32P]ATP and polynucleotide kinase before extension on the mRNA template with reverse transcriptase. The cDNA was fractionated in a polyacrylamide-urea gel, purified and sequenced using the chemical method. The sequence complementary to the polyhedrin AUG is indicated by the vertical line and arrowhead at the side of the gel.
Short communication

(Possee, 1986), using a nick-translated (Rigby et al., 1977), cloned copy of the 3' end of the AcMNPV polyhedrin gene (pAcHV; Possee, 1986) to detect homologous sequences (Fig. 3). The size of the AcMNPV polyhedrin mRNA was estimated to be 1-15 kb which agrees closely with the previously published value of 1-2 kb (Smith et al., 1983). However, the PfMNPV polyhedrin mRNA, with a size of 1-0 kb, was significantly smaller. To determine the transcription initiation site for the PfMNPV polyhedrin mRNA an approach similar to that employed for the accurate 5' mapping of the AcMNPV polyhedrin mRNA was used (Howard et al., 1986). An oligonucleotide primer (5'-GTACGTCCCAACGACGG-3') was synthesized which corresponded to the complement of the coding sequence from +25 to +41 of the PfMNPV polyhedrin gene (Fig. 2; positions 605 to 621). This was radiolabelled at the 5' end with [γ-32P]ATP and polynucleotide kinase (Howard et al., 1986) and used to prime copying of the mRNA with reverse transcriptase. Sufficient cDNA was produced to permit sequencing of the transcript and thus determine the 5' end of the PfMNPV polyhedrin RNA. The sequencing gel is shown in Fig. 4 and is clearly readable up to a T nucleotide at the 3' end of the cDNA, followed by two unreadable bands in all four lanes. If the T nucleotide is taken as the point of transcription initiation then the mRNA start site is located at a position 44 nucleotides upstream of the PfMNPV polyhedrin ATG (Fig. 2). This is in the central region of the consensus transcription initiation motif highlighted by Rohrmann (1986) and is only one nucleotide upstream from the site previously determined for AcMNPV polyhedrin mRNA (Howard et al., 1986).

A number of baculovirus polyhedrin and granulin genes have been cloned and sequenced. These include AcMNPV (Hooft van Iddekinge et al., 1983). Bombyx mori MNPV (Iatrou et al., 1985), OpMNPV (Leisy et al., 1986a), Op single nucleocapsid NPV (Leisy et al., 1986b), Trichoplusia ni GV (Akiyoshi et al., 1985) and Pieris brassicae GV (Chakerian et al., 1985). The polyhedrin genes have a high degree of similarity as do the two granulin genes. It is also notable that even though NPVs and GVs do not share significant sequence similarity as judged by hybridization analyses (Smith & Summers, 1982), they do show similarity at the nucleotide and amino acid level when sequence data are compared (Rohrmann, 1986). In this same study a 12 nucleotide consensus was highlighted at the 5' end of all the polyhedrin and granulin genes which had been sequenced; it also features in the 5' region of the two pl0 genes sequenced (AcMNPV, Kuzio et al., 1984; Lübbert & Doerfler, 1984: OpMNPV, Leisy et al., 1986c). A sequence very similar to the late consensus has also been found spanning the mRNA start site used by the AcMNPV 39K gene at 18 h p.i. (Guarino & Summers, 1986). This is of particular significance because transcription of the 39K gene early in infection is mediated by a different site closer to the ATG codon. Therefore the very late consensus appears to be intimately associated with the temporal regulation of both late and very late genes. Furthermore, a similar sequence has been identified upstream of the arginine-rich (basic protein) gene of AcMNPV (Wilson et al., 1987). Although the sequences found in the 39K and basic protein genes do not agree precisely with the consensus proposed by Rohrmann (1986) it is noticeable that the central ATAAG motif is preserved in both of these examples. When the 5' ends of late and very late mRNAs have been accurately mapped using primer extension analysis or sizing of S1-protected probes against DNA sequencing ladders the conclusions have invariably been that the transcripts initiate within this motif (AcMNPV polyhedrin, Howard et al., 1986; OpMNPV polyhedrin, Leisy et al., 1986a; AcMNPV basic protein gene, Wilson et al., 1987; OpMNPV p10, Leisy et al., 1986c; PfMNPV polyhedrin, Fig. 2). Wilson et al. (1987) proposed that the differences in the sequences flanking the ATAAG motif might account for the differential temporal regulation of the basic protein and polyhedrin genes. However, in the PfMNPV polyhedrin gene described here the first A of this pentanucleotide region is replaced with a G. This was confirmed by sequencing both strands of DNA. The G nucleotide also occurs in a similar position in the M. brassicae MNPV polyhedrin gene promoter (Cameron & Possee, 1989). This baculovirus is very similar to PfMNPV, sharing 70% homology at the DNA level when compared using dot blot hybridization analysis in 50% formamide (Possee & Kelly, 1988). The difference in the consensus did not significantly alter the position of transcription initiation from the polyhedrin gene of PfMNPV; the mRNA was found to initiate only one nucleotide
further upstream from the site utilized in AcMNPV polyhedrin (Howard et al., 1986). This suggests that the central TAAG sequence of the consensus may be the most important feature for transcription initiation of late and very late baculovirus genes. It will be interesting to determine whether the other late and very late genes of PfMNPV have the same consensus transcription initiation sequence and to investigate the effect of altering single nucleotides within the central pentanucleotide.

The photographic skills of C. D. Hatton are gratefully acknowledged.

REFERENCES


(Received 1 July 1988)