Characterization of the Genetic Organization of the HindIII M Region of the Multicapsid Nuclear Polyhedrosis Virus of Orgyia pseudotsugata Reveals Major Differences among Baculoviruses

By ADRIAN F. GOMBART, GARY W. BLISSARD AND GEORGE F. ROHRMANN

Department of Agricultural Chemistry, Oregon State University, Corvallis, Oregon 97331-6502, U.S.A.

(Accepted 6 March 1989)

SUMMARY

A region including the 4 kb HindIII M fragment of the multicapsid nuclear polyhedrosis virus (MNPV) of Orgyia pseudotsugata (OpMNPV) genome was sequenced, transcriptionally mapped, and compared to the homologous region in the MNPV of Autographa californica (AcMNPV). Five open reading frames (ORFs) oriented in the same direction were identified and were found to be expressed at late times post-infection. The mRNAs from the five ORFs were found to coterminate at a single site downstream of ORF 5. The conserved late gene promoter/mRNA start site sequence (ATAAG) was present upstream of all the ORFs, but did not appear to be the major site of mRNA initiation for two of the ORFs as determined by primer extension analysis. These data indicated that use of this signal for transcriptional initiation may vary between different ORFs. The predicted amino acid sequences for the five ORFs of AcMNPV and OpMNPV were compared and amino acid homologies of 26 to 72% were observed. The comparison revealed a number of major differences in the genomes of the two viruses. A putative transposable element of 634 nucleotides was found to be inserted into the previously reported AcMNPV ORF 1 sequence. In addition, it was found that a region corresponding to the 4 kb HindIII K/EcoRI S/hr5 region of AcMNPV was not present in the OpMNPV genome.

INTRODUCTION

Baculoviruses comprise a diverse group of pathogens infectious for arthropods, particularly lepidopteran insects. Over 500 species of insects have been reported to be infected with baculoviruses, which include members of the occluded subgroups A (nuclear polyhedrosis viruses, NPVs), B (granulosis viruses) and the non-occluded subgroup C (Martignoni & Iwai, 1986). Although all baculoviruses have a double-stranded, supercoiled DNA genome packaged in a rod-shaped enveloped nucleocapsid, the sizes of the genomes vary considerably (88 to 153 kb) (Burgess, 1977; Schafer et al., 1979), and very limited DNA sequence homology is evident between baculovirus isolates (Smith & Summers, 1982; Rohrmann et al., 1982). In addition, there is considerable diversity in the viral infectivity spectrum for host insects with many baculoviruses demonstrating a very narrow host range (see Crook, 1981). Despite the evident diversity of baculoviruses, little is known about the genetic relationships existing between different members. A series of investigations have compared the relatedness of the genomes of the multicapsid NPV of Orgyia pseudotsugata (OpMNPV) and the MNPV of Autographa californica (AcMNPV). It was initially observed that the two viruses possess over 24% DNA sequence homology when DNA–DNA hybridization was done under non-stringent conditions (Rohrmann et al., 1982). Subsequent hybridization of cosmid clones representing different regions of the OpMNPV genome to blots of AcMNPV DNA demonstrated that the genomes of the two viruses are organized in a collinear manner (Leisy et al., 1984). Recently, Bicknell et al. (1987) reported that the OpMNPV genome lacked the homologous repeat sequences which have been implicated as enhancers of delayed early genes in AcMNPV (Guarino & Summers, 1986).
In this report, we focus on the HindIII M region of the OpMNPV genome which comprises 4 kb of DNA located upstream of the p26-p10-p74 gene complex. We recently demonstrated that an open reading frame (ORF) in this region encodes a polyhedral envelope (PE) protein (Gombart et al., 1989) which is homologous to the AcMNPV PE protein recently described by Whitt & Manning (1988) and may be bound by thiol linkages to the PE. The gene encoding the OpMNPV PE protein demonstrated 58% amino acid identity to the AcMNPV ORF 3 gene reported at 83 map units by Oellig et al. (1987, 1989). The area surrounding the AcMNPV ORF 3 is well characterized and contains a series of five ORFs in the same orientation which appear to be transcribed as an overlapping set of mRNAs which coterminate near the end of the fifth ORF. In the investigations outlined in this report, we describe the sequence and transcriptional organization of the HindIII M region of OpMNPV.

METHODS

Insect cell lines and virus. The Lymantria dispar cell line and the production of OpMNPV have been described previously (Quant-Russell et al., 1987). For the work described in this paper, cell monolayers were infected with second passage OpMNPV at an m.o.i. of $\geq 5$.

Cloning and DNA sequencing. The HindIII M region was sequenced using appropriate restriction enzyme fragments and exonuclease III deletion mutants produced using the method of Henikoff (1984). Both M13 and plasmid sequencing reactions were performed by the chain termination method of Sanger et al. (1980) using [α-35S]dATP (New England Nuclear) and Sequenase (United States Biochemical). The plasmids pUC18, pUC19 and the M13 strains mp18 and mp19 were used for cloning and sequencing. A vector pBS(−) (Stratagene Cloning Systems), modified to contain a BglII site in the polylinker, was also used. This modified vector is 3216 nucleotides (nt) in length.

RNA isolation. Isolation of total and polyadenylated RNA from OpMNPV-infected L. dispar cells for Northern hybridization, primer extension and S1 nuclease analysis was described previously (Blissard et al., 1989).

Northern hybridization analysis. Northern blotting of viral mRNAs, production of cRNA transcripts, prehybridization and hybridization of the blots were carried out as described by Blissard et al. (1989).

Primer extension and S1 nuclease analysis. In order to identify transcription start sites for ORFs 2 to 5, 17 residue oligonucleotides complementary to the mRNA for each ORF were synthesized and used in primer extension analysis. The location of each primer is shown in Fig. 2. Primer extension analyses were performed as described by Blissard & Rohrmann (1989).

In order to identify the 3' termini of the HindIII M transcripts, S1 nuclease protection experiments were performed using probes specific for transcripts running through ORFs 2, 3 and 5 (Fig. 4d) using methods described by Flavoro et al. (1980).

RESULTS AND DISCUSSION

Nucleotide sequence of the HindIII M region

Restriction fragments and exonuclease III deletion mutants from the HindIII M region (Fig. 1) were cloned into M13 and pBS (−) and sequenced (Fig. 2). From the sequence analysis, we identified five ORFs (ORFs 1 to 5) (Fig. 1, 2) in the same orientation and with limited intergenic regions ranging from 4 to 43 nt. The ORFs varied in size from 309 to 1272 nt encoding proteins of 11.5K to 47.3K (see Table 1). The ORF 1 coding region begins in the HindIII E region and terminates in the HindIII M region at nt 626 (Fig. 1, 2). ORF 3 encodes a PE protein and is described in detail elsewhere (Gombart et al., 1989). The 5' flanking region of each ORF contains the TAAG consensus sequence at various distances from the translation start (ATG) (Fig. 2, 3). This sequence has been implicated in specifying the initiation and affecting the levels of transcription of baculovirus late genes or both (Rohrmann, 1986; Possee & Howard, 1987). The nucleotides at positions −3 and +4 surrounding the translation start codon of ORFs 1 to 3 contain a purine base (Fig. 3) indicating that they conform to Kozak's rules (PuNNATGPu) for efficient eukaryotic translation initiation (Kozak, 1986). ORFs 4 and 5 have a purine at the −3 position and a pyrimidine at the +4 position. A single 3' processing/polyadenylation signal (AATAAA) was found in the HindIII M region at nt 3842, approximately 20 nt upstream of the termination codon of ORF 5 (Fig. 2).
OpMNPV HindIII M gene organization

Fig. 1. Location of the OpMNPV HindIII M region. (a) HindIII restriction enzyme map of the OpMNPV genome indicating the location of polyhedrin (PH) and p39 genes. (b) Restriction map of the HindIII M and Q region indicating the location of ORFs 1 to 5 and the genes for p26, p10 and p74. Abbreviations: A, Asp718; Bg, BglII; C, ClaI; H, HindIII; K, KpnI; P, PstI; S, SalI; Ss, SstI; X, XhoI.

Table 1. Comparison of baculovirus proteins*

<table>
<thead>
<tr>
<th>ORF</th>
<th>No. of amino acids</th>
<th>Mr × 10⁻³</th>
<th>DNA</th>
<th>Amino acid</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 1</td>
<td>OpMNPV</td>
<td>193</td>
<td>21.2</td>
<td>ND‡</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>199</td>
<td>22.1</td>
<td>64</td>
</tr>
<tr>
<td>ORF 2</td>
<td>OpMNPV</td>
<td>103</td>
<td>11.5</td>
<td>69</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>106</td>
<td>12.1</td>
<td>72</td>
</tr>
<tr>
<td>ORF 3</td>
<td>OpMNPV</td>
<td>297</td>
<td>32.4</td>
<td>62</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>322</td>
<td>36.5</td>
<td>58</td>
</tr>
<tr>
<td>ORF 4</td>
<td>OpMNPV</td>
<td>228</td>
<td>25.9</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>219</td>
<td>25.1</td>
<td>26</td>
</tr>
<tr>
<td>ORF 5</td>
<td>OpMNPV</td>
<td>424</td>
<td>47.3</td>
<td>58</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>419</td>
<td>48.3</td>
<td>53</td>
</tr>
<tr>
<td>p10</td>
<td>OpMNPV</td>
<td>92</td>
<td>10</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>94</td>
<td>10.1</td>
<td>41</td>
</tr>
<tr>
<td>p26</td>
<td>OpMNPV</td>
<td>229</td>
<td>26</td>
<td>55</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>240</td>
<td>26</td>
<td>47</td>
</tr>
<tr>
<td>p74§</td>
<td>OpMNPV</td>
<td>158</td>
<td>–</td>
<td>71</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>138</td>
<td>–</td>
<td>70</td>
</tr>
<tr>
<td>PH</td>
<td>OpMNPV</td>
<td>245</td>
<td>29.9</td>
<td>80</td>
</tr>
<tr>
<td></td>
<td>AcMNPV</td>
<td>245</td>
<td>28.7</td>
<td>90</td>
</tr>
</tbody>
</table>

* For references see legend of Fig. 7.
† The sources for the complete ORF 1 sequence are described in the legend of Fig. 6.
‡ This comparison was not done because an uninterrupted ORF 1 DNA sequence from a single strain of AcMNPV is not available.
§ The nucleotide sequences for the p74 ORF are not completed at present and the partial predicted amino acid sequences are compared.

Mapping of the 3' termini of the HindIII M transcripts

To determine the site of termination of mRNAs from the HindIII M region, three different restriction fragments were used for mapping the mRNA termini of transcripts running through ORFs 2, 3 and 5 (Fig. 4d). For ORF 2, the 3' end-labelled 3560 nt Asp718 probe yielded an S1 nuclease-protected fragment of approximately 3200 nt (Fig. 4a). For ORF 3, the 3' end-labelled 5857 nt BglII fragment (vector plus insert) yielded an S1 nuclease-protected fragment of approximately 2360 nt (Fig. 4b). These results suggested that the mRNAs from ORFs 2 and 3
were terminating together downstream of ORF 5 (Fig. 4d). Therefore, it appeared that OpMNPV produced a series of overlapping, coterminal mRNAs similar to those observed by Oellig et al. (1987) in AcMNPV. To determine the 3' termination site for the overlapping mRNAs from the region precisely, a clone containing the 477 nt SalI–HindIII fragment which included 143 nt of the 3' region of ORF 5 (Fig. 2) was 3' end-labelled, and the resulting 3670 nt fragment (vector plus insert) (Fig. 4d) was used in S1 protection experiments. Four S1 nuclease-protected fragments of 178 to 181 nt were detected at 24 to 48 h post-infection (p.i.) (Fig. 4c). From these three sets of S1 nuclease protection experiments, the 3' termini for the HindIII M ORF 1 to 5 transcripts were mapped to four nt at 3900 to 3903 (Fig. 2). The only 3'
OpMNPV HindIII M gene organization

Fig. 2. Nucleotide sequence of the OpMNPV HindIII M region. The putative promoter element (ATAAG) upstream of ORFs 1 to 5, the polyadenylation signal (AATAAA) (nt 3842) and the mRNA termination site (nt 3900 to 3903) are underlined. The mRNA start sites are indicated with the relevant nt underlined and the translational start (ATG) sites for ORFs 2 to 5 and the p26 gene downstream of ORF 5 are indicated. The predicted amino acid sequence of each ORF is given below the nucleotide sequence. Major restriction enzyme sites and the location of the primers used in the primer extension analysis are underlined.

Processing/polyadenylation consensus sequence (AATAAA) in the HindIII M region is located within ORF 5 about 50 nt upstream of the 3' mRNA termination site. A related sequence (AATAAT) is located 16 nt downstream of ORF 5 is indicated. The predicted amino acid sequence of each ORF is given below the nucleotide sequence. Major restriction enzyme sites and the location of the primers used in the primer extension analysis are underlined.

Primer extension analysis to locate transcription initiation sites

To locate the transcriptional start sites of the mRNAs, 17 residue DNA oligonucleotides complementary to sequences near the 5' ends of ORFs 2 to 5 were synthesized (for location and sequence of the oligomers, see Fig. 2) and used for primer extension analysis of total RNA from infected and uninfected L. dispar cells. The exact location of each transcriptional start site was

\textit{Fig. 2.} Nucleotide sequence of the OpMNPV HindIII M region. The putative promoter element (\(\overline{\text{ATAAG}}\)) upstream of ORFs 1 to 5, the polyadenylation signal (AATAAA) (nt 3842) and the mRNA termination site (nt 3900 to 3903) are underlined. The mRNA start sites are indicated with the relevant nt underlined and the translational start (ATG) sites for ORFs 2 to 5 and the p26 gene downstream of ORF 5 are indicated. The predicted amino acid sequence of each ORF is given below the nucleotide sequence. Major restriction enzyme sites and the location of the primers used in the primer extension analysis are underlined.

Processing/polyadenylation consensus sequence (AATAAA) in the HindIII M region is located within ORF 5 about 50 nt upstream of the 3' mRNA termination site. A related sequence (AATAAT) is located 16 nt downstream of ORF 5 is indicated. The predicted amino acid sequence of each ORF is given below the nucleotide sequence. Major restriction enzyme sites and the location of the primers used in the primer extension analysis are underlined.

\textit{Primer extension analysis to locate transcription initiation sites}

To locate the transcriptional start sites of the mRNAs, 17 residue DNA oligonucleotides complementary to sequences near the 5' ends of ORFs 2 to 5 were synthesized (for location and sequence of the oligomers, see Fig. 2) and used for primer extension analysis of total RNA from infected and uninfected L. dispers cells. The exact location of each transcriptional start site was
A. F. GOMBART, G. W. BLISSARD AND G. F. ROHRMANN

Fig. 3. Nucleotide sequence of the 5' flanking region containing the putative regulatory sequence (~TAAG) of each ORF. A 14 nt sequence containing the invariant ~TAAG sequence found upstream of baculovirus late genes is set off by spaces in each sequence. The upstream location of the first nucleotide of each ~TAAG sequence is indicated above each sequence and the mRNA start sites within this region are underlined. The ATG translational start for each ORF is underlined with the -3 and +4 nucleotide positions indicated. The consensus sequence is derived from baculovirus hyperexpressed genes (Rohrmann, 1986). The ~TAAGs of ORFs 3 and 5 only weakly specified initiation as compared to other ~TAAGs within the HindIII M region. The precise mRNA start site for ORF 1 was not determined.

Table 2. Transcript size determined by 5' and 3' mapping

<table>
<thead>
<tr>
<th>Location of primer extension product (nt)</th>
<th>Location of primer extension product</th>
<th>Downstream ORF* product size (nt)†</th>
<th>Predicted transcript size (nt)‡</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORF 2 ~675 60-61</td>
<td>~10 624-625</td>
<td>ORF 1 3900</td>
<td>ORF 2 3280</td>
</tr>
<tr>
<td>ORF 3 404-405 276, 257</td>
<td>624-625, 753, 772 154-155, 71-81 874-875, 948-958</td>
<td>ORF 2 3280, 3140, 3120 ORF 3 3030, 2950</td>
<td>ORF 4 240, 2300, 2110, 2050</td>
</tr>
<tr>
<td>ORF 4 322, 301, 122-123 60-61</td>
<td>1593, 1614, 1792-1793, 1854-1855</td>
<td>ORF 5 1530-1670</td>
<td>ORF 5 1320</td>
</tr>
<tr>
<td>ORF 5 ~386-550 35-36</td>
<td>~2066-2233 2583-2584</td>
<td>ORF 1 3900</td>
<td>ORF 2 3280</td>
</tr>
</tbody>
</table>

* Because of the overlapping nature of the HindIII M transcripts, some primer extension products which were primed within ORFs 2 and 3 extended through the preceding ORF and therefore mapped to the 5' flanking sequence of the upstream ORF.
† Because of the heterogeneity at the 5' and 3' ends of the mRNAs the sizes are rounded off to the nearest 10 nt.
‡ These sizes and locations are estimated from long gel runs and are not determined to the nucleotide.
§ The higher Mr ORF 5 transcripts are derived from estimates from longer electrophoresis runs of the primer extension samples from Fig. 5a.

established by comparing the size of the primer extension products with M13 mp18 sequencing ladders. The results of the primer extension analyses are shown in Fig. 5(a to e) and summarized in Table 2. For accurate sizing of larger primer extension products, gels were run for longer periods of time (data not shown). Using an oligonucleotide complementary to the 5' region of ORF 2 for primer extension, we identified the transcription start site for both ORF 2 and ORF 1. The primer extension product for ORF 1 was not precisely measured but was estimated to be approx. 675 nt (Fig. 5a) which corresponds to an mRNA start site at approx. nt 10 (Fig. 2) and places it near the GTAAG sequence. This indicates that at least a portion of the message for ORF 1 is transcribed through ORF 2. In addition, a major set of primer extension products of 60 to 61 nt (Fig. 5a) corresponding to mRNA start sites at nt 624 to 625 are found within the ~TAAG located immediately upstream of ORF 2 (Table 2, Fig. 2, 3). Primer extension from within ORF 3 resulted in products corresponding to the major ORF 2 and ORF 3 start sites indicating that at least a portion of the ORF 2 mRNA is transcribed through ORF 3. One set of the ORF 3 mRNA start sites corresponds to five primer extension products (71 to 81 nt) initiating at nt 948 to 958 (Fig. 2, Table 2) located near the ORF 3 ~TAAG (Fig. 3). Major
Fig. 4. Determination of the location of the 3' ends of transcripts from ORFs 2, 3 and 5 by S1 nuclease protection assays. (a) S1 protection analysis of the 3' end of mRNAs from within ORF 2. (b) S1 protection analysis of the 3' end of mRNAs from within ORF 3. (c) S1 protection analysis of the 3' end of mRNAs from within ORF 5. Numbers at top of gel indicate h.p.i.; M indicates mock-infected cells; P indicates untreated radioactively labelled probe. The external lanes on (a) and (b) are radioactively labelled markers. The sizes of relevant markers are indicated between (a) and (b). The numbers to the left of (a) and to the right of (b) indicate the size of the probe (vector plus insert) used (upper number) and the size of the protected fragment (lower number). In (c) an M13 mp18 sequencing ladder was used to determine the size of the protected fragments. The numbers to the right indicate their sizes. The autoradiogram shown in panels (a) and (b) was derived from a 1% agarose gel blotted to GeneScreen Plus. Panel (c) represents an autoradiogram of an 8% polyacrylamide, 7 M-urea gel. (d) Strategy for mapping the 3' ends of transcripts from the HindIII M region. Solid lines under the ORFs indicate the three probes used in the 3' S1 analysis. The wavy line beneath each solid probe line indicates the protected fragments. The numbers at the left indicate the size of the probe (this number and the diagram indicate the viral component of the probe only) and protected fragments and asterisks indicate the position of the labelled end. Abbreviations: same as in Fig. 1 and Hp, HpaI; Pv, PvuII; Hc, HinclI.

All probes used for S1 nuclease protection analysis were 3' end-labelled with [α-32P]dATP or [α-32P]dCTP by using the Klenow fragment of Escherichia coli DNA polymerase (Maniatis et al., 1982). For the ORF 2 probe, a pUC19 recombinant containing the entire HindIII M fragment was cleaved at the Asp718 site located 269 nt upstream of the ORF 2 termination codon at nt 950 (Fig. 2, Fig. 4d) and at a downstream Asp718 site located in the pUC19 polylinker. The 3560 nt (3522 nt insert, 38 nt vector) Asp718 fragment was gel-purified and 3' end-labelled. For the ORF 3 probe, a 2682 nt BglII-HindIII fragment (nt 1518 to 4200; Fig. 2, Fig. 4d) was subcloned into the BglII/HindIII sites of a pBS(−) vector modified to contain a BglII site in its polylinker. The recombinant was digested with BglII, 3' end-labelled and gel-purified. This resulted in a 5857 nt fragment (3175 nt vector plus 2682 nt insert) labelled at the BglII site which is located 369 nt upstream from the termination codon of ORF 3 (Fig. 2, Fig. 4d). For the ORF 5 probe, a 477 nt SalI-HindIII fragment containing 143 nt of the 3' region of ORF 5 (Fig. 2, Fig. 4d) was subcloned into pBS(−) cut with SalI and HindIII. The subclone was digested with SalI and 3' end-labelled resulting in a 3670 nt fragment (3193 nt vector plus 477 nt insert) (Fig. 4d).
Fig. 5. Primer extension and Northern analysis of HindIII M mRNA. (a) Primer extension analysis to determine mRNA start sites upstream of ORFs 2 to 5. For size markers, an M13 mp18 sequencing ladder (order ACGT) is present to the right of each primer extension experiment. The numbers to the left indicate the size of the major primer extension products. The numbers at the top indicate the time (h p.i.) and M indicates total RNA from a mock-infected sample. The samples were analysed on an 8% polyacrylamide, 7 M-urea gel. (b) Northern blot analysis of transcripts from the ORF 5 region. The numbers to the left indicate the location of labelled standards, and the numbers on the right indicate the sizes of major transcripts. This analysis was done using poly(A)* RNA collected at 24 h p.i. electrophoresed on 1.25% agarose, 6% formaldehyde gels. The Northern hybridization probe was produced by cloning a HindIII restriction fragment from within ORF 5 (shown cross-hatched) into the
primer extension products of ORF 4 were localized to four regions on the sequence and the primer extension products of 60 to 61 nt correspond to mRNA starts sites at nt 1854 to 1855 (Fig. 2) on the HindIII M sequence and are located at the T and A within the ORF 4 ATAAG (Fig. 3). ORF 5 had major primer extension products of 35 to 36 nt corresponding to initiation at nt 2583 to 2584 (Fig. 2). Only a very weak start site at the G in the ATAAG for ORF 5 was observed. No primer extension products were detected for any of the primers when annealed to mRNA from uninfected cells.

Although the mRNAs from the five ORFs in the HindIII M region coterminate downstream of ORF 5, the regulation of the initiation of transcription appears to be complex. The sequence ATAAG is located upstream of each of the five ORFs in the HindIII M region. In two of the genes (ORFs 3 and 5), however, the major mRNAs do not appear to initiate within this sequence (Fig. 2, 5). Initiation of mRNA transcription of ORF 3 occurs at a number of positions near its ATAAG, which is in contrast to initiation at one or two contiguous nt within the ATAAG motif as is often observed (e.g. ORFs 2 and 4, Fig. 3). This suggests that there may be functional constraints preventing high levels of transcription initiation from the ORF 3 ATAAG. Similarly the ATAAG upstream of ORF 5 appears to be only a minor initiation site (Fig. 5). The highest relative concentration of primer extension products of ORFs 3, 4 and 5 was in higher M_r products. This was especially evident in ORF 3, where only a small fraction of the products initiated near ORF 3, and in ORF 5, which also showed a high concentration of higher M_r products ranging in size from 386 to over 550 nt (Fig. 5).

Northern blot analysis of HindIII M transcription

To investigate the transcriptional expression of the HindIII M region further, a HincII fragment from within ORF 5 was subcloned into plasmid pBS(−) (see Fig. 5c). A labelled cRNA transcript (complementary to the mRNA encoding ORF 5) was synthesized and hybridized to Northern blots of poly(A)+ RNA isolated from OpMNPV-infected cells at 24 h p.i. (Fig. 5b). The Northern blot was exposed for various times to allow identification of bands of differing intensity. The following summarizes the size in nucleotides of the transcripts identified by the ORF 5 cRNA probe and the probable corresponding ORF immediately downstream of the 5’ end of the transcript as predicted from the 5’ and 3’ mapping data (Table 2, Fig. 4, 5): 3900 (ORF 1), 3200 (ORF 2), 2900 (ORF 3), 2400 and 2000 (ORF 4) and 1300 to 1600 (ORF 5). The transcripts in the 1500 to 1600 range are likely to represent the broad range of higher M_r ORF 5 primer extension products (Fig. 5b, Table 2). The data from these experiments confirmed the presence of transcripts of sizes similar to those determined from 5’ and 3’ mapping (Fig. 5, Table 2).

Time course of mRNA expression

Primer extension analysis was done on ORF 3 (encoding the PE protein; Gombart et al., 1989) with total RNA isolated from 0 to 72 h p.i. (Fig. 5a). Specific mRNA was initially observed at 24 h p.i. and was present in high concentrations until 48 h p.i. but was at low levels at 60 and 72 h p.i. This corresponds to the appearance of the PE protein as observed by immunoblot and immunofluorescence studies of infected cells (Gombart et al., 1989). The other ORFs appeared to be expressed in a similar manner with no primer extension products observed at 12 h p.i. and strong expression observed at 24 h p.i.

Comparison of OpMNPV HindIII M region with the homologous region in AcMNPV

In a previous study (Gombart et al., 1989), we demonstrated that the OpMNPV 32K PE-associated protein and the AcMNPV ORF 3 (Oellig et al., 1987) shared 58% amino acid
sequence identity. It was also observed that the predicted AcMNPV PE protein has a series of 16 Arg-Ser repeats, not present in the OpMNPV ORF, making the AcMNPV protein significantly larger. Comparison of the OpMNPV and AcMNPV predicted amino acid sequences from ORFs 1, 2, 4 and 5 demonstrated sequence similarities varying from 26 to 72% (Fig. 6 and Table 1). Other related proteins derived from OpMNPV and AcMNPV sequence data are also included in Table 1. They indicate that ORF 4 at 26% similarity is the least conserved, whereas polyhedrin at 90% similarity is the most highly conserved.

In addition to the Arg-Ser repeats in ORF 3 of AcMNPV described above, a number of other major differences between the two viruses were observed (Fig. 7). Amino acid sequence comparison between the two ORF 1 sequences revealed the existence of a 634 nt insert 1 nt after the codon for amino acid residue 101 (Fig. 6) in AcMNPV. Analysis of this insert revealed 4 bp direct repeats (TTAA) flanking 14 bp inverted repeats on either end of the inserted DNA. This motif of inverted repeats flanked by direct repeats is characteristic of transposable elements...
found in both eukaryotic and prokaryotic systems (Lewin, 1985). The direct repeat sequence is commonly a duplication of viral/host sequence upon insertion of foreign DNA. This particular sequence (TTAA) has been found duplicated at a number of sites where insect DNA has been inserted into a baculovirus genome (Fraser et al., 1985; Carstens, 1987; Beames & Summers, 1988). This putative transposable element located in ORF 1 contains the EcoRI site (Fig. 7) which causes the restriction map alterations characteristic of this strain of AcMNPV (Oellig et al., 1987). A search for ORFs within the nucleotide sequence of the putative transposable element indicated that none existed although a small or spliced gene cannot be ruled out.

Such transposable elements appear to be derived from repeated DNA in the host insect cell (Beames & Summers, 1988). Although they appear to lack ORFs, it is evident that they could have important influences on viral evolution. In the AcMNPV strain used by Oellig et al. (1987), the transposon contains the sequence ATAAAG which has been implicated as a regulatory element of baculovirus late genes (Rohrmann, 1986; Possee & Howard, 1987). Oellig et al. (1987) report a transcript originating near this sequence within the transposon. The first ATG sequence contained within such an mRNA could result in the translation of the 3' half of the ORF 1 polypeptide. In addition, the sole long terminal repeat remaining after the excision of a retrotransposon in AcMNPV contained two bidirectionally oriented ATAAAG sequences which appear to cause the initiation of late transcripts (Friesen et al., 1986). Therefore, in addition to disrupting viral genes, the insertion of transposable elements could also cause them to be expressed in an aberrant manner. Most transposable elements reported in the AcMNPV genome have been detected because they produce a 'few polyhedra' (FP) phenotype which is detectable by examination of plaques. The FP phenotype is often associated with alterations in the HindIII I fragment of the AcMNPV (Beames & Summers, 1988). The detection of such an element in ORF 1 suggests that the incorporation of transposable elements into baculovirus genomes may occur over a greater region of the genome than the FP detection system indicates. Such mutants would not normally be evident because of the lack of a suitable detection system.

When the OpMNPV HindIII M and the contiguous p26–p10–p74 region was aligned with sequences for the similar region of AcMNPV, it was observed that the OpMNPV genome lacks approximately 4 kb of DNA corresponding to the AcMNPV HindIII K/EcoRI S/hr5 region (Fig. 7). This region is located between ORF 5 and the p26 gene in AcMNPV. In OpMNPV, the p26 gene and ORF 5 are separated by an intergenic region of 224 bp (Fig. 2). In order to determine whether a region homologous to the AcMNPV HindIII K/EcoRI S region exists elsewhere in the OpMNPV genome, two hybridization experiments were performed using labelled AcMNPV HindIII K and EcoRI S fragments. The two AcMNPV fragments were labelled to high specific activity (Feinberg & Vogelstein, 1984) and were hybridized under non-stringent conditions (30%, formamide) (Maniatis et al., 1982) to blots of digested OpMNPV cosmid DNA and total viral genomic DNA. No hybridization was detected for either probe (data not shown). This suggests that regions homologous to the AcMNPV HindIII K and EcoRI S fragments do not exist in the OpMNPV genome.

In AcMNPV, the HindIII K/EcoRI S/hr5 region contains, in addition to hr5, two ORFs encoding predicted proteins of 35K and 94K (Friesen & Miller, 1987). These ORFs appear to be expressed as early genes and lack the baculovirus ATAAAG putative late gene regulatory sequence. Their mRNA is initiated downstream of a TATA-like consensus sequence. Baculoviruses probably have a number of groups of genes the functions of which are essential for their replication and survival in the environment (genes involved in DNA replication, viral structure, assembly and the like) as well as genes which are not absolutely essential but confer a selective advantage. The bacteriophage T4, which has a genome size similar to baculoviruses (160 kb), codes for over 140 proteins. Of these, nearly half have been found to be non-essential for phage production (Lewin, 1985). Many of the non-essential genes encode proteins involved in metabolism of DNA precursors and in the regulation of translation. The two ORFs in the AcMNPV HindIII K/EcoRI S region may represent such non-essential genes; these may confer a selective advantage to the virus, but without them other baculoviruses such as OpMNPV can survive. Therefore, the large difference in genome sizes of occluded baculoviruses (88 to 153 kb) (Burgess, 1977; Schafer et al., 1979) may reflect the incorporation of genes from insects or other
viruses and may give the virus a selective advantage but such genes may not be essential for its replication.

We thank Dr Peter Faulkner for making available AcMNPV sequence data prior to its publication. We would also like to thank Rebecca Russell, Dr Margot Pearson and Christian Gross for their valuable technical advice and for critically reading this manuscript and Jack Kramer and Reg McParland of the Center for Gene Research and Biotechnology at OSU for assistance with DNA sequence analysis and primer synthesis. This project was supported by grants from the USDA (87-CRCR-1-2380), NIH (AI 21973) and NIH (A10 7886-01) to G. W. R.

This is Technical Report No. 8665 from the Oregon State University Agricultural Experiment Station. The nucleotide sequence data have been assigned accession number d00366 by EMBL, GenBank and DDBJ.

REFERENCES


WEAN, A., QIN, J., RANKIN, C. & WEAVER, R. R. (1986). Nucleotide sequence of a portion of the Autographa californica nuclear polyhedrosis virus genome containing the EcoRI site-rich region (hr5) and a new open reading frame overlapping the pl0 gene. Virology 67, 2565–2570.


(Received 26 October 1988)