Journal of General Virology (1995), **76**, 2933–2940. *Printed in Great Britain*

Identification and characterization of an early gene in the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus

David S. Bischoff and James M. Slavicek*

USDA Forest Service, Northeastern Forest Experimental Station, Forest Sciences Laboratory, 359 Main Road, Delaware, OH 43015, USA

The *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus (Ld*M*NPV) gene encoding G22 was cloned and sequenced. The *G22* gene codes for a 191 amino acid protein with a predicted M_r of 22000. Expression of G22 in a rabbit reticulocyte system generated a protein with an M_r of 24000, in close agreement with the molecular mass predicted from the nucleotide sequence. G22 is not significantly homologous to any known protein, nor is a G22 homologue present in the *Autographa californica M*NPV (Ac*M*NPV). Temporal expression studies indicated that the *G22* gene was transcribed at readily detectable levels in the presence of cycloheximide. Transcripts were detected immediately after the virus adsorption period and throughout the infection cycle. The early transcriptional start sites of *G22* map to a sequence that resembles a subset of RNA polymerase II promoters/start sites that are found upstream of *Drosophila melanogaster* developmental and retrotransposon genes which lack TATA box motifs. Several consensus late baculovirus promoter/mRNA start site sequences (ATAAG) were identified upstream of the *G22* gene start codon.

Introduction

Nuclear polyhedrosis viruses (NPVs) are members of the *Baculoviridae*. These viruses produce two morphological forms, a budded virus form and a virus form that is occluded into a paracrystalline matrix termed a polyhedron. During the early stages of infection, budded virus is produced that infects a variety of cell types and is thought to bring about a systemic infection in the insect. During the later stages of the infection process the occluded form is produced and is released upon death of the insect (for review see Blissard & Rohrmann, 1990). Baculoviruses have large double-stranded DNA genomes ranging from approximately 80 to 165 kbp which have the potential to encode more than 150 proteins. Genes are expressed in a transcriptionally regulated temporal cascade. Early gene products are expressed by the host transcription and translation machinery, and can also be transcriptionally *trans*-activated by viral gene products (Fuchs *et al.*, 1983; Huh & Weaver, 1990; Glocker *et al.*, 1992). Examples of virus *trans*-activation have been seen in the *Autographa californica* multinucleocapsid nuclear polyhedrosis virus (Ac*MNPV*). Ac*M*NPV IE-0, for example, activates several promoters including *IE-1* and *39K* (Kovacs *et al.*, 1991). Early gene products control the expression of late and very late viral genes. Late gene expression occurs concomitant with DNA replication (Kelly & Lescott, 1981). These genes are transcribed by a virus-induced α-amanitin-resistant RNA polymerase which recognizes a specific late promoter sequence (Guarino & Summers, 1988; Rohrmann, 1986). Late gene products are required for nucleocapsid and polyhedron formation.

The *Lymantria dispar M*NPV (Ld*M*NPV) is pathogenic to the gypsy moth, a serious forest and urban tree-defoliating pest in the north-eastern United States. The genome of Ld*M*NPV is approximately 162 kbp in length with a G+C content of 60% (McCarthy *et al.*, 1979) compared to a genome of 133 kbp with a G+C content of 42% in Ac*M*NPV (Harrap & Payne, 1979). The genes present in the genomes of these viruses are organized in a similar array within short segments; however, overall the organization is not collinear (Bischoff & Slavicek, 1994; Bjornson & Rohrmann, 1992*a*, *b*; Riegel *et al.*, 1994). Phylogenetic analyses of the polyhedrin gene sequences of baculoviruses suggest that Ld*M*NPV is evolutionarily more distant from Ac*M*NPV than all other *M*NPVs investigated (Zanotto *et al.*, 1993).

The viral genome of Ld*M*NPV has been cloned and mapped with restriction endonucleases (Riegel *et al.*, 1994; Smith *et al.*, 1988), and transcription and translation maps have been generated (Slavicek, 1991). A number of Ld*M*NPV genes that are also present in

* Author for correspondence. Fax +1 614 368 0152. e-mail FSWA/S-J.SLAVICEK/OU-S24LO5A@MHS.ATTMAIL.COM

The accession number for the nucleotide sequence of the Ld*M*NPV *G22* gene is U37131.

Ac*M*NPV have been identified and characterized, including *ecdysteroid UDP-glucosyl transferase* (Riegel *et al.*, 1994), *DNA polymerase* (Bjornson *et al.*, 1992), *polyhedral envelope* (Bjornson & Rohrmann, 1992*a*), *polyhedrin* (Smith *et al.*, 1988), *p39 nucleocapsid* (Bjornson & Rohrmann, 1992*b*) and *viral protein kinase* (Bischoff & Slavicek, 1994).

In this report we extend the characterization of Ld*M*NPV by cloning and characterizing the *G22* gene. This is the first early gene identified in Ld*M*NPV that is expressed at high levels in the presence of cycloheximide. The *G22* gene promoter contains a sequence that shows homology to a subset of RNA polymerase II promoters in *Drosophila* (Jarrell & Meselson, 1991) positioned upstream of certain retrotransposon and developmental genes which lack TATA motifs.

Methods

Cells and virus. L. dispar 652Y cells were grown as monolayers in Goodwin's IPL-52B medium supplemented with 6·25 mM-glutamine and 10% fetal bovine serum. Cell cultures were inoculated with Ld*M*NPV clonal isolate 5–6 (Slavicek, 1991) at an m.o.i. of 10. Virus was removed after a 1 h adsorption period and replaced with fresh medium.

RNA isolation and Northern blot analysis. Infected *L. dispar* cells were harvested at various times post-infection (p.i.). Cytoplasmic RNA was isolated as described by Friesen & Miller (1985). For inhibitor studies, cells were treated with cycloheximide at 100 μg/ml for 30 min prior to virus adsorption. Cycloheximide was maintained in the medium throughout the time course. Total RNA was separated on 1·2% agarose gels containing formaldehyde and transferred to nitrocellulose. Northern blotting was performed as described by Mahmoudi & Lin (1989). Probes were radiolabelled with a nick-translation kit (BRL) and [α-^{32}P]dCTP (NEN). A 30 base oligonucleotide that is complimentary to the sequence shown in Fig. 3 (positions 662–692) was end-labelled with [γ-^{32}P]ATP (NEN) and used as a strand-specific probe that hybridized to the *G22* transcript.

Construction of λt11 cDNA library. A cDNA library was constructed using the RiboClone cDNA synthesis system (Promega). Poly(A)$^+$ RNA was isolated and purified from *L. dispar* cells infected with Ld*M*NPV at 7 h p.i. First-strand synthesis was accomplished with avian myeloblastosis virus (AMV) reverse transcriptase using an oligo(dT) primer with an adapter containing a unique *Not*I site. After second-strand synthesis of the cDNA, *Eco*RI linkers were ligated to the ends. The cDNAs were then cloned into λgt11 after digestion with *Eco*RI and *Not*I.

Viral DNA isolation and Southern blot analysis. Non-occluded virus from plaque-purified Ld*M*NPV 5–6 was isolated from infected *L. dispar* cells, and used as a source of genomic DNA for Southern blot analysis. Medium was decanted from the cells, and cellular debris was removed by centrifugation at 550 *g* for 10 min. Virus was pelleted by centrifugation at 104000 *g* for 45 min at 4 °C. The pellet was resuspended in 0·1 × TE at 4 °C overnight. One volume of 2 × DNA extraction buffer (20 mM-Tris–HCl pH 7·5, 120 mM-NaCl, 20 mM-EDTA pH 8·0, 2% SDS and 40 μg/ml proteinase K) was added, and the solution was incubated for 1 h at 50 °C. The solution was adjusted to 1% Sarkosyl and incubated for an additional 1 h at 50 °C. Viral DNA was extracted with 1 vol. Tris–HCl-buffered phenol, 2 vols chloroform–isoamyl alcohol (24:1), and then precipitated with 2 vols ethanol. Viral DNA was digested with restriction endonucleases and fractionated on 0·8% agarose–TBE gels. Southern blot analysis was performed on nitrocellulose using nick-translated probes generated as described above.

Sequencing. G22 sequence was obtained on both strands using the dideoxynucleotide sequencing method. M13 vectors were used to generate ssDNA templates which were then sequenced with the Sequenase version 2.0 DNA sequencing kit (USB). Primers for sequencing were synthesized on a 381A DNA synthesizer (Applied Biosystems). Sequence from plasmids was obtained with the *fmol* DNA sequencing system (Promega) using protocols supplied with the kit. [α-^{35}S]dATP was supplied by NEN. Sequence analysis was done using the MacVector program (IBI).

Primer extension mapping of early and late transcripts. Primer extension reactions were performed using the method of Crawford & Miller (1988). Total RNA was isolated from *L. dispar* cells infected with Ld*M*NPV 5–6 at 2, 7 and 72 h p.i. An 18 base primer that is complementary to the sequence shown in Fig. 3 (positions 297–314) was used in the reactions after being end-labelled with [γ-^{32}P]ATP (NEN). The primer was extended using Moloney murine leukaemia virus (M-MLV) reverse transcriptase. Primer extension products were fractionated on 6% polyacrylamide–8 M-urea gels and visualized by autoradiography.

Ribonuclease protection assay (RPA) mapping of early transcripts. Assays were performed with the RPA II ribonuclease protection assay kit (Promega) using the supplied protocols. The probe was generated as follows: a 1·2 kbp *Bam*HI fragment (15·2–16·4 kbp on the viral genome) that contained the *G22* gene was subcloned into the plasmid pBluescript SK(+) (Stratagene) to generate pDB120. This fragment contains two internal *Sty*I sites (at nucleotide positions 430 and 1018, see Fig. 3). Digestion of pDB120 with *Sty*I, end-filling the overhangs with Sequenase (USB), and religation, resulted in a subclone, pDB162, which contains a truncated form of the *G22* gene. A ^{32}P-labelled RNA probe complementary to the *G22* transcript was generated from pDB162 using T3 polymerase (Promega) and [α-^{32}P]CTP (NEN). This probe is 500 nucleotides in length with approximately 450 bases of viral sequence and 50 bases of vector sequence. The probe was hybridized for 20 h at 45 °C with 5 μg of mRNA isolated at 7 h p.i. from uninfected or infected cells. After treatment with RNase A and RNase T1, the protected fragments were fractionated on 6% polyacrylamide–8 M-urea gels and visualized by autoradiography. A DNA sequencing ladder was used to determine the size of the protected fragment.

In vitro transcription and translation of G22. G22 was expressed from pDB120 using the TnT coupled reticulocyte lysate system and T7 RNA polymerase (Promega) using the directions provided with the kit. The expressed protein was labelled by the addition of [^{35}S]methionine (NEN). Reaction products were analysed by SDS-PAGE and autoradiography.

Results and Discussion

Identification and transcriptional expression analysis of G22

A cDNA library generated with poly(A)$^+$ RNA from *L. dispar* cells infected with Ld*M*NPV 5–6 was probed with the *Eco*RV-*G* fragment. Early viral transcripts had previously been mapped to this genomic region (Slavicek, 1991). One cDNA clone (λ7–4) which hybridized to this probe was identified and contained an insert of approximately 750 bp (data not shown).

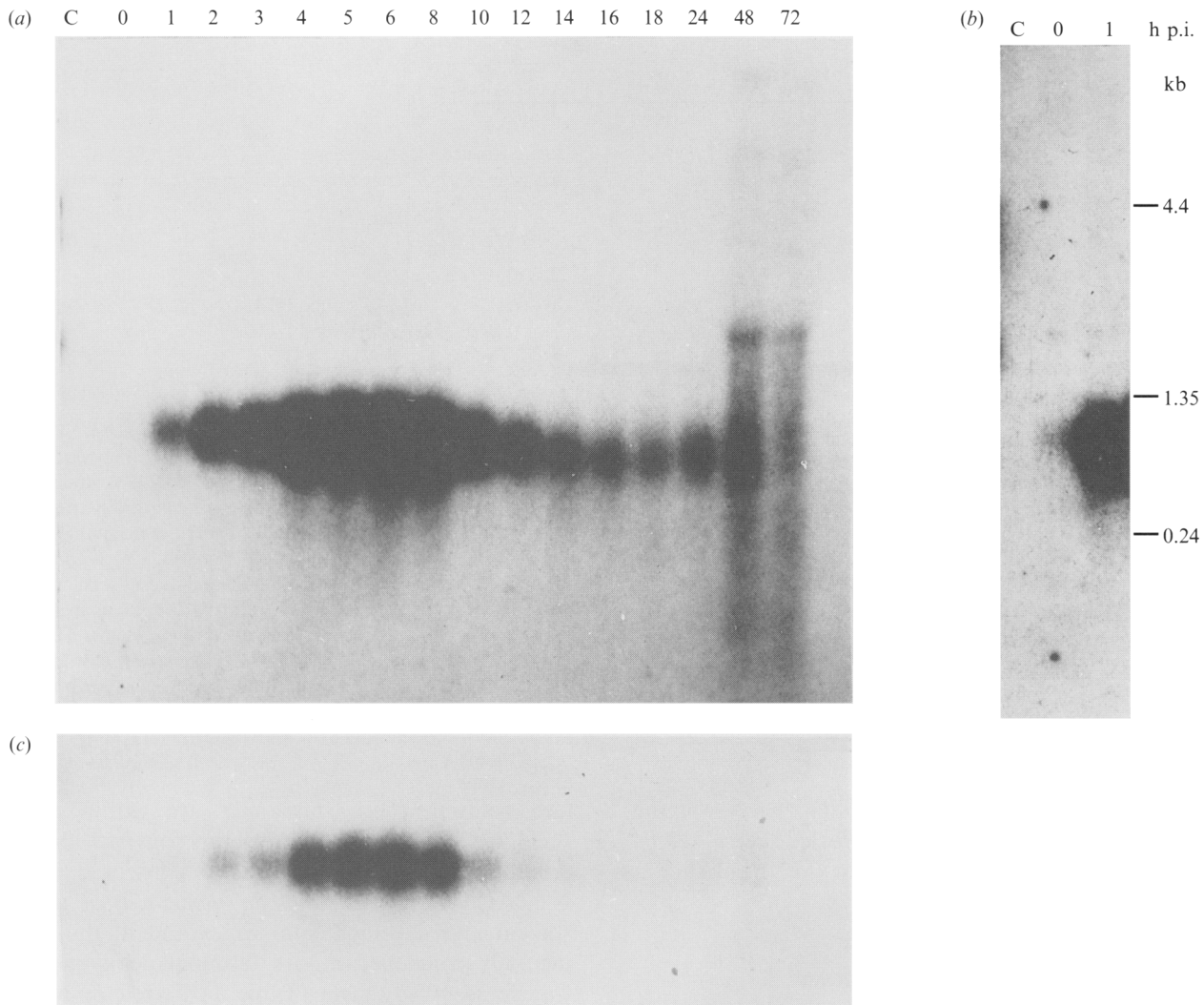

Fig. 1. Expression analysis of the Ld*M*NPV *G22* gene. (*a*) *L. dispar* 652Y cells were infected with 10 $TCID_{50}$ units/cell of Ld*M*NPV 5-6. At the times indicated the cells were harvested and cytoplasmic RNA was isolated. Total RNA (40 μg) was separated by formaldehyde–agarose gel electrophoresis, blotted, and probed with a ^{32}P-labelled *G22* cDNA clone. RNA from uninfected cells was used as a control (lane C). Various exposures of the same Northern blot (*a*) are shown in longer exposure (*b*) and shorter exposure (*c*). RNA size standards are indicated.

The cDNA insert was used as a probe on Northern blots to identify the size of the transcript being expressed. The probe hybridized to a 0·85 kb transcript (Fig. 1). This gene was subsequently designated *G22*, since the gene is present on the *Eco*RV-*G* fragment and is predicted to encode a protein with an M_r of 22000 (see below). The *G22* transcript can be detected throughout the time course of infection (Fig. 1 *a*), beginning immediately after the 1 h period of virus adsorption (Fig. 1 *b*). The level of transcription of *G22* increased from 4 to 8 h p.i. but declined thereafter in virus-infected cells (Fig. 1 *c*). This enhancement of transcription appears to be dependent on the expression of other viral proteins (see inhibitor study below). A *G22* strand-specific probe was also used to confirm that the 0·85 kb transcript corresponded to the *G22* gene. The transcription pattern was identical to the pattern seen using the cDNA probe (data not shown).

Genomic mapping and sequencing of G22

G22 was further localized to the 1·1 kbp *Eco*RV–*Bam*HI fragment located at approximately 15·3–16·4 kbp (9·3–10·0 m.u.) on the viral genome (data not shown). The restriction map of the 1·1 kbp *Eco*RV–*Bam*HI fragment is shown in Fig. 2(*a*). The *G22* cDNA was subcloned into pBluescript and the sequence of the 3′ end of the cDNA was determined. Comparison of this 3′ cDNA sequence with the viral genome sequence indicated that *G22* is transcribed from left to right with

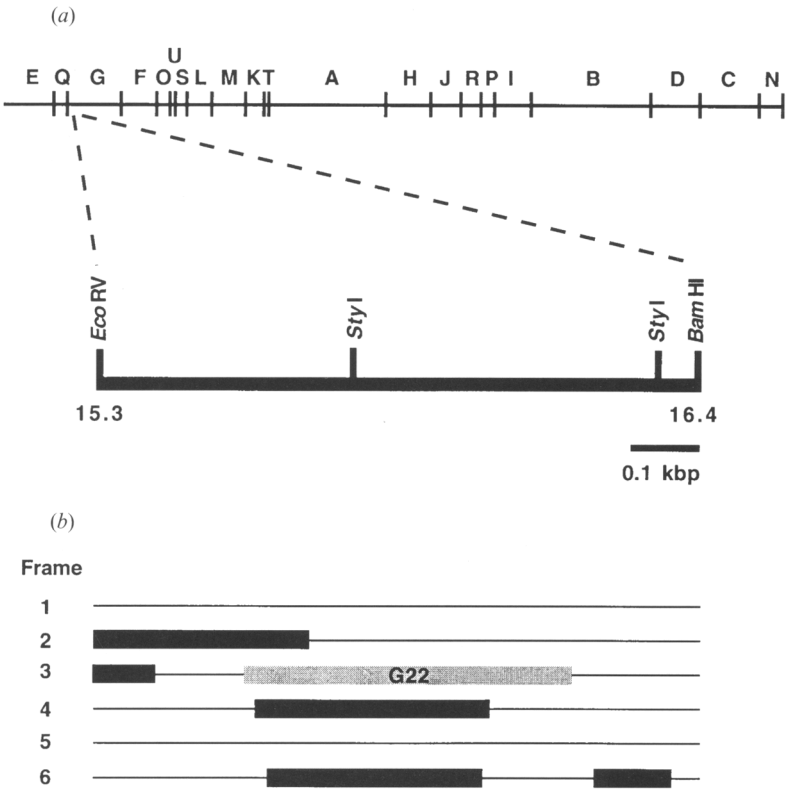

Fig. 2. Genomic location of Ld*M*NPV *G22*, sequencing strategy and ORF analysis. (*a*) *Bgl*II restriction map of the Ld*M*NPV viral genome (Riegel *et al.*, 1994). The enlarged map indicates the 1·1 kbp *Eco*RV-*Bam*HI fragment that encodes G22 at 15·3–16·4 kbp (9·3–10·0 m.u.) on the genome. Restriction sites used in subcloning and sequencing are indicated. (*b*) ORF analysis of this fragment in all six reading frames. The black boxes indicate ORFs that are at least 25 amino acids in length and which begin with an ATG start codon. The shaded box indicates the ORF encoding Ld*M*NPV G22.

respect to the viral genome. Computer analysis of the 1·1 kbp *Eco*RV–*Bam*HI fragment revealed several ORFs that may encode proteins, the largest (572 bp) being in frame 3 (Fig. 2*b*).

Characteristics of the nucleotide sequence

The nucleotide sequence of the 1048 bp *Eco*RV–*Bam*HI fragment and the predicted amino acid sequence of G22 are presented in Fig. 3. The *G22* ORF begins at position 258 and ends at position 831. This gene encodes a 191 amino acid protein with a predicted M_r of 22000. A region upstream of *G22* is approximately 83% GC-rich (residues 70–202). Closer examination of the nucleotide sequence reveals that there are two types of identical direct repeats present within this region: 11 bp repeats beginning at residues 133 and 183, and 13 bp repeats beginning at residues 89 and 118. In addition, there are several areas composed almost entirely of GC residues which contain the sequence GCGCG (starting at residues 76, 89, 105, 118, 166 and 213). GC-rich domains with this core sequence have been identified upstream of early genes in Ac*M*NPV and may function as promoter regulatory elements (Dickson & Friesen, 1991).

A TATA-like promoter was not identified upstream of *G22*. The early transcriptional start sites are at residues 231 and 232, as determined by primer extension mapping (see below). These start sites are within a region (GTCAGTCT) that shows homology to the transcriptional start sites (with the consensus sequence NTCAGTYN) in *Drosophila melanogaster* retrotransposon and developmental genes and in the developmentally controlled mouse terminal deoxynucleotidyltransferase (*Tdt*) gene (Table 1; Jarrell & Meselson, 1991). These promoters represent a subset of RNA polymerase II promoters that lack TATA motifs. Another Ld*M*NPV early gene, *vPK*, also lacks a TATA box and has early start sites within a similar sequence, CTCATTGC (Bischoff & Slavicek, 1994). Several Ac*M*NPV early genes that also lack TATA boxes have been identified, such as *lef-1* (Passarelli & Miller, 1993*a*), *lef*-3 (Li *et al.*, 1993), *lef-6* (Passarelli & Miller, 1994) and *cg30* (Thiem & Miller, 1989). Of these genes, only the start site sequence of *lef-6* (TTACAGTGT) resembles those seen upstream of Ld*M*NPV early genes. Several CGATAA sequences are located upstream of the *G22* transcriptional start sites. These sequences are similar to the GATA consensus sequence (A/T)GATA(G/A) that is recognized by cellular transcription factors (Evans *et al.*, 1988). GATA motifs have been identified upstream of the Ac*M*NPV early genes *gp64* (Kogan & Blissard, 1994), *PE38* (Krappa *et al.*, 1992) and *ME53* (Knebel-Mörsdorf *et al.*, 1993). The consensus late promoter sequence, ATAAG (Rohrmann, 1986), is present at several locations upstream of the *G22* start codon

Fig. 3. Nucleotide sequence of the 1·1 kbp *Eco*RV–*Bam*HI fragment and the predicted amino acid sequence of G22. GC-rich direct repeats are indicated by single (13 bp) and double (11 bp) underlines. Other GC-rich areas that contain the core sequence GCGCG are boxed. Early transcriptional start sites are indicated by arrows. Potential late promoter sequences are shaded. Potential polyadenylation signals are presented in bold type. The position of the poly(A) tail is indicated in bold type with an asterisk.

Table 1. *Comparison of early LdMNPV promoters with the consensus start site for promoters lacking TATA boxes**

Promoter	Start site
Ld*M*NPV *G22*	GTCAGTCT
Ld*M*NPV *vPK*	CTCATTGC
Ac*M*NPV *lef-6*	TTCAGTGT
Mouse *Tdt*	CTCATTCT
gypsy (*mdg4*)	CTCAGTTC
Antennapedia	TTCAGTTG
engrailed	GTCAACTA
Drosophila consensus	NTCAGTYN

* The Table shows nucleotide comparisons of the *G22* start site with those of Ld*M*NPV *vPK* (Bischoff & Slavicek, 1994), Ac*M*NPV *lef-6* (Passarelli & Miller, 1994), mouse *Tdt* (Smale & Baltimore, 1989), *Drosophila gypsy* retrotransposon *mdg4* (Jarrell & Meselson, 1991), *Drosophila* developmental genes *Antennapedia* (Perkins *et al.*, 1988) and *engrailed* (Soeller *et al.*, 1988), and with the *Drosophila* consensus start site sequence (Jarrell & Meselson, 1991).

† Initiation nucleotides are underlined.

(beginning at positions 45, 66, 81 and 110) and may drive expression of *G22* during the late phase of virus replication. Early and late promoters were also identified upstream of Ld*M*NPV *vPK* (Bischoff & Slavicek, 1994). Downstream of *G22* there are several alternative polyadenylation signals at positions 828 and 882 (AATTAA), and at position 957 (ATTAAA).

Characteristics of the protein sequence

G22 could encode a 191 amino acid protein with a calculated M_r of 22000. The predicted amino acid sequence of G22 was compared with other proteins in GenBank at the National Center for Biotechnology Information (NCBI) using the BLAST network service (Altschul *et al.*, 1990). This search revealed that G22 did not exhibit significant homology to any other known protein, nor were any known sequence motifs identified

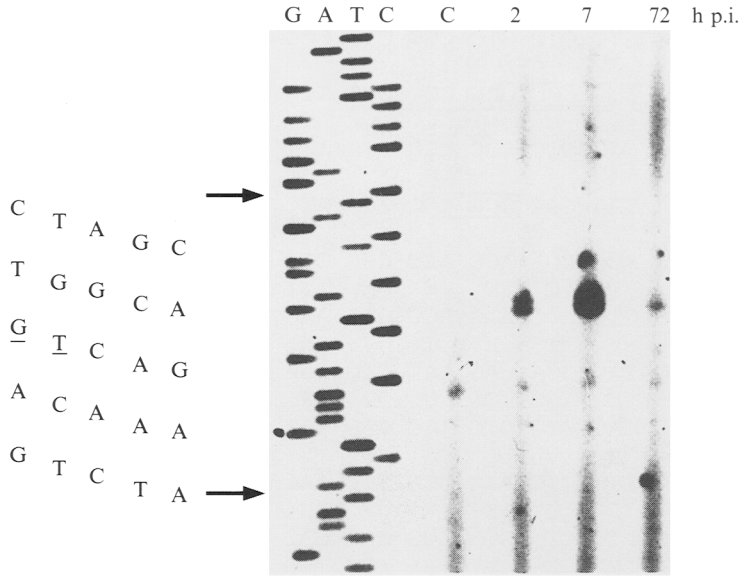

Fig. 4. Mapping the 5′ end of *G22* transcripts. Primer extension analysis of *G22* RNAs isolated at 2, 7 and 72 h p.i. Total RNA (40 μg) was incubated with an end-labelled 18 base primer that is complementary to the *G22* nucleotide sequence at positions 297–314 (Fig. 3). After extension of the primer using M-MLV reverse transcriptase, the extension products were fractionated by PAGE and visualized by autoradiography. RNA from uninfected cells was used as a control (lane C). The sequence ladder was generated using the same primer.

within the predicted G22 amino acid sequence. Recently, the complete nucleotide sequence of the Ac*M*NPV has been reported (Ayres *et al.*, 1994). Although several genes that are present in both Ld*M*NPV and Ac*M*NPV have been identified [*polyhedrin* (Smith *et al.*, 1988), *p39* (Bjornson & Rohrmann, 1992*b*), *vPK* (Bischoff & Slavicek, 1994), *egt* (Riegel *et al.*, 1994) and *DNA polymerase* (Bjornson *et al.*, 1992)], there is no *G22* gene homologue present in Ac*M*NPV.

Mapping of the 5′ and 3′ ends of the G22 transcript

Primer extension reactions were carried out using total RNA isolated from *L. dispar* cells at 2, 7 and 72 h p.i. in order to map the 5′ ends of *G22* transcripts. At 2 and 7 h p.i. transcription initiates at two residues, positioned 26 and 27 bp upstream of the *G22* start codon. At a later time-point (72 h) only one start site for gene transcription was detected, 26 bp upstream of the *G22* start codon (Fig. 4). Transcriptional start sites that originated within the consensus late promoter sequences were not detected.

RPAs were also performed to map the 5′ end of the *G22* transcript. A protected fragment of 204 bp was detected using mRNA from infected cells at 7 h p.i., but not from mRNA isolated from uninfected cells (data not shown). This would correspond to an mRNA start site 28 bp upstream of the *G22* start codon (Fig. 3). This result is in close agreement with the primer extension results and suggests that the *G22* transcript is not spliced.

The 3′ end of the *G22* transcript was mapped by sequencing the λ7-4 cDNA insert (data not shown). The poly(A) tail begins at nucleotide position 975 (see Fig. 3), 13 bp downstream of the ATTAAA polyadenylation sequence. The size of the transcript predicted from the DNA sequence is 815 or 816 nt [not including the poly(A) tail], in close agreement with the approximately 0·85 kb *G22* transcript seen in the temporal expression experiments.

Temporal expression of G22 in the presence of cycloheximide

Recent results indicate that although all early genes appear to be transcribed by the host RNA polymerase II (Fuchs *et al.*, 1983; Huh & Weaver, 1990), differences in the level of transcription of early genes are seen due to regulation of viral gene expression by baculoviral *trans*-regulator proteins (Glocker *et al.*, 1992; Kovacs *et al.*, 1991; Passarelli & Miller, 1993 *b*; Carson *et al.*, 1991). To determine whether viral protein synthesis affected transcription from the *G22* promoter, cells were infected in the presence of the protein synthesis inhibitor cycloheximide (Rice & Miller, 1986). The 0·85 kb *G22* transcript was detected at all time-points in the presence of cycloheximide (Fig. 5*b*). The amount of *G22* transcript present at 2 h p.i. was approximately the same, with or without cycloheximide. In contrast, significantly more *G22* gene transcript was present in the cells without cycloheximide (Fig. 5*a*) than in the cells with cycloheximide at 4 and 6 h p.i. Higher levels of transcription in the absence of cycloheximide may be due to *trans*-activation of the *G22* promoter by other viral proteins or by G22 itself.

In vitro transcription and translation of G22

To determine if *G22* encoded a protein, *G22* was expressed from plasmid pDB120. A band of apparent M_r 24000 was seen after SDS-PAGE and autoradiography (Fig. 6). The M_r of the G22 protein is predicted from the

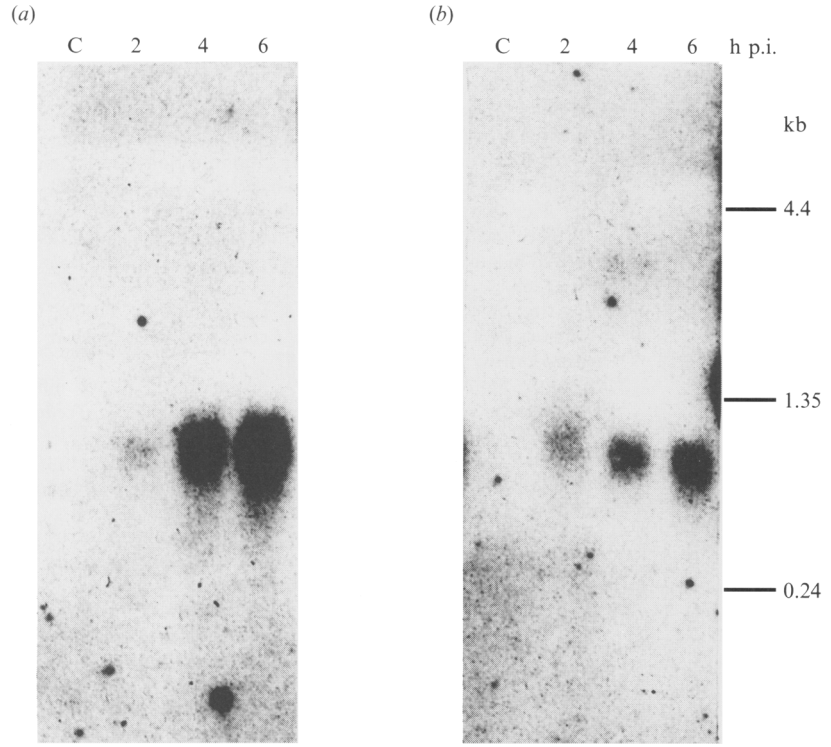

Fig. 5

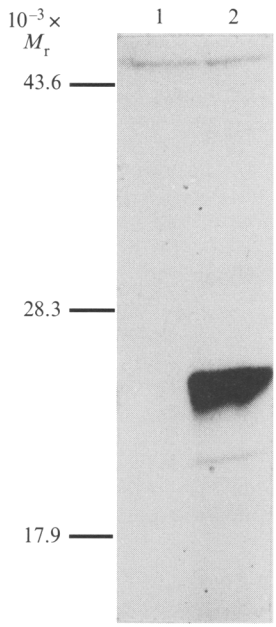

Fig. 6

Fig. 5. Temporal analysis of *G22* transcripts in the presence of cycloheximide. Total RNA was isolated at the indicated times from *L. dispar* cells infected with LdMNPV 5-6 in the absence (*a*) and presence (*b*) of cycloheximide added 30 min prior to the virus adsorption period at a final concentration of 100 μg/ml. The inhibitor was maintained in the medium throughout the time course. RNA from uninfected cells was used as a control (lane C). RNA size standards are indicated.

Fig. 6. SDS–PAGE analysis after expression of *G22* under the control of the T7 promoter. Autoradiograph of G22 expressed in a rabbit reticulocyte transcription and translation system, and labelled with [^{35}S]methionine. Lane 1, pBluescript control plasmid. Lane 2, pDB120 expressing G22. M_r standards are indicated.

nucleotide sequence to be 22000. No radiolabelled band was detected from the parent plasmid.

The G22 transcription initiation site exhibits sequence homology to initiator elements

G22 is the second early gene identified in Ld*M*NPV that contains a start site sequence which exhibits homology to an RNA polymerase II promoter element, termed the initiator, present in the mouse lymphocyte-specific gene *Tdt* (Smale & Baltimore, 1989). Several *Drosophila* TATA-less retrotransposons and developmental genes such as *gypsy*, *B104*, *297*, *Antennapedia* and *engrailed* (Jarrell & Meselson, 1991, and references therein) also contain a sequence that shows homology with the *Tdt* gene initiator. Characteristics of this type of initiator promoter include a consensus promoter sequence, the lack of a TATA motif, and the presence of sequences downstream from the origin of transcription that are necessary for efficient expression. Both the *gypsy* retrotransposon and the *Tdt* gene initiator sequences were shown to be functional since alterations of this sequence resulted in the loss of transcription initiation (Jarrell & Meselson, 1991; Smale & Baltimore, 1989). Recently, it has been shown that the transcriptional start site of the Ac*M*NPV *ie-1* gene (TTCAGTTG) can function as an initiator element (Pullen & Friesen, 1995). Although the *IE-1* promoter region also contains a TATA box, accurate transcription was still detected when the TATA box was deleted. Studies are in progress to determine if the sequence upstream of the Ld*M*NPV *G22* gene with homology to the initiator element functions as an initiator transcription control element. In addition, attempts are in progress to delete the gene in order to investigate the role of G22 in virus replication.

References

Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. G. (1990). Basic local alignment tool. *Journal of Molecular Biology* **215**, 403–410.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M. & Possee, R. D. (1994). The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202**, 586–605.

BISCHOFF, D. S. & SLAVICEK, J. M. (1994). Identification and characterization of a protein kinase gene in the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *Journal of Virology* **68**, 1728–1736.

BJORNSON, R. M. & ROHRMANN, G. F. (1992*a*). Nucleotide sequence of the polyhedron envelope protein gene region of the *Lymantria dispar* nuclear polyhedrosis virus. *Journal of General Virology* **73**, 1499–1504.

BJORNSON, R. M. & ROHRMANN, G. F. (1992*b*). Nucleotide sequence of the p39-capsid gene region of the *Lymantria dispar* nuclear polyhedrosis virus. *Journal of General Virology* **73**, 1505–1508.

BJORNSON, R. M., GLOCKER, B. & ROHRMANN, G. F. (1992). Characterization of the nucleotide sequence of the *Lymantria dispar* nuclear polyhedrosis virus DNA polymerase gene region. *Journal of General Virology* **73**, 3177–3183.

BLISSARD, G. W. & ROHRMANN, G. F. (1990). Baculovirus diversity and molecular biology. *Annual Review of Entomology* **35**, 127–155.

CARSON, D. D., SUMMERS, M. D. & GUARINO, L. A. (1991). Transient expression of the *Autographa californica* nuclear polyhedrosis virus immediate-early gene, IE-N, is regulated by three viral elements. *Journal of Virology* **65**, 945–951.

CRAWFORD, A. M. & MILLER, L. K. (1988). Characterization of an early gene accelerating expression of late genes of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology* **62**, 2773–2781.

DICKSON, J. A. & FRIESEN, P. D. (1991). Identification of upstream promoter elements mediating early transcription from the 35000-molecular-weight protein gene of *Autographa californica* nuclear polyhedrosis virus. *Journal of Virology* **65**, 4006–4016.

EVANS, T., REITMAN, M. & FELSENFELD, G. (1988). An erythrocyte-specific DNA-binding factor recognizes a regulatory sequence common to all chicken globin genes. *Proceedings of the National Academy of Sciences, USA* **85**, 5976–5980.

FRIESEN, P. D. & MILLER, L. K. (1985). Temporal regulation of baculovirus RNA: overlapping early and late transcripts. *Journal of Virology* **54**, 394–400.

FUCHS, L. Y., WOODS, M. S. & WEAVER, R. F. (1983). Viral transcription during *Autographa californica* nuclear polyhedrosis virus infection: a novel RNA polymerase induced in infected *Spodoptera frugiperda* cells. *Journal of Virology* **48**, 641–646.

GLOCKER, B., HOOPES, R. R. & ROHRMANN, G. F. (1992). In vitro transactivation of baculovirus early genes by nuclear extracts from *Autographa californica* nuclear polyhedrosis virus-infected *Spodoptera frugiperda* cells. *Journal of Virology* **66**, 3476–3484.

GUARINO, L. A. & SUMMERS, M. D. (1988). Functional mapping of *Autographa californica* nuclear polyhedrosis virus genes required for late gene expression. *Journal of Virology* **62**, 463–471.

HARRAP, K. A. & PAYNE, C. C. (1979). The structural properties and identification of insect viruses. *Advances in Virus Research* **25**, 273–355.

HUH, N. E. & WEAVER, R. F. (1990). Identifying the RNA polymerases that synthesize specific transcripts of the *Autographa californica* nuclear polyhedrosis virus. *Journal of General Virology* **71**, 195–201.

JARRELL, K. A. & MESELSON, M. (1991). *Drosophila* retrotransposon promoter includes an essential sequence at the initiation site and requires a downstream sequence for full activity. *Proceedings of the National Academy of Sciences, USA* **88**, 102–104.

KELLY, D. C. & LESCOTT, T. (1981). Baculovirus replication: protein synthesis in *Spodoptera frugiperda* cells infected with *Trichoplusia ni* nuclear polyhedrosis virus. *Microbiologica* **4**, 35–57.

KNEBEL-MÖRSDORF, D., KREMER, A. & JAHNEL, F. (1993). Baculovirus gene ME53, which contains a putative zinc finger motif, is one of the major early-transcribed genes. *Journal of Virology* **67**, 753–758.

KOGAN, P. H. & BLISSARD, G. W. (1994). A baculovirus *gp64* early promoter is activated by host transcription factor binding to CACGTG and GATA elements. *Journal of Virology* **68**, 813–822.

KOVACS, G. R., GUARINO, L. A. & SUMMERS, M. D. (1991). Novel regulatory properties of the IE1 and IE0 transactivators encoded by the baculovirus *Autographa californica* multicapsid nuclear polyhedrosis virus. *Journal of Virology* **65**, 5281–5288.

KRAPPA, R., BEHN-KRAPPA, A., JAHNEL, F., DOERFLER, W. & KNEBEL-MÖRSDORF, D. (1992). Differential factor binding at the promoter of early baculovirus gene PE38 during viral infection: GATA motif is recognized by an insect protein. *Journal of Virology* **66**, 3494–3503.

LI, Y., PASSARELLI, A. L. & MILLER, L. K. (1993). Identification, sequence, and transcriptional mapping of *lef-3*, a baculovirus gene involved in late and very late gene expression. *Journal of Virology* **67**, 5260–5268.

MCCARTHY, W. J., MURPHY, T. F. & LANGRIDGE, W. (1979). Characteristics of the DNA from the *Lymantria dispar* nuclear polyhedrosis virus. *Virology* **95**, 593–597.

MAHMOUDI, M. & LIN, V. K. (1989). Comparison of two different hybridization systems in Northern transfer analysis. *Bio/Techniques* **7**, 331–333.

PASSARELLI, A. L. & MILLER, L. K. (1993*a*). Identification and characterization of *lef-1*, a baculovirus gene involved in late and very late gene expression. *Journal of Virology* **67**, 3481–3488.

PASSARELLI, A. L. & MILLER, L. K. (1993*b*). Three baculovirus genes involved in late and very late gene expression: *ie-1*, *ie-n*, and *lef-2*. *Journal of Virology* **67**, 2149–2158.

PASSARELLI, A. L. & MILLER, L. K. (1994). Identification and transcriptional regulation of the baculovirus *lef-6* gene. *Journal of Virology* **68**, 4458–4467.

PERKINS, K. K., DAILEY, G. M. & TJIAN, R. (1988). In vitro analysis of the *Antennapedia* P2 promoter: identification of a new *Drosophila* transcription factor. *Genes and Development* **2**, 1615–1626.

PULLEN, S. S. & FRIESEN, P. D. (1995). The CAGT motif functions as an initiator element during early transcription of the baculovirus transregulator *ie-1*. *Journal of Virology* **69**, 3575–3583.

RICE, W. C. & MILLER, L. K. (1986). Baculovirus transcription in the presence of inhibitors and in nonpermissive *Drosophila* cells. *Virus Research* **6**, 155–172.

RIEGEL, C. I., LANNER-HERRERA, C. & SLAVICEK, J. M. (1994). Identification and characterization of the ecdysteroid UDP-glucosyltransferase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus. *Journal of General Virology* **75**, 829–838.

ROHRMANN, G. F. (1986). Polyhedrin structure. *Journal of General Virology* **67**, 1499–1513.

SLAVICEK, J. M. (1991). Temporal analysis and spatial mapping of *Lymantria dispar* nuclear polyhedrosis virus transcripts and in vitro translation polypeptides. *Virus Research* **20**, 223–236.

SMALE, S. T. & BALTIMORE, D. (1989). The 'initiator' as a transcription control element. *Cell* **57**, 103–113.

SMITH, I. R. L., VAN BEEK, N. A. M., PODGWAITE, J. D. & WOOD, H. A. (1988). Physical map and polyhedrin gene sequence of *Lymantria dispar* nuclear polyhedrosis virus. *Gene* **71**, 97–105.

SOELLER, W. C., POOLE, S. J. & KORNBERG, T. (1988). In vitro transcription of the *Drosophila engrailed* gene. *Genes and Development* **2**, 68–81.

THIEM, S. M. & MILLER, L. K. (1989). A baculovirus gene with a novel transcription pattern encodes a polypeptide with a zinc finger and a leucine zipper. *Journal of Virology* **63**, 4489–4497.

ZANOTTO, P. M. A., KESSING, B. D. & MARUNIAK, J. E. (1993). Phylogenetic interrelationships among baculoviruses: evolutionary rates and host associations. *Journal of Invertebrate Pathology* **62**, 147–164.

(*Received 16 January 1995; Accepted 1 August 1995*)