1994

Identification and characterization of the ecdysteroid UDPglucosyltransferase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus

Christopher I. Riegel
USDA Forest Service

Carita Lanner-Herrera
Swedish University of Agricultural Science

James M. Slavicek
USDA Forest Service

Identification and characterization of the ecdysteroid UDP-glucosyltransferase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus

Christopher I. Riegel, Carita Lanner-Herrera† and James M. Slavicek*

USDA Forest Service, Northeastern Forest Experiment Station, Forestry Sciences Laboratory, 359 Main Road, Delaware, Ohio 43015, U.S.A.

We have located, cloned, sequenced and characterized the ecdysteroid UDP-glucosyltransferase gene (*egt*) gene from the baculovirus *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus.(LdMNPV), which is specific for the gypsy moth (*L. dispar*). The *egt* gene from the related baculovirus *Autographa californica* multinucleocapsid nuclear polyhedrosis virus (AcMNPV) disrupts the hormonal balance of the host larva by galactosylating ecdysone, which prevents moulting. The location of the LdMNPV *egt* gene, determined by hybridization analysis using a cloned coding segment of the AcMNPV *egt* gene, was mapped to between 79·1 and 80·2 map units on the viral genome. This region contains an open reading frame of 1464 nucleotides capable of encoding a 55K polypeptide. This predicted protein exhibits a 42% amino acid identity with the AcMNPV *egt* polypeptide. Transcripts of the *egt* gene were analysed by Northern blot and primer extension. The *egt* gene is transcribed from approximately 12 to 48 h, and maximally at about 16 h post-infection. Transcription occurred in the presence of aphidicolin, a viral DNA synthesis inhibitor, but not in the presence of cycloheximide, a protein synthesis inhibitor. Therefore the LdMNPV *egt* gene is classified as a delayed early gene. The *egt* gene is transcribed in a clockwise direction with respect to the circular map, and transcription initiates at a single site. Comparisons between the two baculoviral *egt* proteins and mammalian UDP-glucuronosyltransferases reveal areas which are conserved between the mammalian and baculoviral genes, as well as areas that are only conserved in the viral *egt* proteins. The LdMNPV protein sequence appears to include a signal peptide, which would allow the protein to be secreted into the haemolymph.

Introduction

Lymantria dispar multinucleocapsid nuclear polyhedrosis virus (LdMNPV) is a member of the *Baculoviridae* (subgroup A). Members of this group of insect viruses with lepidopteran hosts have received considerable attention because several of these insects are significant pests in agriculture and forestry. One species with particular relevance to forestry is the LdMNPV which is pathogenic to the gypsy moth (*L. dispar*), a significant defoliator of over 300 species of trees and shrubs. This virus has been used on a limited basis to control the gypsy moth since it was registered by the U.S. Forest Service under the name Gypchek.

Nuclear polyhedrosis viruses have a two-stage life cycle (reviewed by Blissard & Rohrmann, 1990). Early after infection budded virus is released which gives rise to a systemic infection in affected insects. Later in the course of infection viral nucleocapsids are occluded in a polyhedral protein matrix. Polyhedra are relatively stable in the environment after the death of infected insects and are infectious when consumed by susceptible larvae. Polyhedra can be formulated and applied aerially making them potential biological control agents for agricultural and forest insect pests (reviewed by Wood & Granados, 1991).

In contrast to the prototype baculovirus *Autographa californica* MNPV (AcMNPV), LdMNPV is not well characterized. Restriction endonuclease maps (Smith *et al.*, 1988; this report) and transcription and translation maps (Slavicek, 1991) of LdMNPV isolates have been generated. These studies have shown that the genome of LdMNPV, ranging from approximately 160 kb to 165 kb, is significantly larger than the genomes of most subgroup A baculoviruses (reviewed in Harrap & Payne, 1979). The genome of LdMNPV is further distinguished from other subgroup A baculoviruses by its high G + C

content of approximately 60 % in comparison with a range of 37 to 54 % for other baculoviruses (Harrap & Payne, 1979). These differences suggest that gene composition and organization within the LdMNPV genome are unique among subgroup A baculoviruses.

To characterize further the LdMNPV genome and initiate studies on the virus life cycle and interaction with its host we have used cloned genes from AcMNPV to identify their homologues within LdMNPV. One gene of particular interest is the homologue to the AcMNPV ecdysteroid UDP-glucosyltransferase (*egt*) gene. AcMNPV *egt* expressed *in vitro* catalyses the transfer of glucose and galactose to ecdysteroids (O'Reilly & Miller, 1989, 1991; O'Reilly *et al.*, 1992), and when produced in *Spodoptera frugiperda* larvae catalyses conjugation of ecdysteroids with galactose (O'Reilly *et al.*, 1992). Insect larval moulting and pupation are inhibited when infected with wild-type AcMNPV, whereas these processes are initiated in larvae infected with *egt* gene minus virus strains (O'Reilly & Miller, 1989, 1991). The inhibition of larval moulting in wild-type virus-infected insects extends the length of time for which the larvae feed, thereby increasing the amount of progeny virus produced. In contrast, larvae infected with an *egt* gene minus virus exhibited reduced feeding and earlier mortality (O'Reilly & Miller, 1991). A similar inhibition of larval moulting and pupation by wild-type LdMNPV infection has been noted in the gypsy moth (Burand & Park, 1992).

In this report we describe the identification, genomic mapping and nucleotide sequence of the LdMNPV ecdysteroid UDP-glucosyltransferase gene. In addition, we have characterized the transcriptional expression of the LdMNPV *egt* gene in cell culture. Our results indicate that the *egt* gene is expressed early after infection. The genomic location of the LdMNPV *egt* gene was found to be distinct from that of AcMNPV *egt*, suggesting that the arrangement of the genes of these viruses is not collinear.

Methods

Maintenance of cells and virus. All virus growth was carried out in Ld652Y cells. Cells were grown in Goodwin's IPL52B medium (JRH Biosciences) supplemented with 10 % fetal bovine serum (Hyclone) and 6·0 mM-glutamine (Gibco) at 27 °C. LdMNPV strains A21, A21-2 and 5-6 were used for these studies (Slavicek *et al.*, 1992).

Mapping LdMNPV isolate 5-6 and cloning egt. A cosmid genomic library of isolate LdMNPV 5-6 was generated by partial digestion of genomic DNA with *Pst*I or *Cla*I and the fragments were cloned into pHC79 (Hohn & Collins, 1980). Six overlapping clones that covered the entire viral genome were isolated and restriction-mapped according to standard procedures. The *egt* gene was localized in isolate 5-6 by low stringency (50 % formamide, 5 × SSC, 31 °C) hybridization of a blot of *Bam*HI, *Hin*dIII, *Eco*RI and *Eco*RV fragments of genomic DNA with a probe derived from the AcMNPV *egt* clone pBsBCEP (kindly provided by L. Miller; O'Reilly & Miller, 1990) which contained only

egt coding sequences. A 1·5 kbp *Pst*I fragment from LdMNPV isolate 5-6 which hybridized with the pBsBCEP was cloned. This homologous probe was then used to locate the *egt* gene in the LdMNPV isolate A21-2. The *egt* gene was mapped in isolate A21-2 by probing a genomic Southern blot of the LdMNPV strain A21-2 doubly digested with *Hin*dIII and *Bam*HI. It was further localized by digesting the purified 9·6 kbp *Hin*dIII–*Bam*HI fragment containing the *egt* gene with *Nhe*I and repeating the probing carried out above. The gene was excised from the strain A21-2 genome using the *Nhe*I site at 128·3 kbp and the *Hin*dIII site at 133·2 kbp and subsequently cloned into the multiple cloning sites of pUC18 and the Bluescript plasmids SK− and SK+ (Stratagene) using standard cloning techniques.

Sequencing methods. The nucleotide sequence of the *egt* gene was determined in both strands. Due to the high GC ratio found in the *egt* gene a variety of methods were used to generate the sequence of the LdMNPV *egt* gene. Both double-stranded and single-stranded (using the Bluescript phagemids) sequencing were carried out using Sequenase (USB). Plasmids and phagemids were grown in DH5αF′ cells (Stratagene). Single-stranded phagemid DNA was packaged using R408 helper phage (Stratagene). Particularly difficult regions were sequenced using the fmol *Taq* polymerase sequencing kit (Promega) and single-stranded DNA. Sequence was generated using both commercially available M13 and Bluescript sequencing primers, and specific 18 nucleotide internal primers synthesized on an ABI model 381A DNA synthesizer. A series of *Pst*I and *Sal*I subclones were made in the Bluescript plasmids and single-stranded DNA was generated and sequenced. For some of these subclones it proved necessary to create deletion mutants using an Exo III/Mung Bean nuclease kit (Stratagene) to remove interfering secondary structures.

Measuring RNA expression. Tissue culture flasks (25 cm^2) were seeded with 1×10^6 *L. dispar* 652Y cells. The cells were infected with LdMNPV isolate A21-2 at 10 TCID$_{50}$ units per cell for 1 h at 27 °C. At that time the inoculum was replaced with 5 ml of fresh medium. Cells were harvested at 0, 8, 12, 16, 24, 48 and 72 h post-infection (p.i.) (counted from the end of the 1 h adsorption period). Total cytoplasmic RNA was isolated following the procedure of Friesen & Miller (1985). RNA was separated on a 20 cm 1·2 % agarose gel containing formaldehyde by overnight electrophoresis at 30 V, and the gel was blotted. A 700 bp *Pst*I fragment probe, located between 130·9 and 131·6 kbp which hybridizes to the 3′ end of the *egt* coding sequence, was generated by nick translation (BRL). This probe was hybridized to the blot and washed according to the phosphate buffer procedure of Mahmoudi & Lin (1989). The blot was exposed to Kodak XAR-5 film in cassettes containing X-Omatic intensifying screens (Kodak). Inhibitor studies were carried out using the method of O'Reilly & Miller (1990). Cells infected with LdMNPV strain A21-2 were grown in the presence of the DNA replication inhibitor aphidicolin or the protein synthesis inhibitor cycloheximide. Total cytoplasmic RNA from these cells and from control cells containing no inhibitor was harvested at 16 h p.i. and the RNA was gel-fractionated as above. The gel was blotted and hybridized with a *Bst*XI–*Stu*I 1 kbp ^{32}P-labelled probe which hybridizes to the *egt* coding sequence. The blot was exposed to film as above.

Determining the 5′ end of the mRNA. The 5′ end of the *egt* mRNA was mapped by primer extension using the method of Crawford & Miller (1988). Total cytoplasmic RNA generated by the above method was used. An 18 nucleotide-long internal oligonucleotide, located between 98 and 115 bases downstream from the initiation codon, was 5′ end-labelled with ^{32}P and used for the primer extension reaction with Moloney murine leukaemia virus reverse transcriptase. The same oligonucleotide was used as a sequencing primer in single-stranded sequencing reactions carried out using a complete clone of the *egt* gene as the template. The primer extension and sequencing reactions were then electrophoresed on an 8 % polyacrylamide–urea sequencing gel at
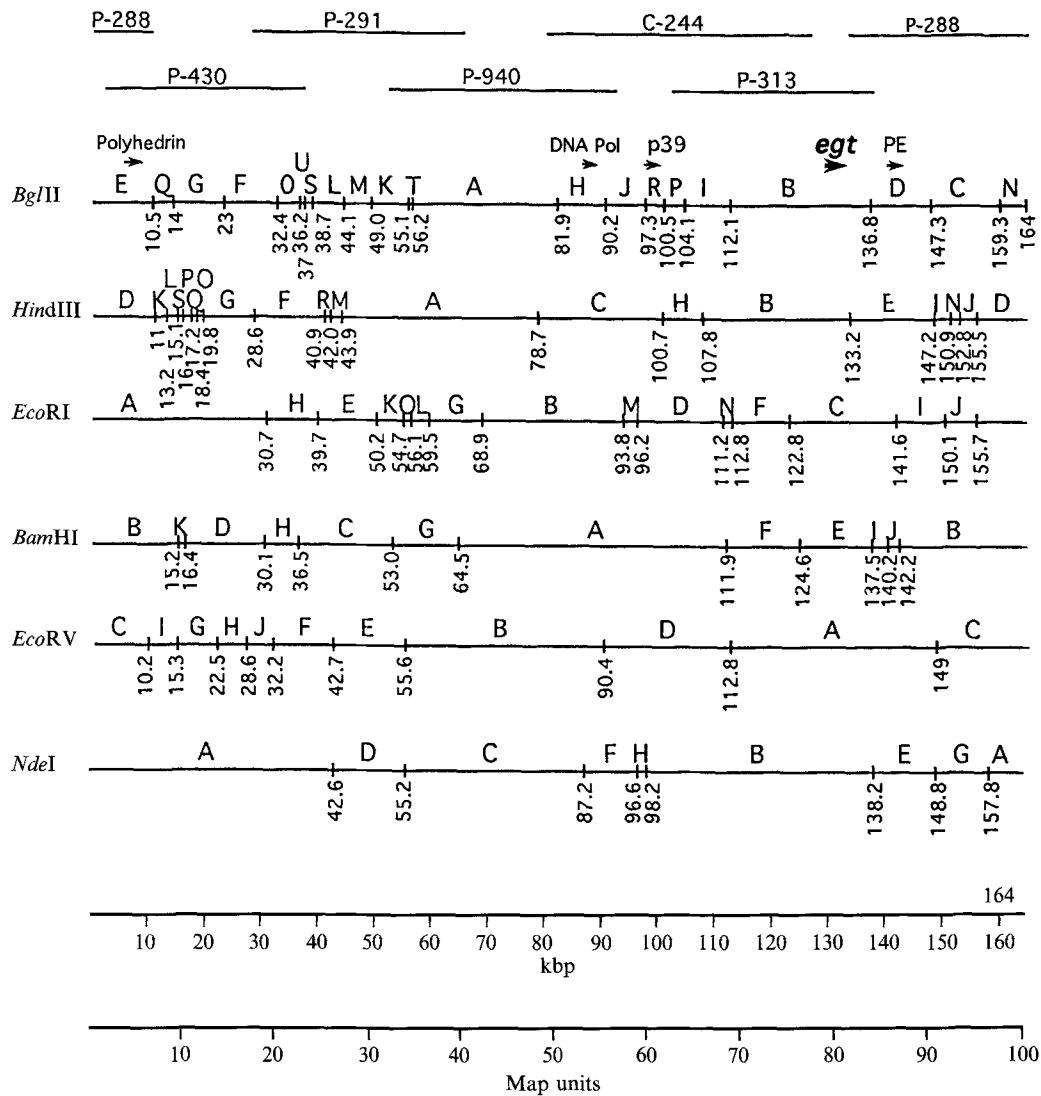
Fig. 1. A restriction map of the LdMNPV isolate 5-6. Shown above the map are the overlapping cosmid clones used to generate the map. Below the map are scales in both kilobase pairs and map units. The location of the polyhedrin gene is indicated on the map, which is oriented according to Vlak & Smith (1982). The *egt* gene is located between 124·6 and 133·2 kbp (76·0 and 81·2 m.u.). Also shown are the locations of the DNA polymerase gene (DNA Pol), the p39-capsid gene (p39), and the polyhedral envelope gene (PE).

1750 V until the xylene cyanol FF dye had migrated three-quarters of the way down the gel (IBI STS 45 gel rig, wedge spacers). The gel was dried and exposed to Kodak XAR film in cassettes containing Kodak X-Omatic intensifying screens. The primer extension reaction was compared to the sequencing ladder generated with the same primer to determine the initial base of the RNA.

Results

Mapping the egt gene

LdMNPV isolate 5-6 was obtained from Gypchek, and plaque-purified four times prior to cloning in the cosmid clone pHC79 (Hohn & Collins, 1980) after partial digestion with *Pst*I or *Cla*I. Six overlapping cosmid clones were isolated that cover the entire genome, and they were used to generate restriction endonuclease maps

for the enzymes *Bgl*II, *Hin*dIII, *Eco*RI, *Bam*HI and *Nde*I (Fig. 1). The genome of isolate 5-6 was found to be 164 kbp in length, which is similar to the length of LdMNPV isolate g (Smith *et al.*, 1988). Comparison of the restriction maps of LdMNPV isolates 5-6 and g indicates that the locations of the *Bgl*II and *Eco*RV sites are essentially identical. Isolate 5-6 was found to contain an additional *Hin*dIII, *Bam*HI and *Eco*RI site at 9·2, 32·3 and 91·5 map units (m.u.) respectively, and lacked a *Hin*dIII and *Bam*HI site, at 46·5 and 75·1 m.u., respectively, in comparison to isolate g. To identify an LdMNPV *egt* homologue, genomic DNA of isolate 5-6 was digested separately with *Bam*HI, *Eco*RI, *Eco*RV and *Hin*dIII; the resulting fragments were separated by agarose gel electrophoresis and transferred to nitro-
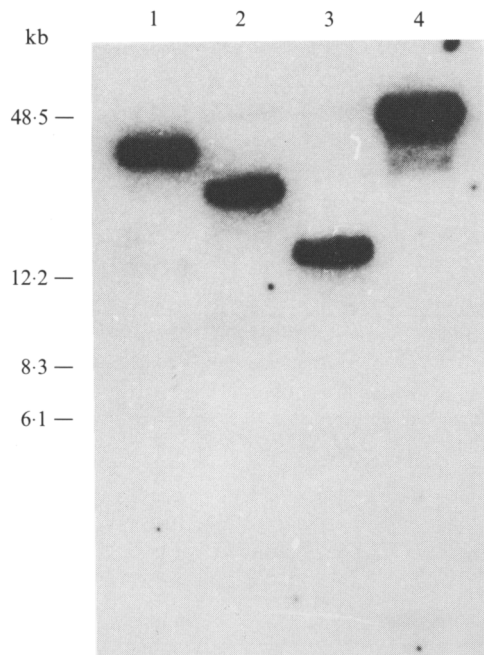
Fig. 2. Identification and genomic mapping of the LdMNPV *egt* gene. LdMNPV isolate 5-6 was digested with *Hin*dIII, lane 1; *Eco*RI, lane 2; *Bam*HI, lane 3; *Eco*RV, lane 4. The fragments were separated by agarose gel electrophoresis, transferred to nitrocellulose and hybridized with a ^{32}P-labelled AcMNPV clone of *egt*. The positions of DNA markers are indicated.

cellulose membranes. The blots were probed with the ^{32}P-labelled AcMNPV *egt* clone pBsBCEP (which contains only *egt* coding sequences located from 8·65 to 9·6 m.u. on the AcMNPV genome), under low-stringency hybridization conditions (50% formamide, 5 × SSC, 31 °C). Hybridization was detected to the *Hin*dIII B fragment, the *Eco*RI C fragment, the *Bam*HI E fragment and the *Eco*RV A fragment with the AcMNPV *egt* gene (Fig. 2). Hybridization to these fragments localized a putative LdMNPV *egt* homologue to the region of 124·6 to 133·2 kbp (76·0 to 81·2 m.u., Fig. 1). The fragment between the *Bam*HI site at 124·6 kbp and the *Hin*dIII site at 133·2 kbp was cloned into pUC18 to generate pEGT8·6. A 1·5 kbp *Pst*I subfragment of this clone was shown to hybridize to pBsBCEP (data not shown). Since isolate 5-6 is a few-polyhedra mutant (Slavicek *et al.*, 1992), and future studies on the efficacy of an *egt* minus virus are planned, the location of the *egt* gene was mapped in the LdMNPV isolate A21-2 (a plaque-purified line derived from A2-1, a wild-type isolate). The restriction endonuclease maps of isolates 5-6 and A21-2 are nearly identical, and are the same in the genomic region containing the *egt* gene. We localized the *egt* gene to the region between the *Bam*HI and *Hin*dIII sites at 124·6 kbp (76 m.u.) and 133·2 kbp (81·2 m.u.) respectively, in strain A21-2 by probing with the biotinylated 1·5 kbp *Pst*I subfragment from pEGT8·6 (see above). A

4·9 kbp fragment located between an *Nhe*I site at 128·3 kbp (78·2 m.u.) and the *Hin*dIII site at 133·2 kbp was shown to contain the *egt* gene by hybridization analysis as described above. This fragment was cloned into pUC18 in the *Xba*I and *Hin*dIII sites of the multiple cloning site to generate pEGT4·9.

Nucleotide sequence of the egt gene

The 4·9 kb clone was mapped with restriction enzymes, and the LdMNPV *egt* gene was further localized to the region from 129·7 to 131·6 kbp by hybridization (Fig. 3a). Sequence analysis of this region revealed an open reading frame (ORF) of 1464 nucleotides which could code for a protein of 488 amino acids (Fig. 3b) which exhibits a 42% amino acid identity with the AcMNPV *egt* gene product. An additional 21% of the amino acids, although not identical, are functionally similar. Comparing the locations of the restriction sites found by sequencing within the coding region to the restriction map generated above (data not shown) revealed that the *egt* gene is transcribed in the forward (clockwise) direction with respect to the orientation of the circular LdMNPV genome. The complete nucleotide sequence of the coding region of the *egt* gene is shown in Fig. 3(c). The coding region starts with nucleotide 1. A putative TATA box was found starting at nucleotide −65. We also found a tandem pair of potential polyadenylation sites starting at nucleotides 1530 and 1534. These sites are 65 and 69 nucleotides downstream of the translation termination codon respectively. The predicted protein sequence was also determined. It contains 488 amino acids and is shown in Fig. 3(c). A potential signal protein cleavage site is located between amino acids 18 and 19. The Ala at position −3 and Ser at position −1 (with respect to the cleavage site) are common in signal peptides. The sequence upstream of the cleavage site shows a high probability of being a signal peptide based on the prediction method of von Heijne (1986). An additional open reading frame (ORF B) of 840 nucleotides was identified in the −3 reading frame (Fig. 3b), that could code for a protein of 280 amino acids. The putative ATG codon of ORF B is in a poor context (CAGATGC; Kozak, 1986). The area around the putative ATG codon reveals neither a TATA box nor the ATAAG consensus late baculovirus promoter (Rankin *et al.*, 1988). ORF B did not exhibit significant homology with sequences in the GenBank database.

RNA mapping

The 5′ end of the *egt* transcript was mapped by the primer extension method. An 18 nucleotide antisense primer was generated that covers the region from
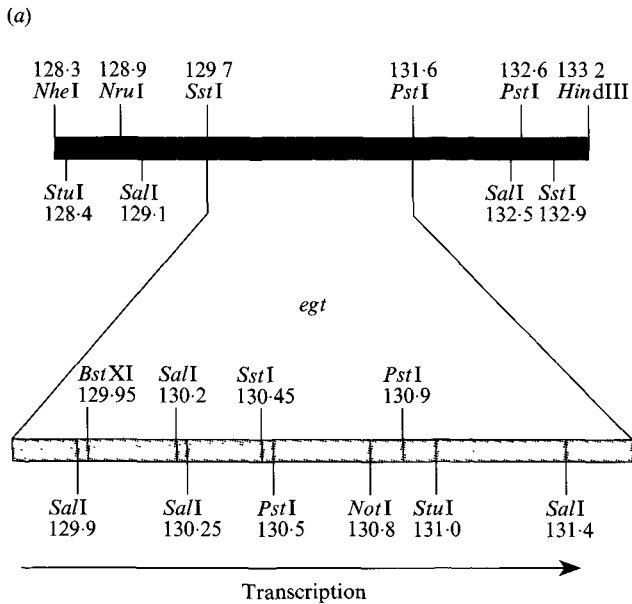
(a)

Fig. 3. Restriction map of the LdMNPV 128·3 to 133·2 kbp region, and ORF distribution and nucleotide sequence from 129·7 to 131·4 kbp. (a) Restriction map of the *egt* gene region of the LdMNPV isolate A21-2 between 78 and 82 m.u. An enlargement of a segment of this region shows the actual *egt* gene. The direction of transcription is shown below. (b) ORF map of the *egt* gene region derived from the sequence. Stop codons are indicated by vertical lines in each of the six reading frames. (c) Nucleotide sequence of the *egt* gene. The derived amino acid sequence is shown below the nucleotide sequence. Transcriptional control regions (a TATA box and tandem polyadenylation sites) are underlined. The translation start and stop codons are also underlined. The transcription initiation site is marked with an asterisk.

(b)
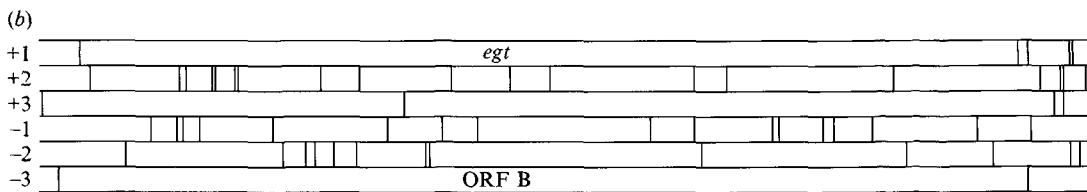
nucleotides 98 to 115 in the coding sequence (Fig. 3c). It was annealed to total RNA isolated from *L. dispar* 652Y cells after infection with LdMNPV isolate A21-2, and extended with reverse transcriptase. The products were electrophoresed alongside a sequencing ladder generated using the same primer (Fig. 4). We found the 5′ end of the *egt* transcript to be located at nucleotide −53. This site is 12 nucleotides downstream of the putative TATA box.

Temporal expression

Fig. 5(a) shows the temporal expression pattern of the *egt* gene. Total cellular RNA from A21-2-infected Ld652Y cells was generated at 0, 8, 12, 16, 24, 48 and 72 h p.i. Messenger RNA for *egt* was detected using a probe generated by nick translation of the 0·7 kbp *Pst*I fragment between 130·9 and 131·6 kbp. This probe is homologous to the 3′ end of the coding sequence for *egt*. The RNA found was approximately 1700 nucleotides long. We first detected the RNA at 12 h p.i. Maximal expression occurred at approximately 16 h p.i. after which time the expression decreased until it had totally disappeared by 72 h p.i. Only a single RNA band was detected at each time point and all of these were the same size.

Fig. 5(b) shows the expression of the *egt* gene at 16 h p.i. in the presence of the protein synthesis inhibitor cycloheximide, and in the presence of the DNA replication inhibitor aphidicolin. The *egt* gene is expressed in the presence of aphidicolin but not in the presence of cycloheximide and may therefore be classified as a delayed early gene.

Discussion

The LdMNPV *egt* gene maps between 78·7 and 79·6 m.u. (Fig. 2). In AcMNPV the gene maps at 9·0 m.u. (O'Reilly & Miller, 1990). The different genomic locations of the *egt* genes in LdMNPV and AcMNPV suggest that gene rearrangement has occurred during the evolution of these virus species and that the genomes of AcMNPV and LdMNPV are not collinear in gene arrangement. In contrast, the spatial arrangement of the DNA polymerase, p39 capsid, and polyhedral envelope genes is conserved between LdMNPV and AcMNPV (Bjornson & Rohrmann, 1992a, b; Bjornson et al., 1992). However, the CG30 (Thiem & Miller, 1989) and SLP genes (Wu & Miller, 1989) that are located downstream of the p39 capsid and the DNA polymerase genes, respectively, in AcMNPV do not appear to be similarly located in LdMNPV.

(c)

```
  -99  GTCCGCGGCGACCCATCTCCGCGAGCGCGTGGCCTATATAAATATGGGACGCGCGGTCGGCCTTCACTCGTCGAC
```
```
                                    1
  -24  GATTCGACTCGCCGGAACGACACCATGACCGCATATTTGATTGTTTTTTGTTTGTGCTGCTGGTCGGCGGCGCGC
                                   M  T  A  Y  L  I  V  F  C  L  C  C  W  S  A  A  R
```
```
   52  TCCGCCAATATCCTGGCCTATTTTCCCACCCCCTCGTACAGCCACCAATTGGTGTTCAGGGCGTACGTGGAACTG
         S  A  N  I  L  A  Y  F  P  T  P  S  Y  S  H  Q  L  V  F  R  A  Y  V  E  L
```
```
  127  TTGGCGGAGCGGGGCCACGCCGTCACGGTGATCAGGCCGCTGACGCGCGTCGATTTCAATCGCAACGCCGGCAAC
         L  A  E  R  G  H  A  V  T  V  I  R  P  L  T  R  V  D  F  N  R  N  A  G  N
```
```
  202  TTGACCACGATAGATTTGGACGGAGACGGCCTGCTGCTTTTGATGAAGGCGTCGACGACGCATCGCAAGAGAGGC
         L  T  T  I  D  L  D  G  D  G  L  L  L  L  M  K  A  S  T  T  H  R  K  R  G
```
```
  277  ATCGTCGCGGACACGGACACGGTCACCGCGGACAACTACGAAGCGCTGGTGCGCATGGTCGACCGCCAGATTCAC
         I  V  A  D  T  D  T  V  T  A  D  N  Y  E  A  L  V  R  M  V  D  R  Q  I  H
```
```
  352  TCGGAGCCCTTTCAACGCCACCTGAAGAGCGCCCGTCGCGGCTACGACCTGCTGGTGGTCGAGGCGTTCGTGGAC
         S  E  P  F  Q  R  H  L  K  S  A  R  R  G  Y  D  L  L  V  V  E  A  F  V  D
```
```
  427  TACGCGCTGATCGCGTCGCACCTGTTCGGAGACGTGCCCGTCGTGCAGATCAGCTCGGGCCACGCGACCGCCGAG
         Y  A  L  I  A  S  H  L  F  G  D  V  P  V  V  Q  I  S  S  G  H  A  T  A  E
```
```
  502  AACTTTGAGACGATGGGCGCGACGAGTCGCCACCCCCGCTACTATCCCAACCTGTGGCGCTTCAATTTCGGCCCG
         N  F  E  T  M  G  A  T  S  R  H  P  R  Y  Y  P  N  L  W  R  F  N  F  G  P
```
```
  577  CTGAGCGTGTGGGACGGCGTGCGCGAGCTGTACACCGAGCTCAGGCTGCAGCGCGAGTTCGGCCTGCTGGCGGAT
         L  S  V  W  D  G  V  R  E  L  Y  T  E  L  R  L  Q  R  E  F  G  L  L  A  D
```
```
  652  CGACAGGACGCGTTGCTGAAGCGCCGCTTCGGCCCCGGAGGCGCCGGGGCTGCGCGAGTTGCGCTCGCGCGTCCGC
         R  Q  D  A  L  L  K  R  R  F  G  P  E  A  P  G  L  R  E  L  R  S  R  V  R
```
```
  727  CTGCTGTTTGTGAACGTGCACTCCGTCTTCGACAACAACCGGCCGGTGCCGCCGAGCGTGCAGTATTTGGGCGGT
         L  L  F  V  N  V  H  S  V  F  D  N  N  R  P  V  P  P  S  V  Q  Y  L  G  G
```
```
  802  CTGCATCTGCACGATCGACGCGCGGAGCCCCTGTCGGAGGCCGTGGCCCGATTCTTGGACGAGTCGCGCCGCGGA
         L  H  L  H  D  R  R  A  E  P  L  S  E  A  V  A  R  F  L  D  E  S  R  R  G
```
```
  877  GTCGTGTACGTCAGCTTCGGCTCGGGCCTCGCCACGGAGGACATGGACGCGGACATGGCGGCCGCGCTGCTCGAC
         V  V  Y  V  S  F  G  S  G  L  A  T  E  D  M  D  A  D  M  A  A  A  L  L  D
```
```
  952  GCGTTCAAGATGATGCCCTACGACGTGCTGTGGAAGCACGACGGGCGCGTGGACGGCTTGACCATTCCCGCCAAC
         A  F  K  M  M  P  Y  D  V  L  W  K  H  D  G  R  V  D  G  L  T  I  P  A  N
```
```
 1027  GTGTTCGTCCAGAAATGGTTCGCCCAGTTCGAGGTGCTGCAGCACAAAAACGTCAAGGCCTTCGTCACGCAGGCG
         V  F  V  Q  K  W  F  A  Q  F  E  V  L  Q  H  K  N  V  K  A  F  V  T  Q  A
```
```
 1102  GGCGTGCAATCGACGGACGAGGCGGTCGAGAATCTGGTGCCGCTCGTGGGCGTGCCGCTCATGGGAGACCAGGCG
         G  V  Q  S  T  D  E  A  V  E  N  L  V  P  L  V  G  V  P  L  M  G  D  Q  A
```
```
 1177  TTCAACGCGCACCGATACGTCGAGCTGGGCATCGGCGTCGCCCTGGACGCGACGCGCCTCACCGCGGCCGACCTG
         F  N  A  H  R  Y  V  E  L  G  I  G  V  A  L  D  A  T  R  L  T  A  A  D  L
```
```
 1252  GCGCGCGCCGTCGAGCAGGTGACCTCGGACCGCGCCTATCGAGAGAACCTGGAGCGCTTGCGCCGTCTGTTGCGA
         A  R  A  V  E  Q  V  T  S  D  R  A  Y  R  E  N  L  E  R  L  R  R  L  L  R
```
```
 1327  CACCAGTGCGCGAGCCCGACGCACAAGGCGGTTTGGTACACGGAGCACGCGCTGCGCCGCGACGGCGACGCCCTC
         H  Q  C  A  S  P  T  H  K  A  V  W  Y  T  E  H  A  L  R  R  D  G  D  A  L
```
```
 1402  AAAACCAAAGCGGCCAACGTGGACTACGCCGAGTATTGCATGTCGACTTGTTGGCGCCCCTGCTGAGCGTGTCGC
         K  T  K  A  A  N  V  D  Y  A  E  Y  C  M  S  T  C  W  R  P  C  end
```
```
 1477  TCATGAGTCATCTGCACTCGCTGATTCGAATGTTCGTTTGGTGATTGCGCAATAATAAATAAATTTTAATGAACC
```
```
 1552  GAATTTTATTTTCCGCTGACGCAATGCCGCCCACCGACC  1590
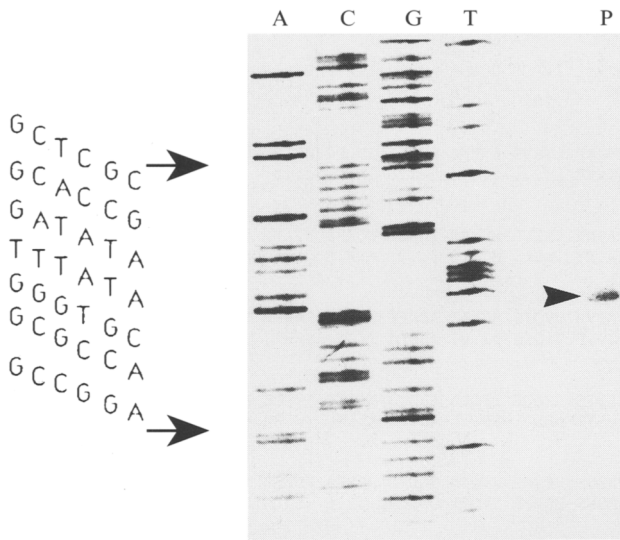```

Fig. 3(c). For legend see p. 833.

Fig. 4. Mapping the 5′ end of the LdMNPV *egt* RNA transcript. On the right of the sequencing gel is a primer extension reaction (P) carried out on total RNA, using a specific internal primer from the *egt* sequence, with reverse transcriptase; the product is marked with an arrowhead. On the left is a sequencing ladder generated using the same primer. The nucleotide sequence of part of this region is shown on the left.
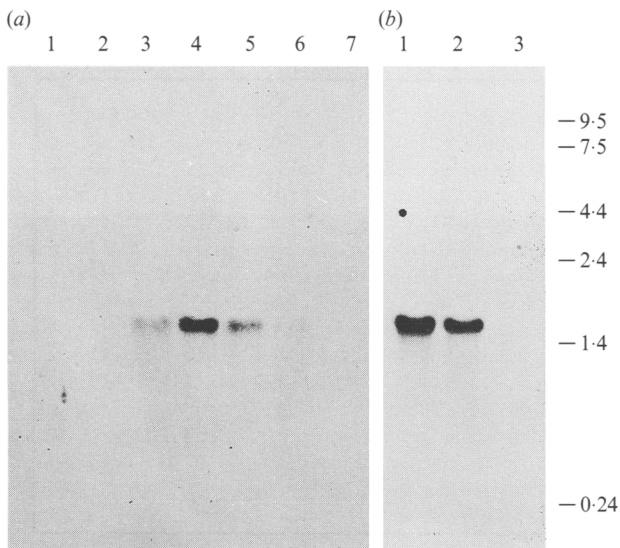


Fig. 5. Northern blot hybridization analysis of LdMNPV *egt* RNA. (*a*) Temporal expression pattern of the *egt* gene. Total cellular RNA was harvested at various times p.i., 0 h, lane 1; 8 h, lane 2; 12 h, lane 3; 16 h, lane 4; 24 h, lane 5; 48 h, lane 6; 72 h, lane 7. This was electrophoresed, blotted and probed with an LdMNPV *egt* DNA probe. On the right are RNA size standards. (*b*) Results of *egt* expression in the presence of aphidicolin (lane 2) and cycloheximide (lane 3) at 16 h p.i. Lane 1 contains RNA from virus-infected cells that were not treated with aphidicolin or cycloheximide.

The LdMNPV *egt* RNA is expressed as a single RNA (Fig. 5*a*). In contrast, the AcMNPV *egt* RNA is expressed as two RNAs with a common 5′ terminus (O'Reilly & Miller, 1990). LdMNPV *egt* RNA tran-

scription is likely to be controlled by a TATA box 12 nucleotides upstream of the transcription start site. Analysis of the 34 nucleotides upstream of the TATA box failed to reveal a GTGT motif similar to that reported by O'Reilly & Miller (1990) to precede (starting 15 bp upstream) the TATA box in AcMNPV. LdMNPV *egt* RNA was first detected at 12 h p.i. and was present until 48 h p.i. (Fig. 5*a*). Maximal synthesis occurred at approximately 16 h p.i. In contrast, AcMNPV *egt* is expressed by 3 h p.i. and decreases in expression until 12 h p.i. By 24 h p.i. the transcripts are no longer detectable (O'Reilly & Miller, 1990). The *egt* gene was expressed in the presence of aphidicolin but not in the presence of cycloheximide (Fig. 5*b*). Consequently, *egt* expression requires viral protein synthesis but not viral DNA replication and *egt* is therefore considered a delayed early gene. In contrast, AcMNPV *egt* is an immediate early gene (O'Reilly & Miller, 1990). One possible reason for this difference is that gypsy moth larvae take longer to develop (30 or more days; Hough & Pimentel, 1978) and therefore have longer periods of time between their moults (about 6 days) than do the *Spodoptera frugiperda* larvae infected with AcMNPV (about 3 days between moults; O'Reilly & Miller, 1991). This difference would lessen the pressure to maintain *egt* as an immediate early gene, as the virus would not need to interfere with the larval moult quite as rapidly in the gypsy moth. Tandem polyadenylation sites (AATAAA-TAAA) were found starting 63 and 67 nucleotides downstream of the end of the translation termination codon.

The LdMNPV *egt* sequence encodes a polypeptide with a predicted $M_r$ of 55000 and a predicted isoelectric point of 7·6. The LdMNPV *egt* gene exhibits a G+C value of 65% which is slightly higher than the overall G+C composition of the entire genome (60%). The AcMNPV *egt* ORF encodes a polypeptide of $M_r$ 57000, with a predicted isoelectric point of 9·8 (O'Reilly & Miller, 1990). A variety of other UDP-glucuronosyltransferases have subunit $M_r$s in the 50K to 60K range (Burchell & Coughtrie, 1989). The LdMNPV *egt* gene shows a 49% nucleotide identity to the sequence of the AcMNPV *egt* gene (O'Reilly & Miller, 1990) within the coding region. Fig. 6(*a*) shows the aligned regions of amino acid identity between the two egt proteins. There is a 42% amino acid identity between these two proteins. In addition, another 21% of the amino acids are functionally similar to the corresponding amino acid in the other protein.

A search of the GenBank sequence database using the algorithm of Wilbur & Lipman (1983, IBI MacVector version 3.04) reveals that the LdMNPV gene has its most significant homology with the AcMNPV *egt* gene. The regions exhibiting the greatest homology were from

amino acids 19 to 52 (65% amino acid identity), 89 to 105 (82%), 133 to 190 (59%), 252 to 270 (95%) and 348 to 391 (59%). The next fourteen most homologous genes are all UDP-glucuronosyltransferases, with optimized homology scores ranging from one-sixth to one-half that given for the AcMNPV gene. Those with the greatest homology include human 3,4-catechol oestrogen UDP-glucuronosyltransferase (Ritter *et al.*, 1990), rat bilirubin UDP-glucuronosyltransferase (Sato *et al.*, 1990) and murine UDP-glucuronosyltransferase (Kimura & Owens, 1987) which exhibit 24, 23 and 23% amino acid identity respectively with LdMNPV *egt*. Fig. 6(*b*) shows the protein sequences of AcMNPV *egt* and these three mammalian UDP-glucuronosyltransferases aligned with the LdMNPV *egt* sequence (by the MacVector program). Amino acids identical in all five proteins are shown boxed. Interesting areas of homology from these genes are shown in Fig. 6(*c*). The regions between amino acids 19 and 52 (inclusive, on the LdMNPV egt protein), 252 and 270, and 348 and 391 are highly conserved in all five proteins. These regions may be involved in some function common to all UDP-glucuronosyltransferases (e.g. they may be involved in the active site). One further area of interesting homology is the conserved PxxP at amino acids 257 to 260. These prolines, along with the third proline found in the region at 259 in the baculovirus proteins or 255 in the mammalian proteins, may be involved in the conformation of the active protein. Interestingly, the regions between amino acids 89 to 105 and 155 to 195 are very highly conserved (82% and 59% identical respectively) between the two baculoviruses, yet few or no amino acids are conserved in these regions between all five proteins. There is also no apparent homology in the region from 89 to 105 amino acids amongst the three mammalian proteins. These regions may be involved in something specific to the baculovirus proteins (e.g. substrate recognition). Conversely, the region between amino acids 271 and 347 exhibits lower identity than the overall AcMNPV/LdMNPV identity, while in this region the mammalian proteins exhibit a rather high degree of identity with each other. This may indicate that this region is important to the function of the mammalian proteins, but not necessary for the baculovirus proteins. Another area of conservation in the mammalian proteins is between amino acid residues 484 and 500 (with respect to the LdMNPV coding sequence), i.e. after the end of the LdMNPV protein. Both Ritter *et al.* (1990) and Kimura & Owens (1987) point out this region as a putative membrane-spanning region. In contrast to the membrane-bound mammalian proteins, examination of the amino terminus of LdMNPV egt reveals a potential secretion signal sequence and cleavage site between amino acids 18 and 19. Cleavage at this site would cause the amino terminus of

(*a*)

```
MTAYL IVFCL CCWSA ARSAN ILAYF PTPSY SHQLV FRAYV ELLAE RGHAV
MT... .1..L ...tA ..aAN ILA.F PTPaY SH.iV yk.Yi E.LAE k.H.V


TVIRP LTRVD FNRNA GNLTT IDLDG DGLLL LMKAS TTHRK RGIVA DTDTV
TVvkP^L.... ..... GNiT. In.D. ..... ...A^Sa..RK RGvVs DTDTV


TADNY EALVR MVDRQ IHSEP FQRHL KSARR GYDLL VVEAF VDYAL IASHL
TA.NY .gLi. M...Q ..... -.R.L .a... tfDLv VVEAF .DYAL v.gHL


FGDVP VVQIS SGHAT AENFE TMGAT SRHPR YYPNL WRFNF GPLSY WDGVR
y...P ViQIa pG.g. AENFd T.GA. aRHP. ..PNi WR.NF ...-- ----.


ELYTE LRLQR EFGLL ADRQD ALLKR RFGPE APGLR ELRSR VRLLF VNVHS
nv.TE mRL.. EF.iL An..n ALLK. .FGPn tPti. .LR.k V.LL. 1NlHp


VFDNN RPVPP SVQYL GGLHL HDRRA EPLSE AVARF LDESR RGVVY VSFGS
iFDNN RPVPP SVQYL GG^iHL....p .t^LS..i... mn.Sk .G.iY VSFGS


GLATE DMDAD MAAAL LDAFK MMPYD VLWKH DGRVD GLTIP ANVFV QKWFA
si.T. ....e ....L intFK^l..Y. iLWK.^D..V. .iTlP ANV.. Q.WF.


QFEVL QHKNV KAFVT QAGVQ STDEA VENLV PLVGV PLMGD QAFNA HRYVE
Q..VL QHK.m .AFiT QgGlQ SsDEA lE..i PmV.l PmMGD Q.y.A Hk..q


LGIGV ALDAT RLTAA DLARA VEQVT SDRAY RENLE RLRRL LRHQC ASPTH
LGva. ALDt. .vss. qL..A indV. .n^.tYk..m. .L..L i.Hd. At.p.^^


KAVWY TEHAL RRDGD ALKTK AANVD YAEYC MSTCW RPC
KAi.f TEr.i R.... s.... s.... .An.. .S..y ...^^ ^^^^^ ^^^
```

Fig. 6. Comparison of the LdMNPV *egt* protein sequence to the AcMNPV *egt* protein sequence and mammalian UDP-glucuronosyltransferases. (*a*) Protein sequence comparison between the protein encoded by the LdMNPV *egt* gene and the AcMNPV *egt* preprotein. The complete LdMNPV amino acid sequence is shown. Beneath the LdMNPV sequence is shown the AcMNPV sequence. Identical amino acids are shown as upper case letters. Functionally similar amino acids are shown as lower case letters. Dots (.) indicate dissimilar amino acids. Carets (^) indicate inserted amino acids. Dashes (-) indicate missing amino acids. (*b*) Aligned amino acid sequences of LdMNPV *egt* protein (L), AcMNPV *egt* (A), rat bilirubin UDP-glucuronosyltransferase (R), murine UDP-glucuronosyltransferase (M) and human 3,4-catechol oestrogen UDP-glucuronosyltransferase (H). The alignment was carried out by the MacVector program using the method of Wilbur & Lipman (1983). The numbers on the right of the figure refer to the amino acid position within the LdMNPV *egt* protein. Boxed amino acids are conserved between all five proteins. (*c*) Map of the homologous regions of the five proteins above. The numbers at the top of the figure refer to the amino acid position with respect to the LdMNPV *egt* protein. To the left are the two proteins compared in each bar. To the right is the overall amino acid identity between the two proteins. Beneath each region the percentage amino acid identity between the protein in question and LdMNPV egt protein is shown. In the top bar areas which have an identity 10% or more greater than the overall identity of 42% are darkened. The next three bars compare LdMNPV *egt* with the three mammalian proteins from (*a*). Regions that showed significant identity in the AcMNPV/LdMNPV comparison and which have retained high identity are darkened. The bottom bar is a compilation showing the percentage of absolute identity between all five proteins within each region. Those regions which have an identity 10% or more greater than the overall identity of 8% are darkened.

(b)

```
L            MTAYLIVFCL CCWSAARSAN ILAYFPTPSY SHQLVFRAYV ELLAERGHAV        50
A            MTILCWLALL STLTAVNAAN ILAVFPTPAY SHHIVYKVYI EALAEKCHNV
R  MGLCAPLRGLSGLLLLLCA LPW AE GGK VL VFPMEG  SHWLSMRDVV RELHARGHQA
M     MPGKWISALLLLQIS CCFRSVKCGK VL VWPM EF  SHWMNIKIIL DELVQRGHEV
H.    MSVKWTSVILLIQLSF  CFSSGNCGK VL VWA AEY SHWMNIKTIL DELIQRGHEV

L  TVIRP LT  RVDFNRNA G NLTT I  DL    DGDGL     LLLMK A STTHRKR  GIVAD TDTV 100
A  TVVKPKLF  AYSTKTYC G NITE I  NA    DMSVE     QYKKLVANSAMFRKR  GVVSD TDTV
R  VVLAP EV  TVHMKGEDFF TLQT Y  AFPYTKEEYQR     EILGN A KKGFEPQ  HFVKTFFETM
M  TVLRP SAYYVLDPKKSP GLKFET FPTSV     SKDNLENFFIKFVD V WTYEMPR  DTCLS YSPL
H  TVLAS SA  SILFDPNN S SALK IEIYP     TS  LTK  TELEN F IMQQIKRWSDLPKDTFWL

L    TA  DNYEALVR M   VDRQIHSEP F  QRHL   KSARR G YDLLVVEAF VDYALIASHL      150
A    TA  ANYLGLIE M   FKDQFDNIN    VRNL   IANNQ T FDLVVVEAF ADYALVFGHL
R   ASIKKFFDLYAN S   CAALLHNKT L  IQQL   NSS   S FDVVLTDPV FPCGALLAK
M   LQ  NMIDEFSD Y   FLSLCKDVV S  NKELMTKLQES K FDVLLSDPV ASCGELIAEL
H   YF  SQVQEIMS  IFGDITRKFCKDV VSNKKFM  KKVQE SRFDVIFADAI FPCSELLAEL

L  FGDVPVVQIS SGHATAENFE T MGATSRHPR YYPNLWRFNF GPLSVWDGVR              200
A  YDPAPVIQIA PGYGLAENFD T VGAVARHPV HHPNIWRSNF     DDTEA
R  YLQIPAVFFL RSVPCGIDYE A  TQCPKPSS YIPNLLTMLS DHMTFLQRVK
M  L QIPFLYSI RFSPGYQIEK S SGRFLLPPS YVPVILSGLG GQMTFIERIK
H  F NIPFV YS LSFSPGYTFE KHSGGFIFPPS YVBVVMSELT DQMTFMERVK

L  ELYTELRLQR EFGLLADRQ D ALLKRRFGPE APGLRELRSR VRLLFVNVHS              250
A  NVMTEMRLYK EFKILANMS N ALLKQQFGPN TPTIEKLRNK VQLLLLNLHP
R  NMLYPLTLKY ICHLSITPY E SLASELLQRE M SLVEVLSH ASVWLFRGDF
M  NMICMLYFDF WFQMFNDKKWD SFYSEYLGR  PTTLVETMGQ AEMWLIRSNW
H  NMIYVLYFDF WFEIFDMKKWD QFYSEVLGR  PTTLSETMGK ADVWLIRNSW

L  VFDNNRPVPP SVQYLGG LHL HDRRAEP ISE AVARFLDES R RGVVYVSFGS           300
A  IFDNNRPVPP SVQYLGGGIHL VKSAPLTKLSP VINAQMNKS K SGTIYVSFGS
R  VFDYPRPIMP NMVFIGG IN  CVIK KP LSQ EFEAYVNASGE HGIVVFSLGS
M  DLEFPHPTLP NVDYVGG LH  C KPAKP IPK DMEEFVQSSGD HGVVVFSLGS
H  NFQFPHPLLP NVDFVGG LH  C KPAKP LPK EMEDFVQSSGE NGVVVFSLGS

L  GLATEDMDAD MAAALLDAFK  MMPYDVLWKH  DG RVDGLTIP ANVFVQKWFA           350
A  SIDTKSFANE FLYMLINTFK  TLDNYTILWKI DDE VVKNITLP ANVITQNWFN
R   MVSE IPEK KAMEIAEALG  RIPQTLLWRY  TGTRPSNLA K NTILV KWLP
M   MVS NMTEE KANAIAWALA  QIPQKVLWKF  DGKTPA TL G HNTRVYKWLP
H   MVS NMTEE RANVIASALA  QIPQKVLWRF  DGNKPDTLGLN TRLY  KWIP

L  QFEVLQHKNV KAFVIQAGVQ STDEAVENLV PLVGVPLMGD QAFNAHRYVE             400
A  QRAVLRHKKM AAFIITQGGLQ SSDEALEAGI PMVCLPMMGD QFYHAHKLQQ
R  QNDLLGHPKH RAFIITHSGSH GIYEGICNGV PMVMMPLFGD QMDNAKRMET
M  QNDLLGHPKT KAFVIHGGAN GVYEAIYHGI PMIGIPLFGE QHDNIAHMVA
H  QNDLLGHPKT RAFIIHGGAN GIYEAIYHGI PMVGIPLFAD QPDNIAHMKA

L  LGIGVALDAT RLTAADLARA VEQVTSD RAY RENLERIRRL LRHQCAS PTH           450
A  LGVARALDTV TVSSDQLLVA INDVLFNAPTY KKHMAELYAL INHDKATFPPLD
R  RGAGVTLNVL EMTADDLENA LKTVINN KSY KENIMRLSSL HKDRPIE PLD
M  KGAAVALNIR TMSKSDVLNA LEEVIEN PFY KKNAIWLSTI HHDQPMK PLD
H  RGAAVRVDFN TMSSTDLLNA LKRVIND PSY KENVMKLSRI QHDQPVK PLD

L  KAVWYTEHAL RRDGDA LKTK AANVDYAEYC MSTCWRPC                          488
A  KAIKFTERVI RYRHIS RQLY SLKTTAANVP YSNYYMYK SVFSIVMNHL THF
R  LAVFWVEYVM RHKGAPHLRPA AHDLTWYQYH SLDVIGFL LAIVLTVVFI VYKSCAYGCR
M  RAVFWIEFVM RHKRAKHLRPL GHNLTWYQYH SLDVIGFL LSCVATTIVL SV KCLLFIYR
H  RAVFWIEFVM RHKGAKHLRVA AHDLTWFQYH SLDVIGFL LVCVATVIFI VT KCCLFCFW

A
R  KCFGGKGRVKK SHKSKTH
M  FFVKKENKMKN
H  KFARKAKKGKN D
```
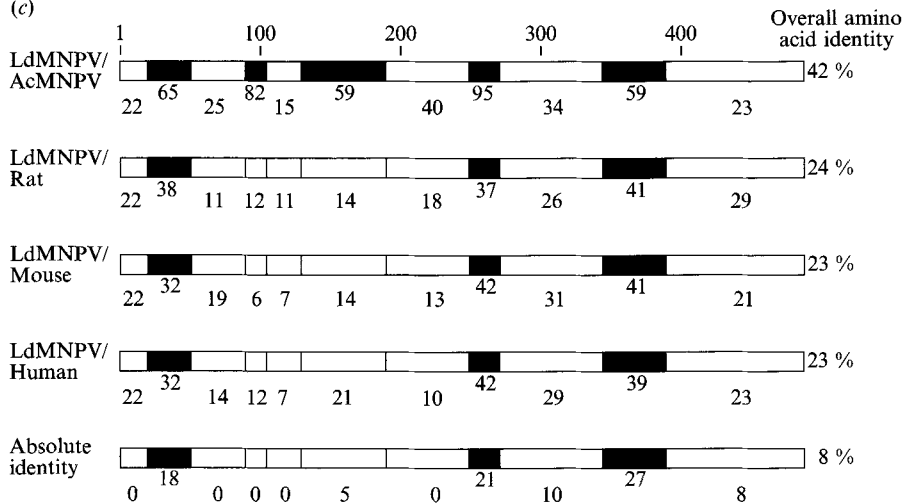
(c)

Fig. 6(*b, c*). For legend see opposite.

the mature protein to have a five amino acid stretch of identity with the amino terminus found by O'Reilly *et al.* (1992) on mature AcMNPV egt protein. A signal sequence is important for baculovirus egts because its proposed mode of action requires secretion into the haemolymph.

The cloning and characterization of the LdMNPV *egt* gene will allow insight into the biological and biochemical mechanisms of action of viral hormonal control. Gaining knowledge of the precise mechanism of action of this viral system may eventually allow its modification and use as part of a more environmentally benign biological control system, both for the gypsy moth and eventually for other important forest and agricultural pests which are vulnerable to baculoviruses. The role of the *egt* gene during viral infections can be assessed through generation of *egt* minus viral strains. In addition, sites within the egt protein necessary for function can be identified through site-directed mutagenesis targeted to the regions of greatest homology.

References

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

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.

BURAND, J. P. & PARK, E. J. (1992). Effect of nuclear polyhedrosis virus infection on the development and pupation of gypsy moth larvae. *Journal of Invertebrate Pathology* **60**, 171–175.

BURCHELL, B. & COUGHTRIE, M. W. H. (1989). UDP-glucuronosyl transferases. *Pharmacology and Therapeutics* **43**, 261–289.

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.

FRIESEN, P. D. & MILLER, L. K. (1985). Temporal regulation of baculovirus RNA: overlapping early and late transcripts. *Journal of Virology* **54**, 392–400.

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

HOHN, B. & COLLINS, J. (1980). A small cosmid for efficient cloning of large DNA fragments. *Gene* **11**, 291–298.

HOUGH, J. A. & PIMENTEL, D. (1978). Influence of host foliage on development, survival, and fecundity of the gypsy moth. *Environmental Entomology* **7**, 97–102.

KIMURA, T. & OWENS, I. S. (1987). Mouse UDP glucuronosyl transferase: cDNA and complete amino acid sequence and regulation. *European Journal of Biochemistry* **168**, 515–521.

KOZAK, M. (1986). Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292.

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

O'REILLY, D. R. & MILLER, L. K. (1989). A baculovirus blocks insect molting by producing ecdysteroid UDP-glucosyl transferase. *Science* **245**, 1110–1112.

O'REILLY, D. R. & MILLER, L. K. (1990). Regulation of expression of a baculovirus ecdysteroid UDP-glucosyl transferase gene. *Journal of Virology* **64**, 1321–1328.

O'REILLY, D. R. & MILLER, L. K. (1991). Improvement of a baculovirus pesticide by deletion of the *egt* gene. *Bio/Technology* **9**, 1086–1089.

O'REILLY, D. R., BROWN, M. R. & MILLER, L. K. (1992). Alteration of ecdysteroid metabolism due to baculovirus infection of the fall army worm *Spodoptera frugiperda*: host ecdysteroids are conjugated with galactose. *Insect Biochemistry and Molecular Biology* **22**, 313–320.

RANKIN, C., OOI, B. G. & MILLER, L. K. (1988). Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene* **70**, 39–49.

RITTER, J. K., SHEEN, Y. Y. & OWENS, I. S. (1990). Cloning and expression of human liver UDP-glucuronosyltransferase in COS-1 cells. *Journal of Biological Chemistry* **265**, 7900–7906.

SATO, H., KOIWAI, O., TANABE, K. & KASHIWAMATA, S. (1990). Isolation and sequencing of rat liver bilirubin UDP-glucuronosyltransferase cDNA: possible alternate splicing of a common primary transcript. *Biochemical and Biophysical Research Communications* **169**, 260–264.

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.

SLAVICEK, J. M., PODGEWAITE, J. D. & LANNER-HERRERA, C. (1992). Properties of two *Lymantria dispar* nuclear polyhedrosis virus isolates obtained from the microbial pesticide Gypchek. *Journal of Invertebrate Pathology* **59**, 142–148.

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

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.

VLAK, J. M. & SMITH, G. E. (1982). Orientation of the genome of *Autographa californica* nuclear polyhedrosis virus: a proposal. *Journal of Virology* **41**, 1118–1121.

VON HEIJNE, G. (1986). A new method for predicting signal sequence cleavage sites. *Nucleic Acids Research* **14**, 4683–4690.

WILBUR, W. J. & LIPMAN, D. J. (1983). Rapid similarity searches of nucleic acid and protein databanks. *Proceedings of the National Academy of Sciences, U.S.A.* **80**, 726–730.

WOOD, H. A. & GRANADOS, R. R. (1991). Genetically engineered baculoviruses as agents for pest control. *Annual Review of Microbiology* **45**, 69–87.

WU, J. & MILLER, L. K. (1989). Sequence, transcription and translation of a late gene of the *Autographa californica* nuclear polyhedrosis virus encoding a 34·8K polypeptide. *Journal of General Virology* **70**, 2449–2459.