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

†Present address: Swedish University of Agricultural Sciences, S-268 00 Svalov, Sweden.

The nucleotide sequence data reported in this paper will appear in the GenBank Sequence Database under accession number U04321.
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 molting 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 molting 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 molting 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 60 mM-glutamine (Gibco) at 27 °C. LdMNPV strains A21, A21-2 and B652Y cells. The cells were infected with LdMNPV isolate A21-2 at 10 TCID₅₀ 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 12% agarose gel containing formaldehyde by overnight electrophoresis at 30 V, and the gel was blotted. A 700 bp PstI fragment probe, located between 130 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 either 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 5X-Nick labelled probe which hybridizes to the egt coding sequence. The blot was exposed to film as above.

**Determination of 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 32P and used for the primer extension reaction with Moloney murine leukemia 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
Characterization of LdMNPV egt gene

Fig. 1. A restriction map of the LdMNPV isolate 5-6. Shown above the map are the overlapping cosmids 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).

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 PstI or ClaI. Six overlapping cosmids clones were isolated that cover the entire genome, and they were used to generate restriction endonuclease maps for the enzymes BgIII, HindIII, EcoRI, BamHI and NdeI (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 BgIII and EcoRV sites are essentially identical. Isolate 5-6 was found to contain an additional HindIII, BamHI and EcoRI site at 9.2, 32.3 and 91.5 map units (m.u.) respectively, and lacked a HindIII and BamHI 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 BamHI, EcoRI, EcoRV and HindIII; the resulting fragments were separated by agarose gel electrophoresis and transferred to nitro-
cellulose membranes. The blots were probed with the
\(^{32}P\)-labelled AcMNPV egt clone pBsBCEP (which con-
tains only egt coding sequences located from 8-65 to 9-6
m.u. on the AcMNPV genome), under low-stringen-
cy hybridization conditions (50 % formamide, 5 x SSC,
31 °C). Hybridization was detected to the HindIII B
fragment, the EcoRI C fragment, the BamHI E fragment
and the EcoRV 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 BamHI site at 124-6 kbp and the HindIII site
at 133-2 kbp was cloned into pUC18 to generate pEGT8-6. A 1-5 kbp PstI 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 BamHI and HindIII sites at
124-6 kbp (76 m.u.) and 133-2 kbp (81-2 m.u.) respect-
ively, in strain A21-2 by probing with the biotinylated
1-5 kbp PstI subfragment from pEGT8-6 (see above). A
4-9 kbp fragment located between an NheI site at
128-3 kbp (78-2 m.u.) and the HindIII 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 XbaI and HindIII sites of the multiple
cloning site to generate pEGT4-9.

Nucleotide sequence of the egt gene

The 4-9 kbp 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. Com-
paring 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
coids 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 nucleo-
tides 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
Characterization of LdMNPV egt gene

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 PstI 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 GTCCCGCGGCAACCATCTCGGGACCCGCGTGGCGCTACTCGTGCAC

-24 GATTGACTCAGGAGACCAGACCATGACCGCATATTTGTTGTTTGTGCTGCTGTTGCTGGGCGCCGGG

MTAYLIVFCLCCWSAAR

52 TCCGCAATATACCTGCGCTTACCCACCACCGCTGCTACAGCCCAATTGGGTTACGGTGAGCTGGAACCTG
SANILAYFPSPSYSHQLVFRAVYVEL

127 TTGGCGGAGGGCGCCACCGGCTACGATGACCGGCGTGATCCGACGACATGCAAGAGAGGG
LAERGHAVTVIPRLTRVDFNRRAN

202 TTAGCCAGCTAGTAGTTGAGCCGAGGCTGCTGTTTTGTAGAAAGGCTGAGACGACATGCAAGAGGG
LTTIDLDGDGLLMKASTTHRRAG

277 ATCTGTCGGAACCCGACCGATCAAGGATGCTGACCTGAGGCGCTGACGTGTTGATGCGAGGGCGCAAGGAG
IVADTDTRDEAVRELVMVRMDRQIH

352 TCAGAGCCCTTTCACGCCACAGAGGAGCCCGTTCGCGCTACGAGTCGTTCTGAGGCTTCGCCTACGATG
SEPFRHLKSAARRGYDLLLVVEAFVD

427 TACGCGTGCTACGACGATCGCTGACCTGCTGAGGAGGCTGCGCGCTGACCTGACGCTGCGACGACGAGCCAG
YALIAHFLFDTVQVQISSGHA

502 AACCTTGAAGCATGGGGGACGAGTGCCACCCGCACTATCCAAACCTGTGCCCTCAATTTCTGGCCCG
NFETMGATSRHPFYYPNLWRNFGP

577 CTAGGCGGCTGAGGCTGCGCTGACCTGACGATGCTGAGGGCTGAGGCTGCGTGGACGACGACGAGGGAG
LSVWDGVRELQLRERFGGLLAD

652 CGACGACGAGCTTTGATCGAGAGGCGCCCTTCGCGGAGGGAGGGGCGGCGTGCCGTAGCTGCGGCTCGCC
RQDALLPEAPFLRELRSVR

727 CTGCTGTTTTGTAAGCTGACTGATCCTGGCTTTGCAGAACAACCCGCGTGGCGGCGAGCTGACATTTTGCGCGT
LLFVNVHSVFDNNRPVPPSVQYL

802 CTGCAACTGACGATGCAGCGCCGAGGCGCCCTTGGCAGAGCCGCGCGCTCAATTTGGGAGGG
LHLLHDDRAPELSEAVARFLDESRRG

877 GTCTGCTGTGCTGATCGCTGCTGCTCGCCGACTCGAGGACGACGACGAGCCATGCACGTGCGCTGAC
VVYVSFGSGLATEDMDMADAAALLD

952 GGCTGACAGATGACTGCTGCGCTGCTGCTGCTGCTGCTGCTGCGCTGAGATTGACCGAGCCAGGG
AFKMMMPYDVLWKHDGRVDGLTIMAN

1027 GTGGTTGCTGAGAATAAATGGTTCGCGCGATGATGCTGAGACAAAAAACGTCAGAGCCTTCTGTACGAGGACGGC
VFVQKQKFQFVLFHFNKVAFVQTQA

1102 GCCGTGCAATCGACGCGGCAGGGCCTGACAGATGTCTGCGTCGCTGCTGCGCTGCTGCTGCTGAGGAGACCCAGG
GVQSTDEAVENLVPYGVPLMGDQA

1177 TTTAAGCCGCCAGCTAGCTGAGCTGCGAGGCTGCGCTGCGCTGAGGCTGCGCTGCGCTGAGGCACTCTG
FNARHRYVELIGIVALDATALATRADL

1252 GCCGGCCCGTCGAGACGCTGACTGCGCCGCTATCGAGAAGCTGAGACGCTGCTGCGCTGCGCTGCTGAGGCA
ARAVEQDREYNERLENPLRRLLR

1327 CACCCAGTGAGCCGCCAGCAGCACAAGCCGGTGGTTGTCAGACAGCGCTGCGCCGGGACCGCAGCCCGGCTTCC
HQCASPTHKAVWYTEHALRDRGDA

1402 AAAAAACCGCGCCAACCTTGAACCTGGCCAATATTGCTGATGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG
KTKAANVDYAEYCCMYSTMCTRCPend

1477 TCATGATCATCTGCACTGCGGATCTGAGATGTTCTGTTGTAGTCGAGAATTATGCAATATATATATAATTAATGAACC

1552 GAATTTTTTTTTCGCTGACGCAATGCGCCACCACTCCGAC 1590

Fig. 3(c). For legend see p. 833.
Characterization of LdMNPV egt gene

The LdMNPV egt RNA is expressed as a single RNA (Fig. 5a). In contrast, the AcMNPV egt RNA is expressed as two RNAs with a common 5' terminus (O'Reilly & Miller, 1990). LdMNPV egt RNA transcription 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. 5a). 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. 5b). 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 identical in all five proteins 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. (a) Alignment of amino acid sequences of LdMNPV egt and the five mammalian 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, 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 LdMNPV egt to be secreted, like the mammalian proteins. 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 LdMNPV egt to be secreted, like the mammalian proteins. TheLdMNPV 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 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 LdMNPV egt to be secreted, like the mammalian proteins. The LdMNPV 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 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 LdMNPV egt to be secreted, like the mammalian proteins.
**Characterization of LdMNPV egt gene**

(b) LdMNPV CWSASAAR 1 YFFTPSY 50 GLVFRAYV ELLAERGHV

MAYLIVFCL 1

100

MTLCHLWALL ATQAANAA 1 LAVFTYPA 100 GVIVHYVKE ELAERECHV

MGLCAPALGSSLLXXCA LPW EE GGK V VPFCEG CHLSEKMDLV BEHMARKDA

500

M PACGVSSVALLQIIIS CCFRSVCKK 2 WFM EP ENNHIKIL DELVORPHV

H MENVWTVSIVLLQSF CSGSNCCG W VWA AY SNNNHKIL DELVORPHV

1000

L TAPL LT VIVFRKCA G NIT 1 DL DOGGL LLLMK A STHRRK GVVR TTVV

2000

A TAPKRL EFYSTYGC G NITE 1 NA DNSVQ GFRKLWASMFRK GHSV TTVV

3000

R VWEAP EV TVHMGEDFF TQGT 3 AYFTKKEEYQR EILGN A KCFEQQ PKVTFTEFM

4000

H TAPL SATIVLDFQERP GIAFIT FTTSVSK 4 SKDNLHFFIFV E VFTDMPR DSCLS YSPL

5000

TAPL EA SILFDPPN S BALK IEHD T E LTK TELEN F IQSQTKMDSDPDTWFL

6000

L TA DNYALVR M VDQHISEDF F QRHL KSARR G VLLVWEA VYDALASHL

7000

A TA ANYGLIE M KFDQFDNIN 5 VNL IANQH T DLYVWEA ADAYLTFGH

8000

R ASIKRETLUS S QAHLHETL L IQQL NSS S STVLIIPF FPFGLLAK

9000

M LQ GNNEPDYD Y FLSLCKDVS S NNEIKMLOQ S KSMLAPDVP ARMELIAX

10000

R TF SQOQINS IFGODTRKFCEDY VSSKFRM KKVQE SRTAFHAD FCSELLEAEL

11000

L FGAAVQYQIS SGHATAAFN T MGATSRHF YTVRNLKPFN FPLSVMDFV

12000

A YDFAAVDIQA GSVGLAESF T VGANAHFV YHNLRNIFN LG DTMET

13000

R YLGDAFVFL RVSPCGIDYE A TOCFKPFSS YIPNNLNTLS DMHTYLRQW

14000

M L QCPYLYSI KFSFYQIEK S SGRFLLEF YTVSLSGL GMOTPMIERK

15000

F NIVF VS ISFPOGTFE KNSSGITFPEP YTMKSELT GOMTPMVERK

16000

H KLYTELEGR EFCDLADQ D ALKRKFGO AGLRLSREV VAILFVNHVS

17000

A WNVNTRKVK IFKILAMS N ALLQLQGDFN TPIEKLKNI VCQLLINLHP

18000

R NMLYPYTLKY ICHLSITFY E SIALLQRME S MLEVLSH AGWPGDFGD

19000

M NNCMLXVDF WTCMOXMKKX KDSTESLGR PTTELVQNC AEMGJKSW

20000

H NVMYLVDFWF WEFIDMKKX GYFGEVLGR PTTTLSEMDK ADWLRINNW

21000

L VTQNRRDFVF SVOYLER LNHIDRFAEP EEE ARAFPFDRE E SLYYVS

22000

A IFDNNHRYFT SVOYLELRHL KSAPLFEP LVIAPWF E SLYYVS

23000

R VTNREHFTVF NIVRHEL IHK VQIK FP EGFATNYQEK E EVVYVS

24000

A DFLPHPLFHL NVQVLS LL C KAPF D EVVYVS

25000

H DQFHPHPLFHL NVQVLS LL C KAPF D EVVYVS

26000

L GLAYEDMD MALALLAFK HMBQYVEFN DG RVQLTTP ANYQVFAA

27000

A ELDPSIFANE FLNYIFFVTPK IDNITLFN CPF VYNNMIFL ANYQVFAA

28000

R MVSE IFKPE KAMEIMAEFG RIPQTFLH YTOFSPNL A KTVK LIPL

29000

M MVS KNTIP KANAAMALA QIPQVMDP DOKPFA TL G BVTRVYFAA

30000

H MVS NMTIE RANVISAALA QIPQVMDP DOKPFA TL G BVTRVYFAA

31000

L RYYVFLVFL RYFYVLQ SVEAVNLV KVLVMDQ RFAFLHRYBE

32000

A DRYVYFLVFL RYFYVLQ SVEAVNLV KVLVMDQ RFAFLHRYBE

33000

R QNLLLPYRA KVTYQYESMT IGGIVQVSF ILVSDKMHE

34000

M QNLLLPYRA KVTYQYESMT IGGIVQVSF ILVSDKMHE

35000

R RGQFMTF NRTUWANG TKYIBGGHYT GMSGFQEP LQDHYVIAAY

36000

M RGQFMTF NRTUWANG TKYIBGGHYT GMSGFQEP LQDHYVIAAY

37000

R RYLKQadxK RAYDGMCTK S VFGDFVQV GQSGDHN

38000

M RYLKQadxK RAYDGMCTK S VFGDFVQV GQSGDHN

39000

R RYLKQadxK RAYDGMCTK S VFGDFVQV GQSGDHN

40000

M RYLKQadxK RAYDGMCTK S VFGDFVQV GQSGDHN

41000

L LQSGVALDAT RLTAQAPL MVCASD RAY RKLHLEEL RLHOCAS PRM

42000

A LRKAAADTV TYSQDQLYQP IMOLAFHPE KRNAEAL INREDVPFD

43000

R LRAKADTV TYSQDQLYQP IMOLAFHPE KRNAEAL INREDVPFD

44000

M LRAKADTV TYSQDQLYQP IMOLAFHPE KRNAEAL INREDVPFD

45000

H LRAKADTV TYSQDQLYQP IMOLAFHPE KRNAEAL INREDVPFD

46000

L KAKWTVFV TMSOJGLLD LMKIND NPY KWHKDRER QOQPKQ PRD

47000

A KAKWTVFV TMSOJGLLD LMKIND NPY KWHKDRER QOQPKQ PRD

48000

R KAKWTVFV TMSOJGLLD LMKIND NPY KWHKDRER QOQPKQ PRD

49000

M KAKWTVFV TMSOJGLLD LMKIND NPY KWHKDRER QOQPKQ PRD

50000

H KAKWTVFV TMSOJGLLD LMKIND NPY KWHKDRER QOQPKQ PRD

51000

(c) Overall amino acid identity

<table>
<thead>
<tr>
<th>LdMNPV</th>
<th>AcMNPV</th>
<th>Overall amino acid identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>65</td>
<td>42%</td>
</tr>
<tr>
<td>25</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>82</td>
<td>95</td>
<td></td>
</tr>
<tr>
<td>59</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>40</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td>49</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>39</td>
<td></td>
</tr>
<tr>
<td>39</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>23</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LdMNPV</th>
<th>Rat</th>
<th>Overall amino acid identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>38</td>
<td>24%</td>
</tr>
<tr>
<td>11</td>
<td>11</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>42</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>41</td>
<td>29</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>23</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LdMNPV</th>
<th>Mouse</th>
<th>Overall amino acid identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>32</td>
<td>23%</td>
</tr>
<tr>
<td>19</td>
<td>14</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>41</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LdMNPV</th>
<th>Human</th>
<th>Overall amino acid identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>32</td>
<td>23%</td>
</tr>
<tr>
<td>14</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>21</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LdMNPV</th>
<th>Absolute identity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>0</td>
<td>15</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
</tr>
<tr>
<td>0</td>
<td>21</td>
</tr>
</tbody>
</table>

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.

The authors wish to thank Dr Lois Miller for her kind gift of the plasmid containing the egt gene from AcMNPV. We would also like to thank Martha Fikes for her excellent technical assistance and Mary Ellen Kelly for photographic assistance.

References


(Received 28 June 1993; Accepted 29 October 1993)