

## Sequence and *in vivo* transcription of *Lacanobia oleracea* granulovirus *egt*

Ian Smith and Clare Goodale

Central Science Laboratory, Ministry of Agriculture, Fisheries and Food, Sand Hutton, York YO4 1LZ, UK

We have determined the nucleotide sequence and located the major *in vivo* transcript termini of the *Lacanobia oleracea* granulovirus (LoGV) *egt* gene. The open reading frame encodes a 460-amino acid polypeptide having extensive sequence similarity to ten nucleopolyhedrovirus (NPV) ecdysteroid UDP-glucosyltransferase (EGT) proteins; the degree of similarity is particularly high within several previously identified EGT 'domains', and eight invariant amino acid residues are conserved. A phylogenetic tree, constructed by the neighbour joining method, showed LoGV EGT to be the most highly diverged of the eleven baculovirus sequences compared. Database searching revealed that part of a published DNA sequence from *Cryptophlebia leucotreta* granulovirus appears to encode the N-terminal

region of EGT, and the relative genomic locations of the *egt* and *granulin* genes in that virus were compared with their positions in LoGV. In infected *L. oleracea* larvae, *egt* is transcribed predominantly as a 1.6 kb mRNA. Primer extension analysis suggested that the major *egt* 5' transcription terminus is located within a baculovirus late gene promoter motif (GTAAG), in contrast to the early gene promoter contexts determined by others for three NPV *egt* mRNA 5' ends. An early transcriptional start site is also used in LoGV *egt* expression, but at a much lower level. The 3' terminus of *egt* mRNA was identified by sequencing DNA fragments generated by rapid amplification of cDNA ends, and is located 58–62 nucleotides beyond the translation stop codon.

### Introduction

While genetic modification offers a clear prospect of improved pesticidal performance in the case of several nucleopolyhedroviruses (NPVs) (reviewed by Bonning & Hammock, 1996), progress along these lines with members of the other baculovirus genus, the granuloviruses (GVs), has generally been markedly slower. Historically, this discrepancy is attributable in large part to the availability of insect cell lines permissive for a number of NPV isolates, in which detailed molecular biological analyses and genetic modification projects can now be undertaken with moderate ease. In contrast, considerable efforts to develop cell lines which are stably permissive for GV replication have generally met with little success (see Crook, 1991), although Winstanley & Crook

(1993) have recently reported stable maintenance of *Cydia pomonella* lines in which *C. pomonella* GV (CpGV) can replicate fully.

Of 634 baculovirus isolates recorded in a recent survey (Murphy *et al.*, 1995), perhaps 30 can be propagated, and thus manipulated, in extant cell lines. To circumvent the obstacle of having to develop a cell line for each baculovirus in which genetic modification might be deemed advantageous, we are interested in the possibility of deriving a simple genetic modification protocol that could be applied generically to any isolate.

We are using a Scottish isolate of a GV pathogenic for the tomato moth, *Lacanobia oleracea* (Crook *et al.*, 1982). This virus is designated LoGV, and two of its genes are under study: *granulin* and *egt*. The *granulin* gene (I. Smith, unpublished) encodes a major viral occlusion body protein, and is homologous to the NPV *polyhedrin* gene and thus liable to possess a strong promoter (reviewed by Crook, 1991) which could be linked to a foreign gene to ensure abundant expression. The *egt* gene, which has now been identified in several NPVs, encodes a secreted enzyme, ecdysteroid UDP-glucosyltransferase (EGT), which interferes with host larval development (O'Reilly

**Author for correspondence:** Ian Smith.

Fax +44 1904 462111. e-mail i.smith@csl.gov.uk

This paper is dedicated to the memory of Norman Crook.

The EMBL accession number of the sequence reported in this paper is Y08294.

& Miller, 1989, 1991; reviewed by O'Reilly, 1995). Deletion of *egt* from *Autographa californica* multicapsid NPV (AcMNPV) has been shown to yield a faster-acting genotype (O'Reilly & Miller, 1991), and the construction of an *egt*-negative genotype into which a foreign gene is inserted may be an attractive and feasible proposition for genetic improvement of baculovirus pesticides.

This paper presents the results of experiments to characterize the LoGV *egt* gene and its transcription in infected tomato moth larvae.

## Methods

■ **Virus and insects.** Purified occlusion bodies of a Scottish isolate of LoGV (Crook *et al.*, 1982) were kindly provided by Norman Crook (Horticulture Research International, Wellesbourne, UK). LoGV was propagated in fifth-instar *L. oleracea* larvae from a colony established at the Central Science Laboratory in 1991 (Corbitt *et al.*, 1996). The virus used for the DNA cloning experiments was derived from an individual larval cadaver among a group of insects infected with a 30% mortality dose of LoGV. It is therefore likely (but, statistically, by no means certain) to comprise a single genotype (Huber & Hughes, 1984); its genomic DNA restriction profiles were free of visible submolar bands in ethidium bromide staining (Smith & Crook, 1988) and were indistinguishable from those of wild-type LoGV (Crook *et al.*, 1982). Stocks of this virus are designated LoGV-S1 and constitute the genotype referred to throughout this work as LoGV.

■ **General nucleic acid techniques.** Many of the experiments described here involved the use of established methods, for which descriptions and source references are given by Sambrook *et al.* (1989).

■ **Viral DNA.** DNA was extracted from LoGV occlusion bodies from single or small numbers of diseased larvae using protocols described by Smith & Crook (1988), substituting for glycerol gradient centrifugation a 1 h centrifugation through 40–65% (w/w) sucrose. Viral restriction fragments were cloned using pBluescript II SK(–) (Stratagene) and *Escherichia coli* DH5 $\alpha$  (Life Technologies). *Xho*I restriction fragments were cloned randomly, while fragments generated by digesting LoGV genomic or cloned DNA with other restriction endonucleases were first isolated by electrophoretic separation in agarose gels and recovery with a GeneClean kit (BIO 101).

■ **DNA sequence determination and analysis.** Double-stranded templates were used throughout, and were sequenced with Sequenase 2.0 (Amersham) and [ $\alpha$ -<sup>35</sup>S]dATP (Amersham), using primers flanking the pBluescript multiple cloning site or designed *ad hoc* and purchased from Cruachem. Sequencing templates were isolated either from small-scale cultures of *E. coli* DH5 $\alpha$  by the boiling method of Wang *et al.* (1988), or from larger-scale cultures using reagents purchased from Qiagen.

DNA sequence was read digitally from autoradiographs using the HIBIO system (Hitachi) and analysed with DNASIS software (Hitachi). Phylogenetic analysis was carried out as described by van Wezel *et al.* (1997), using the program CLUSTAL W (Thompson *et al.*, 1994) to generate an unrooted tree by the neighbour joining method (Saitou & Nei, 1987) with 1000 bootstrap replicates. TreeView (R. D. M. Page, University of Glasgow) was used to display the tree. Accession numbers for the sequences used to compile Fig. 2 are as follows: AcMNPV, M22619; *Bombyx mori* NPV (BmNPV), L33180; *Buzura suppressaria* single-capsid NPV (BsSNPV), U61154; *Choristoneura fumiferana* (Cf)MNPV, U10441; Cf-DEF (defective isolate of CfMNPV), U10476;

*Lymantria dispar* (Ld)MNPV, U04321; *Mamestra brassicae* (Mb)MNPV, U41999; *Orgyia pseudotsugata* (Op)MNPV, D17353; *Spodoptera littoralis* (Sl)MNPV, X84701. The unpublished sequence of *Spodoptera exigua* MNPV (SeMNPV) EGT was kindly provided by Douwe Zuidema (Wageningen Agricultural University, The Netherlands).

■ **RNA isolation and Northern hybridization.** Total RNA was extracted from larvae infected at early fifth instar with LoGV, when the insects began to show visible signs of infection, and also from uninfected larvae of the same age. Extraction from larval tissue (primarily fat body) was carried out using reagents purchased from Stratagene, following the supplier's protocols. RNA was quantified spectrophotometrically and stored under ethanol at –80 °C. Polyadenylated RNA was isolated from total RNA preparations using a kit (Pharmacia) and quantified as above. Electrophoresis of duplicate RNA samples (10  $\mu$ g each) was carried out in formaldehyde–agarose gels, after which RNA was either visualized by ethidium bromide staining or transferred to Hybond-N membranes (Amersham) by capillary blotting. Transcripts of *egt* were detected by Northern hybridization using <sup>32</sup>P-labelled DNA probes comprising either a 0.92 kbp *Hind*III fragment located within the *egt* coding region (see Results), or an oligonucleotide (5' CCGGCCAAGATAGCTTATCG 3') complementary to nucleotide positions 570–550 of the sequence shown in Fig. 1. Hybridization was carried out at 65 °C and the final wash conditions were 65 °C, 0.1  $\times$  SSC, 2  $\times$  15 min.

■ **Primer extension.** An oligonucleotide primer (20 pmol of the oligonucleotide described above) was labelled at its 5' end with [ $\gamma$ -<sup>32</sup>P]ATP (Amersham) and polynucleotide kinase. The labelled primer was purified away from unincorporated isotope by passing the reaction products over a Nick Column (Pharmacia). An aliquot of the eluant containing ~2 pmol primer was added to 5  $\mu$ g total RNA and incubated at 65 °C for 5 min, after which the sample was cooled to 40 °C over a period of 20 min to allow the primer to anneal. Annealed primers were extended with reverse transcriptase (Superscript; Life Technologies) and dNTPs in React 4 buffer (Life Technologies) at this temperature for 30 min. An aliquot of the sample was then loaded on to a conventional DNA sequencing gel alongside a sequence ladder generated by extension of the same labelled primer by Sequenase, following the supplier's protocol and using as a template a cloned 675 bp *Eco*RI fragment encompassing the 5' region of the *egt* gene (see Results).

■ **RACE.** Rapid amplification of cDNA ends (RACE) was carried out as described by Frohman (1990). Briefly, poly(A)<sup>+</sup> RNA from infected larvae was reverse transcribed and amplified by PCR using a generic 'adapter' primer and an *egt* cDNA-specific primer (5' GAACATGTTGAAACGCC 3', positions 1069–1085; Fig. 1). One aliquot of the products was examined by agarose gel electrophoresis, and others were subjected to re-amplification using either the same primer pair described above or a nested *egt*-specific primer (5' TTCGATACCTTATGCCG 3', positions 1360–1376), in conjunction with the adapter primer.

The sequence of the initial RACE product approaching the position corresponding to the poly(A) addition site was determined using a PCR product sequencing kit (Amersham) with a third *egt*-specific primer (5' GTTTTATAACGTGAGGCC 3', complementary to positions 1593–1576).

## Results

### Localization and nucleotide sequence of LoGV *egt*

Low-stringency hybridization of a <sup>32</sup>P-labelled 1.14 kbp *Bst*EII–*Xho*I fragment, constituting the greater part of the SIMNPV *egt* ORF (I. Smith & F. R. Hunter-Fujita, unpublished),

<i>EcoRI</i>	GAATTCACGCGCGCCGAAACGAAAATAATTAAGCATCACAGCGTACAAACTCAAGTTCGGCGCGGTGGACAGGCTCTCGCAAAAGCTGTTGTAGAACGCC	100
	GTCCAGCTCTTTAGTGTACATGTTCAAAACTTTTCAAAGTGTGGGTGCGCAGTATAGACAAACCCCGCGTTGACTATCAACGTTGACATAGTATGTCGCT	200
<i>BglII</i>	CTTTTTTAAACTGTTTCGAGATCTTTAGACGCGTTCATTTCTTGCACTTCGAGATCGTCGTGTCTTAACGCCGCGGCTCCATTGTAAGAGTGAATAC	300
<i>BamHI</i>	TGGATCCAAACGAACGACGAGAGTGTCTCAGAGTCTATTTTTAAAGATGACCGCTTCGAGCCATGAGTATATAAACCAACATTGGGGACTCGACCTCA	400
	TTCTTTTAATCGACACAGTAAAAATAGGCCAAAATGTTTATATCTATCTACTTTTAGCCTTGGCCGTTGAACGTATTTTGTGTGCCAACATTTTGTGTG	500
	M F I S I L L L A L A V E R I L C A N I L C V	23
<i>ClaI</i>	TTTTCCCCTCTGCGTACAGTCACCAATCTGTTTTTTCGGCTTACATCGATAAGCTATCTTGGCCGGACACAACGTGACGGTGATTACACCTATGCC	600
	F P T P A Y S H Q S V F S A Y I D K L S W A G H N V T V I T P M P	56
<i>EcoRI</i>	GCCTGCGGTCGATCACGTACACCAAGTGTGTAGCTCGCTATCTGTGCATTATTTTAAACAATCTAATCAAGAATCTACAATGATCAAGAAACGAGGAGTG	700
	R A V D H V H Q V V S S L S V H Y F N N L I K N S T M I K K R G V	89
	GTCGCTGACGAAACCCGTCACAAAGGAAAATATATGGGACTGATTAATTTGGTGGCGCATGAAATAAAAAGTCTAATGTCACCCGCTCTACTCAGAA	800
	V A D E T T V T K E N Y M G L I N L V A H E I K S P N V T R L L R N	123
<i>HindIII</i>	ACAAGGGTAAACAGTTCGACTTAATGTTTGTGAAGCTTATGTCAAGTATATCTTGGTGTGGAGCCATCTACGACGCGCCGGTATCCAGTTCTCGTC	900
	K G N K F D L I V C E A Y V S Y I L V F G A I Y D A P V I Q F S S	156
	GGGCTACGCCATACCTGAAAACCTTTGAAAACGTGGGGCGGAGGTAGCTAGAAAACCATCAAGCACCCAAACATTGGCGGTCTGATTTTCAGCAAGTCT	1000
	G Y A I P E N F E T V G G E V A R N H I K H P N I W R S D F S K S	189
	AATTTGAACAACCTGATGACGAAAACTATTGAAAAACGAATGGGCTCTATTGAAAAAGGAACAAGAGAATGTTGAAACGCGACTTTGGATACCACC	1100
	N F E Q L M T E N Y L K N E W A L L E K E Q E N M L K R D F G Y H H	223
<i>PstI</i>	ACGACATGTGCCAAGTCTAGGGTCTTAATGTGTTTATTAAAGTTCCTGCAAGTGTTCGATAATAACAGAGATGTTTCTAACAATATCCAGTACTT	1200
	D M C Q L K S R V L M L P I N V P A V F D N N R D V S N N I Q Y L	256
<i>EcoRI</i>	GGGTGGAATTCACCTCAAAAACCAAGAACAGTGCAGGACTCACGTTTGTCTCGTTTATGGAAGCACCACATTATAGTTTACGCTAGCTTCGGATCG	1300
	G G I H L K K P R T V R D S R L L S F M E K H H I I V Y A S F G S	289
<i>ClaI</i>	GGCATCGATGTACTTAACATGGACGCTAATCTTATCGCGGAGTTTGTGCAAGTGTTCGAATTCGATACCTTATGCGGTGTGTGGAAAGTGGATTTCGTTCA	1400
	G I D V L N M D A N L I A E F V R V F N S I P Y A V L W K V D S S I	323
	TACACCTCAAAACATAACATTTTCATCAACGTCACACGAGTCAATGTTCCCGCAACGAGACGTTTTAAATCACCCGCACATAAAAAGTTTTTCATCACGCA	1500
	H L K H N I S S N V H T Q S W F P Q R D V L N H P H I K V F I T Q	356
	GGGTGGTGTCAATCGACGGATGAAGCAGTTAATAGCGGTGTTCCATGATCGGGATTCTATAATGGGCGATCAGTTTATAACGTGAGGCGTTACACA	1600
	G G V Q S T D E A V N S G V P M I G I P I M G D Q F Y N V R R Y T	389
	GAATTTGGGTATCGGAGAAAAAGTAAACATATTCGCTTGGAAAGAGGGGTTGGATCGCAAAATTAAAAAGTGGTGACAAACAAATCGTACGAGCTAA	1700
	E L G I G E K V N I L R L E E E G L D R K I K N L V H N K S Y E L N	423
<i>HindIII</i>	ACATCAAGCGGTTAAATCTGTTTATTTCTGACACACCGGTTAAACCTCTCGAAAAGCTTTGTGGTTCCACCAACTATGTGTTAAGAAACAAGGACGCTAT	1800
	I K R L N L F I S D T P V K P L R K A L W F T N Y V L R N K D A I	456
	AGACAAGTTTAAAGTAAACAAAGAATTTATATTATATACTTTTATTATTATAAAATATATGTATCATAAAATTTAAAAAGTGTGCGTTTTATTTTTTGT	1900
	D K F K *	460
	GATGTAATAAGCACCAAAGCCAAACACCAAACCCACCGTCACAATTACCACGGTGCAGCATATTAATAGGATAGGTACATTTACACCCGCGCACAAGGG	2000
	TTTTTCTTTATATTTTCGTAATAAGGAACTTCGTGTCTCCCGTCAAACTTCTCCGTTTGTTCGTCATCAATTCGCTCAAGGCGTAGGTACGATTT	2100
<i>SalI</i>	GTGTCGAC	2108

Fig. 1. Nucleotide and amino acid sequence of the LoGV *egt* gene and its product. The DNA sequence of the 2108 nt region is shown and sites for the commoner restriction enzymes indicated. The *egt* ORF begins at position 434, and the deduced EGT amino acid sequence is presented below the coding region with an asterisk at the stop codon TAA. Underlined are a baculovirus late gene transcription start motif (beginning at position 287), a TATA box (369) and the poly(A) addition site of *egt* mRNA (1874). Hooked arrows represent the 5' termini of *egt* mRNAs detected *in vivo*, their tails denoting the position of the first base for each of two mRNAs (positions 289 for the major transcript and 399 for the minor) identified in this work.

to a Southern blot of LoGV DNA cut with assorted restriction enzymes yielded a series of autoradiographic bands corresponding, generally, to a single LoGV fragment in each restriction profile. Among these were *XhoI*-C (~ 13 kbp) and a 0.92 kbp *HindIII* fragment, the former of which was cloned

and found by restriction mapping to contain the latter (data not shown). By subcloning this *HindIII* fragment from the clone containing *XhoI*-C and sequencing its ends, similarities to two appropriately separated sections of the AcMNPV EGT amino acid sequence published by O'Reilly & Miller (1989) were

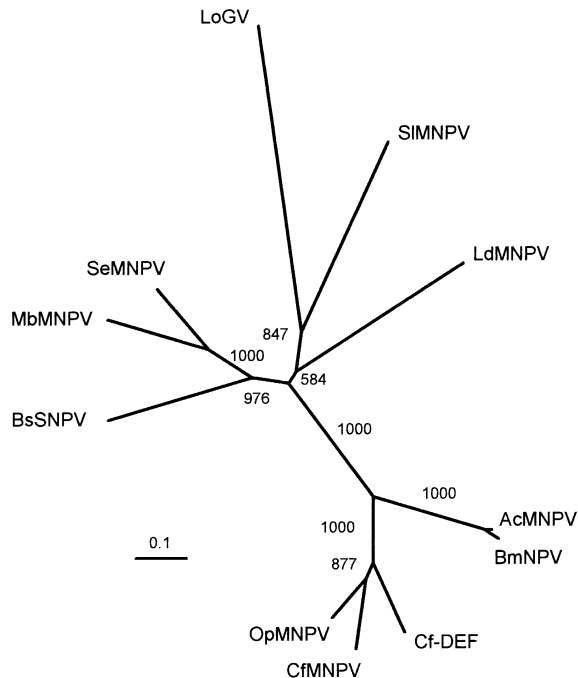


Fig. 2. Phylogenetic analysis of deduced baculovirus EGT sequences. The unrooted tree was constructed by neighbour joining using CLUSTAL W (see Methods). Numbers indicate the support, per 1000 bootstrap replications, for each internal branch with which they are juxtaposed. Branch lengths reflect sequence divergence, the scale bar indicating 0.1 amino acid distance units per residue. Accession numbers of the sequences used in this comparison are given in Methods.

readily detected. These preliminary experiments established the location and orientation of the LoGV *egt* ORF, as well as its likely boundaries, within *XhoI*-C.

Various restriction fragments within *XhoI*-C (between the *SalI* and *EcoRI* sites shown in the upper part of Fig. 3) were subcloned and partially sequenced, and gaps were closed by extension into undetermined regions from sequencing primers designed *ad hoc*. The 2108 bp of LoGV DNA thus sequenced was found to encompass a 1380 nt ORF (Fig. 1), as well as several hundred bp of flanking sequence which included portions of two predicted ORFs oriented inversely to the central ORF. Neither of the deduced amino acid sequences of these partial ORFs showed appreciable similarity to sequences in the databases, but the 460 amino acid sequence derived from the central ORF displayed extensive similarities to several NPV EGT sequences.

### EGT sequence comparisons

An alignment of all available deduced baculovirus EGT sequences, which incorporates our LoGV data, has recently been presented by Hu *et al.* (1997) in their analysis of BsSNPV *egt*, and will not be reiterated here. These workers also constructed a phylogenetic tree for the EGTs using maximum parsimony. As shown in Fig. 2, application of a different methodology to the same data set yielded a tree having two

highly supported NPV groupings (comprising, in clockwise order on Fig. 2, AcMNPV–OpMNPV and BsSNPV–SeMNPV) and a third, weakly supported grouping consisting of LoGV, SIMNPV and LdMNPV. A comparison of branch lengths shows that LoGV is the most diverged isolate (assuming that EGT relationships accurately reflect overall virus genome/proteome relationships), and the good support for placing SIMNPV as its nearest, albeit distant, neighbour may mean that we were fortunate in our choice of probe for detecting the *egt* homologue in LoGV.

An unexpected match with part of a sequence from *Cryptophlebia leucotreta* GV (ClGV) DNA (Jehle & Backhaus, 1994) was revealed by database searching. When translated, a portion of the first 115 nt of the 2231 bp sequence published by these authors matched perfectly with the deduced amino acid residues 18–38 of LoGV EGT (Fig. 3). This identity begins at what is anticipated by analogy with AcMNPV EGT (O'Reilly *et al.*, 1992) to be the N terminus of the mature, secreted EGT polypeptides. During the course of genomic mapping studies, LoGV *XhoI*-C (which contains *egt*) was found to be adjacent to *XhoI*-D (which contains *granulin*) (I. Smith, unpublished). As shown in Fig. 3, the LoGV *egt* and *granulin* genes are separated by ~ 8 kbp of DNA; this contrasts with ClGV, in which a single ORF is located in the ~ 1.1 kbp between the two.

### Detection of LoGV *egt* mRNA by Northern blot hybridization

When the 0.92 kbp *HindIII* fragment referred to earlier in this section (nucleotides 835–1755 of Fig. 1) was labelled and hybridized to a blot of RNAs isolated from infected *L. oleracea* larvae, a prominent band corresponding to a ~ 1.6 kb species appeared in each lane on the resultant autoradiograph (Fig. 4). Minor bands at ~ 2.1, 2.8 and 3.4 kb were also present, at variable intensities relative to those of the 1.6 kb bands, and no hybridization was detectable to RNA isolated from uninfected larvae. Comparable results were obtained when this experiment was repeated with an aliquot of the end-labelled oligonucleotide that was used for primer extension analysis (see below) acting as the probe (data not shown). These findings indicate that the major LoGV *egt* transcript in infected *L. oleracea* larvae is a 1.6 kb mRNA.

### Primer extension analysis

To locate the 5' terminus of LoGV *egt* mRNA, cDNA was synthesized from a <sup>32</sup>P-labelled oligonucleotide annealed to RNA from infected larvae. This reaction yielded four discrete products (Fig. 5), two of which were discounted on the grounds that they either terminated prematurely downstream of the *egt* start codon (the strong band at +11) or were also synthesized from uninfected larval RNA (the bands at –26). Of the two remaining infected larva-specific products, the minor terminated at a position 35 nt upstream, and the major

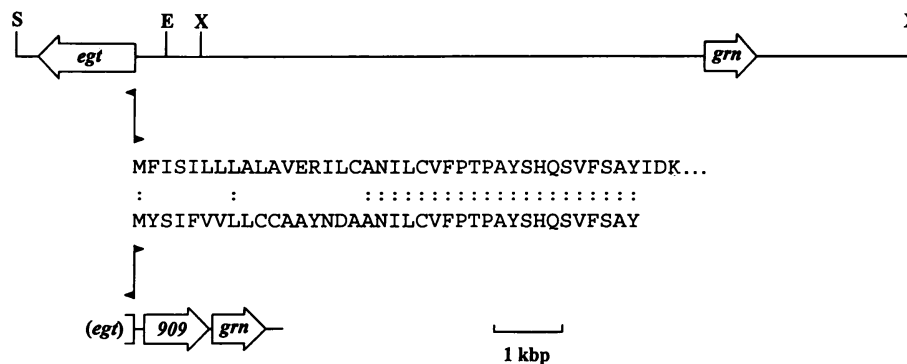


Fig. 3. Relative locations of *egt* and *granulin* genes in LoGV and CIGV. Regions of the genomes of LoGV (top line) and CIGV (bottom line) encompassing the two genes are shown to the same scale. For LoGV, the region comprises the right-hand end of *XhoI*-C and the adjacent *XhoI*-D fragment, which contains the *granulin* (*grn*) gene in the position and orientation indicated. The restriction sites shown are for *Sall* (S), *EcoRI* (E) and *XhoI* (X). The bottom line summarizes the CIGV DNA sequence data of Jehle & Backhaus (1994), showing ORF 909 (909) adjacent to *granulin*. At the left-hand end of this line is a partial ORF deduced, by conceptual translation of part of the strand complementary to nucleotides 115–1 of Jehle & Backhaus's (1994) sequence, to encode the N-terminal region of CIGV EGT. The amino acid sequence of this partial ORF is displayed in the lower line of the central part of the figure, aligned below the corresponding section of LoGV EGT. Colons between the two amino acid sequences denote identical residues.

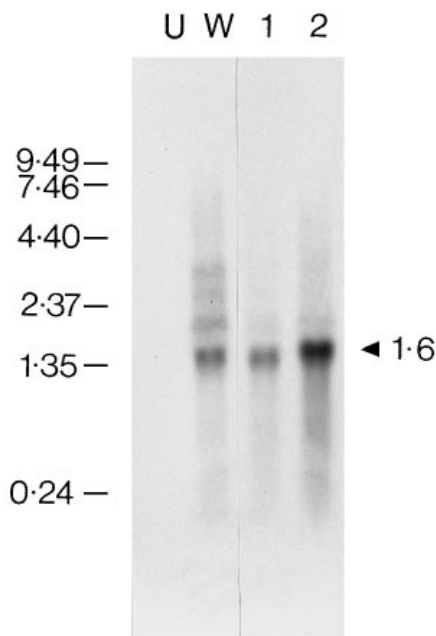


Fig. 4. Northern analysis of LoGV *egt* mRNA. Each lane of the 1% agarose gel from which this autoradiograph was obtained contained 10 µg of total RNA from the tissues of a single insect. RNA was isolated from an uninfected larva (lane U), from a larva infected with wild-type LoGV (lane W), and from two larvae infected with LoGV-S1 (lanes 1 and 2). The positions of RNA size markers (Life Technologies; sizes shown are in kb), separated in the same gel in the unshown lane adjacent to lane U, are indicated, as is that of the major hybridizing species in each infected larval RNA sample.

145 nt upstream, of the *egt* start codon. These results suggest that a low level of *egt* transcripts initiating 25 nt downstream of a TATA box motif coexists with a much higher level of

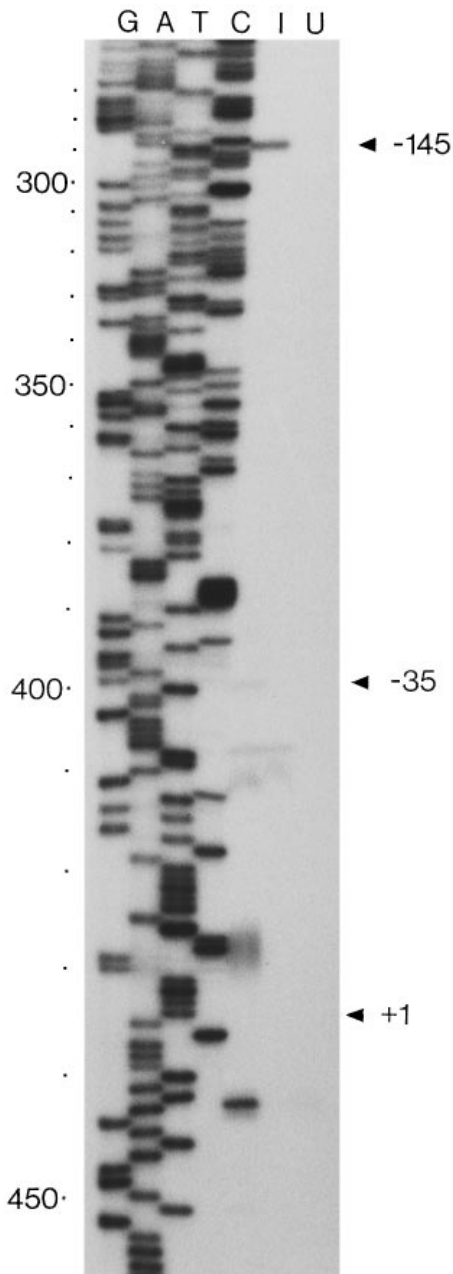
transcripts initiating within a GTAAG late promoter motif (see Fig. 1) among the gross larval RNA population at this (late) stage of virus infection.

#### RACE analysis of the 3' end of *egt* mRNA

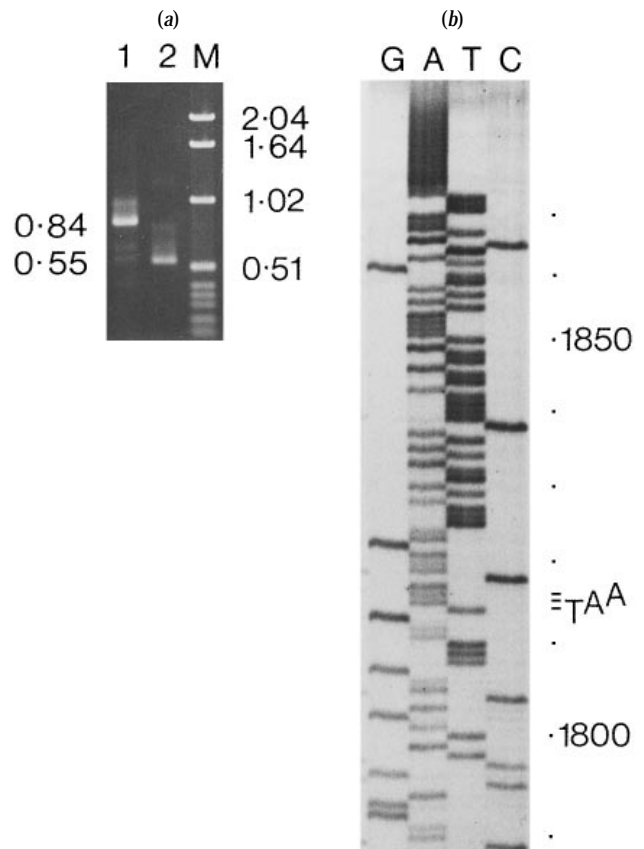
The site of poly(A) addition, constituting the 3' terminus of *egt* mRNA, was located by RACE analysis using poly(A)<sup>+</sup> RNA from infected *L. oleracea* larvae. An initial amplification using an *egt*-specific oligonucleotide yielded a product of 0.84 kbp (data not shown), from which the fragments shown in Fig. 6(a) were obtained following a further amplification with either the same primer, or a second primer which should anneal to *egt* cDNA 291 nt downstream of the first (see Methods). The difference in size between products generated from these re-amplifications was estimated to be 0.29 kbp, suggesting that both were bona fide *egt* mRNA-derived species and that the site of poly(A) addition would be located approximately at position 1875 of Fig. 1, about 60 nt beyond the *egt* stop codon.

This prediction was confirmed by directly sequencing an aliquot of the first RACE reaction product (Fig. 6b). The poly(A) tract was seen to initiate after the last base of a triplet of T residues which, in the genomic DNA sequence (Fig. 1), is followed by a run of five As. Although this sequence confounds attempts to define the exact site of poly(A) addition, the result showed that the poly(A) tail is added at a position 58–62 nt downstream of the *egt* stop codon.

In conjunction with the primer extension data described above, these experiments suggest that the size of the major LoGV *egt* transcript, excluding the poly(A) tract, is ~ 1585 nt, which agrees well with the 1.6 kb estimate obtained from Northern blot hybridization (Fig. 4).



**Fig. 5.** Primer extension analysis of LoGV *egt* mRNA. Products of reverse transcription reactions extended from a  $^{32}\text{P}$ -labelled *egt*-specific oligonucleotide (complementary to positions 570–550 of the sequence shown in Fig. 1; see Methods) were denatured and separated in a sequencing gel prior to autoradiography. The templates used consisted of 5  $\mu\text{g}$  total RNA from an uninfected larva (lane U) or from a larva infected with LoGV (lane I). Lanes G, A, T and C contain the products of a standard sequencing reaction using the same oligonucleotide primer, with a cloned *EcoRI* fragment encompassing the 5' region of the *egt* gene (positions 1–675 in Fig. 1) acting as the template. The positions of nucleotides in Fig. 1 to which this sequence is complementary are indicated to the left of the sequencing reaction lanes. Arrowheads mark the positions, relative to the first base of the *egt* start codon (+1), of the 5' termini of minor (–35) and major (–145) *egt* transcripts, inferred from the sizes of primer extension products arising from infected larval RNA and visible as discrete bands corresponding to sites upstream of the start codon.



**Fig. 6.** RACE analysis of the 3' terminus of LoGV *egt* mRNA. (a) Aliquots of the products of PCR amplification of infected larva-derived cDNA using an *egt*-specific primer (see Methods) were re-amplified using either the same primer (lane 1), or a second primer designed to anneal 291 nucleotides further downstream (lane 2), in conjunction with a common adapter primer (Frohman, 1990; see Methods). The products of these re-amplifications were separated in 0.8% agarose alongside a '1 Kb DNA ladder' (Life Technologies) (lane M), some of whose fragment sizes (in kbp) are shown to the right. The estimated sizes of the major re-amplification products in lanes 1 and 2 are shown to the left. (b) An aliquot of the RACE reaction product was subjected to direct sequence analysis (see Methods) using an oligonucleotide primer designed to anneal to the strand complementary to that shown in Fig. 1 at positions 1576–1593. A portion of the autoradiograph obtained after electrophoretic separation is shown, with nucleotide positions corresponding to those in Fig. 1 indicated to the right along with those of the *egt* stop codon.

## Discussion

When fifth (penultimate) instar *L. oleracea* larvae are infected within their first 24 h with a lethal dose of LoGV occlusion bodies, they almost invariably remain as fifth instar larvae after uninfected controls have moulted into the sixth instar, and subsequently die without moulting to the sixth instar (I. Smith, unpublished data). This observation is reminiscent of the findings of O'Reilly & Miller (1989), who showed that most fourth instar *S. frugiperda* larvae infected with AcMNPV failed to moult to the fifth instar and, furthermore, that this inhibition was attributable specifically to *egt* expression. Burand *et al.* (1996) have also reported that inhibition of fourth to fifth instar

moulting in *L. dispar* larvae infected with LdMNPV is a consequence of expression of the *egt* gene, characterized by Riegel *et al.* (1994).

Three of the ten NPV *egt* genes have now been shown to encode UDP-sugar-conjugating activity resembling that of AcMNPV EGT (O'Reilly & Miller, 1989), (MbMNPV, Clarke *et al.*, 1996; LdMNPV, Burand *et al.*, 1996). In a fourth case (Toister-Achituv & Faktor, 1997), EGT activity accompanies SIMNPV infection in cell culture, although this has not been formally shown to ensue from expression of the *egt* gene described by Faktor *et al.* (1995). In view of the extent of sequence similarity among baculovirus EGTs, including LoGV EGT (see below), we are confident that expression of the gene described here is responsible for the disruption in development of LoGV-infected tomato moth larvae mentioned above. It should be possible to establish this by, for example, transferring *egt* from LoGV to an AcMNPV genome from which the resident *egt* gene has been deleted.

Hu *et al.* (1997) have recently presented an alignment of baculovirus EGTs, incorporating our LoGV sequence, in which ten regions (I–X) showing a high degree of sequence conservation were identified. Eight invariant residues noted by O'Reilly (1995) occur also in the LoGV polypeptide. The predicted length of LoGV EGT, 460 amino acids, is somewhat shorter than those of the NPV proteins; the extreme C-terminal region X defined by Hu *et al.* (1997) is missing in LoGV and this, together with a deletion between regions I and II in a part of the molecule exhibiting considerable variability among the NPV EGTs, accounts largely for the truncated size of LoGV EGT.

The tree shown in Fig. 2 differs slightly from that published by Hu *et al.* (1997). In both cases support for the position of LdMNPV, which is the most notable discrepancy between the two constructions, is weak. A more robust location of this isolate will not be forthcoming without the acquisition of more viral EGT sequences. Our tree does underscore the high degree of divergence (within the NPVs) of the LdMNPV and SIMNPV EGT sequences, which mirrors a recent polyhedrin-based analysis (Clarke *et al.*, 1996), and it also bolsters considerably the clustering of BsSNPV, MbMNPV and SeMNPV which was supported only weakly in the analysis of Hu *et al.* (1997).

The size of the predominant LoGV *egt* mRNA is comparable to those of *egt* transcripts in LdMNPV (Riegel *et al.*, 1994) and CfMNPV (Barrett *et al.*, 1995), as well as to that of the major AcMNPV *egt* transcript (O'Reilly & Miller, 1990). The occurrence of three larger, minor LoGV mRNAs, transcribed from the same strand as the major 1.6 kb message, may reflect imperfect termination of *egt* transcription beyond a non-canonical (AACAAA) polyadenylation signal overlapping the stop codon (Fig. 1), in which case these larger mRNAs would be expected to have heterogeneous 3' termini. While no markedly larger products were in evidence after RACE amplification (Fig. 6a) to substantiate this idea, it is probable that the combination of their low abundance and large size,

relative to the prominent 1.6 kb species, would make them poor competitors in the RACE protocol used here. What is clear, however, is that the major *egt* transcript occurring in LoGV-infected larvae comprises a 145 nt 5' flanking sequence, a 1380 nt ORF, and an approximately 60 nt 3' flanking region which, together with the poly(A) tract, constitute the approximately 1.6 kb species dominating the Northern blot profiles seen in Fig. 4.

Our primer extension data suggest that LoGV *egt* is transcribed mainly from a late promoter. We are not aware of any examples of NPV *egt* genes in which expression occurs from a late transcription site, nor of any other example in the baculoviruses in which homologues of a particular gene are expressed with substantially different temporal profiles in different isolates. Dual early and late transcription initiation sites, on the other hand, have been revealed in several NPV genes expressed in cell culture. The *gp64* envelope glycoprotein gene, for instance, is transcribed under such an arrangement from the genomes of both OpMNPV (Blissard & Rohrmann, 1989) and AcMNPV (Jarvis & Garcia, 1994), with lower levels of *gp64* mRNA arising from an early start site being superseded by higher levels initiating from late sites. At the level of the whole insect, used here in the absence of a permissive cell line, cells are liable to be infected asynchronously so that no such early-to-late progression can be observed. However, the presence of LoGV *egt* transcripts arising from both early and late initiation sites, and at the appropriate relative levels, is compatible with the notion that this GV gene may be expressed in a sequential fashion resembling *gp64* expression in the NPVs.

The *egt* gene is transcribed as an early gene in both AcMNPV and LdMNPV, and Riegel *et al.* (1994) have proposed that a slight delay in LdMNPV *egt* expression relative to that initially reported for AcMNPV *egt* might be tolerated because there is a longer intermoult period in *L. dispar* larvae (~ 6 days) than in the *S. frugiperda* larvae (~ 3 days) used as the host for AcMNPV in the studies of O'Reilly & Miller (1989, 1991). In *L. oleracea* the duration of the fifth instar is 5 days (Corbitt *et al.*, 1996), comparable to that reported for late-instar *L. dispar* by Riegel *et al.* (1994), and yet LoGV *egt* mRNA bears the hallmarks of a late transcript. It may be that, in natural infections, rapid expression of *egt* is unimportant, and thus that EGT would attain levels required for biological efficacy within an acceptable time irrespective of whether the gene was transcribed from an early or a late promoter. Suggesting that inhibition of the earliest larval moult may be counter-productive to AcMNPV in terms of yield, Engelhard & Volkman (1995) have shown that in another AcMNPV host, *Trichoplusia ni*, moulting to fifth instar following infection of larvae at early fourth instar is not inhibited at all under what the authors refer to as 'biologically relevant doses' of ingested occlusion bodies.

Nonetheless, it is noteworthy that in LoGV *egt* expression apparently receives a 'boost' supplied by the late promoter

which the NPV genes apparently do not require. Might this signify a GV/NPV distinction? An examination of the sequence upstream of the partial CIGV *egt* ORF of Jehle & Backhaus (1994) reveals a candidate late promoter motif (TTAAG) located 200 nt before the start codon, but this is situated within *ORF 909*, and several TATA box motifs occur in the 151 nt region separating the start codons of *egt* and *ORF 909*. It is not clear, then, whether the late transcription start site used in LoGV *egt* transcription is peculiar to this virus, or whether a late start site will turn out to be a common feature of GV *egt* genes.

The degree of sequence conservation between the deduced mature N termini of LoGV and CIGV EGTs is striking (although it is doubtful whether this will be sustained throughout the lengths of the two polypeptides), but the relative genomic locations of *egt* and *granulin* are quite different in the two genomes (Fig. 3). In CpGV, the genes occur in the same relationship as they do in CIGV (N. E. Crook, personal communication). The partially defined ORF upstream of LoGV *egt* bears no strong resemblance to *ORF 909* of CIGV, and it will be pleasing to establish whether an *ORF 909*-like gene occurs elsewhere within (or, if not, outside) the ~8 kbp separating *egt* and *granulin* in LoGV; and, conversely, if ORFs located within this 8 kbp are conserved, in sequence if not in location, in CIGV and CpGV.

We are indebted to Norman Crook for providing LoGV occlusion bodies, for preliminary sequence analysis, and for reviewing the manuscript. We thank Peter Young for carrying out the phylogenetic analysis of EGT sequences shown in Fig. 2; Christine Conyers for sequencing RACE products; and Iris Fernandes for photography. David O'Reilly and Jenny Cory kindly let us see their MbMNPV *egt* manuscript prior to publication. Our work is supported by the UK Ministry of Agriculture, Fisheries and Food.

## References

- Barrett, J. W., Krell, P. J. & Arif, B. M. (1995). Characterization, sequencing and phylogeny of the ecdysteroid UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis viruses isolated from *Choristoneura fumiferana*. *Journal of General Virology* **76**, 2447–2456.
- Bonning, B. C. & Hammock, B. D. (1996). Development of recombinant baculoviruses for insect control. *Annual Review of Entomology* **41**, 191–210.
- Blissard, G. W. & Rohrmann, G. F. (1989). Location, sequence, transcriptional mapping, and temporal expression of the gp64 envelope glycoprotein gene of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. *Virology* **170**, 537–555.
- Burand, J. P., Park, E. J. & Kelly, T. J. (1996). Dependence of ecdysteroid metabolism and development in host larvae on the time of baculovirus infection and the activity of the UDP-glucosyl transferase gene. *Insect Biochemistry and Molecular Biology* **26**, 845–852.
- Clarke, E. E., Tristem, M., Cory, J. S. & O'Reilly, D. R. (1996). Characterization of the ecdysteroid UDP-glucosyltransferase gene from *Mamestra brassicae* nucleopolyhedrovirus. *Journal of General Virology* **77**, 2865–2871.
- Corbitt, T. S., Bryning, G., Olieff, S. & Edwards, J. P. (1996). Reproductive, developmental and nutritional biology of the tomato moth, *Lacanobia oleracea* (Lepidoptera: Noctuidae) reared on artificial diet. *Bulletin of Entomological Research* **86**, 647–657.
- Crook, N. E. (1991). Baculoviridae: subgroup B: comparative aspects of granulosis viruses. In *Viruses of Invertebrates*, pp. 73–110. Edited by E. Kurstak. New York: Marcel Dekker.
- Crook, N. E., Brown, J. D. & Foster, G. N. (1982). Isolation and characterization of a granulosis virus from the tomato moth, *Lacanobia oleracea*, and its potential as a control agent. *Journal of Invertebrate Pathology* **40**, 221–227.
- Engelhard, E. K. & Volkman, L. E. (1995). Developmental resistance in fourth instar *Trichoplusia ni* inoculated with *Autographa californica* M nuclear polyhedrosis virus. *Virology* **209**, 384–389.
- Faktor, O., Toister-Achituv, M. & Kamensky, B. (1995). Identification and nucleotide sequence of an ecdysteroid UDP-glucosyltransferase gene of *Spodoptera littoralis* multicapsid nuclear polyhedrosis virus. *Virus Genes* **11**, 47–52.
- Frohman, M. A. (1990). RACE: rapid amplification of cDNA ends. In *PCR Protocols*, pp. 28–38. Edited by M. A. Innis, D. H. Gelfand, J. J. Sninsky & T. J. White. London: Academic Press.
- Hu, Z. H., Broer, R., Westerlaken, J., Martens, J. W. M., Jin, F., Jehle, J. A., Wang, L. M. & Vlak, J. M. (1997). Characterization of the ecdysteroid UDP-glucosyltransferase gene of a single nucleocapsid nucleopolyhedrovirus of *Buzura suppressaria*. *Virus Research* **47**, 91–97.
- Huber, J. & Hughes, P. R. (1984). Quantitative bioassay in insect pathology. *Bulletin of the Entomological Society of America* **30**, 31–34.
- Jarvis, D. J. & Garcia, A. (1994). Biosynthesis and processing of the *Autographa californica* nuclear polyhedrosis virus gp64 protein. *Virology* **205**, 300–313.
- Jehle, J. A. & Backhaus, H. (1994). The granulin gene region of *Cryptophlebia leucotreta* granulosis virus: sequence analysis and phylogenetic considerations. *Journal of General Virology* **75**, 3667–3671.
- Murphy, F. A., Fauquet, C. M., Bishop, D. H. L., Ghabrial, S. A., Jarvis, A. W., Martelli, G. P., Mayo, M. A. & Summers, M. D. (editors) (1995). *Virus Taxonomy. Sixth Report of the International Committee on Taxonomy of Viruses*. Vienna & New York: Springer-Verlag.
- O'Reilly, D. R. (1995). Baculovirus-encoded ecdysteroid UDP-glucosyltransferases. *Insect Biochemistry and Molecular Biology* **25**, 541–550.
- 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-glucosyltransferase 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 armyworm *Spodoptera frugiperda*: host ecdysteroids are conjugated with galactose. *Insect Biochemistry and Molecular Biology* **22**, 313–320.
- 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.
- Saitou, N. & Nei, M. (1987). The neighbor joining method: a new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* **4**, 406–425.
- Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

- Smith, I. R. L. & Crook, N. E. (1988).** *In vivo* isolation of baculovirus genotypes. *Virology* **166**, 240–244.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994).** CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **22**, 4673–4680.
- Toister-Achituv, M. & Faktor, O. (1997).** Transcriptional analysis and promoter activity of the *Spodoptera littoralis* multicapsid nucleopolyhedrovirus ecdysteroid UDP-glucosyltransferase gene. *Journal of General Virology* **78**, 487–491.
- van Wezel, G. P., White, J., Young, P., Postma, P. W. & Bibb, M. J. (1997).** Substrate induction and glucose repression of maltose utilization by *Streptomyces coelicolor* A3(2) is controlled by *malR*, a member of the *lacI-galR* family of regulatory genes. *Molecular Microbiology* **23**, 537–549.
- Wang, L.-M., Weber, D. K., Johnson, T. & Sakaguchi, A. Y. (1988).** Supercoil sequencing using unpurified templates produced by rapid boiling. *BioTechniques* **6**, 839–843.
- Winstanley, D. & Crook, N. E. (1993).** Replication of *Cydia pomonella* granulosis virus in cell cultures. *Journal of General Virology* **74**, 1599–1609.

---

Received 23 September 1996; Accepted 17 October 1997