Characterization, sequencing and phylogeny of the ecdysteroid UDP-glucosyltransferase gene from two distinct nuclear polyhedrosis viruses isolated from *Choristoneura fumiferana*

John W. Barrett,1,2 Peter J. Krell2 and Basil M. Arif1*

1 Natural Resources Canada, Canadian Forest Service, Sault Ste. Marie, Ontario P6A 5M7 and 2 Department of Microbiology, University of Guelph, Guelph, Ontario N1G 2W1, Canada

The ecdysteroid UDP-glucosyltransferase (*egt*) gene isolated from a plaque-purified isolate of *Choristoneura fumiferana* multinucleocapsid nuclear polyhedrosis virus (CfMNPV) was compared to its homologue from a defective MNPV virus (CfDEF) present in wild-type virus populations infecting the eastern spruce budworm, *C. fumiferana*. The *egt* genes were located in the same relative position within the virus genomes and their genomic location and arrangement were similar to that found in *Autographa californica* MNPV (AcMNPV) and *Orgyia pseudotsugata* MNPV (OpMNPV). The genes encoded 491 and 494 amino acid open reading frames respectively, and were 67% identical at the amino acid level and 74% identical at the nucleotide level. Transcripts of the *egt* of CfMNPV peaked around 12 h post-infection (p.i.) and disappeared after 36 h p.i. Transcripts of the *egt* of CfDEF peaked between 6 and 9 h p.i. and were not detected 24 h p.i. The *egt* from CfMNPV was more similar to the partially sequenced *egt* identified from OpMNPV, at the nucleotide and amino acid levels, than it was to the *egt* from the CfDEF, AcMNPV, *Bombyx mori* NPV, *Lymantria dispar* MNPV or *Spodoptera exigua* MNPV. Phylogenetic analysis of *egt* supported the baculovirus evolution scheme suggested by polyhedrin sequence analysis.

Introduction

In insects, the moulting hormone ecdysone initiates and regulates a cascade of genes expressed during moulting. Recent studies have revealed that baculoviruses have evolved mechanisms to delay normal lepidopteran larval moulting. The ecdysteroid UDP-glucosyltransferase (*egt*) gene, first identified in *Autographa californica* multinucleocapsid nuclear polyhedrosis virus (AcMNPV), catalyses the conjugation of ecdysone and UDP-glucose (O'Reilly & Miller, 1989), and results in the clearing of ecdysone from the haemolymph. The direct result is a delay in moulting so that virus-infected larvae feed longer and cause greater foliar damage. Interruption of the AcMNPV *egt* coding region increased the virulence, but did not reduce the infectivity of the virus. Insects infected with the modified AcMNPV *egt* virus fed less and died more rapidly than those infected with unaltered AcMNPV (O’Reilly & Miller, 1991). These advantages suggested that alteration in the expression of *egt* in other baculoviruses, either through deletion or inactivation, might also provide a more effective biological control insecticide.

The eastern spruce budworm, *Choristoneura fumiferana*, is an economically important forest pest and is responsible for the destruction of approximately 10.1 million hectares of forest per year in Canada (Anonymous, 1994). To date, there is no effective and economically viable biological control agent against this devastating pest. At present, formulated *Bacillus thuringiensis* (*B.t.*) is used to control forest lepidopteran pest species. However *B.t.* has the potential to impact on non-target lepidopteran species and has short persistence times (Cunningham & van Frankenhuyzen, 1991). Baculoviruses are naturally occurring pathogens which are generally insect specific. It has been demonstrated that genetic modification of baculoviruses can improve their effectiveness as pest control agents (Wood & Granados, 1991). Manipulation of the CfMNPV *egt* appears to be one viable approach to improve baculoviruses against the spruce budworm. We report the identification and characterization of the *egt* genes...
identified in CfMNPV and in a defective virus, designated CfDEF (Arif et al., 1994), found in wild-type CfMNPV populations in the budworm larvae. Restriction enzyme analysis, DNA sequencing and comparison of viral structural proteins demonstrated that CfMNPV and CfDEF are two distinct viruses and not variants of one (Arif et al., 1994).

Methods

Viruses and cells. CfMNPV, a plaque-purified virus has been described previously (Arif et al., 1984). Seed cultures of FPMI-CF203, hereafter referred to as CF-203, FPMI-CF70 (CF-70), and FPMI-CF124T (CF-124T) were the generous gift of Dr S. S. Sohi (Natural Resources Canada, Forest Pest Management Institute, Sault Ste. Marie, Ontario, Canada). CF-203 cells were maintained in SF900 medium (Gibco/BRL Life Technologies) supplemented with 10% fetal bovine serum (FBS) (Gibco). CF-70 and CF-124T cells were grown in Grace's medium with 10% FBS and 0.25% tryptose broth.

Detailed EM studies revealed that, by itself, CfDEF cannot pass through the larval gut barrier. CfMNPV and CfDEF differed markedly in genomic restriction patterns and in their structural polypeptide profiles (Arif et al., 1994).

Identification of egt gene. The egt gene from CfMNPV was identified by Southern blot hybridization to a 1.1 kb BsrEI/XhoI heterologous probe extracted from the Ptrl-G fragment from AcMNPV (O'Reilly & Miller 1990). The egt gene isolated from the CfMNPV virus was identified by hybridization to a 550 bp PCR fragment from the 5' region of the CfMNPV egt. HindIII-digested fragments of CfMNPV were subcloned into pUC19 and HindIII-digested fragments of CfDEF were cloned into pT7T3. Sub-cloned DNA was sequenced using the Pharmacia dye deoxynucleotide kit to confirm the presence of the gene. Complete sequencing of both strands for both viruses was performed by the DNA Sequencing Core Facility at Queens University (Kingston, Ontario, Canada). Sequence data were compiled and analysed using MacVector 4.1.4 (International Biotechnologies).

RNA extraction and Northern blot analysis. The temporal pattern of CfMNPV egt transcription was examined by Northern blot analysis of total RNA samples extracted from CF-203 cells infected with CfMNPV. CF-203 cells were infected at an m.o.i. of 20 with CfMNPV extracellular virus. Transcription of the CfDEF virus was examined by Northern blot analysis of total RNA samples extracted from CF-70 cells infected with CfDEF at an m.o.i. of 10. At the appropriate times post-infection (p.i.) the cells were harvested and total RNA was extracted (Chomczynski & Sacchi, 1987). Total RNA was separated on 37% formaldehyde gels (Sambrook et al., 1989) and blotted onto Photogene membrane (Gibco/BRL). Probes were radio labelled using the BRL random primer labelling kit (Gibco/BRL). Northern blot hybridizations were carried out at 42 °C in 50% formamide and washes were done at 65 °C in 0.1× SSC and 0.1% SDS.

Evolutionary and phylogenetic analysis of the egt genes. The CfMNPV and CfDEF egt nucleotide and amino acid sequences were aligned using the CLUSTAL program (Higgins & Sharp, 1988). Sequences for phylogenetic analysis were aligned with MacVector and imported into PAUP (Swafford, 1990). Phylogenetically informative sites were determined and used in PAUP to construct the most parsimonious phylogenetic tree. The relatively small number of available egt sequences (seven) meant that an exhaustive search algorithm could be employed to evaluate all possible trees and produce the single most parsimonious tree. Based on the low nucleotide identity observed between CfMNPV and LdMNPV egt genes we chose to define LdMNPV egt as the outgroup. The bootstrap method, included in PAUP, was used to test the variability of the phylogeny produced.

Glucosyltransferase activity. The activity of the egt genes was confirmed by the conjugation of glucose and [3H]ecdysone from extracts of CF-124T cells infected with either CfMNPV or CfDEF. This procedure differs from that of O'Reilly & Miller (1990) in that larger starting volumes and longer incubation times were used.

CF-124T cells were infected with either CfMNPV or CfDEF and incubated at 28 °C for 5 days before harvesting. The cells were scraped off the flask into the medium and were disrupted by several passes through a Dounce homogenizer. The debris was pelleted and the supernatant loaded into a centrprep 30 tube (Amicon) to concentrate proteins larger than 30 kDa from an initial 15 ml of media to 1 ml. A 500 µl aliquot of this supernatant was incubated for 12 h at 37 °C with 250 µl of 100 mM-MgCl2, 250 µl of 0.1 M-Tris–maleate, 50 µl unlabelled ecdysone, 50 µl UDP-glucose and 1 mCi of [3H]ecdysone (Du Pont). This conjugation mix was added to a Sep-pak C18 column (Waters) and collected with 100% methanol. The solubilized conjugated ecdysone was evaporated to dryness in a SpeedVac and then resuspended in a total of 50 µl 60% ethanol in H2O. The samples were then applied to silica gel GF TLC plates (Analtech) and separated in organic solvent (1-butanol, acetone, glacial acetic acid, ammonia (30%) and water) (70:50:18:1.5:60) for 3.5 h. The plate was dried and then sprayed with En3Hance (New England Nuclear) and a film was placed on top of the TLC plate and left overnight at −70 °C.

Results

Identification of the egt gene

The entire CfMNPV egt coding region was located on a 6.7 kb HindIII-G fragment (Fig. 1a). A 3.6 kb PvuII/HindIII fragment from HindIII-G was subcloned into pUC18. The sequence of the egt (accession number U10441) within this fragment contained a 1473 bp ORF (Fig. 2) encoding a putative 491 amino acid protein (Fig. 3). This protein included a 17 residue proposed signal peptide. There was a possible TATA signal located 58 nt upstream of the start codon. There was no evidence of a consensus polyadenylation sequence (AATAAA) within 327 nt of the end of the ORF. The closest possibility for a polyadenylation site was the sequence ATAAA 81 nt downstream of the stop codon.

The coding region of the CfDEF egt overlapped two HindIII fragments (Fig. 1b) and was composed of 1482 nt (Fig. 2, accession number U10476) encoding a putative 494 amino acid protein including a 17 residue proposed signal peptide (Fig. 3). A single TATA box was located 69 nt upstream of the start codon and a polyadenylation signal 30 nt downstream of the stop codon.

A sequence immediately upstream of the egt coding region was similar to the late expression factor 1 (lef-1) gene identified in AcMNPV (Passarelli & Miller, 1993). Conservation of the promoter regions between the egt
genes of the two viruses was hard to evaluate because of the location of the *lef-1* gene 68 nt upstream of the *CfMNPV egt* start site and in the opposite orientation. Although the region was highly conserved the similarity could reflect conservation of *lef-1* rather than of promoter elements.

**Northern blot analysis**

A single transcript of approximately 1.6 kb was observed for CF-203 cells infected with *CfMNPV* (Fig. 4). Time course analysis indicated that for the *CfMNPV*-infected cells, an *egt* transcript was first observed 9 h p.i., peaked around 12 h p.i., decreased, then disappeared after 36 h p.i. In contrast, a 1.65 kb transcript from *CfDEF*-infected CF-70 cells was faintly evident by 3 h p.i., had peaked between 6 and 9 h p.i. and the transcript disappeared after 12 h p.i. (Fig. 4). Although not visible in Fig. 4, the 3 h p.i. and 12 h p.i. samples were detected on highly over-exposed filter autoradiograms.

**Evolutionary relationships of egt genes**

Comparison of the nucleotide and amino acid sequences between the *CfMNPV* and *CfDEF egt* genes indicated a 74% identity between the two Cf nuclear polyhedrosis viruses at the nucleotide level and 67% identity at the amino acid level (Table 1). The percentage identity between the *egt* of *CfMNPV* and that of *OpMNPV* was higher (76.9% at the nucleotide and 79.4% at the amino acid level) than that observed between *CfDEF* and *CfMNPV*. This comparison, however, was based on only a partial sequence of the *OpMNPV egt* (Pearson et al., 1993) which represents approximately half of the coding region. This may indicate that, at least for the region of *egt* compared, *CfMNPV* and *OpMNPV* are more closely related and have shared a more recent ancestor than have *CfMNPV* and *CfDEF*. Pairwise comparison at the nucleotide level between most *egt* sequences range from 56% to 79.6% identity. The only exception was the 96% identity between *AcMNPV* and *Bombyx mori NPV* (BmNPV) which may be host range variants of the same virus (Crozier et al., 1994).

Our phylogenetic analysis of the available *egt* sequences, including nucleotide and amino acid sequences, revealed that at least two major groups of baculoviruses have evolved separately (Fig. 5). One group was composed of *CfMNPV*, *CfDEF*, *OpMNPV*, *AcMNPV* and BmNPV and the other group included *Spodoptera exigua MNPV* (SeMNPV) and *Lymnantria dispar MNPV* (LdMNPV). Divergence within the main group indicated that the line resulting in *AcMNPV* and BmNPV split from the common ancestor before *CfDEF*, *CfMNPV* and *OpMNPV* evolved.

**Glucosyltransferase activity**

Our assay revealed that EGT protein from cells infected by either virus was functionally active (Fig. 6). Extracts from uninfected CF-124T control cells did not exhibit this enzymic activity. Interestingly, EGT activity was observed in the uninfected controls of CF-70 and CF-203 cells (not shown).

*CfDEF* infection of CF-124T cells did not result in the production of polyhedral inclusion bodies (PIBs). Cells infected with *CfMNPV*, however, produced PIBs 60 h p.i. EGT activity in CF-124T cells infected with *CfDEF* confirmed that early gene transcription and expression were achieved. The block that prevents production of viral progeny was thus likely to be at a later stage of virus replication. A time course analysis of EGT production in CF-124T cells infected with either *CfMNPV* or *CfDEF*
Fig. 2. For legend see opposite.
**Baculovirus egt genes**

**Fig. 2.** Nucleotide sequence of the *egt* genes from CfMNPV and CfDEF. The initiation and termination codons are in bold letters. Some possible regulatory elements (i.e., TATA box, early gene regulatory element and polyadenylation sequences) are underlined. The early gene promoter (CAGT) is double underlined. Stars indicate identical nucleotides; dashes represent gaps introduced to produce a better alignment.

indicate that EGT is active from 12 h p.i., and at a lower m.o.i. active up to 10 days p.i. (data not shown).

**Discussion**

The *egt* ORFs from the two viruses are approximately the same size, 491 a.a. for CfMNPV and 494 a.a. for CfDEF, and both are smaller than the 506 a.a. reported for AcMNPV. The missing sequences occur within the extreme 3' end of the coding region. However these genes are larger than the 488 a.a. reported for a comparable homologue from LdMNPV (Riegel *et al.*, 1994a). Although there is no evidence of the traditional TATA box for CfMNPV *egt*, there is a putative TATA box (AATAAA) located -58 nt from the translation start codon. In addition, the conserved recognition start site CAGT (Blissard *et al.*, 1992) identified from early genes of baculoviruses is located at -19 nt. The putative upstream regulatory element found in early genes of baculoviruses is located at -19 nt. The putative upstream regulatory element found in early genes of baculoviruses is located at -19 nt. The putative upstream regulatory element found in early genes of baculoviruses is located at -19 nt.
Fig. 3. The aligned putative amino acid sequences of the egt genes from the CfMNPV and CfDEF baculoviruses. Possible glycosylation sites are underlined. The signal peptide is in bold. Stars indicate identical residues.

Fig. 4. Northern blot analysis of the egt transcripts from total RNA from cells infected with CfMNPV and CfDEF. Numbers at the sides refer to transcript size based on an RNA ladder.

Table 1. Comparison of the egt genes from various baculoviruses*

<table>
<thead>
<tr>
<th></th>
<th>CfMNPV</th>
<th>CfDEF</th>
<th>AcMNPV</th>
<th>BmNPV</th>
<th>OpMNPV</th>
<th>SeMNPV</th>
<th>LdMNPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>CfMNPV</td>
<td>1</td>
<td>60</td>
<td>119</td>
<td>178</td>
<td>238</td>
<td>298</td>
<td>418</td>
</tr>
<tr>
<td>CfDEF</td>
<td>1</td>
<td>61</td>
<td>121</td>
<td>181</td>
<td>240</td>
<td>300</td>
<td>360</td>
</tr>
<tr>
<td>AcMNPV</td>
<td>1</td>
<td>62</td>
<td>122</td>
<td>182</td>
<td>241</td>
<td>301</td>
<td>361</td>
</tr>
<tr>
<td>BmNPV</td>
<td>1</td>
<td>63</td>
<td>123</td>
<td>183</td>
<td>242</td>
<td>302</td>
<td>362</td>
</tr>
<tr>
<td>OpMNPV</td>
<td>1</td>
<td>64</td>
<td>124</td>
<td>184</td>
<td>243</td>
<td>303</td>
<td>363</td>
</tr>
<tr>
<td>SeMNPV</td>
<td>1</td>
<td>65</td>
<td>125</td>
<td>185</td>
<td>244</td>
<td>304</td>
<td>364</td>
</tr>
<tr>
<td>LdMNPV</td>
<td>1</td>
<td>66</td>
<td>126</td>
<td>186</td>
<td>245</td>
<td>305</td>
<td>365</td>
</tr>
</tbody>
</table>

* The numbers above the diagonal are the percentage identity of corresponding egt genes at the nucleotide level. The numbers below the diagonal are the percentage identity of the corresponding egt genes at the amino acid level.
† Comparisons using OpMNPV are based on only 879 bases and 293 amino acid residues.

Blissard et al. (1992) noted that with modification of the CAGT element the correct transcript was synthesized but with significant loss of transcriptional activity. We have not yet compared the levels of egt activity in cells infected with CfMNPV and CfDEF to determine if that correlates with upstream CAGT elements.
The transcript size of the CfMNPV egt was 1.60 kb whereas the predicted egt transcript from CfDEF was 1.65 kb. Although similar in size, both transcripts are smaller than the 1.8 kb and 1.7 kb egt transcripts observed for AcMNPV (O'Reilly & Miller, 1990) and LdMNPV (Riegel et al., 1994a) respectively. We observed only single size class transcripts for CfMNPV and CfDEF egt genes. There was no indication of the two transcripts observed by O'Reilly & Miller (1990). The larger 3.0 kb message, a product of incorrect transcriptional processing, that encompassed AcMNPV egt and the DA26 gene was not present for either virus. In addition, we did not find any sequence similarity for AcMNPV DA26 within the regions sequenced downstream of the CfMNPV or CfDEF egt genes, so the absence of the larger transcript may not be surprising. The difference in temporal appearance of egt transcripts for CfMNPV and CfDEF may reflect slight differences in transcriptional regulation in the two different cell lines used.

The phylogenetic relationships of the egt sequences were examined by DNA and protein parsimony analysis. In both situations phylogenetic trees were constructed using either the complete egt data set for CfMNPV, CfDEF, AcMNPV, BmMNPV, LdMNPV and SeMNPV and these were compared with the partial sequence from OpMNPV or only about the first 800 nt of each species were used in the analysis. Under all conditions the structures of the phylogenetic trees were the same (data not shown). The only difference was that the length of the OpMNPV branch was longer for the analysis of the complete sequence sets used, to account for the missing data.

Comparisons of the nucleotide and amino acid sequences of the various egt genes indicate that, based on sequence similarity and phylogenetic analysis, the egt genes fall within two groups. A similar grouping of these baculoviruses was obtained using polyhedrin sequence analysis (Barrett unpublished data; Zanotto et al., 1993). This suggests that egt was an ancestral gene and was present before baculovirus species divergence.

Bootstrap resampling (100 replicates) of the data indicated that this egt phylogeny (Fig. 5) was highly specific and showed little variability. The only area of variation was the organization of CfMNPV, CfDEF and OpMNPV. Ninety-two times out of 100 OpMNPV egt and CfMNPV egt grouped together, with CfDEF egt splitting off from the ancestral sequence [(OpMNPV, CfMNPV), CfDEF]. The other eight times OpMNPV egt grouped with CfDEF egt and CfMNPV egt has split earlier [(OpMNPV, CfDEF), CfMNPV]. This phylogeny, with OpMNPV and CfDEF grouped together, required an extra ten steps (1219) more than the single most parsimonious tree (1209 steps) that resulted in the phylogeny presented in Fig. 5.

When the amino acid sequences were aligned, there were certain regions which were conserved among all the EGT sequences (Fig. 7). Riegel et al. (1994a) identified five regions with more than 59% identity between LdMNPV and AcMNPV. When the putative EGT amino acid sequences of CfMNPV, CfDEF and OpMNPV (partial) were added to the alignment it was possible to confirm and more narrowly define the conserved domains within the EGT coding region (Fig. 7).

One conserved domain (I) was located immediately downstream of the proposed cleavage site between the
Fig. 7. For legend see opposite.
signal and mature peptides. A motif, A/G-A/V-X-A/T, located immediately upstream of the cleavage site, and the strong identity within the mature peptide immediately located immediately upstream of the cleavage site, and signal and mature peptides. A motif, A/G-A/V-X-A/T, located immediately upstream of the cleavage site, and the strong identity within the mature peptide immediately downstream of the cleavage site, probably ensure that correct protease recognition and processing occurs. Identity within the first 15 residues of the mature peptide, A-N/S-I-L-A/V-L/Y-L/F-P-T-P-A/S-Y-S-H-H/Q, ranges from 100% between CfMNPV and CIDEF to 72% between CfMNPV and LdMNPV.

The largest and most highly conserved region (IV) within the protein sequence is located between positions 237–268 (according to CfMNPV alignment) and is 100% identical in CfMNPV, CIDEF and OpMNPV and 90% identical to the same region in AcMNPV. In LdMNPV the identity falls to 77%. Within this region the motif, P-V-P-P previously identified in AcMNPV and LdMNPV is 100% between CfMNPV and CIDEF to 72% between CfMNPV and LdMNPV.

The authors thank H. Lauzon for the pUC19/HindIII-G clone from CfMNPV, P. Jamieson for the pT7T3/Hind III-T and pT7T3 HindIII-W clones from CIDEF and S. Sohi for seed cultures of CF-203, CF-124T and CF70 cells. Thanks also to S. Maeda (BmNPV egt) and to J. Vlak and R. Mans (SemNPV egt) for sharing their sequence data before publication. Special thanks to J. Turgeon, R. Kothapalli, X. Li and H. Lauzon, for reviewing earlier versions of this manuscript. This research was supported by NSERC strategic grant STR010482.

References


(Received 14 March 1995; Accepted 9 June 1995)