Identification and Characterization of a Putative Baculoviral Transcriptional Factor IE-1 from *Choristoneura fumiferana* Granulovirus

Kianoush Khajeh Rashidan, Nasha Nassoury, Abderrazzak Merzouki and Claude Guertin*

Institut National de la Recherche scientifique-Institut Armand-Frappier, Québec, Canada

A gene that encodes a protein homologue to baculoviral IE-1 was identified and sequenced in the genome of the *Choristoneura fumiferana* granulovirus (ChfuGV). The gene has an 1278 nucleotide (nt) open-reading frame (ORF) that encodes 426 amino acids with an estimated molecular weight of 50.33 kDa. At the nucleotide level, several *cis*-acting regulatory elements were detected within the promoter region of the *ie-1* gene of ChfuGV along with other studied granuloviruses (GVs). Two putative CCAAT elements were detected within the noncoding leader region of this gene; one was located on the opposite strand at −92 and the other at −420 nt from the putative start triplet. Two baculoviral late promoter motifs (TAAG) were also detected within the promoter region of the *ie-1* gene of ChfuGV. A single polyadenylation signal, AATAAA, was located 18nt downstream of the putative translational stop codon of *ie-1* from ChfuGV. At the protein level, the amino acid sequence data that was derived from the nucleotide sequence in ChfuGV IE-1 was compared to those of the *Cydia pomonella* granulovirus (CpGV), *Xestia c-nigrum* granulovirus (XcGV) and *Plutella xylostella* granulovirus (PxGV). The C-terminal regions of the granuloviral IE-1 sequences appeared to be more conserved when compared to the N-terminal regions. A domain, similar to the basic helix-loop-helix like (bHLH-like) domain in NPVs, was detected at the C-terminal region of IE-1 from ChfuGV (residues 387 to 414). A phylogenetic tree for baculoviral IE-1 was constructed using a maximum parsimony analysis. A phylogenetic estimation demonstrates that ChfuGV IE-1 is most closely related to that of CpGV.

Introduction

The baculoviridae are large and complex DNA viruses that primarily infect arthropods, especially insects that belong largely to the orders *Lepidoptera*, *Dieptera*, and *Hymenoptera*. The Baculoviridae are comprised of two genera: the nucleopolyhedroviruses (NPVs) and the granuloviruses (GVs). Granuloviruses have been characterized at the cytological level, and their application for pest control in agriculture and forestry has been fairly successful (Crook, 1991; Moscardi, 1999). Spruce budworm, *Choristoneura fumiferana,* is the most devastating coniferous tree pest in eastern Canada and the United States. *Choristoneura fumiferana* granulovirus (ChfuGV) are being considered as an alternative biological insecticide due to their specificity for their insect host.

Baculovirus genes are expressed in a transcriptional cascade in which each successive phase is dependent on the expression of genes during the previous phase. Basically, there are presently four commonly established phases in this cascade. These phases are the immediate early (IE), delayed early (DE), late (L), and very late (VL). During the early phases, the molecular environment and other aspects of the intracellular environment of the cell are altered in preparation for the replication and expression of viral DNA. Early virus promoters are composed of multiple *cis*-acting regulatory components. These promoters require host RNA polymerase II for proper transcription (Nissen and Friesen, 1989; Theilmann and Stewart, 1991; Rodems and Friesen, 1993; Guarino and Dong, 1994; Pullen and Friesen 1995a). DE genes need the presence of regulatory proteins that are transcribed during the immediate early phase in order to be efficiently transcribed (Guarino and Summers, 1986). Baculoviral genes that encode *trans*-activating factors have been identified. These regulatory proteins are able to detect

*To whom correspondence should be addressed.
Tel: 450-687-5010; Fax: 450-686-5533
E-mail : claude.guertin@inrs-iaf.uquebec.ca

enhancer elements that are located in several regions of baculoviral genome. They also contribute to the speed and facility with which the virus is able to direct the host cell metabolic machinery to transcribe particular genes that are necessary for viral replication.

Immediate early protein IE-1 is the principal transcriptional regulator known in *Autographa californica* multiple nucleopolyhedrovirus (AcMNPV) and *Orgyia pseudotsugata* MNPV (OpMNPV) (Hayakawa *et al.*, 2000). Transcription factors are commonly known as structurally-complex proteins that contain various functional components that are associated with DNA binding, protein oligomerization, phosphorylation, activation, and other activities. Transcription factors fall into different groups, such as zinc fingers, helix-turn-helix (HTH), helix-loop-helix (HLH), or basic leucine zippers. These proteins share a short domain that is associated with DNA binding or oligomerization (Altshul *et al.*, 1992; Lewin, 1997). IE-1 protein in baculovirus is detected throughout infection (from IE to VL phases); this involvement suggests that the presence of IE-1 is essential for a productive infection (Ribeiro *et al.*, 1994; Choi and Guarino, 1995; Okano *et al.* 1999). The presence of homologous regions (*hr*s) in the *cis* position of the promoter of DE genes enhances the IE-1 transactivation effect (Guarino and Summers, 1986; Choi and Guarino, 1995). Studies conducted on IE-1 from AcMNPV, the prototype subgroup I baculovirus, demonstrate that IE-1 oligomerization is a key factor for IE-1 transcriptional stimulation (Olson *et al.*, 2001). These data also indicate that IE-1 needs to be oligomerized as a dimer in order to turn into an active protein that is capable of binding to the 28 bp palindromic repeats of the *hr*s. This was revealed using *ie-1* defective mutants of AcMNPV. These mutants lose their oligomerization activity and hence suffer the loss of IE-1 *hr*-dependant transactivation (Olson *et al.*, 2001). The 28-mer palindrome is the minimal requirement for the enhancer activity of *cis* linked viral promoters (Guarino and Dong, 1994; Rodems *et al.*, 1997; Kremer and Knebel-Morsdorf, 1998; Leisy and Rohrmann, 2000; Massari and Murre, 2000; Olson *et al.*, 2001). Evidence demonstrates that it is essential for IE-1 dimers to interact through cooperative binding with both palindromic half sites of the 28-mer palindrome in order to stimulate the *hr* enhancer activity (Rodems *et al.*, 1997; Kremer and Knebel-Morsdorf, 1998; Leisy and Rohrmann, 2000). The model that is presented for the DNA binding by IE-1 proposes that the interaction of the IE-1 dimer across the 28-mer axis of symmetry makes simultaneous contact with both half sites (Olson *et al.*, 2001). This contact is a vital parameter that can eventually lead to transcriptional enhancement.

To date, *ie-1* homologues have been reported from 10 NPVs and 3 GVs; however, no significant information about the granuloviral *ie-1* genes is available in literature. This paper is the first major report that concentrates on the analysis of a granuloviral *ie-1* gene and its deduced amino acid sequence.

Materials and Methods

In vivo **production and purification of virus** *Choristoneura fumiferana* fourth-instar larvae were infected by ChfuGV using a virus-contaminated artificial diet (Forté *et al.*, 1999). Progeny viruses were then extracted from the infected larvae as previously described (Bah *et al.*, 1997).

DNA extraction, cloning, and sequencing ChfuGV DNA was isolated from purified-enveloped nucleocapsids of the virus. To isolate the enveloped nucleocapsids, occlusion bodies were solubilized under alkaline conditions (0.5 M sodium carbonate pH 10.5 for 60 min on ice), and undissolved granules were separated by centrifugation (5,000 g for 10 min). The supernatant was then layered on a sucrose gradient (10% to 50%) and centrifuged at 25,000 g (Beckman SW-41) for 60 min at 4°C. Enveloped nucleocapsids were then collected, and after being washed in distilled water, centrifuged at 100,000 g (Beckman SW-41) for 60 min at 4°C, then resuspended in a TE buffer (10 mM-Tris-HCl, 1 mM-EDTA, pH 7.4). The purity and integrity of the enveloped nucleocapsids were examined by transmission electron microscopy. For DNA extraction, the purified-enveloped nucleocapsids of ChfuGV were digested in 250 µg/ml proteinase K in the presence of 0.5% SDS, for 1 h at 37°C. Digested proteins were extracted once with saturated phenol by adding equal volumes of phenol to the preparation, twice with one volume of phenol: chloroform (1 : 1 v/v) and once with an equal volume of chloroform: isoamylalcohol (24 : 1 v/v). The DNA that was present in the aqueous phase was precipitated with 2.5 volume of 100% ethanol in the presence of 0.1 volume sodium acetate (3 M, pH 5.2) for 3 h at 4°C, then pelleted at 16,000 g for 30 min. The purified DNA was dried and dissolved in a TE buffer. The viral genome was digested by *Bam*HI, and the fragments were ligated into pSK-Bluescript (Stratagene, La Jolla, USA). During the sequence analysis of the 11 kb *Bam*HI restriction fragment of the ChfuGV genome, a homologue of the baculoviral *ie-1* gene was detected. The sequencing was performed using specific primers in both directions by primer walking (Sanger *et al.* 1977).

Computer analysis The sequences were analyzed by the Sequencer program version 4.0.5 (Gene Codes Corp., Ann Arbor, USA) and the MacVector program version 4.5.0 (Eastman Kodak). The nucleotide sequence and its predicted amino acid sequence were compared to homologues in GenBank/EMBL and SWISSPROT by using BLAST algorithm (Altshul *et al.*, 1992). CLUSTALX (1.81) (Thompson *et al.*, 1997) was used for multiple amino acid sequence alignments. BOXSHADE was used for similarity shading and scoring among the aligned sequences. The alignment was used as the input to construct the phylogenetic tree with Branch and Bound search settings of PAUP 4.0b4a (Swofford, 2000). Amino acid composition and isoelectric points were calculated using Peptide Statistic (http://web.umassmed.edu/cgi-bin/biobin/pepstats) and ProtParam (http://www.expasy.ch/tools/protparam.html) tools. Hydropathy plots were obtained using Kyte and Doolittle hydropathy values (−4.5 to 4.5) via Nixon web tools (http://www.bmb.psu.edu/nixon/webtools.html).

Results and Discussion

During sequencing of the ChfuGV genome library, a putative gene that encodes a protein homologue to the baculoviral IE-1 protein was identified. The *ie-1* gene from ChfuGV has a 1278 nucleotide open-reading frame (ORF) that encodes 426 amino acids with a predicted molecular mass of 50.33 kDa (Fig. 1). The amino acid composition in terms of non-polar (ACFGILMPVWY), polar (DEHKNQRST), basic (HRK), and acidic residues (DE) indicates a content of non-polar (48.4%), polar (51.6%), acidic (11%) and basic (16.7%). The protein is particularly rich in leucine (10.3%) and lysine (10.1%) residues. Table 1 presents the amino acid composition, predicted molecular mass, and isoelectric points of all of the known baculoviral IE-1 proteins.

The upstream region of the *ie-1* gene from ChfuGV, along with all other granuloviruses, was studied with the aim of detecting all possible *cis*-acting control elements within this region. Previous studies on the promoter region of the *ie-1* gene from AcMNPV pointed out that these elements were located from −34 to −546 nt of the start triplet (Pullen and Friesen, 1995a; Leisy and Rohrmann, 2000). To identify all of the *cis*-acting DNA elements that regulate the ChfuGV *ie-1* gene expression, along with other known granuloviral *ie-1* genes, we examined a region that extends from nucleotide −1 to −800 within the noncoding leader region of all granuloviral *ie-1* genes. The outcome revealed the presence of several *cis*-acting regulatory elements within the promoter region of the *ie-1* gene of ChfuGV and other studied granuloviruses.

When these regions in granuloviruses were compared to the same regions in NPVs, no significant preservation was observed between the studied sequences from the two genera. Furthermore, an alignment of all of the granuloviral *ie-1* gene noncoding leader regions revealed that, although the similar motifs were identified in these regions, still only a small conservation level was observed among them (data not shown).

The tetranucleotide CAGT is the consensus baculovirus early-transcriptional start site of numerous baculovirus genes (Blissard *et al.*, 1992; Pullen and Friesen, 1995b). Current evidence suggests that the CAGT motif is an initiator element that is required for proper *ie-1* transcription during AcMNPV infection. This motif plays a critical role in accelerating the expression of viral genes, such as *ie-1*, that are vital in launching a productive infection. These data suggest that, in AcMNPV, the CAGT motif affects the rate by which *ie-1* is transcribed and the deletion of sequences, including CAGT, that is eliminated early in the *ie-1* transcription. In AcMNPV, the nucleotides that surround the CAGT motif are adequate for appropriate transcription in a TATA-independent mode (Pullen and Friesen, 1995b). Although no CAGT motif was detected in the *ie-1* promoter region of ChfuGV, we detected a similar motif (TCATT) that was centered at position −482. The motif (A/C/T)CA(G/T)T is a proposed consensus for arthropod transcriptional initiator elements. A large group of

arthropod polymerase II promoters contains one or more copies of the pentanucleotide TCAGT, or one of its cognates

Fig. 1. Nucleotide sequence and predicted amino acid sequence of *ie-1* gene from ChfuGV. All possible *cis*-acting elements located within the noncoding leader region are in underlined italic. Putative polyadenylation signal is in underlined bold, astrix (*) shows the stop codon.

Table 1. Amino acid compositions (%), isoelectric points and molecular mass of ChfuGV and other baculoviral IE-1 proteins. Peptide Statistic and ProtParam tools were used for calculation

	ChfuGV	CpGV	XcGV	PxGV	AcMNPV	OpMNPV	BmNPV
(%)							
Non polar	48.4	43.9	45.2	47.6	46.4	46.6	46.7
Polar	51.6	56.1	54.8	52.4	53.6	53.4	53.3
Acidic	11.0	13.1	10.7	13.7	11.5	12.3	11.8
Basic	16.7	17.4	16.9	17.0	12.9	14.6	12.3
(%)							
Ala	2.1	3.9	3.9	2.5	5.3	6.6	5.7
Arg	4.0	6.1	3.9	4.6	2.2	5.2	2.1
Asn	8.0	8.0	6.4	9.4	9.5	8.6	9.9
Asp	6.3	5.3	6.6	6.9	5.3	6.4	5.7
Cys	1.4	1.6	1.2	2.3	1.5	0.9	1.5
Gln	3.8	5.3	4.1	2.8	4.0	4.8	3.9
Glu	4.7	7.8	4.1	6.9	6.2	5.9	6.2
Gly	3.3	2.3	1.7	2.0	2.4	3.4	2.6
His	2.6	3.7	5.2	2.5	2.7	1.8	2.2
Ile	8.7	4.3	7.2	6.6	4.6	3.6	4.6
Leu	10.3	7.8	8.9	9.2	7.7	8.2	7.7
Lys	10.1	7.6	7.9	9.9	7.9	7.7	8.0
Met	3.5	4.1	3.1	3.8	1.9	3.4	2.1
Phe	5.4	4.3	4.3	4.1	5.8	4.8	5.7
Pro	0.9	2.0	4.3	2.5	3.1	3.4	3.3
Ser	6.3	6.1	8.7	5.3	9.3	7.5	8.9
Thr	5.9	6.1	7.9	4.1	6.5	5.5	6.3
Trp	0.9	1.0	1.2	0.8	0.2	0.4	0.2
Tyr	5.4	4.9	4.3	5.6	6.4	4.8	6.3
Val	6.3	7.6	5.0	8.1	7.4	7.1	7.2
Isoelectric Point	9.17	8.23	8.61	8.19	6.09	8.28	5.79
Molecular Mass(kDa)	50.33	57.92	56.42	46.60	66.88	64.31	66.92

TCATT, ACAGT, and GCAGT. The high conservation level of this motif at the promoter region of arthropods implies the significance of these promoter elements (Cherbas and Cherbas, 1993). The presence of this motif in ChfuGV, as well as CpGV and XcGV, suggests the importance of this motif as a transcriptional initiator element in granuloviruses.

The only TATA element in the *ie-1* promoter region of ChfuGV is located 203nt upstream of the start codon triplet. Potential TATA motifs were observed in the *ie-1* promoter region of all of the GVs that were studied; however, only two of them (PxGV and CpGV) had these elements located within a reasonable distance from the putative ATG translation initiation codon. A CCAAT element was located 92 nt upstream of the start triplet of the *ie-1* gene in ChfuGV on the opposite strand. Another CCAAT element was also located at position −420 nt. CCAAT elements were also detected in the promoter region of PxGV. Two late motifs (TAAG) were observed at positions −368 and −508 from the putative start triplet. The late motifs were detected in all of the available GV promoter regions. A single polyadenylation signal, AATAAA, was located 18nt downstream of the putative translational stop codon of ChfuGV *ie-1* (Fig. 1).

The multiple sequence alignment of ChfuGV and other granuloviral IE-1 proteins revealed that the C-terminal regions of these proteins are more conserved when compared to the N-terminal regions (Fig. 2). This finding confirms the data from nucleopolyhedral IE-1 (Olson *et al.*, 2001; Wang *et al.*, 2001). Available data on IE-1 proteins from NPVs propose that the C-terminal regions of these proteins play a critical role in DNA binding; the N-terminal regions of IE1, which are rich in acidic residues, have been hypothesized to be an acidic activation domain (Theilmann and Stewart, 1991; Rodems *et al.*, 1997; Forsythe *et al.*, 1998; Leisy and Rohrmann, 2000; Olson *et al.*, 2001). Two stretches of amino acid sequences are well-conserved in the C-terminal regions of all granuloviral IE-1 proteins. They are ^{332}[R/H][Y/F]R[L/I]NCF[K/R]^{339} and ^{345}[L/V]W[I/V]NS[M/I]V^{351}. Another well-conserved amino acid sequence, ^{393}KL[V/I]IRY[I/V]L^{400}, was detected in all of the GVs, except PxGV. Although PxGV kept almost the same pattern, the second, fourth, and the eighth residues were found to be shifted to F, M, and M, respectively.

The deduced amino acid sequence of ChfuGV IE-1 was compared to other known baculoviral IE-1s. Results revealed a non-significant similarity between granuloviral and

Table 1. Continued

	EpNPV	LdMNPV	SeMNPV	TnSNPV	HaNPV	CfNPV
(%)						
Non polar	46.03	50.53	44.26	41.68	45.84	44.46
Polar	53.96	49.47	55.74	58.32	54.16	55.54
Acidic	12.59	15.19	12.47	11.77	10.89	12.68
Basic	13.31	15.37	15.47	15.43	16.04	14.11
(%)						
Ala	4.9	8.8	4.1	4.1	3.8	3.9
Arg	4.1	6.2	3.5	3.0	4.7	4.6
Asn	10.6	4.9	8.7	13.1	9.7	8.2
Asp	5.8	6.2	7.1	7.3	4.7	5.9
Cys	0.9	2.5	0.8	1.2	0.6	0.9
Gln	3.8	3.9	3.5	3.7	3.9	5.2
Glu	6.8	9.0	5.3	4.5	6.2	6.8
Gly	2.2	3.0	3.1	2.0	2.4	2.9
His	2.2	1.4	2.9	1.9	3.0	2.0
Ile	4.9	4.6	7.6	7.7	6.4	5.2
Leu	8.8	8.0	6.4	6.2	7.9	7.5
Lys	7.0	7.8	9.1	10.6	8.3	7.5
Met	4.0	2.7	4.1	2.4	3.8	2.9
Phe	4.7	4.2	4.5	4.3	4.2	5.2
Pro	3.6	4.4	3.6	4.3	4.7	3.0
Ser	6.3	6.4	9.2	8.4	8.6	8.9
Thr	7.4	3.7	6.3	6.0	5.0	6.4
Trp	0.5	0.9	0.3	0.3	0.3	0.4
Tyr	5.0	4.6	4.1	3.2	4.2	5.2
Val	6.7	6.9	5.7	5.8	7.6	7.5
Isoelectric Point	5.92	5.85	7.64	8.92	9.17	6.46
Molecular Mass(kDa)	64.63	65.08	81.94	84.52	76.49	64.95

nucleopolyhedral IE-1 proteins (Table 2). Even the hydrophilicity plot analysis demonstrated no significant structural conservation between the IE-1 proteins and the GVs and NPVs (Fig. 3). Despite the lack of significant homology between amino acid sequences of the IE-1 proteins of GVs and NPVs, the IE-1 protein from ChfuGV shares a conserved domain (residues 387 to 414) with nucleopolyhedral IE-1s (Fig. 4). This conserved domain was originally discovered within the C-terminal region of NPVs IE-1 proteins. Each domain contains two amphipathic á-helices that are led by a group of basic residues, and forms a bHLH-like domain. The HLH protein family of transcription factors is characterized by highly-conserved DNA binding, dimerization, and transactivation motifs, and contains very important elements that have been identified in different living organisms that range from yeast to humans (Atchley *et al*., 2000). Despite the high levels of similarity inside the conserved domain, the HLH proteins usually exhibit considerable sequence divergence outside this region. In AcMNPV IE-1, the existence of hydrophobic residues within the HLH-like domain in C-terminus is essential for IE-1 oligomerization and contributes to IE-1 stability (Rodems *et al*., 1997; Slack and Blissard, 1997). It seems that the hydrophobic surface of

the amphipathic helices is directly involved in the homophilic interaction between IE-1 monomers. The basic residues that precede the HLH-like domain of baculoviral IE-1 can be regarded as elements that are involved in DNA-binding activity. These residues may interact through ionic interactions with the negatively-charged phospho-sugar backbone during binding to the DNA molecule.

A highly-conserved stretch of five amino acids could be marked within this domain in most of the known baculoviral IE-1 proteins, apart from CpGV, XcGV, and PxGV. The conserved sequence is [V/L][V/L]KL[V/I/L] (residues 391 to 395 in ChfuGV IE-1). Similar sequences could also be detected in CpGV (VSKLV), XcGV (V-KLI) and PxGV (V-KFV) (Fig. 4). A multiple sequence alignment of deduced amino acid sequences of granuloviral and nucleopolyhedral IE-1 proteins revealed that, beside the basic helix-loop-helix like (bHLH-like) domain, no significant conservation could be detected among IE-1 proteins from the two genera of *Baculoviridae* family. The IE-1 proteins from ChfuGV and CpGV showed a high degree of similarity and shared several distinct motifs within both the C and N-terminals. Five such motifs were located within the N-terminal ([63]N[K/H]F[R/K]TTT[69], [71]HMF[I/V]C[75], [82]YVKNER[Y/F][88], [102]GNA[F/

(A)

```
ChfuGV_    1  ------------------------------MKSEMDISEVKTGRGAAYLFNLTSIIIIM
CpGV_      1  --------------MEVETTMYLEDQIMTTNDSENELDGDLEEEARGASHTFANQNFVKLV
XcGV_      1  MDEFHTLSNIDSQMQLLDAHYLQNPTLNTDQEENLDDLELSDDLCDSDGTKLKEKQGVSKE
PxGV_      1  --------------------MSDLLTTRDSEDE-DCDDIEEP--------------
```

```
ChfuGV_   30  SILDNIVNPDK------------------TESVQETQFQDEDDKLTIFRWLNHFRTTTF
CpGV_     48  NNPREQQQPPREQQQPSELNRIFLCP-INYNKITTINDDADEDCVSENYKWFNKFKTTTC
XcGV_     61  SSSLFKTMPDSLLPFNKRSFVLPENTNALSTSSSSKQHQWPAPDLNTCRWLSMYTTTSN
PxGV_     22  ------------------------SPAKQPRDNNGAVEDWILKYKTTLI
```

```
ChfuGV_   71  HMFICNWDGIKYVKNERYCEDVYLRKYHQVNGNAYELLLKDCKFFITTKLIQHGKIFSFP
CpGV_    107  HMFVCHSR-VEYVKNERFPAEVYMQKYHELYGNAFELTIKDAKFYITTRLLQLLKVDF--
XcGV_    121  HMVICHID-VIYVRNIRFGSALYLDNYHTKYGNTYEIHILSV-----------------
PxGV_     47  NMFVCHDS-FEYYSLVRHPSSVYMKMFYRQYGSAYELHINHIKVYVTEKCLEKNGVGKNR
```

```
ChfuGV_  131  SEKVMKHVDIKI--------------
CpGV_         --------------------------
XcGV_         --------------------------
PxGV_    106  YPNESVQVNYPYSEQVFLLGLKVDVK
```

(B)

```
ChfuGV_    1  -----------------------------VT--PLLQIIMKDGNSLHQHIVEALINTVF
CpGV_      1  ---------------MWHPAESLIKHVDSVTSARILNLRMRDGYTVQQHIIEKLFYVLL
PxGV_      1  ----------------------------------GVTDRLTIQAAIVLQFNLINV
XcGV_      1  KLYVTSTLLKINNSPNDLYPNDMVEKIKDAAAISYLLNLCTKNGQTVMTQLILLFHSINL
```

```
ChfuGV_   29  IGISNNSVGVVKRRMLADNSAKRINKITNRYSWNSEKIKIRNLCESN-------------
CpGV_     45  IGLGPRDISAVMERINSNKATERCLAIQSKYSWNSPERKIKNVFCRNSTVASVTANINGS
PxGV_     22  EGLDNETINMVKKNLLNATIILNRDKKRKRPPNEDCTDPIAKLFEVS-------------
XcGV_     61  ISLSTEQIDSILNTLPIFVIHRLLAHMKKNYSWKSELYKISKIKTIP-------------
```

```
ChfuGV_   76  -------------DVQSLFDIMYKNGGYYK-----ETQDFTLDYLKFIACFNYNVQRLKTD
CpGV_    105  NSNTTCHTTNNKSIEKMFELLYKNGNTPR-----FIEDKELAHELYYRNFHQNVCQLKKH
PxGV_     69  ----------------ARLEYR-----------VEDKEYTHQSYYNIFNMAIN----N
XcGV_    108  ------------GQQTHIQLLLKASTYPRNKPLEYRKDQSYVHSRYVQEFYDAMQPTNTT
```

```
ChfuGV_  119  IIPHVDNISCTKMLVKNIAALLYKIENLRDMQELKDIKGNEQDVDDFLRISLNYPMGDVI
CpGV_    160  VTDYMDKVSNVRLLIKDLAMVLYKHEQKKELQNLKNEYGTEADVEEFNRLSVAHPRGDVI
PxGV_     96  NINVQDNVSNLTWYFKKIVNVTCDSPS-EELLNFKSFDCDYTTIKKFIMMSRMNKKLNVI
XcGV_    156  TAFTPDPVIHVKWLFSAIVKSVIVNPP-MSIFRLKIFDSTPNDIAEFLRASKSHPLANVI
```

```
ChfuGV_  179  FNMKTKGTNSQRYRLNCFKIDQVYLWINSMVY---KISKFDLESVILKYKTGTHYIISFK
CpGV_    220  FNMKVLDTNTQRYRINCFRMDSVHVWVNSMVYS--DVQQFNLKKMIQRHRWGTHHILQFD
PxGV_    155  LHTKVKYVNKLIYRLNCFKLCDDHVWINSIVLNKISKKLDLENIIFKEKYGSHYIINLK
XcGV_    215  LHTRAPEIGKEHERLNCFKLNKIHVWINSMVFDKIQHKRVDLKQIILNVAWGEHHIISFQ
```

```
ChfuGV_  236  YYTNTMLSKLHSEVVKLVIRYILSKREYSLLENDIKIN-NKLLYKCLIFS
CpGV_    278  YMMNSMESKLHAEVSKLVIRYILSRRSFDLLQNDCSK--LKLSYKKIVYE
PxGV_    215  YMNNQELRIVHMETVKFVMKYIMERRDLALLKNDIDCI-SKMYFKCIRF-
XcGV_    275  YIHNQKLKQMHIETVKLIIRYILERRNFIKLAKDMARCPRKIKYDYIIF-
```

Fig. 2. Alignment of the N-terminal (a) and C-terminal (b) of granuloviral IE-1 protein. The alignment was processed using CLUSTAL X software and BOX SHADE tool. Two shading levels were set: gray for more than 50% and black for 100% identity. Accession numbers for the sequences used are as follow: CpGV AAK70674, XcGV CQ9PZ32; PxGV Q9DW20.

Y]E[L/M][107], and [113]KF[F/Y]ITT[K/R]L[I/L]Q[122]) while seven were found in the C-terminal region ([222][L/M]F[D/E][I/L][M/ L]YKNG[230], [303]DV[D/E][D/E]F[L/M]R[I/L]S [L/V][312], [317]GDV[L/V]FNMK[324], [331]QRYR[L/I]NCF[K/R][I/M]D[341],

Table 2. Comparison of deduced amino acid sequence of ChfuGV IE-1 with that of three GVs and ten NPVs IE-1 proteins. Pair wise identity values (%) between taxa are shown

|	CpGV	XcGV	PxGV	HaNPV	HzNPV	EpNPV	BmNPV	AcM-NPV	SeNPV	LdM-NPV	TnSNPV	OpM-NPV	CfNPV
ChfuGV	39.4	24.5	23.4	13.3	12.9	14.3	14.2	14.0	12.5	14.1	12.7	13.0	13.4
CpGV		24.9	22.1	13.1	12.6	11.1	12.0	11.7	11.5	11.2	12.0	10.8	11.3
XcGV			26.9	10.0	10.1	10.8	13.7	12.8	11.5	11.9	12.3	11.2	10.8
PxGV				13.8	13.7	12.6	13.3	12.8	12.1	12.9	11.0	12.6	12.3
HaNPV					98.0	26.0	25.9	26.1	27.0	28.0	22.3	25.9	25.8
HzNPV						26.4	26.0	26.2	27.5	28.1	22.7	25.9	26.1
EpNPV							48.8	49.4	23.7	25.3	22.8	67.0	66.9
BmNPV								95.9	25.3	25.5	21.5	47.5	47.6
AcMNPV									24.8	24.8	21.5	47.9	47.5
SeNPV										25.2	22.4	24.3	24.1
LdMNPV											24.4	25.4	25.5
TnSNPV												22.0	23.1
OpMNPV													74.4
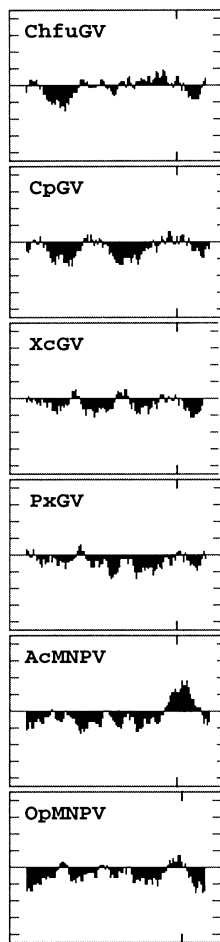
Fig. 3. Hydrophilicity profiles from C-termial regions of all known granuloviral and two nucleopolyhedral IE-1 proteins. Above the axis (+) denotes hydrophilic regions and below the axis (-) indicates hydophobic regions.

^{345}[L/V]W[I/V]NSMVY^{352}, ^{383}M[L/M]SKLH^{388}, and ^{393}KLVIRY[I/V]LS[K/R]R^{403}).

A phylogenetic analysis of the IE-1 proteins is shown in Fig. 5. The tree was produced via maximum parsimony to estimate the evolutionary relationship between baculoviral IE-1 proteins. These analyses show a clear division between the GV and NPV IE-1 proteins. Two main branches for NPVs were found in this tree. However, due to the limited number of known granuloviral IE-1 proteins, the tree cannot present a complete picture of this genus. ChfuGV is positioned beside CpGV in the same cluster, supported by a 99% bootstrap value.

This study provides further information about the importance of the basic helix-loop-helix like domain in baculoviral IE-1 proteins. It also demonstrates that the presence of several potential *cis*-acting regulatory elements within the upstream region of the *ie-1* gene is a common
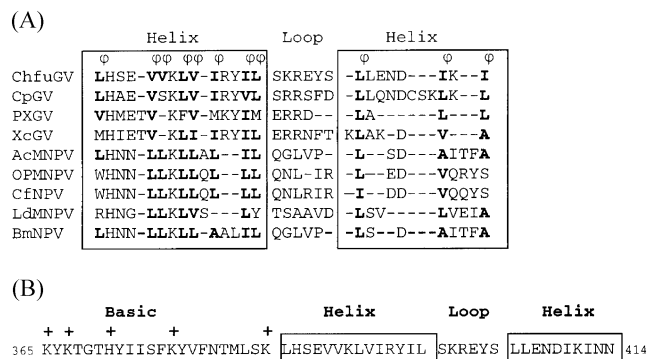

Fig. 4. (a) HLH-like domain of IE-1. ChfuGV IE-1 residues 387 to 414 were aligned with equivalent IE-1 residues from NPVs and other GVs. Conserved hydrophobic (φ) residues are illustrated. (b) Basic residues (+) preceding the HLH-like domain of ChfuGV IE-1 are depicted
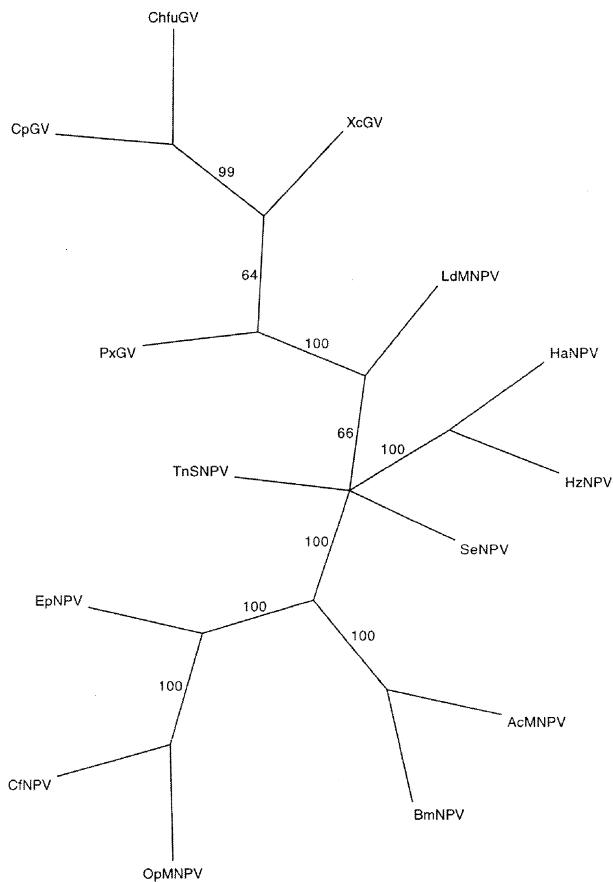
Fig. 5. Phylogenetic analysis performed on baculoviral IE-1 proteins. The unrooted tree was produced by a Branch and Band search using PAUP4.0b4a. Bootstrap values (100 replicates) are shown. Tree was constructed according to the following settings: 1-uninformative characters ignored. 2-Branch-and-bound search options: an initial upper bound: compute via stepwise; b keep minimals only; c collapse zero-length branches ON; MULPARS ON; e addition sequences: furthest. Accession numbers for the sequences used are as follow: CpGV AAK70674, XcGV CQ9PZ32; PxGV Q9DW20; HaNPV AAK96269; SlNPV AAL01702; HzNPV 010619; AcMNPV P11138; BmNPV P33245; EpNPV AAK85693; OpMNPV 22114; CfNPV P41716; SeNPV Q9J811; LdMNPV 036454; TnSNPV AAL14888.

property in all known baculoviral IE-1 proteins. It will be of interest to determine the cellular location of the *ie-1* gene product during the process of infection.

References

Altshul, S. F., Gish, W., Miller, W., Meyers, E. W. and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410.

Atchley, W. R., Wollenberg, K. R., Fitch, W. M., Terhalle, W. and Dress, A. W. (2000) Correlations among amino acid sites in bHLH protein domains: an information theoretic analysis. *Mol. Biol. Evol.* **17**, 164-178.

Bah, A., Bergeron, J., Arella, M., Lucarotti, C. J. and Guertin, C. (1997) Identification and sequence analyses of the granulin gene of *Choristoneura fumiferana* granulovirus. *Arch. Virol.* **142**, 1577-1584.

Blissard, G. W., Kogan, P. H., Wei, R. and Rohrmann, G. F. (1992) A synthetic early promoter from a baculovirus: roles of the TATA box and conserved start site CAGT sequence in basal levels of transcription. *Virology* **190**, 783-793.

Cherbas, L. and Cherbas, P. (1993) The arthropod initiator: the capsite consensus plays an important role in transcription. *Insect Biochem. Mol. Biol.* **23**, 81-90.

Choi, J. and Guarino, L. A. (1995) A temperature-sensitive IE1 protein of *Autographa californica* nuclear polyhedrosis virus has altered transactivation and DNA binding activities. *Virology* **209**, 90-98.

Crook, N. E. (1991) Baculoviridae: subgroup B: comparative aspects of granulosis viruses; in *Viruses of Invertebrates*, Kurstak, E. (ed.), pp. 73-110, Marcel Dekker, New York.

Forsythe, I. J., Shippam, C. E., Willis, L. G., Stewart, S., Grigliatti, T. and Theilmann, D. A. (1998) Characterization of the acidic domain of the IE1 regulatory protein from *Orgyia pseudotsugata* multicapsid nucleopolyhedrovirus. *Virology* **252**, 65-81.

Forté, A. J., Guertin, C. and Cabana, J. (1999) Pathogenicity of a granulovirus towards *Choristoneura fumiferana*. *Canadian Entomologist* **131**, 725-727.

Guarino, L. A. and Dong, W. (1994) Functional dissection of the *Autographa californica* nuclear polyhedrosis virus enhancer element hr5. *Virology* **200**, 328-335.

Guarino, L. A. and Summers, M. D. (1986) Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. *J. Virol.* **57**, 563-571.

Hayakawa, T., Rohrmann, G. F. and Hashimoto, Y. (2000) Patterns of genome organization and content in lepidopteran baculoviruses. *Virology* **278**, 1-12.

Huh, N. E. and Weaver, R. F. (1990) Categorizing some early and late transcripts directed by the *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol.* **71**, 2195-2200.

Kool, M., Ahrens, C., Goldbach, R. W., Rohrmann, G. F. and Vlak, J. M (1994a) Identification of genes involved in DNA replication of the *Autographa californica* baculovirus. *Proc. Nat. Acad. Sci. USA* **91**, 11212-11216.

Kool, M., Voeten, J. T. M., Goldbach, R. W., Tramper, J. and Vlak, J. M. (1993) Identification of seven putative origins of *Autographa californica* nuclear polyhedrosis virus DNA replication. *J. Gen. Virol.* **74**, 2661-2668.

Kool, M., Voeten, J. T. M. and Vlak, J. M. (1994b) Functional mapping of regions of the *Autograha californica* nuclear polyhedrosis virus. *J. Gen. Virol.* **75**, 3345-3352.

Kovacs, G. R., Choi, J., Guarino, L. A. and Summers, M. D. (1992) Functional dissection of the *Autographa californica* nuclear polyhedrosis virus immediate-early 1 transcriptional regulatory protein. *J. Virol.* **66**, 7429-7437.

Kremer, A. and Knebel-Morsdorf, D. (1998) The early baculovirus

he65 promoter: On the mechanism of transcriptional activation by IE1. *Virology* **249**, 336-351.

Leisy, D. J. and Rohrmann, G. F. (2000) The *Autographa californica* nucleopolyhedrovirus IE-1 protein complex has two modes of specific DNA binding. *Virology* **274**, 196-202.

Lewin, B. (1997) Genes VI. Oxford University Press, New York.

Massari, M. E. and Murre, C. (2000). Helix-loop-helix proteins: regulators of transcription in eukaryotic organisms. *Mol. Cell Biol.* **20**, 429-440.

Moscardi, F. (1999) Assessment of the application of baculoviruses for control of lepidoptera. *Annu. Rev. Entomol.* **44**, 227-289.

Nissen, M. S. and Friesen, P. D. (1989) Molecular analysis of the transcriptional regulatory region of an early baculovirus gene. *J. Virol.* **63**, 493-503.

Okano, K., Mikhailov, V. S. and Maeda, S. (1999) Co-localization of baculovirus IE-1 and two DNA-binding proteins, DBP and LEF-3, to viral replication factories. *J. Virol.* **73**, 110-119.

Olson, V. A., Wetter, J. A. and Friesen, P. D. (2001) Oligomerization mediated by a helix-loop-helix-like domain of baculovirus IE1 is required for early promoter transactivation. *J. Virol.* **75**, 6042-6451.

Pullen, S. S. and Friesen, P. D. (1995a) The CAGT motif functions as an initiator element during early transcription of the baculovirus transregulator ie-1. *J. Virol.* **69**, 3575-83.

Pullen, S. S. and Friesen, P. D. (1995b) Early transcription of the *ie-1* transregulator gene of *Autographa californica* nuclear polyhedrosis virus is regulated by DNA sequences within its 5' noncoding leader region. *J. Virol.* **69**, 156-65.

Ribeiro, B. M., Hutchinson, K. and Miller, L. K. (1994) A mutant baculovirus with a temperature-sensitive IE-1 transregulatory protein. *Virology* **68**, 1075-1084.

Rodems, S. M. and Friesen, P. D. (1993) The hr5 transcriptional enhancer stimulates early expression from the *Autographa californica* nuclear polyhedrosis virus genome but is not required for virus replication. *J. Virol.* **67**, 5776-5785.

Rodems, S. M., Pullen, S. S. and Friesen, P. D. (1997) DNA-dependent transregulation by IE1 of *Autographa californica* nuclear polyhedrosis virus: IE1 domains required for transactivation and DNA binding. *Virology* **71**, 9270-9277.

Sanger, F., Nichlen, S. and Coulson, A. R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**, 5463-5467.

Slack, J. M. and Blissard, G. W. (1997) Identification of two independent transcriptional activation domains in the *Autographa californica* multicapsid nuclear polyhedrosis virus IE1 protein. *Virology* **71**, 9579-9587.

Swofford, D. L. (2000) PAUP Version 4, Sinauer Associates, Sunderland, Massachusetts.

Theilmann, D. A. and Stewart, S. (1993) Analysis of the *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus trans-activators IE-1 and IE-2 using monoclonal antibodies. *J. Gen. Virol.* **74**, 1819-1826.

Theilmann, D. A. and Stewart, S. (1991) Identification and characterization of the IE-1 gene of *Orgyia pseudotsugata* multicapsid nuclear polyhedrosis virus. *Virology* **180**, 492-508.

Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. and Higgins, D. G. (1997) The ClustalX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **24**, 4876-4882.

Wang, W., Leat, N., Fielding, B. and Davison, S. (2001) Identification, sequence analysis, and phylogeny of the immediate early gene 1 of the *Trichoplusia ni* single nucleocapsid polyhedrosis virus. *Virus Genes* **23**, 53-62.