Journal of Biochemistry and Molecular Biology, Vol. 36, No. 5, September 2003, pp. 475-487

JOURNAL OF Biochemistry and Molecular Biology

Choristoneura fumiferana Granulovirus p74 Protein, a Highly Conserved Baculoviral Envelope Protein

Kianoush Khajeh Rashidan[†,]*, Nasha Nassoury[‡], Samia Tazi[†], Paresa N. Giannopoulos[†] and Claude Guertin[†]

[†]Institut national de la recherche scientifique-Institut Armand-Frappier 531 Blvd des Prairies, H7V 1B7, Laval, Canada

[‡]Université de Montréal, 4101 Sherbrook east, H1X 2B2 Montreal, Canada

Received 3 December 2002, Accepted 7 April 2003

A gene that encodes a homologue to baculoviral p74, an envelope-associated viral structural protein, has been identified and sequenced on the genome of *Choristoneura fumiferana* granulovirus (ChfuGV). A part of the ChfuGV p74 gene was located on an 8.9 kb *Bam*HI subgenomic fragment using different sets of degenerated primers. These were designed using the results of the protein sequencing of a major 74 kDa structural protein that is associated with the occlusion-derived virus (ODV). The gene has a 1992 nucleotide (nt) open-reading frame (ORF) that encodes a protein with 663 amino acids with a predicted molecular mass of 74,812 Da. Comparative studies revealed the presence of two major conserved regions in the ChfuGV p74 protein. This study also shows that all of the p74 proteins contain two putative transmembrane domains at their C-terminal segments. At the nucleotide sequence level, two late promoter motifs (TAAG and GTAAG) were located upstream of the first ATG of the p74 gene. The gene contained a canonical poly(A) signal, AATAAA, at its 3' non-translated region. A phylogenetic tree for baculoviral p74 was constructed using a maximum parsimony analysis. The phylogenetic estimation demonstrated that ChfuGV p74 is related the closest to those of *Cydia pomonella* granulovirus (CpGV) and *Phthorimaea operculella* granulovirus (PhopGV).

Keywords: Baculovirus, ChfuGV, p74, Envelope-associated protein, Phylogenetic analysis, Protein analysis.

Introduction

Granuloviruses (GVs) are members of *Baculoviridae*, a family of large rod-shaped enveloped viruses with relatively large double-stranded DNA genomes. *Choristoneura fumiferana* granulovirus (ChfuGV) virion is occluded in a highly organized proteinous matrix that is known as granulin. The infection cycle commences when *Choristoneura fumiferana* (spruce budworm) larvae ingest the occluded virions; subsequently, the proteinous matrix of the occlusion bodies then dissociates under the alkaline pH condition in the midgut and releases enveloped virions that are known as occlusion-derived virus (ODV).

Like other members of the family, ChfuGV produces two distinct forms of virion during a single infection cycle: occlusion-derived virus (ODV) and budded virus (BV). ODVs transmit infection solely from insect to insect, whereas BVs spread infection from infected cells to healthy cells within the infected host (Keddie *et al*., 1989). BVs acquire their envelopes at the plasma membrane of the infected cell (budding) (Blissard and Rohrmann, 1989; Whitford *et al*., 1989). In contrast, ODVs acquire their envelope within the nucleus of the infected cell (Braunagel and Summers, 1994). The attachment of ODVs to microvilli of insect midgut cells followed by the virus penetration is considered vital steps for the initiation of the infection cycle in the host. The protein(s) that could participate in these steps has not yet been identified, but there is evidence suggesting that the ODV penetration is a non-endocytotic process and requires the interaction of the virion envelope proteins with midgut cell surface proteins which eventually leads to membrane fusion (Horton and Burand, 1993).

The ODV phenotype of ChfuGV, like all other known baculoviruses ODVs, has a complex structure. Baculoviral ODV envelopes contain several proteins. These include VP17 (Funk and Consigli, 1993), ODV-E25 (Russell and Rohrmann, 1993), ODV-E35 (Braunagel *et al*., 1996a), GP41 (Whitford and Faulkner, 1993), p74 (Kuzio *et al*., 1989), ODV-E18

The nucleotide sequence data reported in this paper has been submitted to the GenBank nucleotide sequence database and assigned the accession number AAL13071.

*To whom correspondence should be addressed.
Tel: 204-233-0210; Fax: 204-237-3240
E-mail: kiarashidan@juno.com

(Braunagel *et al*., 1996a), ODV-E66 (Hong *et al*., 1994), and ODVP-6E/ODV-E56 (Braunagel *et al*., 1996b, Theilmann *et al*., 1996; Rashidan *et al*., 2002). One or more of these proteins could participate in adsorption, fusion, and the penetration process. In *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV) (prototype of the *Baculoviridae* family), the p74 protein is exposed on the virion surface (Faulkner *et al*., 1997) and has an essential role in the infectivity of the virus (Kuzio *et al*., 1989; Faulkner *et al*., 1997). It has also been shown that although p74 null mutants fail to initiate infection in insect larvae when viruses are fed orally, they are infectious once injected into the haemocoel of the larvae (Faulkner *et al*., 1997). These observations suggest that the p74 protein is essential for the initiation of the infection. Its presence is also an indispensable element for the attachment and/or fusion of the ODV envelope to the membrane of midgut epithelial cells. This evidence places the baculoviral p74 protein in a particular position for consideration as an ODV envelope protein that may play a major role in the initiation of viral infection. It has been suggested that the N terminal of AcMNPV p74 is located outside of the ODV envelope while the C terminal acts as a transmembrane anchor (Faulkner *et al*., 1997). In AcMNPV, the hydrophobic C-terminal region plays a significant role in the proteins localization and its transmembrane anchoring character (Slack *et al*., 2001).

In the present study, we report the identification and characterization of a p74 homologue in ChfuGV. This is the first report devoted to the characterization of a granuloviral p74 protein, as previous reports concentrated on p74 protein homologues in nucleopolyhedrovirus (NPVs).

Materials and Methods

***In vivo* production and purification of virus and DNA extraction** *Choristoneura fumiferana* fourth-instar larvae were infected by ChfuGV using a virus-contaminated artificial diet (Forté *et al*., 1999). The virus isolation from infected larvae that was followed by the DNA extraction was carried out as previously described (Bah *et al*., 1997).

SDS-PAGE and protein microsequencing Enveloped nucleocapsids of ChfuGV were purified as previously described (Rashidan *et al*., 2002). The purity and integrity of the enveloped nucleocapsids were examined by a transmission electron microscope. Purified-enveloped nucleocapsids were disrupted in an equal volume of a Laemmli sample buffer (Laemmli, 1970) in the presence of β-mercaptoethanol (Bio-Rad, Hercules, USA). The samples were then boiled for 5 min and clarified by centrifugation at 13,000 *g* for 5 min before electrophoresis on a 3% stacking/ 12.5% separating sodium dodecyl sulfate-polyacrylamid gel electrophoresis (SDS-PAGE). The proteins were visualized using light Coomassie Brilliant Blue (Invitrogen, San Diego, USA) staining. The molecular mass (*Mr*) of the virion proteins was determined by comparing them to the standards (Bio-Rad).

Gel-purified ChfuGV structural proteins were used for the protein sequencing. Major protein bands from the ChfuGV-enveloped nucleocapsids preparation, including a 74 kDa band, were excised from the gel and subjected to a protein sequencing analysis. Sequencing was performed at the Harvard Microchemistry facility using either microcapillary reverse-phase HPLC nano-electrospray tandem mass spectrometry (μLC/MS/MS) on a Finnigan LCQ quadruple ion trap mass spectrometer or chemical sequencing (Via Edman degradation). The obtained sequences of several stretches of amino acids were used to design the degenerated primers (with the least redundancy) that were used in the PCR reactions on the ChfuGV genomic DNA in order to locate the gene coding for this protein on the ChfuGV genome.

Cloning and DNA sequencing Using different combinations of the degenerated primers, PCR amplifications were performed on ChfuGV genomic DNA. A part of the p74 gene was amplified with a combination of the following pair of degenerated primers, p74FRIII: ACA/G CCA/G AAT/C GCT/C AAT/C AAA and p74-RVV: TTT A/GAT T/CAA G/AGC C/TGA. All of the PCR reactions were carried out using Pwo DNA polymerase (Roche, Indianapolis, USA) and *Taq* DNA polymerase (Amersham Pharmacia Biotech, Uppsala, Sweden) in the buffer that was supplied by the manufacturer that was supplemented with 200 μM of each dNTP and $MgCl_2$ to a final concentration of 2.5 mM. The PCR reactions were subjected to heat for 10 min at 95°C prior to the start of the amplification (30 cycles of 95°C, 30 s; 47°C, 25 s; and 72°C, 40 s). The PCR amplicon was cloned in a PCR2.1 (Invitrogen, San Diego, USA) cloning vector and sequenced. The sequence was then compared to homologues in GenBank/EMBL by using BLAST algorithm (Altshul *et al*., 1990). In order to locate the p74 gene on the ChfuGV genome, subgenomic fragments were generated by *Bam*HI endonuclease and fractioned in 1% agarose gel. The gel-extracted fragments were then used as templates for the PCR reactions using the same set of primers. The PCR reaction on a 8.9 kb ChfuGV *BamH*I fragment yielded an amplicon with the expected size. This 8.9 kb restriction fragment was cloned in a pBluescript-SK^+ cloning vector (Stratagene, La Jolla, USA) and transformed into *E.coli* XL-1 Blue (Stratagene). All of the manipulations (including the restriction enzyme digestion, agarose gel electrophoresis, transformation, colony lifting, and plasmid purification) were carried out according to standard protocols (Sambrook *et al*., 1989). The fragment was sequenced from both

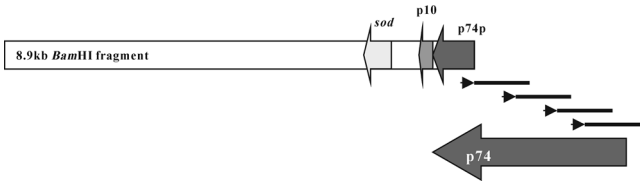


Fig. 1. Schematic presentation of the 8.9 kb BamHI fragment from ChfuGV genomic DNA showing the position and direction of the putative ORFs p74, p10, and *sod*. Only 573nt of p74 *orf* (denoted as p74p, "p" for partial) is on this fragment. The strategy that is employed for sequencing of complete p74 *orf* is presented. The arrows show the position of the primers that were used for the sequencing and the direction of sequencing.

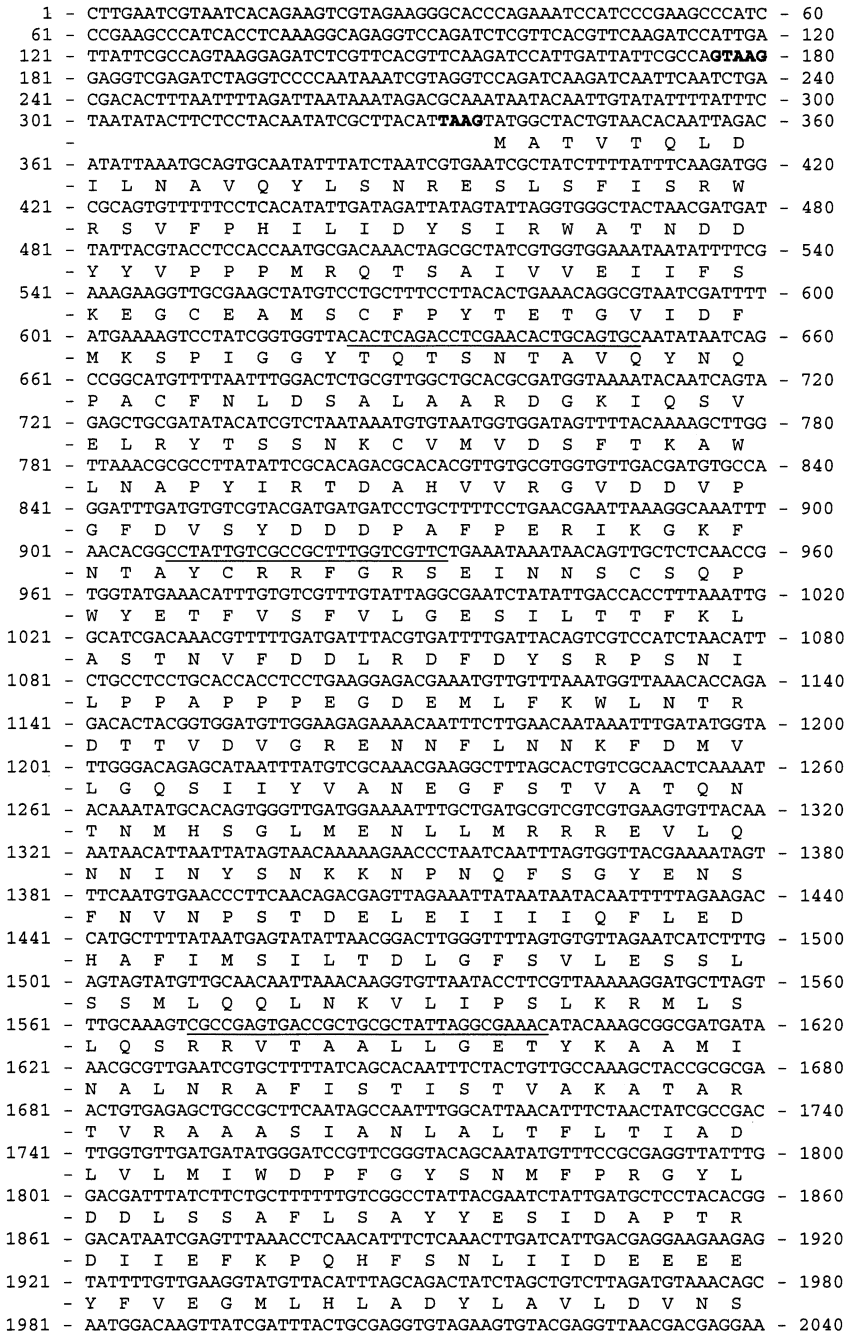


Fig. 2. The figure shows the nucleotide and protein sequence of ChfuGV p74. Two late promoter motifs and a putative polyadenylation signal for p74 gene are in bold. The positions of the degenerated primers are in underlined bold, and the locations of the primers that were used for sequencing are underlined. Asterisk (*) shows the stop codon.

extremities on either strand using the primer walking technique. Automatic sequencing were carried out using an Applied Biosystems automated DNA sequencer model 377XL (Applied Biosystems, Foster City, USA). The first series of sequencing results revealed that the pbam8.9 clone only contained a segment of p74. In order to attain the entire sequence of the p74 ORF, a primer was designed using the preliminary sequencing results. This primer was then employed in the direct sequencing of the ChfuGV genomic DNA. Primer walking on the genomic DNA was conducted until the complete sequencing of the p74 ORF was accomplished.

Computer analysis Sequence data assembly and analysis were performed with the Sequencher software version 4.0.5 (Gene Codes Corp., Ann Arbor, USA) and MacVector program version 4.5.0 (Eastman Kodak, Rochester, USA). The nucleotide sequence and its predicted amino acid sequence were compared to homologues in GenBank/EMBL and SWISSPROT by using BLAST (Altshul *et*

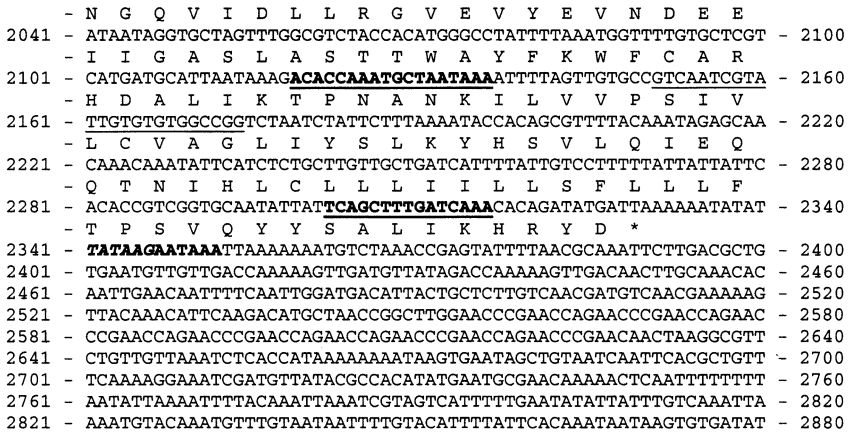


Fig. 2. Continued.

al., 1990). CLUSTALX (1.81) (Thompson *et al*., 1997) was used for the multiple amino acid sequence alignments. BOXSHADE version 3.21 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) with 100 bootstrap replicates. The amino acid composition was calculated using Peptide Statistic (http://web.umassmed.edu/cgi-bin/biobin/pepstats), Pepinfo (www.ebi.ac.uk/servicestmp/1598281014310854.html), and ProtParam (http://www.expasy.ch/tools/protparam.html) Tools. The hydropathy plots were obtained using Kyte and Doolittle hydropathy values (1982) via the EMBOSS Pepwindow program (http://www.ebi.ac.uk/emboss/pepinfo/). The secondary structure prediction method that was used in this study was GORIV (Garnier *et al*., 1996). Potential transmembrane regions were identified using the TMHMMver. 2.0 (Moller *et al*., 2001), Tmpred (Hofmann and Stoffel, 1993), and SOUSI programs (http://sosui.proteome.bio.tuat.ac.jp/sosuimenu0.html). Potential glycosylation and myristoylation sites were predicted using the NetOglyc (Hansen *et al*., 1998) and Proscan (Bairoch *et al*., 1997) programs. The prediction of the coiled-coil region was performed using the COILS program version 2.1 (Lupas *et al*., 1991). Accession numbers for the p74 protein sequences that were used in this study are as follows: *Cydia pomonella* granulovirus (CpGV), AAK70720; *Phthorimaea operculella* granulovirus (PhopGV), AAM70253; *Plutella xylostella* granulovirus (PxGV), AAG27347; *Xestia c-nigrum* granulovirus (XcGV), AAF05191; *Helicoverpa armigera* nucleopolyhedrovirus (HaNPV), AAK96273; *Spodoptera exigua* nucleopolyhedrovirus (SeNPV), AAF33660; *Orgyia pseudotsugata* multicapsid nucleopolyhedrovirus (OpMNPV), AAC59133; *Choristoneura fumiferana* multicapsid nucleopolyhedrovirus (CfMNPV), S29849; *Epiphyas postvittana* nucleopolyhedrovirus (EpNPV), AAK85685; *Bombyx mori* nucleopolyhedrovirus (BmNPV), AAC63805; *Autographa californica* multicapsid nucleopolyhedrovirus (AcMNPV), AAA66768; *Lymantria dispar* multicapsid nucleopolyhedrovirus (LdMNPV), AAC70212; *Spodoptera litura* nucleopolyhedrovirus (SlNPV), AAL01707; *Rachiplusia ou* nucleopolyhedrovirus (RoNPV), AAN28023; *Mamestra configurata* nucleopolyhedrovirus (McNPV), AAM09268; *Culex nigripalpus* baculovirus (CnBV), AAK94152; and *Heliothis zea* virus1 (Hz-1 virus) AAM45758.

Results and Discussion

In this report, we explained the identification and characterization of an ORF from ChfuGV that encodes a 74 kDa protein homologue to a highly conserved envelope protein, termed baculoviral p74. This protein has already been identified in a number of other baculoviruses. These include CpGV (Luque *et al*., 2001), PhopGV, PxGV (Hashimoto *et al*., 2000), XcGV (Hayakawa *et al*., 1999), AcMNPV (Ayres *et al*., 1994), OpMNPV (Ahrens *et al*., 1997), CfMNPV (Hill *et al*., 1993), LdMNPV (Kuzio *et al*., 1999), BmNPV (Gomi *et al*., 1999), SlNPV (Pang *et al*., 2001), SeNPV (Ijkel *et al*., 1999), HaNPV (Zhang and Wu 2001), EpNPV, RoNPV, McNPV (Li *et al*., 2002), CnBV (Afonso *et al*., 2001), and Hz-1 virus (Cheng *et al*., 2002). In a few other BVs, only partial sequences of the p74 protein are known. These include *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus (BsSNPV) (Hu *et al*., 1998), *Trichoplusia ni* single capsid nucleopolyhedrovirus (TnSNPV) (Fielding *et al*., 2002), and *Leucania separata* nucleopolyhedrovirus (LsNPV).

The results of the protein sequencing of a major 74 kDa protein in the ChfuGV-enveloped nucleocapsid yielded several stretches of amino acids. These amino acid sequences were found to be quite similar to a highly conserved ODV-specific envelope protein that is known as p74, after being compared to homologue proteins in GenBank/EMBL using BLAST algorithm (Altshul *et al*., 1990). The ORF coding for the ChfuGV p74 protein was located on the genome of ChfuGV, as described in the Materials and Methods Section.

Figure 2 shows the strategy that was adopted for the sequencing. The sequence and position of the primers that were used to complete the p74 sequence are presented in Fig. 2. The ORF is 1992 nt potentially encoding 663 amino acids with an estimated molecular mass of 74,812 Da. The 5 non-translated region (NTR) of the ChfuGV p74 gene was analyzed in order to identify all of the possible *cis*-acting elements and possible transcription start sites. No obvious TATA box element was detected, however two late promoter

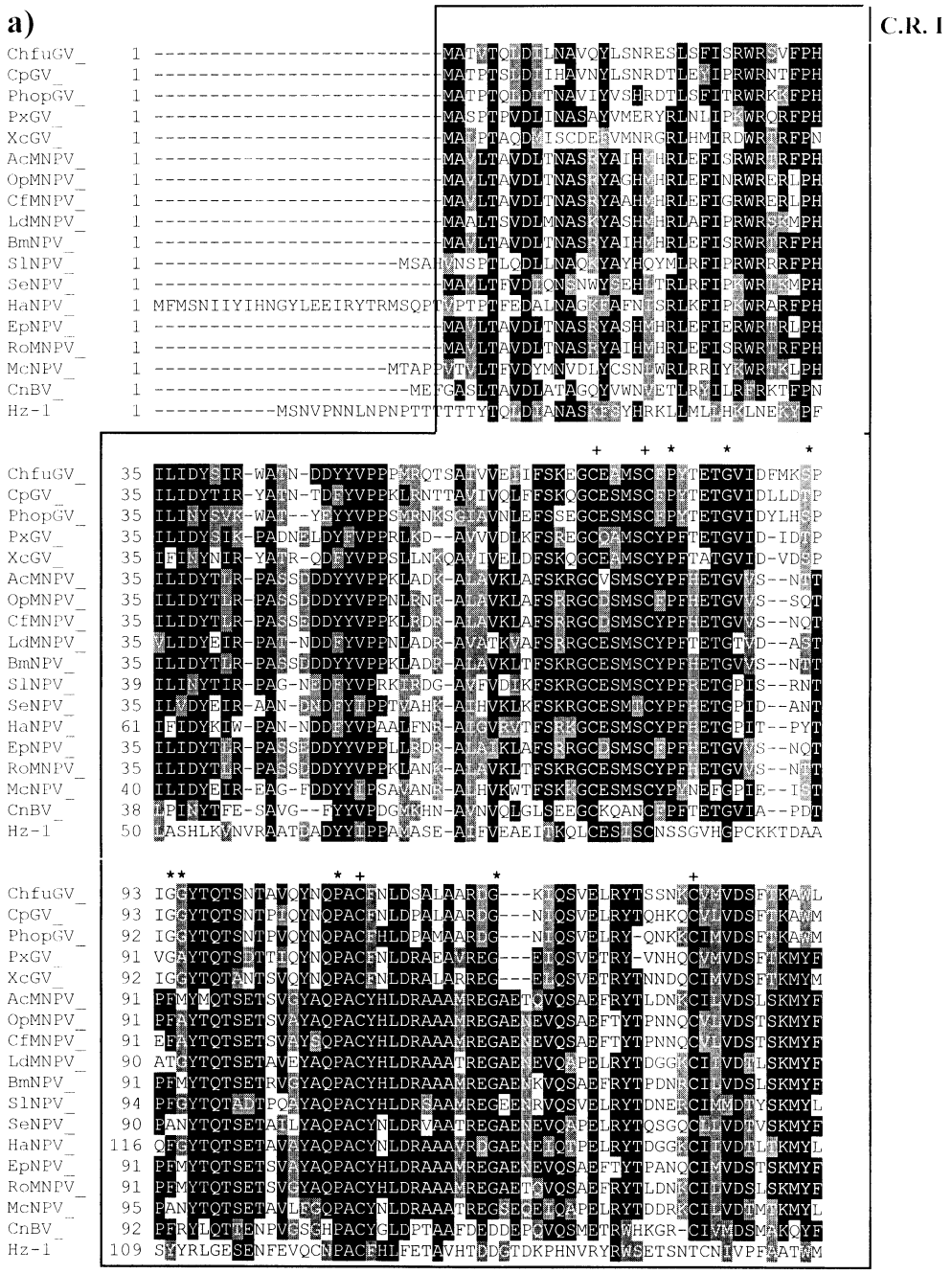


Fig. 3. (a) Amino acid alignments of deduced p74 protein of ChfuGV along with other p74 protein homologues that were studied (see Materials and Methods for the accession numbers). Conserved proline, glycine are depicted by (*), and conserved cysteine residues are illustrated by (+). Two highly conserved regions (C.R.) are shown within two boxes as C.R.I and C.R.II. (b) The figure shows the position of the conserved cysteine residues and their spacing patterns in the baculoviral p74 protein homologues.

motifs (TAAG and GTAAG) were located at 5 nt and 161 nt upstream of the first ATG of the p74 gene. The TAAG motif is known as a strong promoter of the late genes in baculoviruses. The gene contained a canonical poly (A) signal, AATAAA, at 3 NTR at the position +16 nt downstream from the termination signal. The nucleotide sequence surrounding the translational start codon (**A**GTATG**G**) was conformed to Kozaks rule for efficient eukaryotic translation initiation with the presence of two purine bases at the +4 and -3 positions (Kozak, 1986) (Fig. 2).

The deduced amino acid composition [in terms of non-polar (ACFGILMPVWY), polar (DEHKNQRST), basic (HRK), and acidic residues (DE)] indicates non-polar (54.75%), polar (45.25%), acidic (11.01%), and basic (9.50%)

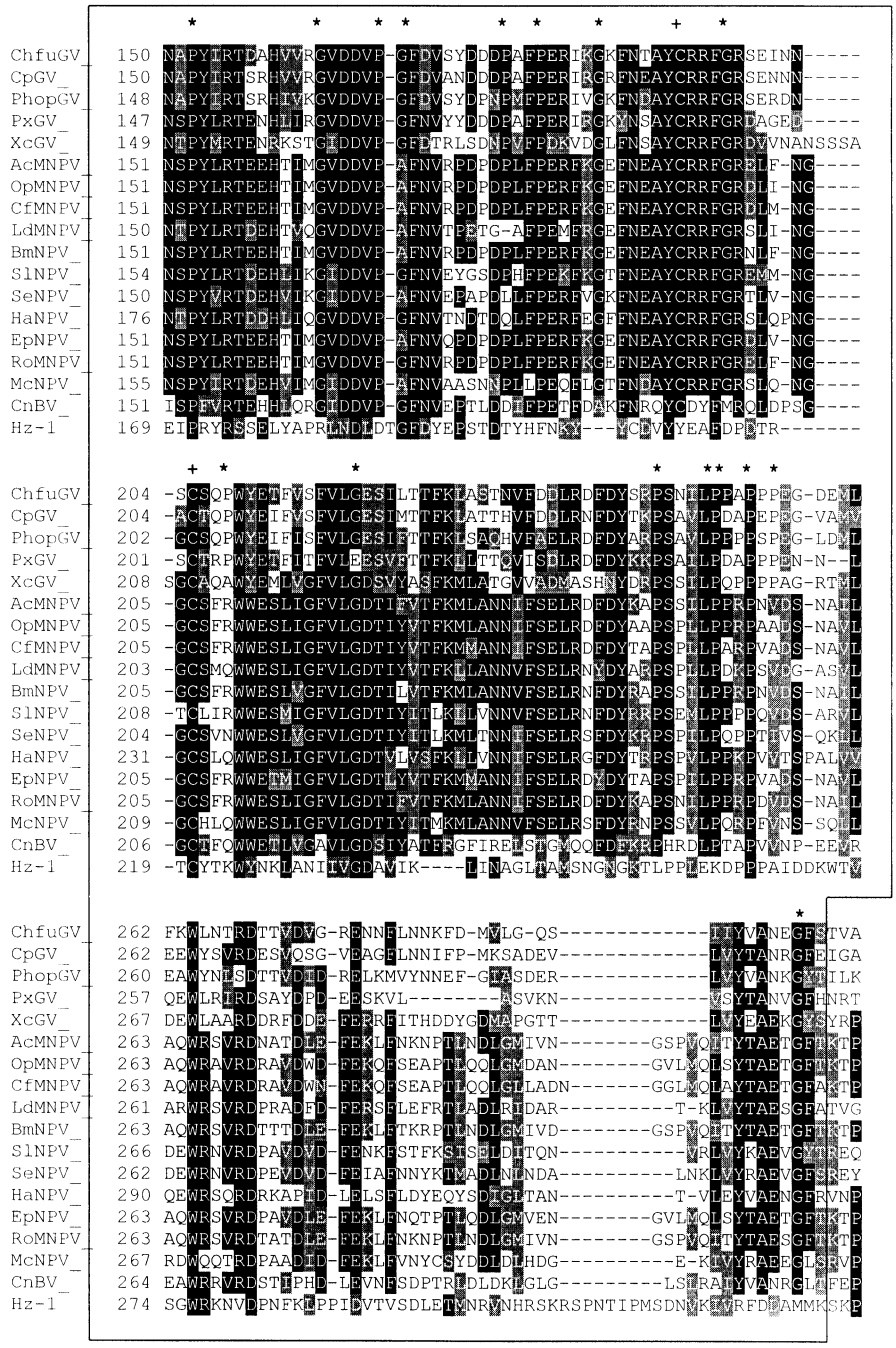


Fig. 3. Continued.

contents. The protein is particularly rich in leucine (9.20%), serine (8.60%) and alanine (7.54%) residues. A study on the amino acid composition of the ChfuGV p74 protein also revealed the presence of a high number of glycine and proline residues (26 and 29 respectively), together they shape 8.3% of the composition of p74 in ChfuGV. In all of the other p74 homologues that were studied, matching high numbers of glycine and proline residues were also observed. The number of glycine and proline residues is as follows: in GVs 26 to 35 and 27 to 33; in NPVs 22 to 34, and 25 to 32; in CnBV 45 and 39; and in the Hz-1 virus 26 and 33. The N-terminal two-third portion of the ChfuGV p74 homologue contains most of these glycine and proline residues (14 and 18 residues respectively). Due to the absence of a side chain proline (with its side-chain back-boned to amide nitrogen of the main chain) and glycine, the is always result the conformation of a protein. Both amino acids are also associated with surface loops. The unhindered rotational freedom of glycine can result in an increased loop flexibility that leads to smooth progress in the conformational changes of the protein. This observation

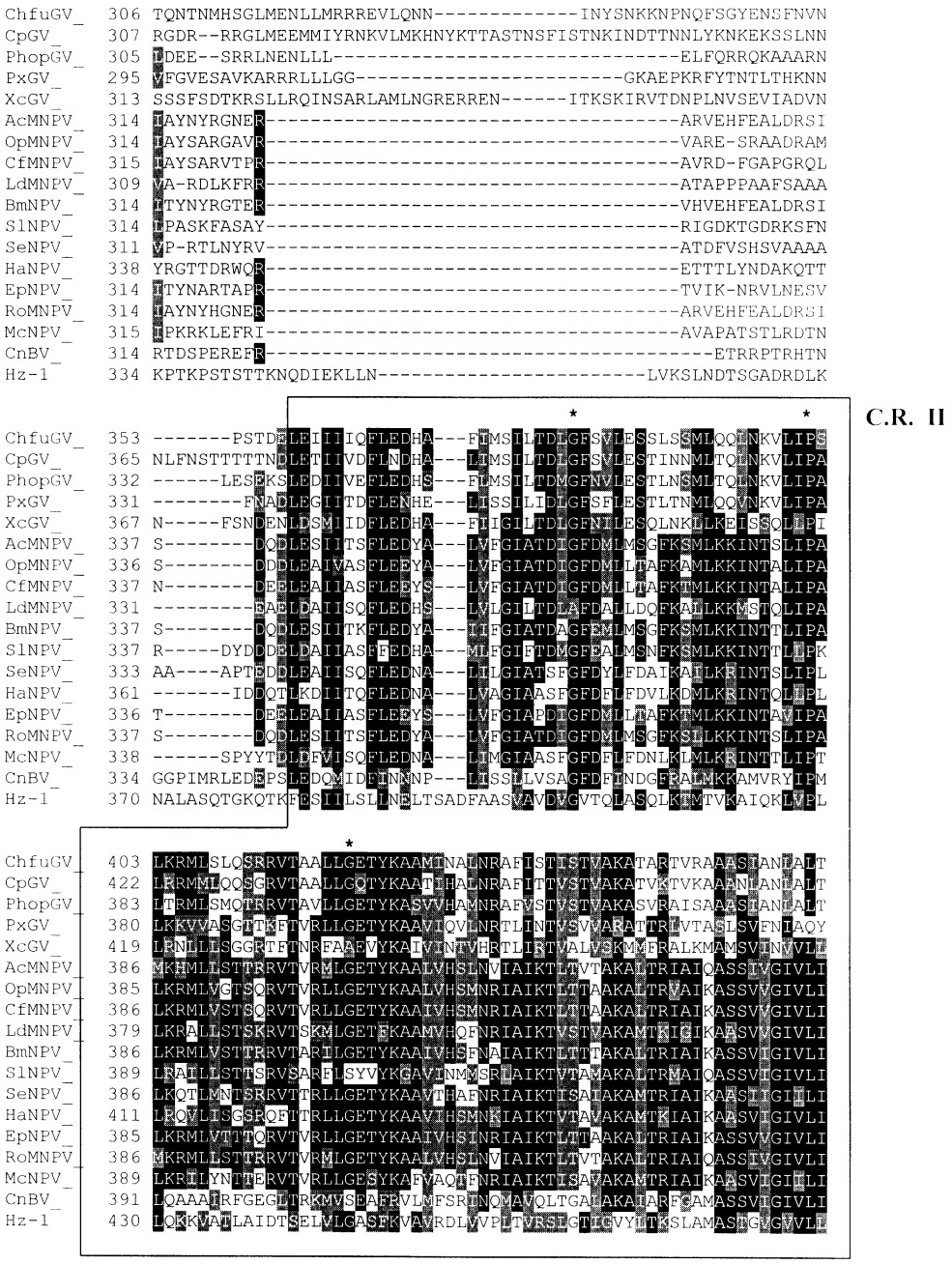


Fig. 3. Continued.

suggests that the N terminal of the ChfuGV p74 protein can be characterized as a flexible structure with a high number of turns and surface loops. The multiple alignment of the ChfuGV p74 protein with those of other BVs and Hz-1 virus revealed that several glycine and proline residues were conserved in almost all of the baculoviral p74 proteins (Fig. 3a). This high level of proline and glycine residue preservation may imply their significance in the creation of a functional conformation in baculoviral p74.

Another remarkable aspect of the ChfuGV p74 protein was the presence of an 8 cysteine-residue cluster. A multiple sequence alignment of ChfuGV p74 with those of other baculoviruses revealed that all of the viruses preserved six of these cysteine residues. Also, the arrangement as well as the position of these residues is astonishingly conserved in each and every one of the p74 homologues that were studied (Fig. 3a and b). The seventh and eighth cysteine residues, initially detected in the ChfuGV p74 protein, were also detected in the CpGV, PxGV, SlNPV, HaNPV, and Hz-1 viruses. This level of preservation of the cysteine residues, which are perhaps involved in the creation of disulfide bridges, implies that the presence and arrangement of these residues are a matter of a

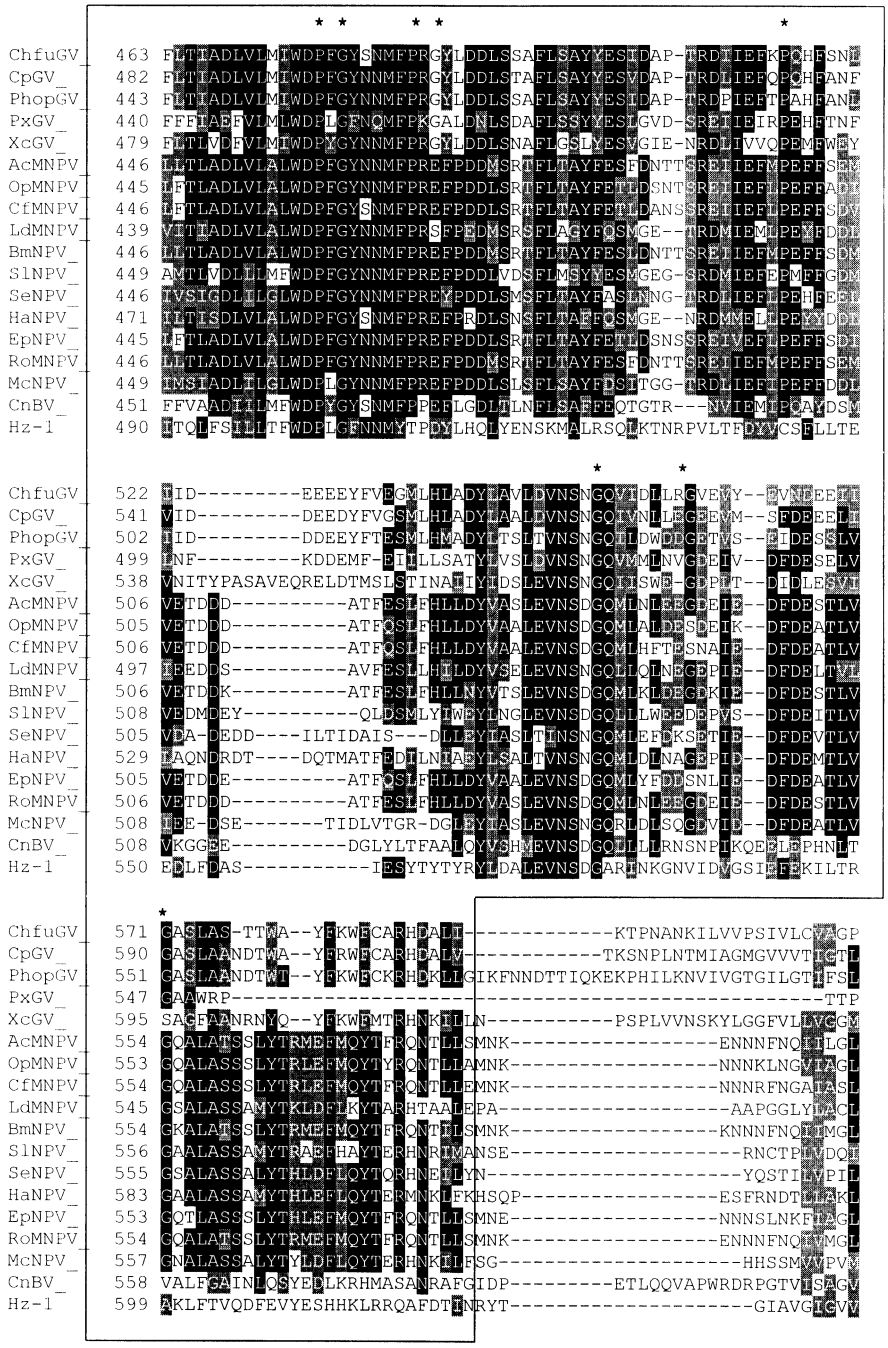
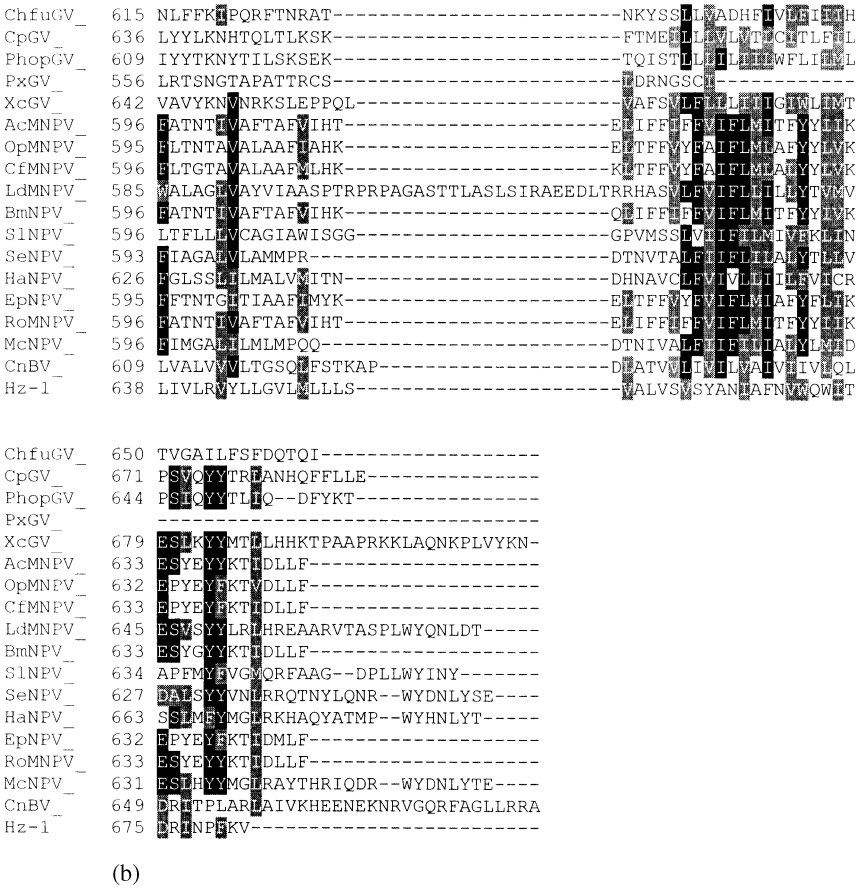


Fig. 3. Continued.

great importance for the task of the p74 protein in all baculoviruses, as well as the Hz-1 virus. Also, without them, the protein may be unable to fold properly in order to form a functional protein. The high degree of conservation, as well as the position preservation of cysteine, glycine, and proline residues in all of the baculoviral and Hz-1 p74 proteins, may imply the important fact that the p74 protein plays a crucial role in the life cycle of all baculoviruses. The removal of these conserved properties may eventually lead to the elimination of the virus in their evolutionary pathway.

The size of the baculoviral p74 proteins varies in the range between 578 aa in PxGV to 710 aa in XcGV. The deduced amino acid sequence of ChfuGV P74 was compared to other known baculoviral p74 proteins. The results revealed a considerable similarity between the ChfuGV p74 sequence and other p74 homologues (Table 1). ChfuGV p74 has a 63.7% amino acid sequence identity to its closest relatives, which are CpGV and PhopGV. The Hz-1 virus p74 was the most distant, showing only a 20.9% amino acid sequence identity.



(b)

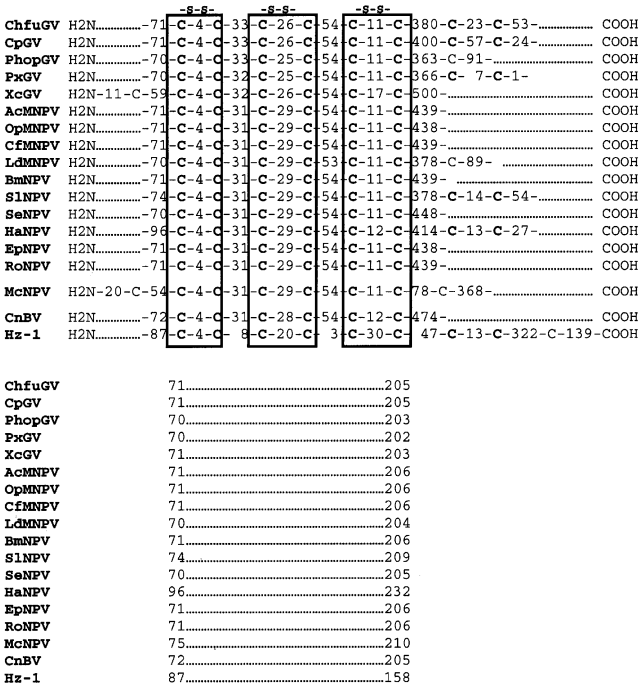


Fig. 3. Continued.

The one-third C-terminal of the ChfuGV p74 protein in ChfuGV is rich in non-polar residues (60.63%). The predicted hydrophobic profile of ChfuGV p74 demonstrated the existence of two highly hydrophobic regions within this segment of the ChfuGV p74 protein. Similar hydrophobic regions were also identified in the C-terminal regions of all of

Table 1. Comparison of conceptual amino acid sequences of ChfuGV p74 with those of other baculoviruses studied here

	CpGV	PpGV	PxGV	XcGV	AcMNPV	CfMNPV	LdMNPV	BmNPV	SlNPV	SeNPV	HaNPV	EpNPV	CnBV	Hz-1
ChfuGV	63.7	63.7	51.7	41.9	46.5	41.6	42.6	45.6	40.8	41.6	41.6	41.4	29.9	20.9
CpGV		63.7	53	44	43.3	42.3	43.7	43.2	41.6	41.7	41.5	41.6	29.9	21.3
PpGV			50	43.9	43.6	42.4	43	44	43.1	41.7	41	43.2	29.5	22.7
PxGV				44.3	46.5	45.7	42.9	45.6	42.7	42.3	44.3	45.1	32.3	19.1
XcGV					39.6	36.6	39.2	40	40.3	38	39.1	36.8	28.9	20
AcMNPV						78.3	57.8	91	52.8	54.9	51.9	78.1	33.8	21.9
CfMNPV							59	77.3	52.8	55.3	52	86	34	20.3
LdMNPV								58.1	51.2	58	58.7	57.8	32.7	19.4
BmNPV									53.6	55.4	52.5	77.2	34.3	20.3
SlNPV										50.5	50.3	52	32	22.1
SeNPV											54.7	54.8	31.2	20.3
HaNPV												51.2	31.9	20.3
EpNPV													33.5	22.1
CnBV														19.4

Pair wise identity values (%) between taxa are shown.

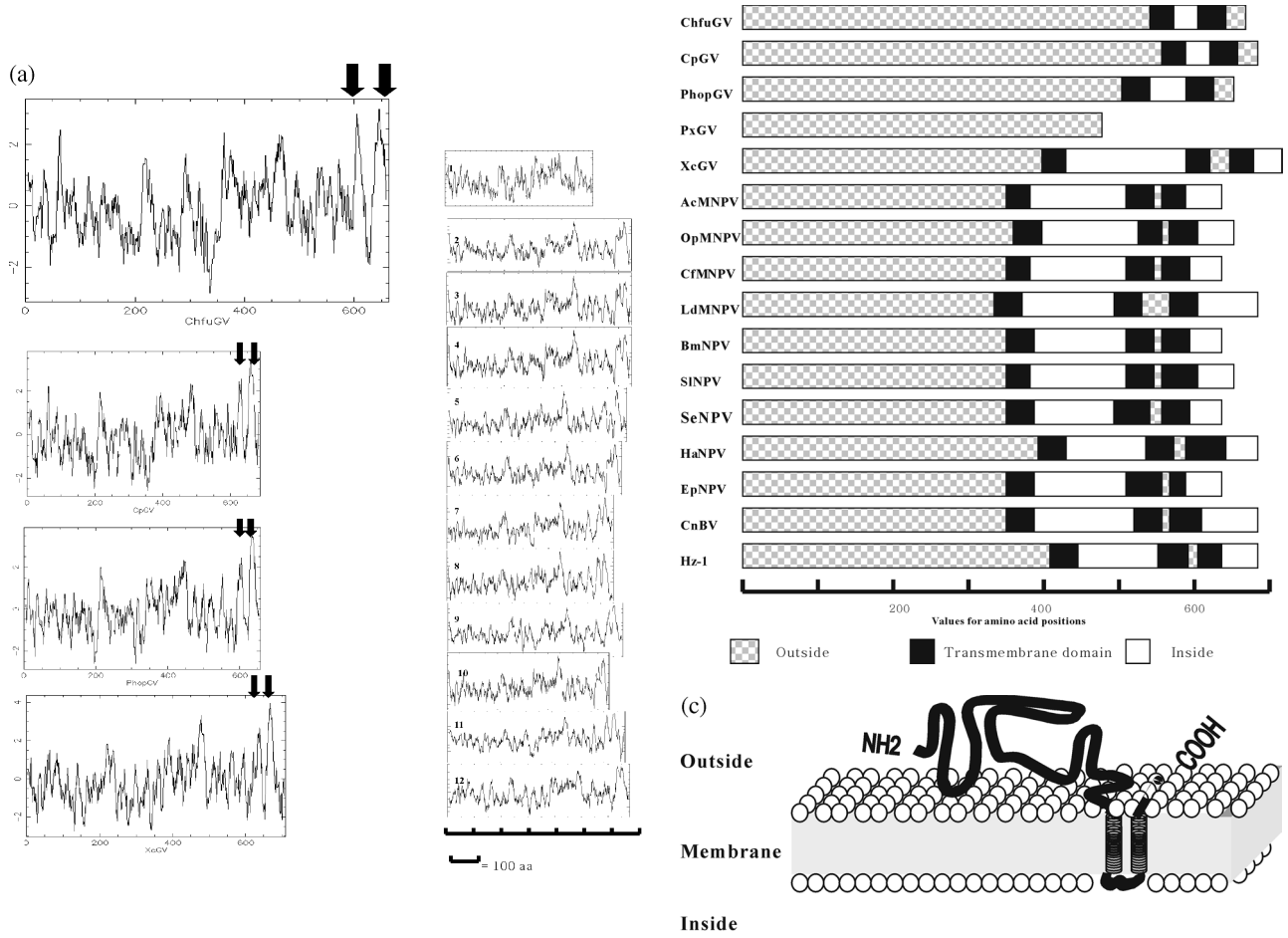


Fig. 4 Structural analysis of p74. (a) Predicted hydrophobic profiles of ChfuGV p74 and those of other GVs, NPVs, CnBV and Hz-1 virus. Arrows indicate two C-terminal hydrophobic domains. Profiles were generated using a Pepwindow hydropathy analysis (7 residue average; Kyte and Doolittle, 1982). Positions of the amino acids are shown on the X-axis and hydrophobicity values are shown on the Y-axis. The list of baculoviral p74s in the right column is as follows: 1, PxGV; 2, AcMNPV; 3, OpMNPV; 4, CfMNPV; 5, LdMNPV; 6, BmNPV; 7, SlNPV; 8, SeNPV; 9, HaNPV; 10, EpNPV; 11, CnBV; 12, Hz-1 virus. (b) Demonstration of transmembrane, outside and inside domains of p74 proteins. (c) Schematic presentation of ChfuGV p74. A computer analysis (TMHMM-ver. 2) of the amino acid sequence predicts that the N-terminal of the protein is located outside of the envelope, followed by two transmembrane regions which results in an external C-terminus. Each transmembrane domain spans about 20 amino acids and is shown as a cylinder.

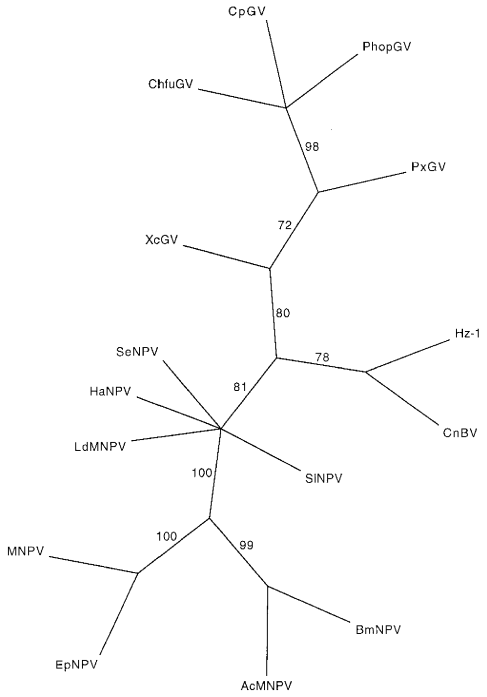


Fig. 5. Phylogenetic analysis performed on baculoviral p74 proteins. The unrooted tree was produced by a Branch and Band search using PAUP4.0b4a. Bootstrap values (100 replicates) are shown. The tree was constructed according to the following settings: 1- uninformative characters ignored. 2- Branch- and -bound search options: an initial upper bound: compute via stepwise; keep minimals only; collapse zero-length branches ON; MULPARS ON; addition sequences: furthest. The accession numbers for the protein sequences that were used are shown in the Material and Methods Section.

the other p74 homologue proteins that were studied (Fig. 4a). Various computer-assisted prediction models that were used in this study predict that ChfuGV is a membrane protein with two membrane-spanning regions within its C-terminal region. These transmembrane domains were located between amino acid residues 600 to 620 and 633 to 656. Despite the low sequence conservation at the C-terminal, almost all of the baculoviral p74 homologues (excluding PxGV) have conserved these two transmembrane regions at approximately parallel positions (Fig. 4b). The lack of transmembrane domains in PxGV, as Slack and his associates also suggested, might be due to a recent evolutionary change in PxGV p74 (Slack *et al*., 2001).

Besides these two regions, an additional transmembrane region was positioned almost in the center of the p74 proteins in all of the NPVs, CnBV, Hz-1 viruses as well as XcGV (Fig. 4b). All of the computer-assisted models that were used in this study predicted that the ChfuGV p74 N-terminal is exposed outside of the ODV envelope, while a part of the C-terminal is anchored within the envelope membrane through two highly hydrophobic regions. This arrangement places the last seven amino acid residues of ChfuGV p74 outside the envelope membrane (Fig. 4c). A similar arrangement is shared by two other GVs that were studied (CpGV and PhopGV). In XcGV, the entire NPVs group, CnBV and Hz-1, the N-terminal of p74 proteins are also exposed outside, although the presence of three transmembrane regions directs the C terminal end toward the inside of the ODV envelope (Fig. 4b).

ChfuGV p74 has two potential *N*- glycosylation N-{P}-[ST]-{P} sites that are located at residues 202 and 332. Glycosylation is important for the proper folding or targeting of some polypeptides (Darvey, 1989). Two *N*-myristoylation (G-{EDRKHPFYW}-X-X-[STAGCN]-{P}) sites were also observed at residues 314 and 571.

A close look at the multiple alignments of the p74 amino acid sequences revealed the existence of two major conserved regions (Fig. 3a). The first region began from amino acid residue number 22 and was prolonged up to residue 302 (in the ChfuGV p74 protein). The second conserved region was located from residues 357 to 593. All of the conserved cysteine residues were found within the first region, while the conserved proline and glycine residues resided within either one of the two conserved regions. Areas situated outside of these two preserved boxes demonstrated high levels of sequence divergence. Several conserved motifs were detected within these two major conserved boxes in all of the baculoviruses. These motifs were ^{32}[F/L/M]P[H/N][I/V][L/F][I/V][D/N]Y^{39}, ^{48}[D/E][Y/F][Y/F][V/I]PP^{53}, ^{71}GC[E/Q/D/V][A/S]M[S/T]C[F/Y]P[Y/F][T/H/R/N][E/A]TG[V/P/T][I/V]^{86}, ^{107}QPAC[F/Y][N/H]LD^{114}, ^{150}N[A/S/T]PY[I/M/L/V]RT^{156}, ^{163}G[V/I]DDVP^{168}, ^{191}AYCRRFGR^{198}, ^{216}FVL[G/E][E/D][T/S][I/V/L]^{222}, ^{473}[I/L/F]WDP[F/L]G[Y/F][N/S]N[N/Q]MFP[R/K]^{485}. The ChfuGV, CpGV, and PhopGV p74 proteins shared several highly conserved stretches of amino acids. These highly conserved stretches were located at the following positions: residues 1 to 9, 91 to 132, 138 to 184, 205 to 228, 372 to 388, 391 to 407, 413 to 426, 430 to 513, 522 to 530, and 534 to 553. The longest was a stretch of amino acids that covered 83 amino acid residues, beginning with residue 430 in the ChfuGV p74 protein. Interestingly, all of these highly conserved regions were located on the external segment of the protein, which suggests the importance of the conformational preservation of this segment. This level of preservation can usually be seen in the envelope proteins that play a role in the process of virion attachment to the cell surface receptors.

A phylogenetic analysis of the p74 proteins is shown in Fig. 5. A tree was produced via maximum parsimony to estimate the evolutionary relationship between the p74 protein species in baculoviruses and Hz-1 virus. These analyses showed a clear division between the GV and NPV p74 proteins. ChfuGV is positioned alongside CpGV and PhopGV in the same cluster, supported by 98 bootstrap values.

The data presented in this paper offer a broad comparative study on an essential baculoviral envelope protein p74 that may play a significant role during the initiation phase of the baculoviral infection. This places more importance on this protein for the *Baculoviridae* family and related viruses.

Acknowledgments The authors would like to thank Dr. Anne Bruneau (Department of Biology, University of Montreal) for helping us in the phylogenetic studies. We also thank Dr. David Morse (Department of Biology, University of Montreal) for his scientific discussions. This research was funded by the *Ministère des Ressources Naturelles du Québec*, Grant number 0316-240S to CG.

References

Afonso, C. L., Tulman, E. R., Lu, Z., Balinsky, C. A., Moser, B. A., Becnel, J. J., Rock, D. L. and Kutish, G. F. (2001) Genome sequence of a baculovirus pathogenic for *Culex nigripalpus*. *J. Virol*. **75**, 11157-11165.

Ahrens, C. H., Russell, R. L., Funk, C. J., Evans, J. T., Harwood, S. H. and Rohrmann, G. F. (1997) The sequence of the *Orgyia pseudotsugata* multinucleocapsid nuclear polyhedrosis virus genome, *Virology* **229**, 381-399.

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.

Ayres, M. D., Howard, S. C., Kuzio, J., Lopez-Ferber, M. and Possee, R. D. (1994) The complete DNA sequence of *Autographa californica* nuclear polyhedrosis virus. *Virology* **202**, 586-605.

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.

Bairoch, A., Bucher, P. and Hofmann, K. (1997) The PROSITE database, its status in 1997. *Nucleic Acids Res*. **25**, 217-221.

Blissard, G. W. and 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.

Braunagel, S. C. and Summers, M. D. (1994) *Autographa californica* nuclear polyhedrosis virus, PDV, and ECV viral envelopes and nucleocapsids: structural proteins, antigens, lipid and fatty acid profiles. *Virology* **202**, 315-328.

Braunagel, S. C., He, H., Ramamurthy, P. and Summers, M. D. (1996a) Transcription, translation, and cellular localization of three *Autographa californica* nuclear polyhedrosis virus structural proteins: ODV-EI8, ODV-E35, and ODV-EC27. *Virology* **222**, 100-114.

Braunagel, S. C., Elton, D. M., Ma, H. and Summers, M. D. (1996b) Identification and analysis of an *Autographa californica* nuclear polyhedrosis virus structural protein of the occlusion-derived virus envelope: ODV-E56. *Virology* **217**, 97-110.

Chen, X., Zhang, W. -J., Wong, J., Chun, G., Lu, A., McCutchen, B. F., Presnail, J. K., Herrmann, R., Dolan, M., Tingey, S., Hu, Z. -H. and Vlak, J. M. (2002) Comparative analysis of the complete genome sequences of *Helicoverpa zea* and *Helicoverpa armigera* single-nucleocapsid nucleopolyhedroviruses. *J. Gen. Virol.* **83**, 673-684.

Cheng, C. H., Liu, S. M., Chow, T. Y., Hsiao, Y. Y., Wang, D. P., Huang, J. J. and Chen, H. H. (2002) Analysis of the Complete Genome Sequence of the Hz-1 Virus Suggests that it is related to members of the Baculoviridae. *J. Virol*. **76**, 9024-9034.

Darvey, J. (1989) Sorting out the secretory pathway. *Bioassays* **11**, 185-187.

Faulkner, P., Kuzio, J., Williams, G. V. and Wilson, J. A. (1997) Analysis of p74, a PDV envelope protein of *Autographa californica* nucleopolyhedrovirus required for occlusion body infectivity *in vivo*. *J. Gen. Virol.* **78**, 3091-3100.

Fielding, B. C., Khan, S., Wang, W., Kruger, C., Abrahams, R. and Davison, S. P. (2002) The genetic organization of a 2,966 basepair DNA fragment of a single capsid nucleopolyhedrovirus isolated from *Trichoplusia ni*. *Virus, Genes* **25**, 35-43.

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

Funk, C. J. and Consigli, R. A. (1993) Temporal expression and immunogold localization of *Plodia interpunctella* granulosis virus structural proteins. *Virus Res.* **28**, 57-66.

Garnier, J., Gibrat, J. -F. and Robson B. (1996) GOR secondary structure prediction method version IV; in *Methods in Enzymology*, Doolittle, R. F. (ed), vol 266, pp. 540-553.

Gomi, S., Majima, K. and Maeda, S. (1999) Sequence analysis of the genome of *Bombyx mori* nucleopolyhedrovirus. *J. Gen. Virol*. **80**, 1323-1337.

Hansen, J. E., Lund, O., Tolstrup, N., Gooley, A. A., Williams, K. L. and Brunak, S. (1998) NetOglyc: prediction of mucin type O-glycosylation sites based on sequence context and surface accessibility. *Glycoconj. J.* **15**, 115-130.

Hashimoto, Y., Hayakawa, T., Ueno, Y., Fujita, T., Sano, Y. and Matsumoto, T. (2000) Sequence analysis of the *Plutella xylostella* granulovirus genome. *Virology* **275**, 358-372.

Hayakawa, T., Ko, R., Okano, K., Seong, S. I., Goto, C. and Maeda, S. (1999) Sequence analysis of the *Xestia c-nigrum granulovirus* genome. *Virology* **262**, 277-297.

Hill, J. E., Kuzio, J., Wilson, J. A., MacKinnon, E. A. and Faulkner, P. (1993) Nucleotide sequence of the p74 gene of a baculovirus pathogenic to the spruce budworm, *Choristoneura fumiferana* multicapsid nuclear polyhedrosis virus. *Biochim. Biophys. Acta* **1172**, 187-189.

Hofmann, K. and Stoffel, W. (1993) TMbase - A database of membrane spanning proteins segments. *Biol. Chem. Hoppe-Seyler* **374**, 166.

Hong, T., Braunagel, S. C. and Summers, M. D. (1994) Transcription, translation, and cellular localization of PDV-E66: a structural protein of the PDV envelope of *Autographa californica* nuclear polyhedrosis virus. *Virology* **204**, 210-222.

Horton, H. M. and Burand, J. P. (1993) Saturable attachment sites for polyhedron-derived baculovirus on insect cells and evidence for entry via direct membrane fusion. *J. Virol.* **67**, 1860-1868.

Hu, Z. H., Arif, B. M., Jin, F., Martens, J. W., Chen, X. W., Sun, J. S., Zuidema, D., Goldbach, R. W. and Vlak, J. M. (1998) Distinct gene arrangement in the *Buzura suppressaria* single-nucleocapsid nucleopolyhedrovirus genome *J. Gen. Virol.* **79**, 2841-2851.

IJkel, W. F., van Strien, E. A., Heldens, J. G., Broer, R., Zuidema, D., Goldbach, R. W. and Vlak, J. M. (1999) Sequence and organization of the *Spodoptera exigua* multicapsid nucleopolyhedrovirus genome. *J. Gen. Virol*. **80**, 3289-3304.

Keddie, B. A., Aponte, G. W. and Volkman, L. E. (1989) The pathway of infection of *Autographa californica* nuclear

polyhedrosis virus in an insect host. *Science* **243**, 1728-1730.

Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulate translation by eukaryotic ribosomes. *Cell* **44**, 282-292.

Kuzio, J., Jaques, R. and Faulkner, P. (1989) Identification of p74, a gene essential for virulence of baculovirus occlusion bodies. *Virology* **173**, 759-763.

Kuzio, J., Pearson, M. N., Harwood, S. H., Funk, C. J., Evans, J. T., Slavicek, J. M. and Rohrmann, G. F. (1999) Sequence and analysis of the genome of a baculovirus pathogenic for *Lymantria dispar*. *Virology* **253**, 17-34.

Kyte, A. J. and Doolittle, R. F. T. (1982) A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* **157**, 105-132.

Laemmli, U. K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature* **227**, 680-685.

Li, S., Erlandson, M., Moody, D. and Gillott, C. (1997) A physical map of the *Mamestra configurata* nucleopolyhedrovirus genome and sequence analysis of the polyhedrin gene. *J. Gen. Virol.* **78**, 265-271.

Li, Q., Donly, C., Li, L., Willis, L. G., Theilmann, D. A. and Erlandson, M. (2002) Sequence and organization of the *Mamestra configurata* nucleopolyhedrovirus genome. *Virology* **294**, 106-121.

Lupas, A., Van Dyke, M. and Stock, J. (1991) Predicting Coled Coils from Protein Sequences, *Science* **252**, 1162-1164.

Luque, T., Finch, R., Crook, N., O'Reilly, D. R. and Winstanley, D. (2001) The complete sequence of the *Cydia pomonella* granulovirus genome. *J. Gen. Virol.* **82**, 2531-2547.

Moller, S., Croning, M. D. R. and Apweiler, R. (2001) Evaluation of methods for the prediction of membrane spanning regions. *Bioinformatics* **17**, 646-653.

Pang, Y., Yu, J., Wang, L., Hu, X., Bao, W., Li, G., Chen, C., Han, H., Hu, S. and Yang, H. (2001) Sequence analysis of the *Spodoptera litura* multicapsid nucleopolyhedrovirus genome. *Virology* **287**, 391-404.

Rashidan, K. K., Nassoury, N., Giannopoulos, P. N. and Guertin, C. (2002) Identification and characterization of a conserved baculoviral structural protein ODVP-6E/ODV-E56 from *Choristoneura fumiferana* granulovirus. *J. Biochem. Mol. Biol.* **35**, 595-603.

Rohrmann, G. F. (1986) Polyhedrin structure. *J. Gen. Virol.* **67**, 1499-1513.

Russell, R. L. and Rohrmann, G. F. (1993) A 25-kDa protein is associated with the envelopes of occluded baculovirus virions. *Virology* **195**, 532-540.

Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Laboratory Press, New York, USA.

Slack, J. M., Dougherty, E. M. and Lawrence, S. D. (2001) A study of the *Autographa californica* multiple nucleopolyhedrovirus ODV envelope protein p74 using a GFP tag. *J. Gen. Virol.* **82**, 2279-2287.

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

Theilmann, D. A., Chantler, J. K., Stweart, S., Flipsen, H. T., Vlak, J. M. and Crook, N. E. (1996) Characterization of a highly conserved baculovirus structural protein that is specific for occlusion-derived virions. *Virology* **218**, 148-158.

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.

Whitford, M., Stewart, S., Kuzio, J., and Faulkner, P. (1989) Identification and sequence analysis of a gene encoding gp67, an abundant envelope glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol*. **63**, 1393-1399.

Whitford, M. and Faulkner, P. (1993) Nucleotide sequence and transcriptional analysis of a gene encoding gp41, a structural glycoprotein of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **67**, 2427.