Nucleotide sequence and transcriptional analysis of the gp41 gene of Spodoptera frugiperda nuclear polyhedrosis virus

Jaw-Ching Liu and James E. Maruniak*

Entomology and Nematology Department, University of Florida, PO Box 110620, Gainesville, FL 32611-0620, USA

The Spodoptera frugiperda multiple nucleocapsid nuclear polyhedrosis virus (SfMNPV) gp41 structural protein gene was located in the 1.9 kbp EcoRI-S fragment and sequenced. An open reading frame (ORF) of 999 nucleotides was detected that encoded a protein of 332 amino acids. The gp41 gene transcript was detected after 12 h post-infection (p.i.) and remained detectable at 48 h p.i. Two major mRNAs, about 1.6 and 2.8 kb in length, were determined by Northern blot analysis. Primer extension analysis demonstrated that the gp41 gene promoter region contains three transcription start sites. Two of the gp41 gene transcription start sites were located at –42 and –41 nucleotides from the ATG translation start codon within a consensus late transcription start site (TAAG) and another transcription start site was located at –140 nucleotides from the ATG translation start codon for which no consensus motif has been determined. Comparison of SfMNPV gp41 nucleotide and amino acid sequences with the gp41 genes from Autographa californica, Bombyx mori, and Helicoverpa zea NPVs showed 60% homology of nucleotide sequences and 70% similarity of amino acid sequences.

Introduction

Spodoptera frugiperda multiple nucleocapsid nuclear polyhedrosis virus (SfMNPV) is a member of the family Baculoviridae. Baculoviruses have been used as biological control agents for insect pests in agriculture and forestry because of their specificity (Hawtin et al., 1992). SfMNPV is specific for the fall armyworm (S. frugiperda) and shows a potential to control this pest. SfMNPV has a double-stranded DNA genome of approximately 121 kbp. The physical map for a number of restriction endonucleases has been described and shown to have different profiles compared to other NPVs (Loh et al., 1981; Maruniak et al., 1984). However, two regions of DNA homology on the physical maps of SfMNPV (SfMNPV-2) and S. exempta MNPV (SeMNPV-25), an Autographa californica MNPV genomic variant (Brown et al., 1985), have been identified by hybridization under high stringency conditions. One of these two regions contained the polyhedrin gene (Brown et al., 1987); the other region has been identified in the current report to be associated with the gp41 structural protein gene.

Two types of virions are produced during the nuclear polyhedrosis virus life cycle. Those virions found within the viral occlusion bodies (OBs) are termed occluded virions (OVs). They obtain their envelope in the nuclei of infected cells from de novo and the OV envelope is involved in the recognition of host microvilli during infection. The second type of baculovirus virion is the extracellular virus (ECV). The single nucleocapsids bud through the plasma membrane of infected cells and form the ECV (Blissard & Rohrmann, 1990; Granados & Williams, 1986). These virions appear to be specialized for secondary infection of other host cells and contain the virus-encoded envelope glycoprotein which is involved in host cell infection, i.e. gp64 (Keddie & Volkman, 1985; Maruniak, 1979).

The gp41 structural protein was identified as a major OV glycoprotein by metabolic labelling (Maruniak, 1979; Stiles & Wood, 1983). It has also been detected by binding of horseradish peroxidase-linked concanavalin A thus indicating it is glycosylated (Braunagel & Summers, 1994). Furthermore, an O-linked single N-acetylglucosamine covalently bonded to the polypeptide was identified (Whitford & Faulkner, 1992a). Experiments with monoclonal antibodies indicated that gp41 is present only in OV; it appears to be associated with OV but not with purified nucleocapsids or the ECV (Whitford & Faulkner, 1992a; Ma et al., 1993) and the location of the gp41 protein has been predicted to be between the envelope membrane and the capsid ( tegument) of the OV. On the other hand, Braunagel & Summers (1994) indicated that the viral proteins of 40–41 kDa are glycosylated in OV and ECV. However, the monoclonal antibody data suggest that the gp41

* Author for correspondence. Fax +1 904 392 7567. e-mail MARUN@NERVM.NERDC.UFLEDU

Sequence data from this article have been deposited with GenBank under accession no. U14725.
proteins of ECV and OV are different proteins. Recently, the gene encoding gp41 protein was characterized (Ayres et al., 1994; Kool et al., 1994; Ma et al., 1993; Nagamine et al., 1991; Whitford & Faulkner, 1992b) but the biological function of gp41 protein is still unknown.

In this paper, we present the complete nucleotide and translated amino acid sequence of the SfMNPV gp41 gene. The sequences were compared with other known gp41 gene sequences of different baculoviruses to reveal the possible functional domain of the gp41 protein. A possible transcriptional regulation mechanism and the phylogenetic relationship of the gp41 gene among different baculoviruses are discussed in this paper.

Methods

Virus and cell culture. The S. frugiperda MNPV isolate SfMNPV-2 (Maruniak et al., 1984) was propagated in the S. frugiperda Sf9 cell line (Luckow & Summers, 1988). Sf9 cells were maintained in TC-100 medium supplemented with 10% fetal bovine serum (Life Technology) and 50 μg/ml gentamicin at 27°C.

DNA cloning and sequencing. The SfMNPV-2 EcoRI-S DNA fragment was cloned into pGEM-3Z and pGEM-7Zf(+) vectors (Promega), and the subfragments EcoRI-HindIII (0.5 kbp), EcoRI-PstI (0.8 kbp), PstI-EcoRI (1.1 kbp) and Hhal-Hhal (0.7 kbp) were cloned into pGEM-3Z. Exonuclease digested subclones were generated with the Erase-a-Base system (Promega). Sequencing was performed with the dideoxynucleotide chain terminator sequencing method (Sanger et al., 1977) with Sequenase (United States Biochemical Corp.). The oligonucleotide primers were synthesized by the DNA Synthesis Laboratory of the Interdisciplinary Center for Biotechnology Research at the University of Florida.

Computer analysis. The GCG computer program was used for nucleotide and amino acid sequence analysis. The GenBank database was searched using BLAST.

RNA purification. Total cellular RNA was isolated using the guanidine isothiocyanate method (Ausubel et al., 1989) from 3 × 10⁶ Sf9 cells infected with SfMNPV-2 at an m.o.i. of 10 p.f.u./cell. At various times post-infection (p.i.), the cells were lysed in 4 M-guanidine isothiocyanate pH 5.5, 20 mM-sodium acetate, 0.1 mM-dithiothreitol (DTT) and 0.5% Sarkosyl. Cell lysates were layered over 5.7 M-CsCl (0.1 mM-EDTA) and centrifuged at 100,000 g for 24 h in a swinging AH650 rotor (DuPont). The RNA was dissolved in sterile water and ethanol precipitated. After washing the RNA pellet in 70% (v/v) ethanol, the pellet was dissolved in sterile water. The RNA concentration was determined by measuring the UV absorbance at 260 nm.

Northern blot hybridization. A total of 5 μg RNA was denatured with 7% formaldehyde, 50% formamide and 1 x MOPS buffer (0.2 M-MOPS pH 7.0, 50 mM-sodium acetate and 10 mM-EDTA) at 55°C for 15 min. Before electrophoresis, 0.1 vol. of 10 x loading buffer (20% Ficoll 400, 1% SDS, 0.1 mM-EDTA, 0.25% Bromophenol Blue and Xylene Cyanole FF) was added. Total RNA was electrophoresed in a 1% agarose gel (1% formaldehyde and 1 x MOPS buffer) in 1 x MOPS buffer (Maniatis et al., 1982). The separated RNAs were transferred to a Zeta-Probe blotting membrane (Bio-Rad) with 20 x SSC buffer. After transfer, the membrane was air-dried and baked at 80°C for 1 h. The DNA probe containing 50 ng of the SfMNPV-2 EcoRI-S DNA fragment was prepared by the nick-translation method (United States Biochemical Corp.) using 30 μCi [α-32P]dCTP (3000 mCi/mmol). Hybridization was done overnight at 42°C and the blot was rinsed at 42°C with 5% and 1% SDS washing buffer twice each (40 mM-Na2HPO4, pH 7.2, 1 mM-EDTA) as described by the manufacturer (Bio-Rad). The blot was exposed with Kodak X-OMAT film.

Primer extension. A total of 10 μg RNA isolated from the infected Sf9 cells was mixed with 0.5 μg of 20-mer oligonucleotide primer (S GACGTATCGCAACATTGTTG 3). This primer was complementary to the region from 104 to 123 bases downstream of the translation start codon of the SfMNPV gp41 protein gene. The RNA and the primer were incubated at 30°C overnight. The reaction was done in buffer containing 50 mM-Tris–HCl, pH 8.3, 75 mM-KCl, 3 mM-MgCl₂, 10 mM-DTT, 0.12 μM of each deoxynucleotidetri-phosphate, 25 μCi [α-32P]dCTP (3000 mCi/mmol) and 200 units of Moloney murine leukaemia virus reverse transcriptase (Life Technology) for 60 min at 37°C (modified from Ausubel et al., 1989). The reaction was stopped by adding EDTA to a final concentration of 20 mM. The extension products were ethanol precipitated and resolved on a 6% polyacrylamide sequencing gel. A sequence marker was done with dideoxynucleotide chain terminator sequencing reaction by using the same primer with a DNA template containing the SfMNPV EcoRI-S fragment.

Results

Cloning and sequencing of the S. frugiperda EcoRI-S fragment

The S. frugiperda EcoRI-S fragment containing the gp41 structural protein gene was cloned into pGEM-3Z and pGEM-7Zf(+) (Fig. 1a). Specific restriction endonuclease digested subclones and exonuclease III deleted subclones were constructed. The T7 and SP6 promoter primers present in the pGEM vector and several specific oligonucleotide primers were used for sequencing (Fig. 1b).

A major open reading frame (ORF) which contained 999 nucleotides encoded the gp41 gene and was oriented from the right to left according to the conventional physical maps (Fig. 1b) (Maruniak et al., 1984). The complete sequence of SfMNPV EcoRI-S fragment was deposited with GenBank (accession no. U14725). One baculovirus late promoter consensus motif, TAAG (Bissard & Rohrmann, 1990), was found from 39 to 43 nucleotides upstream from the ATG translation start codon. The translation stop codon TGA is followed by 394 nucleotides downstream to the polyadenylation signal AATAA.

Transcription analysis of the gp41 gene

Northern blot analysis of total RNA from infected cells isolated from 3 to 48 h.p.i. is shown (Fig. 2). Two mRNAs of approximately 1.6 and 2.8 kb were detected after 12 h.p.i. and remained detectable at 48 h.p.i. when the SfMNPV EcoRI-S fragment containing the gp41 coding region was used as a probe.

Primer extension analysis was used to identify the transcription start site. A 20-mer oligonucleotide, cor-
Fig. 1. Position of the gp41 gene on the SfMNPV genomic map and sequencing strategy. (a) EcoRI restriction map of the SfMNPV-2 genome. (b) Detailed physical map of the EcoRI-S fragment. The gp41 999 bp open reading frame is indicated by the bold arrow under the map. The small arrows below the map indicate the extension and direction of the sequence using T7 or SP6 primers or specific primers. Those specific primers are indicated by an asterisk (*).

Fig. 2. Northern blot analysis of gp41 gene transcripts. Total RNA was extracted from uninfected Sf9 cells (lane C, the uninfected cell control) and SfMNPV infected Sf9 cells at 3, 6, 12, 24 and 48 p.i. (lanes 3 to 7 respectively). The gp41 gene transcripts were detected with a 3²P-labelled SfMNPV EcoRI-S DNA fragment. Lane M, size marker.

Fig. 3. Primer extension analysis of gp41 gene transcripts. Total RNA extracted from SfMNPV infected Sf9 cells at 48 h p.i. was mixed with the primer 5' GACGTAATCGACACACATTGT 3'. The cDNAs were synthesized using Moloney murine leukaemia virus reverse transcriptase and were separated on a 6% sequence gel. Three transcription start sites were identified (lane P, the primer extensor, product). The TA transcription start sites are within the TAAG motif. The upper T transcription start site is not associated with any known motif. The complementary sequence ladder is shown on the left-hand side in the sequence order G, A, T and C.
responding to the complementary region of the coding sequence from nucleotide 104 to 123, was used. Three transcription start sites were located (Fig. 3). Two of the transcription start sites are located at -42 and -41 nucleotides from the ATG translation start codon within the first T and second A of the TAAG consensus motif (Fig. 3).

Table 1. *Amino acid sequence similarities and nucleotide sequence identities (%) of NPV gp41 proteins*

<table>
<thead>
<tr>
<th></th>
<th>BmMNPV</th>
<th>HzSNPV</th>
<th>SfMNPV</th>
</tr>
</thead>
<tbody>
<tr>
<td>AcMNPV-E2</td>
<td>96</td>
<td>76</td>
<td>72</td>
</tr>
<tr>
<td>BmMNPV</td>
<td>(96)</td>
<td>(60)</td>
<td>(59)</td>
</tr>
<tr>
<td>HzSNPV</td>
<td></td>
<td>(59)</td>
<td>(59)</td>
</tr>
<tr>
<td>SfMNPV</td>
<td></td>
<td></td>
<td>76</td>
</tr>
</tbody>
</table>

* Bold lettering and normal lettering in parentheses denote amino acid sequence similarities and nucleotide sequence identities, respectively.

Another transcriptional start site is located at nucleotide -140 from the ATG start codon for which no consensus motif has been determined (Fig. 3).

Amino acid and nucleotide sequence comparison of SfMNPV with other baculoviruses

The amino acid and nucleotide sequences of the *Spodoptera frugiperda* gp41 gene were compared (Table 1) with those of three other NPV gp41 genes [from *Autographa californica* MNPV (AcMNPV-E2), *Bombyx mori* MNPV (BmMNPV) and *Helicoverpa zea* SNPV (HzSNPV)]. At the nucleotide level, the sequences of the NPVs have an average of 60% identity when compared with each other, except for AcMNPV-E2 and BmMNPV which share a much higher identity (97%). However, at the amino acid level, the predicted polypeptide sequences are more conserved (70% similarity). Kyte–Doolittle (1982) and Goldman (review by Engelman et al., 1986) analyses were performed to compare the distributions of hydro-
philic and hydrophobic domains among the four NPV proteins. Those of AcMNPV-E2 and BmMNPV have almost identical hydrophobicity patterns, while those of SfMNPV and HzSNPV show a similar hydrophobicity pattern overall (Fig. 4). In general, the hydrophobicity profiles of all four NPVs were similar within amino acids 100 to 340 of AcMNPV-E2 and BmMNPV and amino acids 40 to 280 of SfMNPV and HzSNPV (Fig. 4). The predicted amino acid sequences of all four NPVs were compared to show the conserved regions (Fig. 5). Sixteen conserved regions (defined as more than three contiguous amino acids being the same) were found within the whole sequence alignment. Within the 50 to 350 amino acid comparison region, 9 of 14 prolines were conserved among the NPVs.

In addition to the amino acid sequences, the nucleotide sequences of the upstream region from the ATG translation codon of the four NPVs were compared. The alignment sequence around the late gene transcriptional consensus motif from -52 to -46 nucleotides of all four
NPVs was identical (Fig. 6). Another late gene transcriptional motif from -20 to -17 was identified in AcMNPV-E2 and BmMNPV; however, this consensus region in SfMNPV and HzSNPV was changed by one or two nucleotides.

**Discussion**

A unique feature of the NPV life cycle is the production of two virion phenotypes: the occluded virion (OV) and extracellular virus (ECV). The biophysical, biochemical and morphological characteristics of OVs and ECVs are quite different. These structural differences may play a functional role in their biological properties. During viral infection, one of the virus-encoded envelope glycoproteins, gp64, is expressed and involved in host cell infection. gp64 is a component of the virion peplomers which are only detected in ECVs and are essential for entry of ECVs into the cells by adsorptive endocytosis (Keddie & Volkman, 1985). In contrast to gp64, gp41 is associated only with OVs. The gp41 structural protein was found exclusively in enveloped OVs but not in either ECVs or enveloped stripped OVs. Currently, the biological function of gp41 is not known, but it may be involved in facilitating the occlusion of virions in the polyhedra or the infection of host midgut cells according to their biochemical characteristics.

In this study, we present the nucleotide sequence and transcriptional analysis of the SfMNPV gp41 gene. The nucleotide sequence of the SfMNPV gp41 gene shows differing degrees of identity with three other NPVs (AcMNPV-E2, BmMNPV and HzSNPV; Table 1). The nucleotide sequence identities of the gp41 genes of SfMNPV and the other NPVs are low (60%). Similar results have been reported when the DNA homology was compared among NPVs of four different Spodoptera spp. (S. exempta, S. exigua, S. frugiperda and S. littoralis). SfMNPV is considered distantly related to the three other NPVs (20–30%) (Kelly, 1977). A molecular evolutionary study based on polyhedrin gene analysis also suggested that SfMNPV is distantly grouped from the AcMNPV and BmMNPV branch (Zanotto et al., 1993). The results show that SfMNPV diverged earlier from these other NPVs. The DNA Sequences of the gp41 genes of AcMNPV and BmMNPV are, however, almost identical (97%). Comparing these results to those found in the polyhedrin gene analysis suggest that AcMNPV and BmMNPV are very closely related species (Rohrmann, 1986; Van Strien et al., 1992).

When the hydrophilic and hydrophobic profiles of the gp41 protein of SfMNPV were compared with those of other NPVs, SfMNPV gp41 showed a similar pattern overall with HzSNPV. Amino acids 40 to 280 of AcMNPV-E2 and BmMNPV showed a pattern identical with that of amino acids 100 to 340 of HzSNPV and SfMNPV. The high hydrophilicity of the carboxy terminus of the p10 protein has been reported; it displays a functional domain which is exposed at the surface of the protein. The hydrophobic region in the middle of the p10 protein may have a bundling or cross-linking function (Van Oers et al., 1993).

The amino acid sequences of the gp41 protein of the above NPVs were compared to reveal the conserved sequence regions (Fig. 5). These conserved amino acid sequences may constitute an important functional domain. Specificially, the regions containing proline and cysteine may be involved in maintaining the conformation of the gp41 protein. In addition to these conserved regions, the first 50 amino acids in AcMNPV and BmMNPV were identical. Also, the terminal sequences (amino acids 368 to 393) of SfMNPV and HzSNPV were almost identical (Fig. 5). These data suggest that SfMNPV and HzSNPV may have evolved from a common ancestor and that AcMNPV and BmMNPV diverged from another distantly related ancestor.

By Northern blot analysis, two gp41 gene transcripts were found after 12 h p.i. These data confirm previous results indicating that the gp41 gene is transcribed late (Ma et al., 1993; Whitford & Faulkner, 1992b). One of
the transcripts was 1.6 kb and another was 2.8 kb long. According to the DNA sequence, the distance between the gp41 gene transcriptional start and stop sites is 1433 nucleotides. By adding the poly(A) tail [a poly(A) tail usually contain 200 bases], the size of the gp41 gene transcript is calculated to be about 1.6 kb. On the other hand, the 2.8 kb transcript did not fit the transcription termination stop-signal principle. This indicated that the transcription termination signal located 394 nucleotides downstream from the translation stop codon was bypassed. This phenomenon of ignoring the major transcriptional stop signal has been reported both for the gp41 gene (Whitford & Faulkner, 1992b) and the p39 capsid gene of AcMNPV (Thiem & Miller, 1989). The SfMNPV polyhedrin transcriptional stop signal is also not recognized when the gene is inserted into AcMNPV DNA, although this may be caused by the different molecular environment (Gonzalez et al., 1989). Another explanation for the two different size transcripts is that the 1.6 kb transcript was a spliced product from the 2.8 kb RNA. But this explanation is not favoured because the gp41 gene coding sequence did not seem to be separated into two regions. Also, gene splicing is not a common phenomenon in baculoviruses except for IEI and IEO (Kovacs et al., 1991).

By primer extension analysis, the transcription start site for the gp41 gene mRNA of SfMNPV was mapped in the promoter region within the TAAG motif at about nucleotide −42 or −41 (T or A). This motif is conserved in all baculovirus late genes, especially the baculovirus nucleotide −42 or −41 (T or A). This motif is conserved a common phenomenon in baculoviruses except for IEI in which the transcriptional start site is dissimilar to a consensus motif is also found in the AcMNPV p74 gene (Kovacs et al., 1991). Another explanation for the two different size transcripts is that the 1.6 kb transcript was a spliced product from the 2.8 kb RNA. But this explanation is not favoured because the gp41 gene coding sequence did not seem to be separated into two regions. Also, gene splicing is not a common phenomenon in baculoviruses except for IEI and IEO (Kovacs et al., 1991).

By primer extension analysis, the transcription start site for the gp41 gene mRNA of SfMNPV was mapped in the promoter region within the TAAG motif at about nucleotide −42 or −41 (T or A). This motif is conserved in all baculovirus late genes, especially the baculovirus structural proteins (Rohrmann, 1986, 1992; Rankin et al., 1988). However, another transcriptional start site was located at the −140 nucleotide for which no consensus motif has been determined. The phenomenon in which the transcriptional start site is dissimilar to a consensus motif is also found in the AcMNPV p74 gene (Kuzio et al., 1989). Another explanation for the difference could be a non-specific primer hybridization, since the baculoviruses have a large DNA genome.

An unexpected small ORF was located downstream of the −140 nucleotide transcriptional start site, and it may be that this start site is used for a bicistronic transcription. Similar bicistronic transcripts have been reported by Kovacs et al. (1991). A translational regulation mechanism is proposed in that paper since translation of the downstream ORF is more efficient compared to the upstream ORF. The upstream ORF may be used for increasing the translation initiation activity. At the same time, Ooi & Miller (1991) suggested an antisense RNA mechanism for transcriptional regulation which may be used to turn off initiation of a 3.2 kb RNA. In transcription of the gp41 gene, the upstream ORF may be used as a competition inhibitor to control gp41 gene transcription. However, we could not exclude the possibility of the bicistronic model although the upstream transcriptional start site is not a common transcriptional start site for baculovirus late genes.

Kool et al. (1994) sequenced the AcMNPV-E2 EcoRI-C fragment and found an extra G residue, close to the end of the gp41 gene coding region, when comparing the sequence with the data published by Whitford & Faulkner (1992b). These results were recently confirmed by Ayres et al. (1994). The differences in the gp41 gene sequences of AcMNPV may be caused by using a different strain. The results from Kool et al. (1994) and Ayres et al. (1994) not only enlarge the gp41 protein by 65 amino acids but also increase the homology with HzSNPV and SfMNPV at the carboxy-terminal regions (Fig. 5). These data provide new information showing the possible evolutionary path of the gp41 gene and by comparing these data, the evolutionary relationship of baculoviruses may be elucidated.

We appreciate the technical support for DNA sequence analysis and advice for manuscript preparation of Alejandra Garcia-Canedo, and we thank Dr Drion Boucias, Dr Richard Condit and Dr Ernest Hiebert for critical review of this paper. Also, we acknowledge Vivian Chang and the Biological Computing Facility (BCF) of the Interdisciplinary Center for Biotechnology Research (ICBR) at the University of Florida for the computer program support. Florida Agriculture Experiment Station Journal Series, no. R04088.

References


MARUNIAK, J. E. (1979). Biochemical characterization of baculovirus structural and infected TN-368 cell polypeptides, glycoproteins, and phosphoproteins. PhD dissertation, University of Texas, Austin, TX, USA.


(Received 27 September 1994; Accepted 15 February 1995)