Sequence and transcriptional analysis of the ubiquitin gene cluster in the genome of *Spodoptera exigua* nucleopolyhedrovirus

Elisabeth A. van Strien, Bastiaan J. H. Jansen,† Ruud M. W. Mans,‡ Douwe Zuidema and Just M. Vlak

Department of Virology, Wageningen Agricultural University, Binnenhaven 11, 6709 PD Wageningen, The Netherlands

The nucleotide sequence of a 1200 bp DNA fragment of *Spodoptera exigua* nucleopolyhedrovirus (SeMNPV) was determined. This sequence contained a cluster of two open reading frames (ORFs), one coding for a viral ubiquitin (v-ubi) and another with homology to orf2 of *Autographa californica* (Ac) MNPV and *Orgyia pseudotsugata* (Op) MNPV. The v-ubi ORF is 240 nucleotides (nt) long, potentially encoding a protein of 80 amino acids with a predicted molecular mass of 9.4 kDa. The amino acid sequence of the v-ubi gene in SeMNPV has 75% and 81.6% identity with the v-ubi gene of AcMNPV and OpMNPV and approximately 84% with cellular ubiquitins. Northern blot analysis revealed three major small transcripts late in infection, of about 690, 550 and 400 nt long. Primer extension analysis showed that transcription started from within two consensus late promoter elements (TAAG), located at positions -6 and -30. The start site at position -4/-5 precedes the shortest leader reported to date for a baculovirus gene. The other ORF, xb187, was identified in the opposite orientation immediately upstream of the v-ubi gene. This ORF potentially encodes a 22 kDa protein with unknown function and about 60% amino acid similarity to the products of the orf2 genes of AcMNPV and OpMNPV. The SeMNPV xb187 ORF is transcribed late in infection via two transcripts, 1.2 kb and 770nt long. The v-ubi–xb187 gene cluster is located at map unit (m.u.) 89 on the genome of SeMNPV. This is different from the position of an identical cluster in the AcMNPV and OpMNPV genomes, located at relative m.u. 20.

Introduction

Ubiquitins are small proteins, abundantly present in eukaryotic cells and thought to be involved in an array of basic cellular processes, such as cell cycle control, stress response, ribosome biogenesis and cell differentiation (see Finley & Chau, 1991; Ciechanover & Schwartz, 1994 for review). Eukaryotic ubiquitins are 76 amino acids long and highly conserved (96% identity), suggesting an essential function in the life cycle of the cell. Cellular ubiquitins occur free or covalently linked to other proteins and are processed via the ubiquitin pathway. In this pathway ubiquitin monomers are covalently linked by their C-terminal Gly-76 to lysine residues of target proteins and to internal lysines of other ubiquitin molecules (Lys-48) or to form multi-ubiquitin chains. Recently, it has been shown that Lys-63 and Lys-29 of ubiquitin can also be used as acceptors for ubiquitination (Arnason & Ellison, 1994, Johnson et al., 1995). Ubiquitination has been implicated strongly in protein degradation (Doherty & Mayer, 1992).

Ubiquitins also play a role in the life cycle of a wide range of viruses. Host ubiquitins have been found covalently linked to some plant viruses (Dunigan et al., 1988) and free ubiquitins were observed in avian leukosis virus particles (Puttermann et al., 1990). Herpes simplex virus type I encodes a protein (ICP4) that induces polyubiquitin gene expression in the host (Kemp & Latchman, 1988). African swine fever virus contains ubiquinated virion proteins (Hingamp et al., 1995) and codes for a protein that shows homology to enzymes involved in the ubiquitin pathway (Rodriguez et al., 1992). Involvement of the ubiquitin pathway in the propagation of viruses has been suggested (Driscol & Finley, 1992).

Ubiquitin-like genes have been found in the baculoviruses.
**Autographa californica** nucleopolyhedrovirus (AcMNPV) (Guarino, 1990) and *Orgyia pseudotsugata* (Op) MNPV (Russell & Rohrmann, 1993), showing 75% (AcMNPV) and 80% (OpMNPV) amino acid identity with eukaryotic ubiquitins. Failed attempts to delete the gene for virus-coded ubiquitin (v-ubi) from AcMNPV suggest that ubiquitin is essential in the baculovirus life cycle (Guarino, 1990). In AcMNPV both v-ubi as well as host ubiquitin are attached to the inner surface of budded virion membranes by a novel type of phospholipid homologue in the baculovirus life cycle (Guarino, 1990). In contrast to AcMNPV, SeMNPV is restricted to a single host, the beet army worm *S. exigua* (Lepidoptera: Noctuidae). SeMNPV is being applied as a biological control agent against this pest insect (Gelernter & Federici, 1986; Smits & Vlak, 1994). SeMNPV is distantl related to AcMNPV, the baculovirus type species, based on comparisons of polyhedrin (van Strien et al., 1992) and p10 gene sequences (Zuidema et al., 1993). The availability of the complete nucleotide sequence of the AcMNPV genome (Ayres et al., 1994) allows comparison of gene sequences and gene organization among baculoviruses.

In this paper, we report the nucleotide sequence and transcriptional analysis of the v-ubi gene of SeMNPV and a flanking ORF (xb187). The amino acid sequence of v-ubi was compared to those of AcMNPV and OpMNPV and to several eukaryotic ubiquitins. In addition, we compared ORF xb187 with the putative homologous orf2 of AcMNPV (Guarino & Smith, 1990) and OpMNPV (Russell & Rohrmann, 1993). Finally, we matched the relative location of the v-ubi and xb187 ORFs of SeMNPV with those of AcMNPV and OpMNPV.

**Methods**

- **Virus, insects and cells.** The SeMNPV field isolate (SeMNPV/US) (Gelernter & Federici, 1986) was kindly provided as polyhedra by B. A. Federici (Department of Entomology, University of California, Riverside, USA). The polyhedra were propagated in fourth instar *S. exigua* larvae (Smits et al., 1988). Extracellular virus (ECV), used in time-course infection experiments, was obtained from the supernatant of IZD-Se-2109 cells infected with haemolymph obtained from SeMNPV-infected fourth instar larvae. The *S. exigua* cell line (obtained from B. Moeckel, Institute of Zoology, Technical University, Darmstadt, Germany) was maintained in plastic tissue-culture flasks in TNM-FH medium (Hink, 1970) supplemented with 10% fetal bovine serum. ECV titres were determined by the end point dilution method (Vlak, 1979) and expressed as TCIDso.

- **Isolation of total RNA and Northern blot hybridization.** Total RNA for Northern blot analysis and primer extension was isolated from SeMNPV-infected IZD-Se-2109 cells at several time-points post-infection (p.i.), as described by Xie & Rothblum (1991). Total RNA was denatured, electrophoresed in agarose gels and blotted onto Hybond-N nylon membrane (van Strien et al., 1992). To identify v-ubi and ORF xb187 transcripts, the blots were hybridized for 16 h at 65 °C with [α-32P]dATP by using T4 polynucleotide kinase (Gibco-BRL). Riboprobies were generated by *in vitro* transcription (Sambrook et al., 1989) by using T7 or T3 RNA polymerase (Gibco-BRL). Templates used for *in vitro* transcription were generated by cloning of DNA fragments containing either v-ubi or ORF xb187 sequences into pBluescript KS (+). After hybridization, the filters were washed for 5 min with 2 × SSC, 0.5% SDS at room temperature, 30 min with 2 × SSC, 0.1% SDS at 65 °C and 30 min with 0.1 × SSC, 0.1% SDS at 65 °C. The filters were exposed to Kodak XAR film.

- **Primer extension.** To identify the transcriptional start site(s) of SeMNPV v-ubi, 15 ng of an oligonucleotide (5′CGGCGTGATCCTCAGGCCAGG3′), complementary to the v-ubi mRNA was labelled at the 5′ end with [α-32P]dATP by using T4 polydeoxynucleotide kinase (Gibco-BRL) in 50 mM-Tris–HCl, pH 9.5, 10 mM-MgCl2, 5 mM-DTT, 5% glycerol for 45 min at 37 °C followed by heat denaturation at 90 °C for 10 min. The labelled oligonucleotide was purified on a 1 ml Sephadex G-25 column. Labelled primer was added to 2 μg of total infected-cell RNA and the mixture was denatured at 90 °C for 5 min and annealed at 54 °C for 15 min. Reverse transcription was carried out at 48 °C for 1 h in a volume of 15 μl containing 5 mM of each of the dNTPs and 1 μl Superscript reverse transcriptase (Gibco-BRL) in a buffer supplied by the manufacturer. The reaction was stopped by addition of 5 μl 'stop' buffer containing 95% (v/v) formamide, 0.01% xylene cyanol and 0.01% bromophenol blue. Six μl of the reaction mixture was analysed in a 6% polyacrylamide sequence gel, followed by drying and autoradiography.

**Results**

- **Location and sequence of the ubiquitin region on the SeMNPV genome.** Upon characterization of the region upstream of the p10 gene of SeMNPV (Zuidema et al., 1993), a genomic EcoRI bank was screened with a probe, an EcoRI–Xbal fragment derived from the left-hand end of Xbal-H. This resulted in the detection of a 2.2 kb EcoRI fragment overlapping the Xbal site between the fragments Xbal-B and Xbal-H (Fig. 1). Sequences upstream of the 2.2 kb EcoRI fragment were obtained from cosmid COS22, encompassing fragments Xbal-B, -H, -D and -R (J. G. M. Heldens and others, unpublished). A segment of approximately 1200 nucleotides (nt), located around map unit

**Localization and DNA sequencing of the ubiquitin gene region.** Colony filter hybridization of an SeMNPV EcoRI genomic library in pBluescript KS (+) (Stratagene), with an [α-32P]dATP labelled probe was performed essentially as described by Sambrook et al. (1989). Subfragments of a selected fragment were isolated from agarose gels using the freeze–squeeze method (Sambrook et al., 1989) and subcloned into the plasmids pTZ19R (Promega) or pBluescript KS (+) (Stratagene). Sequencing of SeMNPV inserts and cosmids with standard sequencing primers and custom designed primers (Eurogentec) was performed with Taq polymerase, using the chain termination method of Sanger et al. (1977) and an automatic sequencer (Applied Biosystems). Sequences were analysed with the UWCGC computer programs (Derveaux et al., 1984) and DNA and deduced amino acid sequences were compared with the updated GenBank/EMBL, SwissProt and PIR data libraries using the BLAST and FASTA programs.
Fig. 1. Location of the ubiquitin gene region on the genomic map of the SeMNPV genome. The XbaI restriction map of the SeMNPV genome is shown (top). The position of the cosmid (COS22) and the locations and orientation of the polyhedrin, p10, ubiquitin and xb187 genes are indicated (arrows). Detailed physical map of the ubiquitin region (bottom). E, EcoRI; X, XbaI; C, Clal; S, Sall.

Transcriptional analysis of the ubiquitin gene region

Transcriptional activity of the SeMNPV v-ubi gene in insect cells was ascertained by Northern blot analysis of RNA isolated at various times after infection by using a strandspecific probe of the v-ubi gene. Three major transcripts, of approximately 400, 550 and 690 nt, were observed late in infection (Fig. 3a). Overexposure of the hybridized blot showed that the 690 nt transcript was first detectable at 8 h p.i., increased in intensity up to 24 h p.i. and appeared to be less abundant very late in infection. The two smaller transcripts of 400 and 550 nt were present from 12 h p.i. and their concentration continued to increase up to 48 h p.i. The amount of the 550 nt transcript was considerably lower. Longer transcripts were also observed, which appeared to originate further upstream of the v-ubi gene, as determined by an antisense riboprobe overlapping with ORF xb187 (data not shown).

RNA primer extension analysis was performed to determine the transcriptional start site(s) of the v-ubi gene (Fig. 3c). An oligonucleotide, complementary to nucleotides 147-162 with respect to the translational start codon of the v-ubi gene of SeMNPV, was used (Fig. 2). The reverse transcription assay showed two major adjacent stops, at -4/-5 and at -28/-29 nt relative to the translational start codon (Fig. 3c), with a slight preference for the nucleotides at -4 and -28. This places the two transcriptional starts of SeMNPV v-ubi at the adenine residues of the consensus late promoter elements TAAG.

Nucleotide sequence analysis revealed the presence of a canonical poly(A) signal (AAATAA) (Birnstiel et al., 1985) at 258 nt downstream of the translational stop codon of the SeMNPV v-ubi gene (Fig. 2). It is embedded in an A/T-rich region. The predicted size of the v-ubi mRNA is therefore approximately 520 nt, indicating that the largest stretch of this mRNA is not translated into protein. This also suggests that the v-ubi mRNA has a tail of approximately 170 adenine residues to give an mRNA of 690 nt in size.

Northern blot analysis of the xb187 gene (Fig. 3b) shows a major transcript of approximately 770 nt and a minor transcript
Fig. 2. Nucleotide sequence of an 1175 bp region of the SeMNPV genome containing the ubiquitin and xb187 genes. The ubiquitin coding sequence starts at nt 513 and terminates at nt 273; the xb187 gene starts at nt 604 and terminates at nt 1164. Late transcription initiation signals (TAAG) and putative transcription polyadenylation signals (AATAAA) are underlined.

The 770 nt transcript could start at -10 and terminate at the poly(A) signal just downstream of the TAA stop codon, assuming a poly(A) tail of about 150 adenine residues. A v-ubi antisense riboprobe was used to investigate if the longer transcript overlapped with the v-ubi gene. Since no hybridization signal was detected, this implied that both xb187 transcripts were 5′ coterminal (data not shown).

Comparison of the SeMNPV v-ubi and xb187 genes

The amino acid sequence encoded by the SeMNPV v-ubi ORF was compared to those of AcMNPV (Guarino, 1990) and OpMNPV (Russell & Rohrmann, 1993), as well as to the ubiquitin monomer sequences of Drosophila melanogaster.
Fig. 3. Transcriptional analysis of the SeMNPV v-ubi-xb187 gene region. Northern blot analysis of total RNA extracted from uninfected (lane C) and SeMNPV-infected IZD-Se2109 cells 2, 4, 6, 8, 12, 16, 24 and 48 h p.i. for v-ubi (a) and xb187 (b) transcripts. (c) Primer extension analysis of v-ubi transcripts performed with a 16-mer oligonucleotide complementary to the v-ubi RNA, 32P-labelled at the 5’ end. The oligonucleotide was annealed to total RNA from uninfected (lane 1) and SeMNPV-infected (lane 2) cells isolated 48 h p.i. and elongated by reverse transcription. The sizes of the extension products were determined by comparison with a sequence ladder run alongside (lanes C, T, A and G) obtained from an SeMNPV v-ubi containing plasmid clone and the 16-mer oligonucleotide as a sequence primer. Asterisks indicate the position of the 5’-terminal nucleotides.

(Arribas et al., 1986) and S. frugiperda (Guarino, 1990) (Fig. 4a). The baculoviral ubiquitins, including SeMNPV v-ubi, differ from the cellular ubiquitins in their C-terminal extension by one or more amino acids. Processed cellular ubiquitins have a fixed length of 76 amino acids. Therefore, only the first 76 amino acids of SeMNPV v-ubi were considered in calculating
the extent of homology. The homology (identity) of SeMNPV ubiquitin with AcMNPV and OpMNPV ubiquitin is 75·0% and 81·6%, respectively, whereas the homology with D. melanogaster and S. frugiperda ubiquitin is 84·2% and 82·9%, respectively (Table 1). Notably, the ubiquitin protein of *Bombyx mori* (Bm) NPV is 100% identical to its AcMNPV homologue (complete BmNPV sequence is in GenBank/EMBL, accession no. L33180).

The residues that are known to be involved in the formation of ubiquitin–protein complexes in eukaryotes, Lys-29, Lys-48, Lys-63 and Gly-76, are all conserved (Fig. 4; Johnson et al., 1995; Chau et al., 1989; Finley & Chau, 1991; Arnason &
Table 1. Amino acid sequence identity (%) between the ubiquitin proteins of SeMNPV, AcMNPV, OpMNPV and S. frugiperda and D. melanogaster

<table>
<thead>
<tr>
<th></th>
<th>AcMNPV</th>
<th>OpMNPV</th>
<th>Dm</th>
<th>Sf</th>
</tr>
</thead>
<tbody>
<tr>
<td>SeMNPV</td>
<td>75.0</td>
<td>81.6</td>
<td>84.2</td>
<td>82.9</td>
</tr>
<tr>
<td>AcMNPV</td>
<td>85.5</td>
<td>70.3</td>
<td>75.0</td>
<td></td>
</tr>
<tr>
<td>OpMNPV</td>
<td>80.3</td>
<td>78.9</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dm</td>
<td>98.7</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Ellison, 1994; Spence et al., 1995). The altered amino acid residues, at positions 23 and 28, relative to the eukaryotic ubiquitins, are conserved among the baculoviral ubiquitins. The most heterogeneous parts of the SeMNPV ubiquitin are the areas from amino acid residues 15–31 and 53–57. In comparison to eukaryotic ubiquitins, SeMNPV ubiquitin has four additional amino acids at its C terminus. AcMNPV and OpMNPV ubiquitins have 1 and 18 additional amino acids at their C terminus, respectively. The C termini of the other v-ubiquitins do not share homology to SeMNPV ubiquitin (Fig. 4a).

The xb187 gene shows homology (similarity) of 59%, 59% and 61% at the amino acid level with the orf2 genes of AcMNPV, BmNPV and OpMNPV (Fig. 4b), if one allows conserved amino acid substitutions. The amino acid sequence of the putative protein encoded by the xb187 gene is considerably shorter than that of the orf2 genes of AcMNPV, BmNPV and OpMNPV (187 amino acids versus 215, 215 and 219, respectively), truncated by about 30 amino acids at the N terminus. The N-terminal region of the ORF2 and XB187 proteins is very heterogeneous among the four baculoviruses, whereas the C-terminal region displays a higher degree of homology. A search in the updated releases of the GenBank/EMBL database with FASTA and BLAST (Altschul et al., 1990) from the GCG software package did not show significant homologies to other genes.

Discussion

In this report we present the nucleotide sequence and transcriptional analysis of two adjacent ORFs, v-ubi and xb187, on the XbaI-B fragment of the baculovirus SeMNPV.

The general features of the SeMNPV ubiquitin gene seem to be similar to those of other ubiquitins (Fig. 4 and Table 1). The amino acid residues involved in ubiquitination, Lys-29, Lys-48, Lys-63 and Gly-76, are conserved. The additional amino acids at the C termini of the baculoviral ubiquitins may be the remnant of the C-terminal extension of either an ancestral cellular ubiquitin-fusion gene or polyubiquitin gene. However, the residues known to be essential for proteolytic processing, Gly-Gly-X (Lopez-Otin et al., 1989), are conserved and ubiquitin C-terminal hydrolase activity is present in AcMNPV-infected S. frugiperda cells (Moguilevsky et al., 1994). Based on these findings it is likely that the C termini of the baculoviral ubiquitins are processed correctly to yield a ubiquitin of 76 amino acids.

Comparison of the predicted amino acid sequences suggest that SeMNPV ubiquitin is more closely related to OpMNPV than to AcMNPV ubiquitin. The identity with D. melanogaster and S. frugiperda ubiquitin is even greater (Table 1). However, since the variation among the baculoviral ubiquitins is much higher than that observed among cellular ubiquitins, these data show that baculoviral ubiquitins have diverged considerably from the highly conserved cellular ubiquitins and from each other. This may suggest that v-ubiquitins have a special function in baculovirus infection, as for example the association of ubiquitin with ECVs via a phospholipid anchor (Guarino et al., 1995). With these data on v-ubi of SeMNPV, three nucleopolyhedrovirus ubiquitins have been sequenced. Moreover, the identification of a v-ubi gene in the genome of Cydia pomonella granulovirus (N. E. Crook, personal communication), suggests that this gene most likely is preserved in baculoviruses.

Analysis of SeMNPV v-ubi transcription revealed that transcripts started with equal frequency in the two canonical TAAG late promoter elements located at positions –6 and –30 relative to the translational start codon (Figs 2 and 3c). This is unlike the situation in AcMNPV, where the two v-ubi transcripts are present in unequal amounts and have longer untranslated 5′ leaders (Guarino, 1990). The 5′ sequences of the baculovirus v-ubi genes differ considerably (Fig. 5). Only the TAAG consensus promoter sequences are conserved. The transcriptional start site of the SeMNPV v-ubi gene at position –4 is the shortest leader reported to date for a baculovirus gene and is reminiscent of the enhancer gene transcription of Heliothis armigera granulovirus (HaGV) (Roelvink et al., 1995).

**Fig. 5.** Alignment of the 5′ noncoding leader sequence of the v-ubi genes of SeMNPV, AcMNPV and OpMNPV. Putative promoter elements are indicated in bold and transcriptional start sites are indicated by asterisks.
Three major v-ubi transcripts, of about 690, 550 and 400 nt in length were observed in infected cells late in infection. The size of the largest mRNA corresponds with the expected length of a polyadenylated transcript of v-ubi (Fig. 2), assuming a poly(A) tail of approximately 150–200 nt. The shorter transcripts of 550 and 400 nt, which appear later in infection (after 12 h p.i., Fig. 3a), may represent non-polyadenylated v-ubi mRNA. The 550 nt transcript might terminate near the poly(A) signal, whereas the 400 nt transcript might terminate in between the translational stop codon and the poly(A) signal, thereby precluding a poly(A) tail. There are a few previous reports on non-polyadenylated transcripts in baculoviruses. The SeMNPV polyhedrin gene transcripts are not polyadenylated (van Strien et al., 1995). The xbi87 gene of SeMNPV shows homology to the orf2 genes of AcMNPV (Guarino & Smith, 1990) and OpMNPV (Russell & Rohrmann, 1993) and is located next to the v-ubi gene in the opposite orientation in all three viruses. The SeMNPV polyhedrin gene transcripts are not polyadenylated (van Strien et al., 1992). The mRNA of the HaGV enhancin gene is also not polyadenylated and has an equally short leader (Roelvink et al., 1995).

The xbi87 gene of SeMNPV shows homology to the orf2 genes of AcMNPV (Guarino & Smith, 1990) and OpMNPV (Russell & Rohrmann, 1993) and is located next to the v-ubi gene in the opposite orientation in all three viruses. The highest homology is found at the C terminus, suggesting that functional domains are located in this part of the protein.

The xbi87 gene is active as a late gene (Fig. 3b). The presence of two xbi87 transcripts can be explained by the fact that there are two poly(A) signals, one overlapping with the last five nucleotides of the ORF and another 431 nt downstream of the translational stop codon (data not shown). The estimated sizes of the transcripts correspond well with the sequence data, including a presumed poly(A) tail of approximately 200 nt. The putative 5' ends of the mRNAs encoding v-ubi and xbi87 do not overlap in SeMNPV and the intergenic region is much larger than in the genomes of AcMNPV and OpMNPV. In SeMNPV this region contains an ACG repeat of unknown function.

In the baculoviruses analysed so far, the v-ubi and xbi87 genes or their homologues are clustered, but located at different positions in the viral genome. In SeMNPV the v-ubi–xbi87 gene cluster is found around m.u. 89, whereas the same cluster in AcMNPV and OpMNPV is located around m.u. 22 and m.u. 19, respectively (Fig. 6). This supports the idea that the baculovirus genome organization has been rearranged during its evolutionary history (Gombart et al., 1989). Sequence comparisons of the SeMNPV v-ubi–xbi87 gene cluster with AcMNPV and OpMNPV, added to comparisons based on polyhedrin gene (van Strien et al., 1992; Zanotto et al., 1993) and p10 gene (Zuidema et al., 1993) sequences, confirmed the distant relationship of SeMNPV to these two viruses.

Note added in proof. Recently, L. M. O’Reilly and L. A. Guarino (Virology, 218, 243–247, 1996) showed by generating a frame-shift mutation in the AcMNPV ubiquitin gene that this gene is not essential for virus replication in cell culture. However, production of budded virus by this mutant is much reduced.

This research was supported by Grant no. 93/20 from DIARP, the Dutch-Israeli Agriculture Research Program. We thank Magda Usmany and Els Kliinge-Roode for technical support and Basil Arif for critical reviewing of the manuscript.

References


Received 3 January 1996; Accepted 3 May 1996