Nucleotide sequence of the p39-capsid gene region of the *Lymantria dispar* nuclear polyhedrosis virus

Rebecca M. Bjornson and George F. Rohrmann*

Department of Agricultural Chemistry, Oregon State University, Corvallis, Oregon 97331-6502, U.S.A.

A 1-85 kb region, containing an open reading frame (ORF) homologous to the baculovirus p39-capsid gene, was sequenced from the *Lymantria dispar* multicapsid nuclear polyhedrosis virus (LdMNPV) genome. Analysis of the p39-capsid gene demonstrated that it was 39% and 47% identical in amino acid sequence with the homologous genes in the *Autographa californica* and *Orgyia pseudotsugata* MNPVs, respectively. Two late promoter elements located upstream of the p39 gene in the LdMNPV genome are conserved downstream with two other baculoviruses, whereas an ORF located downstream is not conserved.

A major component of the capsid of baculoviruses is a protein of 39K (Russell et al., 1991). The gene encoding this protein has been identified in both the *Autographa californica* multicapsid nuclear polyhedrosis virus (AcMNPV) (Thiem & Miller, 1989a) and *Orgyia pseudotsugata* MNPV (OpMNPV) (Blissard et al., 1989) and the predicted proteins show 59% amino acid sequence identity. The p39 genes of both AcMNPV and OpMNPV appear to be transcribed under the regulation of two late promoter elements. In addition, downstream of the p39 gene in AcMNPV and OpMNPV is a gene, CG30, that encodes a protein with putative zinc finger and leucine zipper domains (Thiem & Miller, 1989b). To identify and characterize the region encoding this gene in the *Lymantria dispar* MNPV (LdMNPV), a SmaI fragment of about 800 nucleotides (nts) from within the OpMNPV p39 gene (Blissard et al., 1989) was hybridized under non-stringent conditions to blots of cosmids encompassing the complete LdMNPV genome (for a description of all materials and methods see Bjornson & Rohrmann, 1992). Hybridization occurred to specific restriction fragments on cosmid D (data not shown) located at about 99 kb (map unit 61) (Fig. 1).

The fragments demonstrating hybridization to the OpMNPV p39 gene were cloned, and a region of 1854 nt was sequenced. An ORF of 1068 nt encoding a predicted protein of 39643 Mr (356 amino acids) was identified (Fig. 1 and 2). Two consensus sequences for the baculovirus late gene promoter (ATAAG at nt 177 and 225; Fig. 2) (Rohrmann, 1986) were located 95 and 133 nt upstream of the translational start site. Another promoter sequence (GTAAG) was found at nt 598 within the ORF and in the opposite direction.

A search of the PIR protein database revealed no significant sequence similarity between the predicted LdMNPV p39 protein and other known proteins except for the homologous proteins from OpMNPV (Blissard et al., 1989) and AcMNPV (Thiem & Miller, 1989a) (Fig. 3). The AcMNPV and OpMNPV p39 amino acid sequences are 59% similar whereas they are 39% and 47% similar, respectively, to the LdMNPV p39 sequence. Particularly striking is the conservation of all eight cysteines in the sequence (Fig. 3) suggesting that disulphide bonds may be involved in the higher ordered structure of p39. Another feature of the sequences is the lack of homology of the C-terminal 65 amino acids; this region of the LdMNPV p39 sequence showed only 11% and 22% sequence identity with the AcMNPV and OpMNPV p39 proteins, respectively (Fig. 3). This region in the AcMNPV and OpMNPV p39 sequence was also relatively variable (38% identity). Despite the lack of sequence similarity in the C-terminal region of the LdMNPV and OpMNPV p39 ORFs, both contain sequences of 14 to 17 amino acids with only alanine and proline residues. The alanine/proline-rich sequences of both viruses are encoded by DNA sequences with a G+C content of 90 to 100%.

Upstream of the p39 genes from AcMNPV (Thiem & Miller, 1989a), OpMNPV (Blissard et al., 1989) and LdMNPV is located a region containing two baculovirus late promoter motifs (ATAAG). A 71 nt region containing these promoter sequences was highly conserved (75 to 86%) between the viruses (Fig. 4a). This suggests that the dual promoter organization may be involved in a conserved regulatory phenomenon or that the sequence may be a component of a highly conserved
Fig. 1. Location of the p39-capsid gene on the LdMNPV genome. (a) A BgII map of the LdMNPV genome that shows the location of the polyhedrin (Smith et al., 1988), DNA polymerase (R. M. Bjornson & G. F. Rohrmann, unpublished); p39-capsid and polyhedron envelope (PE) gene regions (Bjornson & Rohrmann, 1992). The location of selected cosmid clones is shown above the LdMNPV genome map. (b) Detailed map of the p39 gene region showing the location and direction of the LdMNPV p39 ORF. The dashed arrow indicates the location of a possible additional late gene. The cross-hatched bar indicates the region sequenced and analysed in this report.

Fig. 2. Nucleotide sequence of the p39 gene region. Late promoter (LP) consensus sequences (ATAAG, nt 373 and 421) are over- and underlined and selected restriction enzyme sites are underlined. The predicted amino acid sequence is shown below the nucleotide sequence. In addition to exonuclease III-generated deletion mutants, three oligonucleotides corresponding to nt 432 to 451, 1348 to 1325 and 1425 to 1441 were used in sequencing the capsid gene. The DNA was sequenced in both directions.
ORF. The sequence was analysed for ORFs in both directions. The forward direction (the same direction as p39) shows stop codons in all three reading frames in the promoter region (nt 370 to 430; Fig. 2). However when the inverse complement of the LdMNPV p39 region was examined, a late promoter sequence (GTAAG) was identified (nt 598) and an ORF was identified beginning at and overlapping the p39 ATG and reading out to the 5' end of the region sequenced (see Fig. 1b). Similarly, both AcMNPV and OpMNPV possessed an ORF containing highly homologous sequences through the p39 conserved region (Fig. 2). An ORF (termed CG30) containing putative zinc finger and leucine zipper amino acid sequence motifs was identified immediately downstream of the p39 gene.

A homologous gene is also present in the sequence downstream of the OpMNPV p39 gene (Blissard et al., 1989). It was suggested by Thiem & Miller (1989b) that CG30 could form a capsid-DNA-binding fusion protein if frameshifting occurred in the carboxy region of the capsid gene. No homologue of CG30 was observed in the 300 nt upstream of the p39 gene. The sequence extends downstream of the 3' end shown in the figure but no homology is present. The amino acids encoded by the complement of the p39 promoter sequences are underlined. Gaps are indicated by (-), underlined amino acids indicate identity between OpMNPV and LdMNPV, and the conserved cysteines are indicated by (+) above the amino acid. The AcMNPV, OpMNPV and LdMNPV p39 genes. Amino acid identities are indicated by the dots below the AcMNPV sequence, gaps inserted to facilitate alignment are indicated by (-), underlined amino acids indicate identity between OpMNPV and LdMNPV, and the conserved cysteines are indicated by (+) above the amino acid. The AcMNPV and OpMNPV data are from Thiem & Miller (1989a) and Blissard (1989), respectively.

Fig. 3. Comparison of the predicted amino acid sequences of the AcMNPV, OpMNPV and LdMNPV p39 genes. Amino acid identities are indicated by the dots below the AcMNPV sequence, gaps inserted to facilitate alignment are indicated by (-), underlined amino acids indicate identity between OpMNPV and LdMNPV, and the conserved cysteines are indicated by (+) above the amino acid. The AcMNPV and OpMNPV data are from Thiem & Miller (1989a) and Blissard et al. (1989), respectively.

An ORF (termed CG30) containing putative zinc finger and leucine zipper amino acid sequence motifs was identified immediately downstream of the p39 gene in AcMNPV (Thiem & Miller, 1989b). A homologous gene is also present in the sequence downstream of the OpMNPV p39 gene (Blissard et al., 1989). It was suggested by Thiem & Miller (1989b) that CG30 could form a capsid-DNA-binding fusion protein if frameshifting occurred in the carboxy region of the capsid gene. No homologue of CG30 was observed in the 300 nt of 3' flanking sequence following the LdMNPV p39 gene (Fig. 2).

We thank Christian Gross and Doug Leisy for critically reading this manuscript. This project was supported by grants from the USDA (89-37250-4706 and 91-37302-6310). This is Technical Report No. 9806 from the Oregon State University Agricultural Experiment Station.

References


(Received 22 October 1991; Accepted 17 February 1992)