Characterization of the nucleotide sequence of the *Lymantria dispar* nuclear polyhedrosis virus DNA polymerase gene region

Rebecca M. Bjornson, Barbara Glocker and George F. Rohrmann*

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

The DNA polymerase gene of the *Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus (LdMNPV) was cloned and sequenced. The predicted DNA polymerase protein (1113 amino acids, 115.9K) was found to have an amino acid identity of 48% with the corresponding gene of the *Autographa californica* MNPV (AcMNPV). It contains five domains associated with substrate binding, primase interaction, and pyrophosphate hydrolysis and three domains associated with 3'-5' exonuclease activity common to other DNA polymerases. A region with a conserved TATA promoter and a CAGT mRNA start site sequence motif was identified and shown to be transcribed by RNA polymerase II, indicating that the LdMNPV DNA polymerase gene is expressed as an early gene. An open reading frame possibly expressed as a late gene, oriented in the opposite direction and overlapping the N-terminal coding region of the DNA polymerase gene was found in the LdMNPV sequence and was shown to be conserved in the same position in AcMNPV.

**Introduction**

*Lymantria dispar* multinucleocapsid nuclear polyhedrosis virus (LdMNPV) is pathogenic for the gypsy moth (*Lymantria dispar*), a major defoliator of forest and shade trees in the northeastern United States. Although LdMNPV can contribute significantly to the collapse of gypsy moth populations and a LdMNPV virus preparation has been registered for use as an insecticide, this virus is not well-characterized. Recently the nucleotide sequences of two regions of the LdMNPV genome have been reported (Bjornson & Rohrmann, 1992a, b). These encode structural protein genes that are shared with other baculoviruses. One region, 1-85 kb in size, contains an open reading frame (ORF) homologous to the p39-capsid genes of *Autographa californica* MNPV (AcMNPV) and *Orgyia pseudotsugata* MNPV (OpMNPV). The other region, 6-4 kb in size, contains ORFs homologous to the polyhedron envelope protein gene, and to the 3’-terminal half of ORF 1. ORF 1 is a gene located upstream of the polyhedron envelope protein gene in other baculoviruses. These three genes demonstrated amino acid sequence identities of 39%, 27% and 34% with the corresponding predicted proteins in AcMNPV. In addition, the LdMNPV polyhedrin gene has been located and sequenced and shows approximately 80% amino acid sequence identity to AcMNPV polyhedrin (Smith *et al.*, 1988; Chang *et al.*, 1989). Despite the sequence similarity of the LdMNPV genes with their homologues from AcMNPV and OpMNPV, there was no overall pattern of conservation detected between the LdMNPV and AcMNPV genomes. To build on our understanding of both the genome organization of the LdMNPV and the relatedness of baculoviruses to other organisms, we extended these investigations to the characterization of an LdMNPV gene that is present in most organisms, the DNA polymerase (DNAP) gene. A gene encoding DNAP was previously reported for AcMNPV and was observed to be expressed as an early gene (Tomalski *et al.*, 1988). In this report we describe the location, nucleotide sequence, and phylogenetic relatedness of the LdMNPV DNAP gene.

**Methods**

*LdMNPV* cosmids and DNA sequence analysis. LdMNPV cosmids were derived by cloning DNA from LdMNPV clonal isolate CI 5-6 (Slavicek, 1991) into the cosmid vector pHC79 (Hohn & Collins, 1980) as partial digests made using *PstI* or *ClaI*. They were supplied by the U.S. Forest Service Laboratory in Delaware, Ohio. All plasmid subcloning was done in pBlueScribe (Stratagene) modified by the addition of a *BglII* site (Gombart *et al.*, 1989).

Because of the 60% G+C content of LdMNPV DNA (McCarthy *et al.*, 1979), double-stranded plasmid DNA was sequenced using *Taq* polymerase and related reagents (Promega) by the modified dideoxynucleotide chain termination method following the supplier’s instruc-

---

*The nucleotide sequence data reported in this paper will appear in the DDBJ, EMBL and GenBank Sequence Databases under accession number D11476.*

---

0001-0993 © 1992 SGM
Exonuclease III deletion clones were produced according to the method of Henikoff (1987). *Escherichia coli* host strains HB101 and DH5α (Sambrook et al., 1989) were used for plasmid production. The DNA was sequenced in both directions.

The nucleotide and predicted amino acid sequences were analysed with the IntelliGenetics suite of sequence analysis programs (Brutlag et al., 1990).

**Enzymes and isotopes.** Restriction enzymes and DNA-modifying enzymes were purchased from Bethesda Research Laboratories, New England Biolabs and United States Biochemical and were used according to the manufacturers' instructions. All isotopes were purchased from New England Nuclear.

**Southern blot analysis and DNA labelling.** DNA was labelled using the method of Feinberg & Vogelstein (1984). Restriction digests of DNA from an LdMNPV cosmid library and viral genomic DNA were electrophoresed in 0.8% agarose gels and blotted onto Gene Screen Plus (DuPont) according to the manufacturer's instructions. Non-stringent hybridization conditions (30% formamide) were utilized as described in Sambrook et al. (1989).

**Nuclear extract preparation and in vitro transcription reactions.** Preparation of nuclear extracts and analysis of in vitro transcription products were done as described by Glocker et al. (1992) and in the legend to Fig. 4.

**Results and Discussion**

**Location, nucleotide sequence and analysis of the LdMNPV DNAP gene region**

To locate the LdMNPV DNAP gene, a 1.4 kb *PstI* fragment encoding the N-terminal half of the OpMNPV DNA ORF product was used. This fragment had been found to contain the OpMNPV DNAP gene by hybridization of the AcMNPV *PstI* H fragment (which contains the AcMNPV DNAP gene) to a blot of the OpMNPV genome, and by DNA sequence analysis. DNA was radioactively labelled and hybridized to blots of the LdMNPV cosmids using non-stringent hybridization conditions. Hybridization was observed to the *BglII* H fragment, and to *SsrI* fragments of 1.8 and 3.8 kb that are located at about 87.5 kb [54 map units (m.u.)] on the genome (Fig. 1a, b). Restriction enzyme fragments and exonuclease III deletion clones encompassing a 4.8 kb region were subcloned (Fig. 1b) and sequenced (Fig. 2). The sequence was analysed for ORFs in both directions.

![Fig. 1. Location and analysis of the LdMNPV DNAP gene region. (a) A BglII map of the LdMNPV genome that shows the location of the polyhedrin (from Smith et al., 1988), DNAP, p39-capsid and polyhedron envelope (PE) gene regions (Bjornson & Rohrmann, 1992a, b). The location of selected cosmid clones is shown above the LdMNPV genome map. (b) A detailed map of the region containing the LdMNPV DNAP gene and p39-capsid gene. The p39-capsid region is described elsewhere (Bjornson & Rohrmann, 1992b). The crosshatched bar indicates the region sequenced. The arrows indicate the location and orientation of major ORFs. Abbreviations: C, CiaI; E, EcoRI; H, HindIII; N, NdeI; P, PstI; S, SsrI; Sm, SmaI; V, EcoRV; X, XhoI; Xb, XbaI.](image-url)
and three ORFs were identified (Fig. 1 b). Based on their location relative to promoter elements and the sequence context of their ATG translational initiation codons (Kozak, 1986), these ORFs appeared to encode proteins. An additional incomplete ORF (ORF 1) coding for over 150 amino acids (>17K) may begin 5' to the region sequenced. The largest predicted product [3339 nucleotides (nt), 1113 amino acids, 115.9K] showed substantial similarity to other DNAPs (see below). Immediately upstream and in the opposite orientation of the DNAP gene is the second of the ORFs (ORF 2) (573 nt, 191 amino acids, 21.3K) with an ATG in good context (ATT ATG G) and preceded by a late promoter element (ATAAG) (Rohrmann, 1986) (Fig. 1, Fig. 2, nt 1132) (see below). Downstream and in the opposite orientation to the DNAP gene is another ORF (ORF 4) (219 nt, 73 amino acids, 8.5K) with an ATG in good context (AGC ATG G) and preceded by a late promoter element (Fig. 1, Fig. 2, nt 4740).

In AcMNPV, an ORF is located downstream and in the same orientation as the DNAP gene. This ORF has homology with the occlusion body protein (called spheroidin) of an entomopoxvirus pathogenic for *Chorisoneura biennis* (Vialard et al., 1990). This 'spheroidin-like protein' (SLP) gene is also located downstream of the DNAP gene of OpMNPV (C. Gross, personal communication). No homology to this SLP gene was observed downstream of the LdMNPV DNAP gene.

**LdMNPV ORF 2 is conserved in AcMNPV**

The LdMNPV ORF 2 gene (Fig. 1, Fig. 2, nt 1102 to 530) encompasses 573 nt and is preceded by a late promoter element (ATAAG) (Fig. 2, nt 1132). An ORF located in a similar position relative to the DNAP gene and also preceded by a late promoter element (GTAAG) is present in AcMNPV (Tomalski et al., 1988) and probably extends beyond the region sequenced. The putative ORF 2 products show 23% amino acid sequence identity with an additional 9% of conservative substitutions. There is also one region of 12 amino acids with 10 identities and one conservative substitution (data not shown). A search of the PIR protein database with the complete ORF 2 amino acid sequence and the conserved sequence of 12 amino acids (and also the ORF 4 sequence) failed to locate other proteins with a similar sequence. In both viruses the late promoter element is located within 25 nt of the 5' end of the DNAP ORF. In contrast to most genes found in common between AcMNPV and LdMNPV (Bjornson & Rohrmann, 1992a, b), which show only limited conservation of genome location, the position of ORF 2 has been conserved. It may be conserved in this position because it plays a role in the shut-off of DNAP gene transcription. One way this could occur is if the ORF 2 mRNA were to inhibit the translation of DNAP mRNA by hybridizing to its 5' flanking sequence, or by forming dsRNA which may accelerate the degradation of the DNAP mRNA. The presence of the late promoter transcription complex in close proximity to the early promoter could also lead to inhibition of RNA polymerase II transcription. Indeed, an example of transcriptional interference has been reported in AcMNPV in which the onset of transcription of the polyhedrin gene leads to a decline in the levels of a specific mRNA initiated downstream and in the opposite orientation (Ooi & Miller, 1990).

**LdMNPV DNAP promoter is transcribed by RNA polymerase II**

Recently, Glocker et al. (1992) reported that all baculovirus early genes they examined were transcribed by nuclear extracts from uninfected *Spodoptera frugiperda* (Sf9) cells. In addition, most of these genes were trans-activated by AcMNPV-infected Sf9 cell nuclear extracts. Furthermore, these extracts were capable of transcribing RNA polymerase II-dependent promoters from heterologous systems (e.g. the adenovirus major late promoter). Conversely it has also been shown that nuclear extracts from evolutionarily distant hosts (e.g. human cells) were capable of accurately transcribing baculovirus early genes in *vitro* (Hoopes & Rohrmann, 1991). From these studies it was concluded that baculovirus early genes are transcribed by RNA polymerase II. Although baculovirus early genes usually have a TATA promoter element and CAGT mRNA start site consensus sequence (Blissard & Rohrmann, 1989), the AcMNPV DNAP gene promoter region lacks both of these consensus sequences (Fig. 3). Despite its unconventional regulatory region, the AcMNPV DNAP gene is transcribed and trans-activated *in vitro* as an early gene (Glocker et al., 1992). The LdMNPV DNAP gene has both TATA and CAGT consensus sequence motifs in its 5' flanking sequence (Fig. 2, Fig. 3). To determine whether these conserved sequences act as RNA polymerase II promoter elements, a DNA template containing the LdMNPV DNAP gene promoter region (Fig. 4b) was analysed by *in vitro* transcription using Sf9 cell nuclear extracts and compared to the *in vitro* transcription from an AcMNPV DNAP template. Run-off transcripts of the size expected for initiation from the CAGT mRNA start site consensus sequence were observed when the LdMNPV template was truncated by *PvuII* or *BglI* (Fig. 4a, lanes 2 and 4). The AcMNPV DNAP template was transcribed at a relatively lower level suggesting that the AcMNPV DNAP promoter sequence is suboptimal (Fig. 4a, lane 1). In addition it was found that the LdMNPV DNAP template was
Fig. 2. Nucleotide sequence of the LdMNPV DNA polymerase gene region. Late promoter (LP) and early promoter (EP) sequences and their direction are underlined and overlined respectively. The DNAP gene mRNA initiation site, major restriction sites, poly(A) signals and translation start codons of the major ORFs are indicated.

Fig. 3. Alignment of the 5' flanking sequences of the DNA polymerase genes from LdMNPV and AcMNPV. The AcMNPV sequence is from Tomalski et al. (1988). The mRNA initiation sites are indicated by triangles, the TATA boxes and CAGT consensus sequences are underlined where present and the ATG translation initiation codon is shown at the end of the sequences. Vertical lines indicate identity; gaps are indicated by dashes.

Fig. 4. Run-off transcript was sized by electrophoresis alongside a DNA sequencing ladder. The RNA transcribed from the BglI-cut template was found to be 196 to 197 nt in length, corresponding to RNA initiating at the C and A of the CAGT mRNA start site consensus sequence (data not shown). Therefore these data indicate that the consensus sequences located upstream of the LdMNPV DNAP gene are RNA polymerase II promoter elements.

Baculovirus DNA polymerase gene sequence homology

The amino acid sequences of the predicted DNAPs of LdMNPV and AcMNPV are 45% identical (Fig. 5). The LdMNPV sequence also shows 13 to 27% sequence identity with a number of selected DNAPs from other organisms: human DNA polymerase α, 19%; yeast polymerase 1, 13%; herpes simplex virus, 27%; cytomegalovirus, 25%; Epstein–Barr virus, 19%; vaccinia virus, 24%; adenovirus, 16%. In addition to these overall similarities, there are seven regions (I to VII) that are conserved both in sequence and in their relative linear positions of the LdMNPV and AcMNPV sequences.
Fig. 4. In vitro transcription analysis of baculovirus DNA polymerase promoter region. (a) In vitro transcription of the DNA polymerase templates. Lanes 1 and 2, AcMNPV and LdMNPV DNA templates digested with PvuII, respectively; lanes 3 and 4, LdMNPV DNA template digested with BglI. Transcription was by non-infected (lanes 3 and 4) and 4 h p.i. (lanes 1 and 2) AcMNPV-infected S9 extracts. [In]ected cell nuclear extracts were employed to transcribe the AcMNPV DNA template cut with BglI. The following conditions were used for the transcription reactions: DNA concentration, 37.5 ng/ml; extract and incubation, 15 min each at 3° C. (b) Schematic diagram of the DNA polymerase templates. The predicted run-off transcript product is indicated by the arrow above templates cut with either BglI or PvuII, are shown. The size of the predicted run-off transcript product is indicated by the arrow above each template.

spatial arrangement between both these and other DNAs. These domains are believed to be involved with substrate binding, primase interaction, and pyrophosphate hydrolysis (Wang et al., 1989; T. S.-F. Wang, personal communication). Finally, three regions associated with 3'-5' exonuclease activity in other systems (Morrison et al., 1991) are also indicated. The AcMNPV sequence is from Tomalski et al., 1992).

Locations of LdMNPV genes: implications for baculovirus diversity

The large size (163 kb) and the high G + C content (approx. 60%) of the LdMNPV genome relative to more well-characterized baculoviruses (e.g. AcMNPV; 128 kb, 45% G+C) suggest that the LdMNPV genome has
diverged considerably from AcMNPV. Including this investigation, six LdMNPV genes shared with AcMNPV have now been located and sequenced (Bjornson & Rohrmann, 1992a, b; Smith et al., 1988). Two pairs of these genes appear to have a closely linked spatial arrangement. The gene (in this report called ORF 2) that is located upstream and in the opposite direction of the DNAP gene is present in both AcMNPV and LdMNPV. In addition, a truncated form of a gene called ORF 1 that is located upstream of the polyhedron envelope protein gene in AcMNPV and OpMNPV is also present in this position in LdMNPV (Bjornson & Rohrmann, 1992b). In contrast to the linkage of these two pairs of genes, other genes that are spatially conserved in AcMNPV and OpMNPV including the ORFs between and downstream of ORFs 1 and 3 of the polyhedron envelope protein gene region (Gombart et al., 1989), were not found in similar positions in LdMNPV. These were: a gene, CG30, located downstream of the p39-capsid gene (Thiem & Miller, 1989) and the SLP gene found downstream of the DNAP gene (Wu & Miller, 1989; C. Gross, personal communication). Thus, of the six LdMNPV genes related to other baculovirus genes, the linkage of two pairs of these genes is highly conserved, but the overall linear spatial arrangement of the genomic regions containing the polyhedrin, DNAP, p39-capsid and ORF 1 polyhedron envelope protein gene is not conserved.

We thank D. Mosbaugh for critically reading this manuscript and S. Hiremath and J. Slavicek for the LdMNPV cosmids. This project was supported by a grant from the USDA (91-37302-6310). This is Technical Report No. 9858 from the Oregon State University Agricultural Experiment Station.

References


(Received 20 March 1992; Accepted 12 August 1992)