Identification, Sequence, and Transcriptional Mapping of *lef-3*, a Baculovirus Gene Involved in Late and Very Late Gene Expression

YONGHONG LI,¹ A. LORENA PASSARELLI,¹ AND LOIS K. MILLER^{1,2*}

Departments of Genetics¹ and Entomology,^{2*} University of Georgia, Athens, Georgia 30602

Received 2 April 1993/Accepted 4 June 1993

A *trans*-acting gene required for late viral gene expression in transient expression assays was identified in the genome of *Autographa californica* nuclear polyhedrosis virus. A genomic library of *A. californica* nuclear polyhedrosis virus DNA lacking a clone spanning the region from 43 to 48 map units (mu) was unable to activate gene expression from a reporter plasmid when the reporter gene was driven by a baculovirus late or very late promoter in transient expression assays. The genomic region responsible for activating reporter gene expression was further mapped to 43.4 to 45.2 mu of the viral genome. The nucleotide sequence of this region was determined and shown to contain several small open reading frames (ORFs) and one major ORF, named *lef-3*, for late expression factor 3. The *lef-3* ORF was predicted to encode a polypeptide of 385 amino acids and with a molecular mass of 44,529 daltons. No homolog to the *lef-3* ORF was found in existing data bases. Further analysis showed that only the *lef-3* ORF in the region from 43.4 to 45.2 mu was necessary for late gene activation in the assays used. The temporal regulation of *lef-3* transcription was studied by Northern (RNA) blot hybridization; *lef-3* was found to be an early gene that was transcribed primarily as a 2.0-kb mRNA. Primer extension analysis and S1 nuclease protection assays revealed that *lef-3* transcription initiated about 280 bp upstream of the first ATG codon and terminated near a polyadenylation signal, 130 bp downstream of the last codon of the *lef-3* ORF.

Although baculoviruses are now widely used as vectors for the expression of a diversity of heterologous genes, the fundamental nature of the RNA polymerase(s) and other factors involved in transcription from late and very late viral promoters remains largely unknown. On the basis of the differential sensitivity of early and late baculovirus transcription to α -amanitin in nuclear run-on assays (10, 16), it is generally thought that an RNA polymerase switching mechanism is involved in the transition between these two phases of transcription. This view is further supported by biochemical evidence of a virus-induced RNA polymerase activity appearing by the onset of the late phase of infection (8, 43) and by molecular evidence that the primary determinants of late and very late promoters are located at the transcriptional initiation site, invariably a TAAG sequence, and within the sequences specifying the untranslated leader regions (24, 25, 32, 41, 42). The nature of the new polymerase and the nature of the factors that recognize or distinguish these unusual late and very late promoters remain to be determined.

During infection, the transition from the early to the late transcriptional phases is dependent on protein synthesis and DNA replication. In the presence of aphidicolin, an inhibitor of both host and viral DNA polymerase activities (23), early but not late or very late gene transcription is observed (34, 36). A mutant defective in the viral proliferating cell nuclear antigen (*pcna*) homolog is delayed in late gene expression (6), and a mutant defective in the DNA helicase motif of p143 is blocked in late gene expression (21). The basis for this dependence of late gene expression on DNA replication is not known, nor is it known whether this dependence would also be observed in transient expression assays with naked viral and reporter DNAs. Transient expression of a reporter

gene under the control of late and very late promoters requires the expression of many viral genes (27, 28). For the polyhedrin promoter, expression from reporter plasmids is independent of homologous repeat (*hr*) sequences (33), which have been assigned the functions of enhancers of early gene expression (11, 13) and origins of DNA replication (17, 29). Expression from polyhedrin promoter-based reporter plasmids is also known to respond authentically to promoter mutations when the plasmids are cotransfected with intact *Autographa californica* nuclear polyhedrosis virus (AcMNPV) DNA in transient expression assays (25, 33).

We recently developed a method for identifying AcMNPV genes that are required in transient assays for the expression of a reporter gene under the control of the late *vp39* (major capsid protein gene) or very late polyhedrin gene promoters (27). The method uses an overlapping set (library) of 12 clones that covers the entire AcMNPV genome and that can substitute for intact AcMNPV DNA in *trans*-activating the expression of reporter genes under late and very late promoter control. Removal of some of the clones from the library results in the reduction or total elimination of expression from the reporter plasmids. By removing one clone at a time and substituting smaller plasmid clones for the larger clone, we found it possible to identify precisely the genes required for *trans*-activating reporter gene expression.

Thus far, we have identified four genes, *ie-1*, *ie-n*, *lef-1*, and *lef-2*, that are necessary and sufficient to substitute for 2 of the 12 library clones, covering the region from 83.7 to 8.7 map units (mu) of the AcMNPV genome, in supporting transient expression from late and very late reporter plasmids (27, 28). *ie-1* and *ie-n* were previously identified as genes involved in the *trans*-activation of some early viral promoters in transient expression assays (3, 12, 14), so the role of these two genes in the transient expression assay for late gene expression may simply be to activate early gene

* Corresponding author.

expression. However, *lef-1* and *lef-2* have no apparent effect on expression from the early *pcna* (also known as *etl*) promoter, which serves as a control in our assays, and are therefore more likely to be involved in late gene expression as components of viral transcriptional or replication complexes. In this study, we used the transient expression assay for late expression factor identification to define a third gene located between 43.4 and 45.2 mu of the AcMNPV genome. The sequence and transcriptional regulation of this gene (*lef-3*) were determined.

MATERIALS AND METHODS

Cells and virus. *Spodoptera frugiperda* IPLB-SF21 (SF-21) cells (39) were used as host cells for propagating AcMNPV; the L-1 variant of AcMNPV (20) was used as the wild-type strain. The cells were maintained as described before (26) in TC100 medium (GIBCO BRL, Grand Island, N.Y.) supplemented with 10% fetal bovine serum and 0.26% tryptose broth.

Transient expression assays. SF-21 cells (2×10^6 cells per 60-mm plate) were cotransfected with a reporter plasmid and either AcMNPV DNA or cloned segments representing all or most of the viral genome. The reporter plasmids were pETCAThr5 (27), pCAPCAT (37), and phcwt (33), in which the chloramphenicol acetyltransferase (CAT) gene was under the control of early (*etl*), late (capsid protein gene *vp39*), and very late (polyhedrin gene) viral promoters, respectively. For each transfection, 2 μ g of the reporter plasmid was combined with 0.5 μ g of viral DNA or 0.5 μ g of each of the clones from the AcMNPV library. When one of the cloned DNAs was omitted, 0.5 μ g of plasmid or an equivalent amount of a gel-purified DNA fragment was substituted in the transfection mixture. The total amount of DNA used in each cotransfection was 8.0 μ g; salmon sperm DNA was used to balance DNA concentrations whenever necessary. The transfection was performed by the calcium phosphate method (26). The cells were collected 24 h after transfection for the early promoter and 48 h after transfection for both the late and the very late promoters and were disrupted by three repeated freeze (liquid nitrogen)-thaw (37°C) cycles. Cell debris was pelleted, and 1/10 of the supernatant was used to assay for CAT activity as described previously (9, 33).

DNA cloning and sequencing. The DNA fragment corresponding to the AcMNPV genome from the *EcoRI* site at 43.4 mu to the *SstI* site at 45.4 mu (see Fig. 2A) was cloned into the pBluescript SK vector (Stratagene, La Jolla, Calif.). The *EcoRI*-*MluI* fragment of this clone was then subcloned into the *EcoRI* and *SmaI* sites of the pBluescript KS vector (Stratagene) by filling in the *MluI* site of the fragment with the large fragment of DNA polymerase I (Klenow) and ligating the fragment to the vector. The resulting clone, pSDEM2, was used to generate two sets of nested deletions with exonuclease III and mung bean nuclease (15) by use of the *KpnI* or *SstI* sites as deletion block points, in accordance with protocols recommended by Stratagene. pSDEM2 and the sets of deletion clones were sequenced from the T3 and T7 primer sites by the dideoxynucleotide chain termination method (35) with a sequencing kit (US Biochemical, Cleveland, Ohio) and double-stranded DNA plasmid templates (38). Sequences were compiled and analyzed by use of the programs of Pustell and Kafatos (31) and the merger program of IntelliGenetics. The deduced open reading frames (ORFs) within the sequenced region were compared with the trans-

lated sequences from GenBank (release 73) and U-EMBL (release 31) by use of the FASTA program (30) and the BLAST (1) network service at the National Center for Biotechnology Information.

RNA isolation. Total RNA was isolated from mock- or AcMNPV-infected SF-21 cells at various times postinfection (p.i.) by the guanidinium isothiocyanate method (7). A virus inoculum (10 PFU per cell) was allowed to adsorb to SF-21 monolayers (5×10^6 cells per 100-mm plate) at room temperature. The inoculum was removed and replaced with fresh medium after a 1-h adsorption period. The infected cells were then incubated at 27°C; the beginning of the incubation was considered zero time p.i. For inhibition of DNA synthesis, aphidicolin (5 μ g/ml) was included in the medium added to 0 h. For blocking of protein synthesis, cycloheximide (100 μ g/ml) was added to the medium 30 min before inoculation.

Northern (RNA) blot analysis. Total RNA (10 μ g) from mock- or AcMNPV-infected cells was denatured by glyoxalation (22), electrophoretically separated in 1.0% agarose, and transferred to Zeta-Probe blotting membranes (Bio-Rad Laboratories, Richmond, Calif.). The 1-kb DNA ladder (GIBCO BRL) used as the size marker was also glyoxalated. Blots were hybridized at 50°C to a strand-specific RNA probe (hybridization buffer: 50% formamide, 1.5 $\times$ SSPE [1 $\times$ SSPE is 180 mM NaCl, 10 mM Na₂HPO₄, 1 mM EDTA], 0.1 mg of tRNA per ml, and 0.5% salmon sperm (DNA)). For construction of a *lef-3*-specific probe, deletion plasmid pKHD31, generated for sequencing of the *lef-3* gene, was linearized with *ClaI* and transcribed in the presence of [³²P]UTP with T3 RNA polymerase.

Primer extension analysis. A synthetic 23-mer oligonucleotide (5'-GCTCATCAATTTACCACTAATCC-3') near the N terminus of *lef-3* was radiolabeled at the 5' end with T4 polynucleotide kinase, annealed to 10 μ g of total RNA, and extended with Moloney murine leukemia virus reverse transcriptase under the same conditions as those described earlier (6). The extension products were resolved on an 8% polyacrylamide-urea gel along with a DNA sequencing ladder generated from plasmid pSDEM2 by use of the radiolabeled oligonucleotide primer described above.

S1 nuclease protection analyses. The termini of *lef-3* transcripts were determined by S1 nuclease protection analyses (40) with 10 μ g of total cell RNA per reaction. The DNA fragment used as a probe to map the 5' end of *lef-3* was prepared by digesting plasmid pKHD31 (one of the nested deletion plasmids generated for sequencing of the *lef-3* gene) with *ClaI*, dephosphorylating the ends with calf intestinal phosphatase, radiolabeling the 5' ends with T4 polynucleotide kinase, and recleaving the plasmid with *SstI*, a site within the vector. The resulting 984-bp fragment was isolated and gel purified. The DNA fragment used to map the 3' end of *lef-3* was made by digesting plasmid pSDEM2 with *ClaI*, radiolabeling the 3' ends with the Klenow DNA polymerase fragment, recleaving the plasmid with *EcoRI*, and gel purifying the resulting 1,408-bp fragment. DNA-RNA hybridization was carried out with 80% formamide-40 mM piperazine-*N,N'*-bis(2-ethanesulfonic acid) (PIPES)-0.4 M NaCl-1 mM EDTA at 50°C overnight for 5' mapping and 60% formamide-40 mM PIPES-0.4 M NaCl-1 mM EDTA at 52°C overnight for 3' mapping.

Nucleotide sequence accession number. The sequence reported in this paper was deposited in GenBank under the accession number L18873.

FIG. 1. Localization to PstH5 of the gene (*lef-3*) required for late gene expression of AcMNPV. (A) Restriction map of the AcMNPV genome from 33.0 to 57.0 mu and key restriction sites in the region: H, *HindIII*; P, *PstI*; S, *SmaI*; Bg, *BglII*. mu are given under the lines representing the positions of known genes. The three overlapping genomic clones of the library in this region of the genome, PstH1, PstH5, and HC9, are indicated by bars, with the open area representing the region known to be included in the clone and the shaded areas representing additional regions that may be included. (B) Autoradiogram of CAT transient expression assays. SF-21 cells were cotransfected with a reporter plasmid (pETCATHr5, pCAPCAT, or phcwt) and additional viral DNA as indicated above each lane: wt, wild-type AcMNPV DNA; L, entire library of cloned overlapping AcMNPV DNA fragments; L*, AcMNPV genomic library lacking PstH5; 0, reporter DNA only. Cm, chloramphenicol; AcCm, various forms of acetylated chloramphenicol.

RESULTS

Identification of late expression factor gene 3 *lef-3*. The reporter plasmids and clone library used in the transient assays to define AcMNPV genes involved in late gene expression were described previously (27). In brief, three different reporter plasmids were used. All three contained a portion of the *hr5* sequence and the CAT gene under the control of a viral promoter. The early promoter of the viral *pcna* (also known as *etl*) homolog (6) was the promoter driving CAT expression in reporter plasmid pETCATHr5 (27) and served essentially as a control in these experiments; it is dependent on *ie-1* expression and is stimulated by *ie-n* expression (28). The *vp39* promoter, including 462 bp upstream of the translational start codon (36, 37), served as the late promoter driving CAT expression in reporter plasmid pCAPCAT, and the polyhedrin promoter served as the very late promoter driving CAT expression in reporter plasmid phcwt (33).

CAT activity was obtained for all three reporter plasmids

FIG. 2. Fine mapping of *lef-3* within PstH5. (A) Diagram of the *SmaI*-D fragment (40.3 to 49.4 mu) cloned as plasmid pXma17 (a) and the derived subfragments (b to i). The relative CAT activity for each transfection with the plasmid or a subfragment is given as a plus or a minus, meaning that CAT activity was comparable to that of the positive or the negative control, respectively. Restriction sites were *SmaI* (S), *PstI* (P), *EcoRI* (E), *BglII* (Bg), *SstI* (SI), *SstII* (SII), *BamHI* (B), *SphI* (Sp), *MluI* (M), and *HpaI* (H). (B) CAT assays for cotransfections with pCAPCAT and the various DNA fragments, indicated as letters above the lanes. The lanes labeled L - PstH5 (lanes 2 to 11) indicate that PstH5 was omitted from these transfections. The lowercase letters above each lane (a to i) refer to the plasmid or DNA fragment shown in panel A and used to replace PstH5. Lane 2 (0) contains no additional DNA supplement. Lane 1 (L) contains the entire library of overlapping AcMNPV DNA clones. AcCm and Cm are as defined in the legend to Fig. 1.

cotransfected with either wild-type viral DNA or the complete set of overlapping AcMNPV clones (Fig. 1B, lanes 1, 2, 5, 6, 9, and 10). Thus, as previously reported (27), the ability of wild-type AcMNPV DNA to *trans*-activate the reporter plasmids could be basically reconstituted by substituting the entire set of viral DNA fragments for intact viral DNA. When the PstH5 clone, which extends from approximately 37 to 50 mu (Fig. 1A), was omitted from the cotransfections, CAT expression from the early promoter was not substantially affected (Fig. 1B, compare lanes 2 and 4). However, the omission of the PstH5 clone from the cotransfections with the reporter plasmids carrying the CAT gene under late and very late promoter control substantially reduced CAT expression (Fig. 1B, compare lanes 6 and 8 and lanes 10 and 12). Thus, at least one *trans*-activating factor necessary for late and very late gene expression in these assays was present exclusively in the PstH5 clone. The PstH5 clone substantially overlaps two other clones of the genomic library, PstH1 and HC9; the region unique to PstH5 extends from approximately 43 to 48 mu (Fig. 1A).

To delineate further the region containing the PstH5-

FIG. 3. Distribution of translational termination codons and sequence of the region from 43.4 to 45.2 mu of AcMNPV, containing *lef-3*. (A) Restriction map of the region sequenced. Stop codons within all six possible reading frames are indicated below the map by vertical lines. The *lef-3* ORF extends in the counterclockwise direction. (B) Nucleotide sequence and predicted amino acid sequence for the AcMNPV *lef-3* ORF. The nucleotide sequence from *MluI* (45.2 mu) to *EcoRI* (43.4 mu) is presented with numbering to the left of the first nucleotide in each line. The predicted amino acid sequence for *lef-3* is shown below the nucleotide sequence with the standard single-letter symbol for each amino acid. CAGT motifs near the N terminus of *lef-3* are noted with a colon (:) above the sequence. The actual transcriptional initiation site determined empirically is indicated with asterisks above the sequence. The underlined sequence represents the region complementary to the primer sequence used for 5'-end mapping. The potential polyadenylation signal (AATAAA) is doubly underlined. The *ClaI* site used in Northern blot analysis and S1 nuclease assays is underlined with a broken line.

FIG. 4. Assignment of *trans*-activating function to the ORF. (A) Schematic representation of ORFs and AcMNPV portions of plasmids used to replace PstH5 in the CAT assay. ORFs of more than 80 amino acid (aa) residues are represented by open arrows. The relative CAT activity for each cotransfection with pCAPCAT or phcwt and various DNAs is given as a plus or a minus. (B) CAT assay with reporter plasmid pCAPCAT. All cotransfections were done with pCAPCAT as a reporter plasmid. Additional DNAs are indicated above each lane. Lane 1 included the entire clone library (L), while lanes 2 to 5 included clones lacking PstH5. No additional DNA was added in lane 2, while lanes 3 to 5 contained deletion plasmids a through c, diagrammatically shown in panel A. (C) Same as panel B, except that CAT activity was assayed following cotransfections with reporter plasmid phcwt. AcCM and Cm are as defined in the legend to Fig. 1.

specific *trans*-activating function, we first determined that the PstH5 clone could be replaced in these assays by a plasmid, pXma17, which contains the *Sma*I-D fragment of AcMNPV (40.3 to 49.4 mu) and encompasses the entire region unique to PstH5 (Fig. 1A and Fig. 2A). Plasmid pXma17 was equivalent to PstH5 in terms of its ability to *trans*-activate the *vp39* promoter in the presence of the remainder of the library clones (Fig. 2B, compare lanes 1 and 3). Two subfragments of pXma17 extending from the unique *Bgl*III site at 44.6 mu to the *Sma*I site at 40.3 or 49.4 mu (Fig. 2A, rows b and c) were unable to functionally substitute for PstH5 (Fig. 2B, lanes 4 and 5). However, subfragments extending from the *Sst*I site at 40.3 mu to the *Bam*HI (Fig. 2A, row d) or *Sst*II (Fig. 2A, row e) site were able to *trans*-activate late gene expression in these assays (Fig. 2B, lanes 6 and 7). A *Pst*I-*Sst*I subfragment (42.0 to 45.4 mu) (Fig. 2A, row f) as well as an *Eco*RI-*Sst*I subfragment (43.4 to 45.4 mu) (Fig. 2A, row g) also tested positive in transient expression assays (Fig. 2B, lanes 8 and 9). The smallest of these fragments was mapped with various restriction enzymes (Fig. 2A). Further subfragments of this region were tested for their ability to substitute for PstH5 in these assays (Fig. 2B, lanes 10 and 11), and the *trans*-activating region

was localized to the *Eco*RI-*Mlu*I (43.4 to 45.2 mu) segment of AcMNPV (Fig. 2A, row h). A small fragment extending from the *Sph*I site to the *Hpa*I site, however, was unable to restore expression when substituted for the PstH5 clone (Fig. 2B, lane 11).

Nucleotide sequence of *lef-3* and the flanking regions. The *Eco*RI-*Mlu*I fragment that was found to restore late gene expression to the AcMNPV library in the absence of PstH5 was cloned and sequenced. The nucleotide sequence is presented in Fig. 3B. Computer analysis revealed a major ORF in the counterclockwise direction and predicted a polypeptide product of 385 amino acids and with a molecular mass of 44,529 daltons. The predicted initiation codon is in a favorable context (AATATGG) for efficient translation in accordance with Kozak's rules (PuNNATGpu) (18) and other abundantly expressed AcMNPV genes (26). This ORF was designated *lef-3*, for late expression factor 3. The coding region of *lef-3* has a high A+T content (63.8% A+T; 36.2% G+C). No apparent significant homology to *lef-3* was found when data bases (GenBank release 73; U-EMBL release 31) were searched by use of the Genetics Computer Group FASTA program (30) and the BLAST (1) network service at the National Center for Biotechnology Information, nor were any significant known sequence motifs found within the *lef-3* gene product, with the exception of five potential N-glycosylation sites (IntelliGenetics Keybank program, release 9), although no apparent signal sequence was found.

Identification of the ORF within the *Eco*RI-*Mlu*I fragment that is necessary for late and very late gene expression. In addition to *lef-3*, sequence data revealed two smaller ORFs in the minimal *trans*-activating fragment (*Eco*RI-*Mlu*I) that overlapped *lef-3* and could potentially encode polypeptides of 84 and 192 amino acids, as depicted in Fig. 4A. To determine whether one of the smaller ORFs was responsible for *trans*-activating the late promoter, we tested the ability of several deletion plasmids to substitute for PstH5 in transient expression assays. Only the plasmid containing the entire *lef-3* ORF was functionally similar to PstH5 in providing a *trans*-activating factor for the *vp39* promoter (Fig. 4B). This plasmid (Fig. 4A, deletion clone a) was also able to substitute for PstH5 in *trans*-activating the very late polyhedrin promoter (Fig. 4C). The two smaller deletion plasmids (Fig. 4A, clones b and c) were unable to functionally substitute for PstH5 when used together in the cotransfections, indicating that the two smaller ORFs and/or partial ORFs do not act in concert (data not shown). Thus, *lef-3* is necessary for *trans*-activation, although we cannot preclude the possibility that the 84-amino-acid-encoding ORF (Fig. 4A) is necessary for late gene expression in addition to *lef-3*. The data are consistent with prior experiments showing that the disruption of either the *Bgl*III or the *Sph*I site abolished *trans*-activating function (Fig. 2, lanes 4, 5, and 11); these sites are within the translated region and are upstream of the poly(A) signal, respectively, of *lef-3* (Fig. 4A) (see also the 3' mapping data below).

Northern (RNA) blot analysis. To examine the temporal regulation of *lef-3* transcription, we performed a Northern blot analysis with a *lef-3*-specific RNA probe. The uniformly labeled, strand-specific RNA probe extended from nucleotide (nt) 940 to about nt 1840 in an orientation opposite that of *lef-3* mRNA (Fig. 3B and Fig. 5A). As shown in Fig. 5B, a primary virus-specific transcript of approximately 2.0 kb was observed in infected cells at 3 h p.i. The level of this 2.0-kb RNA was highest at 6 h p.i., declined slightly at 12 h p.i., and declined dramatically but remained detectable at 24 and 48 h p.i. When cells were treated with the DNA

FIG. 5. Transcriptional mapping of *lef-3*. (A) Location and nature of probes used in the Northern blot analysis and the 5' and 3' S1 nuclease protection assays, shown in relationship to *lef-3* (represented by an open arrow), and key restriction enzyme sites within the region sequenced. The orientation of the RNA probe used in the Northern blot analysis is indicated by the arrowhead at the 3' end. The positions of radiolabels in 5' and 3' probes are indicated by asterisks. The potential poly(A) site (AATAAA) is indicated by a vertical arrow. The transcriptional mapping data for *lef-3* are summarized by the filled arrow at the top, representing the size, location, and orientation of the *lef-3* transcript. sk vector represents pBluescript sequences. (B) Northern blot analysis of *lef-3* transcripts. Total RNA was isolated from either mock-infected cells (lane M) or AcMNPV-infected cells at 3, 6, 12, 24, and 48 h p.i. (lanes 3, 6, 12, 24, and 48, respectively) and 12 h p.i. in the presence of aphidicolin (lane A) or cycloheximide (lane C). The RNA was denatured, electrophoretically separated, and transferred to a nylon blotting membrane. The membrane was hybridized with the *lef-3* strand-specific RNA probe indicated in panel A. The sizes of DNA markers, in kilobases, are indicated on the left. (C) 5' Mapping of *lef-3* transcripts by S1 nuclease protection analyses. A 984-bp fragment, 5' 32 P labeled at the *Cla*I-generated end as shown in panel A, was hybridized to RNA samples as described for panel B and subjected to digestion with S1 nuclease. The S1 nuclease-resistant products were resolved on a 4.0% polyacrylamide gel and visualized by autoradiography. Molecular weight markers (lane M), in nucleotides, were derived from *Msp*I-digested pUC19. Lane P, undigested probe. The major protected fragment is indicated by an arrowhead. (D) 3' Mapping of *lef-3* transcripts by S1 nuclease protection analyses. The probe used for hybridization to total RNA was a 1,408-bp *Eco*RI-*Cla*I fragment radiolabeled with 32 P at the *Cla*I-generated end as shown in panel A. Molecular weight markers were derived from *Taq*I-digested pUC19. Lane P, undigested probe. The S1 nuclease-resistant fragment is indicated by an arrowhead.

synthesis inhibitor aphidicolin, which inhibits late and very late transcription because of a dependence on DNA replication, the level of the 2.0-kb RNA decreased only a small degree (Fig. 5B, compare lanes 12 and A), indicating the early nature of *lef-3* transcription. Treatment of infected cells with the protein biosynthesis inhibitor cycloheximide significantly elevated the levels of *lef-3* transcripts (Fig. 5B, compare lanes 12 and C), a result also consistent with the classification of *lef-3* as an early gene. A minor transcript,

slightly larger than 2.0 kb, was also seen at all times p.i.; this transcript had a temporal pattern similar to that for the 2.0-kb transcript.

Mapping of the 5' and 3' ends of *lef-3* transcripts. To determine the initiation site(s) for *lef-3* transcription in SF-21 cells, we performed a primer extension analysis and S1 nuclease protection assays. Examination of the sequence near the 5' end of the transcribed region of *lef-3* revealed two CAGT motifs that are frequently found near the transcrip-

FIG. 6. Primer extension analysis of the 5' end of the *lef-3* mRNA. A 5' 32 P-labeled oligonucleotide (5'-GCTCATCAATTTCA CACTAATCCT-3') was hybridized to total RNA from mock-infected cells (M) or from infected cells harvested at the hours p.i. indicated above the lanes, and the RNA was reverse transcribed. The lanes designated A and C between the 12- and 24-h lanes indicated that the RNA was isolated at 12 h p.i. in the presence of aphidicolin (A) or cycloheximide (C). The radiolabeled oligonucleotide was also used as a primer with a DNA template containing *lef-3* and was extended in the presence of dideoxynucleotides ddATP (A), ddTTP (T), ddGTP (G), and ddCTP (C). The primer extension products were resolved on an 8.0% polyacrylamide gel. The transcriptional start site is shown in a region specified by the nucleotide sequence of the DNA strand complementary to the *lef-3* mRNA.

tional start sites of early baculovirus genes (2, 4, 5, 14, 19). One CAGT is located 41 bp upstream of the translational start codon, while the other CAGT is located within the N terminus of the translated region and is followed, 57 bp downstream, by an ATG codon in a favorable context (AAA ATGA) for translational initiation (Fig. 3B). A 23-nt primer, corresponding to the sequence from nt 900 to nt 922, was

thus synthesized and radiolabeled at the 5' end for the primer extension analysis. Three major distinct extension products, none of which mapped to a CAGT motif, were observed (Fig. 6). The longest and most abundant primer extension product corresponded to a transcriptional start site about 280 bp upstream of the first ATG codon of *lef-3*. The temporal pattern for this extension product was consistent with that observed in Northern blots; the most obvious feature was an increase in the level of the product in the presence of cycloheximide (Fig. 6, lane C). The shortest primer extension product would place the transcriptional start site at nt 750, 3 bp upstream of the first CAGT motif. No extension products corresponding to the second CAGT motif were observed in these primer extension reactions when gel electrophoresis was carried out to resolve shorter products (data not shown).

S1 nuclease protection analyses were also performed to map the 5' end of the *lef-3* mRNA. A 984-nt DNA fragment radiolabeled at the *Clai* site was used as a probe (Fig. 5A), and a fragment of about 400 to 450 nts was protected when this probe was hybridized to total RNA and subjected to S1 nuclease digestion (Fig. 5C). The transcriptional start site deduced from the size of this protected fragment corresponds to that defined by the longest primer extension products. No protected fragment of a size corresponding to the shortest primer extension product in the primer extension analysis was observed. The shortest extension product may therefore represent a prematurely terminated product.

For mapping of the 3' end of the *lef-3* mRNA, a 1,408-bp *EcoRI-Clai* fragment 3' labeled at the *Clai*-generated end was used as a probe in S1 nuclease protection analyses of total RNA (Fig. 5A). As shown in Fig. 5D, an S1 nuclease-resistant fragment of about 1,100 to 1,200 nts was found in AcMNPV-infected but not mock-infected cells. The size of this fragment suggests that the AATAAA motif occurring 130 bp downstream of the last codon of *lef-3* is a polyadenylation signal.

Taken together, the 5' and 3' mapping data showed that *lef-3* was transcribed from a 1.6-kb region of the AcMNPV genome. The size was in good agreement with Northern blot data, which revealed a major 2.0-kb transcript.

DISCUSSION

We have found another gene of AcMNPV that is required in transient assays for substantial expression from a reporter gene under late and very late promoter control; removal of this gene from a library of clones covering the AcMNPV genome results in a dramatic reduction in the level of expression from late and very late gene reporter plasmids. This gene, known as the late expression factor 3 gene (*lef-3*), is located in a region of the AcMNPV genome (43.4 to 45.2 mu) that has not been previously characterized in any baculovirus system. It is the fifth gene to be identified as functional in late baculovirus gene expression in this subtractive assay system. Like *lef-1* and *lef-2*, *lef-3* affects expression from the late *vp39* and very late polyhedrin promoters but does not affect expression from the early *pcna* (*etl*) promoter, although we cannot currently eliminate the possibility that it affects expression from other early promoters. It is most likely that *lef-3* is involved in late gene expression, but whether it acts at the level of replication, transcription, or translation has not been determined. Since these three *lef* genes affect late gene expression, it is likely that the effect that we observed on very late gene expression is due to their activation of late genes.

Analyses of *lef-3* transcription showed that *lef-3* is an early gene with a major transcript of 2.0 kb and a minor transcript of slightly larger than 2.0 kb. The steady-state levels of *lef-3* transcripts are highest between 3 and 12 h p.i. and are reduced dramatically by 24 h p.i. This temporal pattern is similar to that observed for *lef-1*. As with other AcMNPV early genes, transcription occurs in the presence of cycloheximide and aphidicolin. The elevation of *lef-3* RNA levels at 12 h p.i. in the presence of cycloheximide but not aphidicolin suggests that the shutoff of *lef-3* transcription requires an early gene product, although there are other possible explanations (e.g., the presence of cycloheximide may increase the stability of *lef-3* RNA).

Both primer extension and S1 nuclease analyses of the 5' end of the *lef-3* transcript indicated that transcriptional initiation occurs approximately 280 nts upstream of the translational initiation codon for the predicted 385-residue gene product. This result was surprising in two respects. Firstly, a 280-nt leader is unusually long for an mRNA, particularly for AcMNPV mRNAs; AcMNPV ORFs that are transcribed with long leaders generally have translated ORFs in the leaders. No ATG-initiated ORFs were found within the *lef-3* leader sequence, nor were stable secondary structures observed by computer analysis. Secondly, one CAGT motif was found approximately 60 bp upstream of the translational initiation codon of the 385-codon ORF, and another, with an upstream TATA box, was found approximately 100 bp downstream of the predicted initiation codon (CAGT motifs have been found at the start sites of some but not all early baculovirus RNAs [2, 4, 5, 14, 19]). However, both S1 nuclease and primer extension data indicated that neither of these CAGT motifs was used as a transcriptional initiation site. The site determined for *lef-3* transcriptional initiation mapped approximately 25 bp upstream of a CT TAT sequence, suggesting that a late transcript may initiate on the opposite strand (i.e., from ATAAG) and overlap the 5' end of the *lef-3* RNA. Whether or how this late transcript might affect *lef-3* expression is not yet known. The overlap should not affect our 5' transcriptional mapping data because the 5' end mapped to similar locations when RNA from cells infected in the presence of cycloheximide or aphidicolin was used in primer extension and S1 nuclease analyses. The nature of the less abundant and slightly longer transcript observed in the Northern blots was not specifically defined in these studies.

It is likely that *lef-3* plays a central role in late baculovirus gene expression. Only a few of the 12 clones of the AcMNPV library have such a strong effect on expression from the late and very late promoters when removed from the transient expression assays (28). The four clones showing the strongest effects upon removal are IE15, BC5, PstH5, and HC10. We have found that *ie-1*, *ie-n*, and *lef-2* are able to substitute for IE15 (27), *lef-1* is able to substitute for BC5 (28), and *lef-3* is able to substitute for PstH5 in the subtractive assay system (this paper). We are currently determining the functional genes mapping to the HC10 region.

ACKNOWLEDGMENTS

We thank S. Hilliard for technical assistance.

This work was supported in part by Public Health Service grant AI23719 from the National Institute of Allergy and Infectious Diseases.

REFERENCES

- Altschul, S. F., W. Gish, W. Miller, E. W. Myers, and D. J. Lipman. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403-410.
- Blissard, G. W., and G. F. Rohrmann. 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.
- Carson, D. D., L. A. Guarino, and M. D. Summers. 1988. Functional mapping of an AcNPV immediately early gene which augments expression of the IE-1 trans-activated 39K gene. *Virology* 162:444-451.
- Carson, D. D., M. D. Summers, and L. A. Guarino. 1991. Molecular analysis of a baculovirus regulatory gene. *Virology* 182:279-286.
- Chisholm, G. E., and D. J. Henner. 1988. Multiple early transcripts and splicing of the *Autographa californica* nuclear polyhedrosis virus IE-1 gene. *J. Virol.* 62:3193-3200.
- Crawford, A. M., and L. K. Miller. 1988. Characterization of an early gene accelerating expression of late genes of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* 62:2773-2781.
- Davis, L. G., M. D. Dibner, and J. F. Battey. 1986. Basic methods in molecular biology. Elsevier Science Publishing, Inc., New York.
- Fuchs, L. Y., M. S. Woods, and R. F. Weaver. 1983. Viral transcription during *Autographa californica* nuclear polyhedrosis virus infection: a novel RNA polymerase induced in infected *Spodoptera frugiperda* cells. *J. Virol.* 48:641-646.
- Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2:1044-1051.
- Grula, M. A., P. L. Buller, and R. F. Weaver. 1981. α -Amanitin-resistant viral RNA synthesis in nuclei isolated from nuclear polyhedrosis virus-infected *Heliothis zea* larvae and *Spodoptera frugiperda* cells. *J. Virol.* 38:916-921.
- Guarino, L. A., and W. Dong. 1991. Transient expression of an enhancer-binding protein in insect cells transfected with the *Autographa californica* nuclear polyhedrosis virus IE-1 gene. *J. Virol.* 65:3676-3680.
- Guarino, L. A., and M. D. Summers. 1986. Functional mapping of a trans-activating gene required for expression of a baculovirus delayed-early gene. *J. Virol.* 57:563-571.
- Guarino, L. A., and M. D. Summers. 1986. Interspersed homologous DNA of *Autographa californica* nuclear polyhedrosis virus enhances delayed-early gene expression. *J. Virol.* 60:215-223.
- Guarino, L. A., and M. D. Summers. 1987. Nucleotide sequence and temporal expression of a baculovirus regulatory gene. *J. Virol.* 61:2091-2099.
- Henikoff, S. 1984. Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* 28:351-359.
- Huh, N. E., and R. F. Weaver. 1990. Identifying the RNA polymerases that synthesize specific transcripts of the *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol.* 71:195-202.
- Kool, M., P. M. M. Van den Berg, J. Tramper, R. W. Goldbach, and J. M. Valk. 1993. Location of two putative origins of DNA replication of *Autographa californica* nuclear polyhedrosis virus. *Virology* 192:94-101.
- Kozak, M. 1986. Bifunctional messenger RNAs in eukaryotes. *Cell* 47:481-483.
- Krappa, R., and D. Knebel-Mörsdorf. 1991. Identification of the very early transcribed baculovirus gene PE-38. *J. Virol.* 65:805-812.
- Lee, H. H., and L. K. Miller. 1978. Isolation of genotypic variants of *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* 27:754-767.
- Lu, A., and E. B. Carstens. 1991. Nucleotide sequence of a gene essential for viral DNA replication in the baculovirus *Autographa californica* nuclear polyhedrosis virus. *Virology* 181:336-347.
- McMaster, G. K., and G. G. Carmichael. 1977. Analysis of single- and double-stranded nucleic acids on polyacrylamide and agarose gels by using glyoxal and acridine orange. *Proc.*

- Natl. Acad. Sci. USA **74**:4835–4838.
23. Miller, L. K., J. E. Jewell, and D. Browne. 1981. Baculovirus induction of a DNA polymerase. *J. Virol.* **40**:305–308.
 24. Morris, T., and L. K. Miller. Unpublished results.
 25. Ooi, B. G., C. Rankin, and L. K. Miller. 1989. Downstream sequences augment transcription from the essential initiation site of a baculovirus polyhedrin gene. *J. Mol. Biol.* **210**:721–736.
 26. O'Reilly, D. R., L. K. Miller, and V. A. Luckow. 1992. Baculovirus expression vectors: a laboratory manual. W. H. Freeman & Co., New York.
 27. Passarelli, A. L., and L. K. Miller. 1993. Three baculovirus genes involved in late and very late gene expression: *ie-1*, *ie-n*, and *lef-2*. *J. Virol.* **67**:2149–2158.
 28. Passarelli, A. L., and L. K. Miller. Identification and characterization of *lef-1*, a baculovirus gene involved in late and very late gene expression. *J. Virol.* **67**:3481–3488.
 29. Pearson, M., R. Bjornson, G. Pearson, and G. Rohrmann. 1992. The *Autographa californica* baculovirus genome: evidence for multiple replication origins. *Science* **257**:1382–1384.
 30. Pearson, W. R., and D. J. Lipman. 1988. Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85**:2444–2448.
 31. Pustell, J., and F. C. Kafatos. 1984. A convenient and adaptable package of computer programs for DNA and protein sequence management, analysis and homology determination. *Nucleic Acids Res.* **12**:643–655.
 32. Qin, J., A. Liu, and R. F. Weaver. 1989. Studies on the control region of the p10 gene of the *Autographa californica* nuclear polyhedrosis virus. *J. Gen. Virol.* **70**:1273–1279.
 33. Rankin, C., B. G. Ooi, and L. K. Miller. 1988. Eight base pairs encompassing the transcriptional start point are the major determinant for baculovirus polyhedrin gene expression. *Gene* **70**:39–50.
 34. Rice, W. C., and L. K. Miller. 1986. Baculovirus transcription in the presence of inhibitors and in nonpermissive *Drosophila* cells. *Virus Res.* **6**:155–172.
 35. Sanger, F., S. Nicklen, and A. R. Coulson. 1977. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* **74**:5463–5467.
 36. Thiem, S. M., and L. K. Miller. 1989. Identification, sequence, and transcriptional mapping of the major capsid protein gene of the baculovirus *Autographa californica* nuclear polyhedrosis virus. *J. Virol.* **63**:2008–2018.
 37. Thiem, S. M., and L. K. Miller. 1990. Differential gene expression mediated by late, very late, and hybrid baculovirus promoters. *Gene* **91**:87–94.
 38. Toneguzzo, F., S. Glynn, E. Leci, S. Muolsness, and A. Hayday. 1988. Use of a chemically modified T7 polymerase for manual and automated sequencing of supercoiled DNA. *BioTechniques* **6**:460–469.
 39. Vaughn, J. L., R. H. Goodwin, G. J. Tompkins, and P. McCawley. 1977. The establishment of two cell lines from the insect *Spodoptera frugiperda* (Lepidoptera: Noctuidae). *In Vitro* (Rockville) **13**:213–217.
 40. Weaver, R. F., and C. Weissman. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure; the 5' termini of 15S beta-globin mRNA precursor and mature 10S beta-globin mRNA have identical map coordinates. *Nucleic Acids Res.* **7**:1175–1193.
 41. Weyer, U., and R. D. Possee. 1988. Functional analysis of the p10 gene 5' leader sequence of the *Autographa californica* nuclear polyhedrosis virus. *Nucleic Acids Res.* **16**:3635–3654.
 42. Weyer, U., and R. D. Possee. 1989. Analysis of the promoter of the *Autographa californica* nuclear polyhedrosis virus p10 gene. *J. Gen. Virol.* **70**:203–208.
 43. Yang, C. L., D. A. Stetler, and R. F. Weaver. 1991. Structural comparison of the *Autographa californica* nuclear polyhedrosis virus-induced RNA polymerase and the three nuclear RNA polymerases from the host, *Spodoptera frugiperda*. *Virus Res.* **20**:251–264.