Identification, Sequence, and Transcriptional Mapping of the Major Capsid Protein Gene of the Baculovirus *Autographa californica* Nuclear Polyhedrosis Virus

SUZANNE M. THIEM AND LOIS K. MILLER*

Departments of Genetics and Entomology, University of Georgia, Athens, Georgia 30602

Received 14 November 1988/Accepted 12 January 1989

The gene encoding the major capsid protein of the baculovirus *Autographa californica* nuclear polyhedrosis virus (AcMNPV) was identified, sequenced, and transcriptionally mapped. The location of the gene was determined by immunological screening of an expression library of AcMNPV open reading frame–β-galactosidase fusions with an antibody raised to virus structural proteins. The DNA sequence of the corresponding region, which mapped within 56.6 and 58.0 map units on the AcMNPV genome, revealed a 1,040-base-pair open reading frame capable of encoding a 39-kilodalton polypeptide. The identity of the polypeptide was determined by Western blot (immunoblot) analysis of purified empty capsids with an antibody raised to the capsid–β-galactosidase fusion protein. The identity of the peptide encoded by the gene was confirmed by immunoprecipitation of an in vitro translation product with RNA selected by hybridization to DNA sequences from the coding region of the gene. Transcripts of the capsid gene were analyzed by Northern (RNA) blots and mapped by nuclease protection and primer extension analysis. The capsid gene is transcribed maximally at 12 and 24 h postinfection but not in the presence of cycloheximide, a protein synthesis inhibitor, or aphidicolin, a viral DNA synthesis inhibitor, and is therefore classified as a late gene. The gene is transcribed in a counterclockwise direction with respect to the circular map. There are three transcriptional start sites, all containing the 5'TAG consensus sequence found at the start site of all late AcMNPV genes.

*Autographa californica* nuclear polyhedrosis virus (AcMNPV) belongs to the family *Baculoviridae* and serves as a model system for the study of the molecular biology of baculoviruses (reviewed in references 8 and 13). Baculoviruses are used as biological pesticides (reviewed in reference 14) and have gained wide use recently as a helper-independent expression vector for foreign genes (22, 25, 26). AcMNPV has a rod-shaped capsid containing a circular double-stranded DNA genome of approximately 128 kilobase. Baculovirus nucleocapsids consist of the DNA condensed with a basic protaminelike core protein within the capsid (40). AcMNPV replicates in the nuclei of permissive insect cells and produces two morphological forms: budded virus (BV) and occluded virus (OV). BV consists of single enveloped nucleocapsids which bud from cells beginning by 10 to 12 h postinfection (p.i.) and serve to spread infection among cells within an individual insect. OV are formed in the nucleus and are composed of enveloped nucleocapsids which are enclosed within a crystalline protein matrix beginning 18 to 24 h p.i. The OV serve to spread infection among insects by an oral route.

The infection process is temporally regulated, with at least three phases of gene expression—early, late, and very late (reviewed in references 7 and 10). Late genes include those involved in virus assembly, while very late genes are involved specifically in the occlusion process. Transcriptional analysis indicates that the various temporal classes of genes are dispersed throughout the genome. Multiple overlapping transcripts with coterminal 5′ or 3′ ends are a common transcriptional motif. There is no evidence of splicing of late transcripts.

Few AcMNPV genes encoding specific products have been identified. Two occlusion-specific genes, polyhedrin (17) and p10 (19), have been located, mapped, and sequenced. Recently, the gene encoding the arginine-rich core protein found in AcMNPV nucleocapsids was located, sequenced, and transcriptionally mapped (47). The transcripts of polyhedrin, p10, and the 6.9-kilodalton (kDa) core protein gene all originate within the common sequence motif ATAAG. A 42-kDa peptide identified immunologically in both morphological forms of AcMNPV (34) has been identified as the major capsid protein (38). In this study, we determined the location, sequence, and transcriptional patterns of the AcMNPV capsid protein gene. A potential transcriptional regulatory mechanism is discussed.

(Received in part by Sh. B. T. [S.M.T.])

**MATERIALS AND METHODS**

**Virus and cells.** The virus used for all studies was AcMNPV L-1 (21). OV was isolated from *Trichoplusia ni* larvae infected per os at the fourth instar. AcMNPV was propagated in *Spodoptera frugiperda* IPLB-SF-21 (SF-21) cells (42) maintained in TC100 medium ( Gibco Laboratories, Grand Island, N.Y.) supplemented with 0.26% tryptose broth, 0.6 μg of amphotericin B per ml, 0.06 μg of penicillin G per ml, 0.27 μg of streptomycin sulfate per ml, and 10% fetal bovine serum. BV was prepared by infecting subconfluent monolayers of cells (10^5 cells per 100-mm plate) with a multiplicity of 0.5 PFU per cell. For time courses, monolayers of cells (10^5 cells per 100-mm plate or 10^6 cells per 35-mm plate) were inoculated with passage 2 AcMNPV at a multiplicity of 20 PFU per cell. The inoculum was removed after adsorption for 1 h at room temperature. Time zero was defined as the time when the inoculum was removed and incubation at 27°C was initiated. To block protein synthesis,
we used cycloheximide at 100 µg/ml throughout the time course beginning with a 30-min pretreatment prior to the addition of inoculum. To block DNA replication, we added aphidicolin to medium at 5 µg/ml following the adsorption period (33).

Construction and screening of an ORF library. A library of potential AcMNPV protein-coding regions was constructed by cloning random AcMNPV DNA fragments into a plasmid, pJS413 (45) (Molecular Genetics, Inc., Minnetonka, Minn.), which facilitates the selection of DNA fragments containing open reading frames (ORFs). Plasmid pJS413 contains the ribosome-binding site and N\(_2\) terminus of the \( \text{cro} \) gene from bacteriophage lambda and the \( \text{lacZ} \) gene from \( \text{Escherichia coli} \) under the control of the \( \text{lac} \) operator and promoter. \( \text{cro} \) and \( \text{lacZ} \) are separated by a multiple cloning site which places the two genes out of translational reading frame with respect to each other, resulting in a Lac\(^{\text{+}}\) phenotype. A DNA fragment containing an ORF results in the production of a trihybrid fusion protein (\( \text{cro-ORF-lacZ} \)) and a Lac\(^{-}\) phenotype if it is inserted into the multiple cloning site such that the translational reading frames of \( \text{cro} \) and \( \text{lacZ} \) are placed back in phase. Random blunt-ended fragments of viral DNA were prepared by sonication and T4 DNA polymerase repair, essentially by the method of Dinninger (46). Fragments 300 to 600 base pairs (bp) long were selected and inserted into the \\( S_{\text{mal}} \) site of pJS413, which was used to transform \( \text{E. coli} \) NF1829, a strain which overproduces \( \text{lac} \) repressor (43). Clones expressing \( \beta \)-galactosidase were selected and screened for the presence of viral DNA by colony hybridization by standard methods (23). The resulting library of clones expressing AcMNPV ORF-\( \beta \)-galactosidase fusions was screened immunologically with antiserum against virus structural proteins by the method of Helfman et al. (15).

**Antibody production.** Antibodies were raised in rabbits by administration of 300 to 500 µg of antigen emulsified in Freund complete adjuvant for initial injections and Freund incomplete adjuvant for subsequent injections by intradermal (41) and subcutaneous routes. Antiserum to virus structural proteins was raised against BV isolated from tissue culture. Antiserum to the selected AcMNPV ORF-\( \beta \)-galactosidase fusion protein CG-4-\( \beta \)-galactosidase was raised against an immunoadfinity column-purified fusion protein. Fusion protein production was induced in log-phase \( \text{E. coli} \) NF1829 cultures harboring the plasmid of interest by the addition of isopropyl-\( \beta \)-D-thiogalactopyranoside (Sigma Chemical Co., St. Louis, Mo.) to 1 mM. Bacterial lysates were prepared, and the fusion proteins were purified on a Proteosorb-\( \text{lacZ} \) immunoadfinity column as described in Promega Biotec Technical Bulletin 026 (Promega Biotec, Madison, Wis.).

**Isolation and purification of virus components.** AcMNPV extracellular enveloped nucleocapsids, BV, were purified from tissue culture supernatants by centrifugation for 30 min at 40,000 \( \times \) g through a 10 to 50% (wt/wt) sucrose gradient in 0.01 M Tris hydrochloride (pH 7.5). The band of enveloped nucleocapsids was removed, diluted, and pelleted by centrifugation for 1 h at 80,000 \( \times \) g. Enveloped nucleocapsids from OV (PDV), nucleocapsids, and capsids were purified from polyhedra by a modification of the method of Tweeen et al. (40). Enveloped nucleocapsids were released from polyhedra by treatment with 0.1 M sodium carbonate for 1 h at room temperature. The solution was neutralized by the addition of 1 M Tris hydrochloride (pH 7.5). After removal of undisrupted polyhedra by centrifugation at 7,500 \( \times \) g for 10 min, the released PDV was layered onto a 10 to 50% (wt/wt) sucrose gradient in 0.01 M Tris hydrochloride (pH 7.5) and centrifuged for 30 min at 40,000 \( \times \) g. Bands of PDV were removed, diluted in 0.01 M Tris hydrochloride (pH 7.5), and pelleted by centrifugation for 1 h at 80,000 \( \times \) g. Detergent solutions used in the preparations of capsids and nucleocapsids included the protease inhibitors phenylmethylsulfonyl fluoride (1 mM), leupeptin (20 µM), and pepstatin (40 µM), all purchased from Fluka BioChemika (Ronkonkoma, N.Y.). To prepare nucleocapsids, we stirred PDV at room temperature for 1 h in 1% (vol/vol) Nonidet P-40-30 mM NaCl-0.01 M Tris hydrochloride (pH 7.5). Nucleocapsids were separated from solubilized envelope proteins by centrifugation through to 10 to 50% (wt/wt) sucrose gradients in 0.01 M Tris hydrochloride (pH 7.5) for 30 min at 40,000 \( \times \) g. The band of nucleocapsids was removed, diluted in 0.01 M Tris hydrochloride (pH 7.5), and pelleted by centrifugation for 1 h at 80,000 \( \times \) g. Empty capsids were isolated by incubating nucleocapsids in 2% Nonidet P-40-0.01 M EDTA-1 mM NaCl-0.01 M Tris hydrochloride (pH 8.5) for 12 h at 37°C. Empty capsids were banded on performed CsCl gradients (1.18 to 1.55 g/cm\(^3\)) in 0.01 M Tris hydrochloride (pH 8.5) by centrifugation at 150,000 \( \times \) g for 2 h. The visible empty capsid band at 1.33 g/cm\(^3\) was removed, diluted to 0.01 M Tris hydrochloride (pH 7.5), and pelleted by centrifugation at 85,000 \( \times \) g for 30 min.

**Polycarylamide gel electrophoresis and Western immunoblotting.** Viral proteins were analyzed by electrophoresis on sodium dodecyl sulfate (SDS)-10% polyacrylamide slab gels by the method of Laemmli (20). Samples were run in duplicate on the same gel. Following electrophoresis, the gel was divided in half. The proteins in one half were stained with Coomassie brilliant blue G, and the proteins in the other half of the gel were electrophoretically transferred to a nitrocellulose filter (39). The filter was probed immunologically with the antiserum raised against the AcMNPV structural ORF-\( \beta \)-galactosidase fusion protein CG-4-\( \beta \)-galactosidase. The immunopositive proteins were visualized with alkaline phosphatase-conjugated goat anti-rabbit immunoglobulin G and Nitroblue Tetrazolium (Sigma) by the method of Blake et al. (3) except that 0.15 M Tris hydrochloride (pH 9.6) was substituted for the barbital buffer.

**Sequencing strategy.** The location of the ORF was determined by probing Southern blots of restricted AcMNPV DNA with radiolabeled plasmid DNA carrying the AcMNPV ORF-\( \beta \)-galactosidase fusion protein (pJCG-4). Plasmids pSTCHX3 and pSTCHX3M were constructed by cloning the AcMNPV HindIII-C-\( X_{\text{hoI}} \)-A fragment spanning 53.5 to 58.9 map units (m.u.) into the HindIII and \( X_{\text{hoI}} \) sites of Bluescript KS(+) and KS(−) cloning vectors (Stratagene, San Diego, Calif.). To sequence the ORF, the AcMNPV fragment was excised from pJCG-4 by cutting, with BamHI and BglII, sites flanking the \( S_{\text{mal}} \) ORF insertion site and inserted into the BamHI site of Bluescript KS(+). A series of overlapping deletions were made in the cloned AcMNPV fragments with exonuclease III (16) and mun bean nuclease. Single-stranded deletion clones were sequenced by the dideoxyribonucleotide-chain termination method (35), using a modified T7 DNA polymerase and a Sequenase kit purchased from U.S. Biochemical Corp. (Cleveland, Ohio). The deduced capsid protein sequence was compared with the National Biomedical Research Foundation data base (release 10.0) and translated sequences from Genbank (release 55), using the IBI-Pustell DNA and protein analysis programs.

**Transcriptional mapping.** Monolayers of AcMNPV-infected SF-21 cells were harvested at various times p.i., and total cell RNA was isolated by the guanidinium isothiocya-
nate method (5). Northern (RNA) blot analysis was performed with 20 μg of total RNA per lane. RNA was denatured by glyoxalation (28), fractionated on 1% agarose gels, and transferred to Zeta- Probe nylon membranes (Bio-Rad, Laboratories, Richmond, Calif.). The blots were probed with strand-specific RNA probes transcribed by T3 and T7 RNA polymerases from the Scal-NarI fragment (57.1 to 57.6 m.u.) of the capsid ORF cloned into Bluescript KS(+) (pSTSNSP).

The 5' and 3' ends of the transcripts were mapped by nuclease protection analysis (44), with either mung bean or S1 nuclease. Probes were generated by radiolabeling the appropriate recombinant plasmid at a single restriction site with either T4 polynucleotide kinase for 5' ends or T4 DNA polymerase for 3' ends. The plasmids were cleaved with a second enzyme and purified on agarose gels to yield probes labeled exclusively at one end. Hybridizations were conducted at 45°C in sodium trichloroacetate acid buffer (27) or at 49°C in 80% formamide–40 mM PIPES (piperazine-N,N'-bis(2-ethanesulfonic acid) (pH 6.4)–0.4 M NaCl–1 mM EDTA. Nuclease-resistant fragments were analyzed on 6% polyacrylamide–7 M urea sequencing gels.

For primer extension analysis, a 15-bp oligonucleotide primer corresponding to the NarI site used in nuclease protection studies was synthesized and labeled at its 5' end with T4 polynucleotide kinase. The primer was annealed to total infected cell RNA isolated at 12 h p.i. under the same conditions used for nuclease protection analysis. The primer was extended with cloned Moloney murine leukemia virus reverse transcriptase (Bethesda Research Laboratories, Inc., Gaithersburg, Md.). A sequencing ladder was generated by using the labeled oligonucleotide to prime a dideoxyribonucleotide-chain termination sequencing reaction from a single-stranded DNA spanning the region.

Hybrid selection and in vitro translation. RNA homologous to the capsid ORF was isolated by the method of Esche and Siegmann (9) as modified by Friesen and Miller (11). A single-stranded phage DNA (30 μg) generated from a Blue- script KS(−) plasmid carrying the capsid ORF Scal-NarI fragment, pSTSNSM, was immobilized on a nitrocellulose filter. Hybridization was conducted at 42°C in 50% formamide–10 mM PIPES (pH 6.4)–0.4 M NaCl–1 mM EDTA for 12 h with 1.5 μg of total RNA isolated from SF-21 cells 12 h p.i. After hybridization, the filters were washed extensively in 1× SSC (0.15 M NaCl and 0.015 M trisodium citrate, pH 7.0)–0.1% SDS–2 mM EDTA, and the selected RNA was eluted by boiling for 1 min in 1 mM EDTA containing 10 μg of yeast tRNA (Sigma). In vitro translation was performed at 30°C for 1 h in a rabbit reticulocyte system (Promega Biotec) in the presence of 50 μCi of [35S]methionine (1,200 Ci/mM; Dupont, NEN Research Products, Boston, Mass.) by the method of Pelham and Jackson (30). Proteins from infected cells were radiolabeled by incubating previously infected cell monolayers (106 cells per 35-mm plate) in methionine-deficient growth medium containing 200 μCi of [35S]methionine per ml. Following 1-h incubations at 6, 12, and 24 h p.i., cells were collected, washed with phosphate-buffered saline (31), and lysed in 1% Nonidet P-40–50 mM Tris hydrochloride (pH 8.0)–150 mM NaCl. Pulse-labeled proteins from infected cells and radiolabeled in vitro translation products were immunoprecipitated with anti-CG-4–β-galactosidase antiserum and Staphylococcus aureus (Sigma) by the method of Kuebler (28). Lysates were preprecipitated once with preimmune serum prior to incubation with immune serum. Infected cell proteins, in vitro translation products, and immunoprecipitates were analyzed by SDS-10% polyacrylamide gel electrophoresis followed by fluorography.

RESULTS

Identification of major capsid gene. To identify specific AcMNPV structural genes, a library of plasmids carrying random AcMNPV ORF–β-galactosidase gene fusions was constructed and immunologically screened with an antiserum raised against BV structural proteins. A colony producing an epitope recognized by the antiserum was identified by its strong reaction in the immunological screen of the library. The plasmid carried by this colony, pJCG-4, contained a 500-bp insert of AcMNPV mapping to the HindIII–C–XhoI-A region (53.5 to 58.9 m.u.) of AcMNPV (see below). The AcMNPV ORF–β-galactosidase fusion protein produced from pJCG-4, CG-4–β-galactosidase, was purified and used to raise antibodies to CG-4 epitopes.

To determine which virus structural protein was recognized by the anti-CG-4 antibody, the structural proteins of enveloped viruses from infected cell culture medium (BV) and pDV from infected insects (OV) were separated on SDS-polyacrylamide gels (Fig. 1, lanes b and c), blotted, and immunologically screened with anti-CG-4–β-galactosidase antiserum (Fig. 1, lanes e and f). The antiserum reacted primarily with the major protein component of both BV and OV enveloped nucleocapsids (Fig. 1, lanes b and c), a
polypeptide of approximately 42 kDa, suggesting that the antiserum recognized the major capsid protein. Owing to the large quantities of the 42-kDa protein in enveloped nucleocapsid lanes (Fig. 1, lanes b, c, e, and f), the Western blot shows a bowing effect in the 42-kDa region. The smaller polypeptides reacting with the anti-CG-4-β-galactosidase antiserum in the blot (lane f) appear to result from proteolytic cleavage of the capsid protein. During this work, we found it necessary to include substantial quantities of protease inhibitors in nucleocapsid preparations to minimize proteolytic cleavage of the capsid protein. Preimmune serum did not react with any viral proteins when tested on similar Western blots (data not shown).

The above results indicate that the anti-CG-4 antibody was recognizing the major capsid of both BV and OV. To confirm that the antiserum was recognizing the major capsid protein, empty capsids were purified and analyzed on SDS-polyacrylamide gels. The 42-kDa protein was the only visible component of purified preparations of empty capsids (Fig. 1, lane d), and this protein reacted with anti-CG-4-β-galactosidase antiserum (Fig. 1, lane g). Thus, the protein recognized by the anti-CG-4-β-galactosidase antiserum is the AcMNPV major capsid protein.

Mapping and sequencing the capsid gene. The location of the capsid ORF was mapped by Southern blot analysis to the region of the AcMNPV physical map corresponding to restriction fragments HindIII-C and XhoI-A (53.5 to 58.9 m.u.). The AcMNPV HindIII-C-XhoI-A fragment (Fig. 2A) was cloned (pSTCHX3 and pSTCHX3M). Since there were few useful restriction sites within the region, the ORF was localized more precisely by hybridizing 32P-labeled pJC1-4 to Southern blots of a series of exonuclease III deletion clones of pSTCHX3 and pSTCHX3M. Once the location of the ORF was defined by the deletion clones, both strands of DNA from the region containing the capsid gene were sequenced with the appropriate deletion clones (Fig. 2B). The sequence revealed an ORF of 1,040 bp capable of encoding a 39-kDa polypeptide (Fig. 3) reading counterclockwise with respect to the circular map of the virus. Several smaller ORFs reading in both directions were observed (Fig. 2C). When the deduced amino acid sequence of the capsid protein was screened against data bases, similar-
FIG. 3. Nucleotide sequence and predicted amino acid sequence of the AcMNPV major capsid gene. The last nucleotide in each line is numbered at the right. Transcriptional start sites (‡ TAAG) are indicated by asterisks. The transcriptional initiation points at positions 141, 357, and 404 are designated by arrows. Potential polyadenylation signals (AATTAAA) are underlined. An AATTAAA sequence (position 1558) at the 3′ end of the short capsid protein gene transcript that may be utilized poorly as a polyadenylation signal is printed in bold face. Arrowheads indicate the nucleotides at positions 491 (Narl) and 1399 (MsrI) that were radiolabeled to prepare probes used for transcriptional mapping.

2012
entities were found with short regions of three viral structural proteins. The greatest similarity noted was to a rotavirus capsid protein (4).

**Northern blot analysis.** Because AcMNPV genes can be transcribed from both DNA strands within the same region (11), we utilized strand-specific probes to monitor transcription from both strands by Northern blot analysis (Fig. 4). A probe complementary to the capsid ORF coding strand hybridized to a major 2.2-kilobase transcript expressed at 12 and 24 h. This RNA was not expressed in infected cells incubated in the presence of aphidicolin or cycloheximide (Fig. 4A). The opposite strand (sense) probe did not hybridize to any transcripts (Fig. 4B), indicating that there is no detectable transcription in a clockwise direction through this region of the capsid ORF.

**Location of 5′ and 3′ ends of capsid transcript by nuclease protection and primer extension analysis.** The 3′ ends of the RNA were mapped by an S1 nuclease protection assay with a probe 3′ end labeled at the MsiI site (Fig. 5A; nucleotide 1399 in Fig. 3). One major protected band of 975 nucleotides was observed at 12, 24, and 48 h p.i. (Fig. 5B). A minor band of 160 nucleotides was also observed at this time. Additional minor bands observed throughout the time course are probably artifacts resulting from incomplete nuclease digestion or nuclease cleavage within A+T-rich regions. A comparison of the DNA sequence with the minor 160-nucleotide protected fragment indicated either a polyadenylation or transcriptional termination site at or near an AATTAAAA sequence at nucleotide 1557 (Fig. 3), the only sequence resembling a consensus polyadenylation signal (2, 46) in this region. Although the 975-nucleotide protected fragment represented the major polyadenylation or termination site for late capsid transcripts, this protected fragment was also observed at 6 and 12 h p.i. in the presence of aphidicolin and cycloheximide, indicating that an early transcript initiates upstream from the MsiI site. There are two AATTAAAA polyadenylation signals upstream from the termination site reflected by the 975-nucleotide fragment. The sequence of this region and the nature of this early transcript are the subject of another study (S. M. Thiem and L. K. Miller, submitted for publication).

To map the 5′ ends of the capsid transcripts, we performed nuclease protection analysis on total infected cell RNA isolated at various times p.i. using a 1,650-nucleotide probe which was 5′ end labeled at the NarI site (Fig. 5; nucleotide 491 in Fig. 3). RNA isolated at 12, 24, and 48 h p.i. protected primarily DNA fragments of 84, 132, and 348 nucleotides (Fig. 5C), although the intensity of the bands corresponding to the 132- and 348-nucleotide protected fragments was reduced at 48 h p.i. Protected fragments were not observed with RNA isolated at 6 h p.i. or at 12 h p.i. in the presence of aphidicolin or cycloheximide, indicating that capsid transcripts are expressed exclusively at late times in viral infection. The additional minor protected fragments are probably due to incomplete mung bean nuclease digestion of the single-stranded regions of the DNA.

For primer extension analysis, a 5′-end-labeled 15-nucleotide synthetic oligonucleotide corresponding to the NarI end of the nuclease protection probe (Fig. 5A) was annealed to RNA isolated at 12 h p.i. and extended with reverse transcriptase. Three extension products were synthesized (Fig. 5D) which corresponded exactly with the protected
FIG. 5. Mapping the 3′ and 5′ ends of the capsid transcripts. (A) Locations of the probes and primer used to map the ends of the capsid transcripts relative to the capsid ORF. The size of each probe is shown below the line, and the radiolabeled ends are indicated by asterisks. At the top of the diagram, the location of the transcript is shown by an arrow, with the 3′ ends indicated by open arrows and the 5′ ends indicated by closed arrows. (B) S1 nuclease mapping of the 3′ ends of the transcript. Total RNA from mock-infected cells (lane M), AcMNPV-infected cells at 6, 12, 24, and 48 h p.i., and AcMNPV-infected cells at 12 h p.i. in the presence of aphidicolin (lane A) or cycloheximide (lane C) was hybridized to the 2,100-bp 3′ probe. The nuclease-resistant DNA-RNA hybrids were denatured and fractionated on a sequencing gel. The sizes (in nucleotides) of the molecular weight standards (lanes MW) and the probe are shown on the left. The predominant protected fragment of 975 bases is designated by the arrow on the right and by a large open arrow in panel A. A minor 160-base fragment is designated by an open arrow to the right of panel B and a small open arrow in panel A. (C) Mapping of the 5′ ends of the transcript by mung bean nuclease digestion and primer extension. A sequencing ladder was made by dideoxyribonucleotide-chain termination sequencing of the region with the 5′-end-labeled oligonucleotide as a primer. For primer extension (lane PE), the 5′-end-labeled oligonucleotide was hybridized to 12-h RNA and extended with reverse transcriptase. For mung bean nuclease analysis, total RNA as described for panel B was hybridized to the 1,650-bp 5′-end-labeled probe and treated with mung bean nuclease. Primer extension products (lane PE) and nuclease-resistant DNA-RNA hybrids (lanes M, 6, 12, 24, 48, A, and C) were denatured and fractionated on a sequencing gel.
fragments observed in the nuclease protection assay (Fig. 5C). A sequencing ladder generated from an appropriate cloned DNA primed with the labeled oligomer revealed that all three transcripts started at the third nucleotide of an ATAAA or GTAAG sequence (Fig. 5C and D).

In vitro translation of a hybrid selected mRNA corresponding to the major capsid protein. To confirm that the ORF encoded a 42-kDa protein, we used a single-stranded DNA containing the SalI-NarI fragment (57.1 to 57.6 m.u.), previously used as a probe for Northern blots (Fig. 4), to hybrid select mRNA isolated at 12 h p.i. The selected mRNA was translated in a cell-free protein synthesis system, and the products were analyzed by SDS-polyacrylamide gel electrophoresis. The translation product was a 42-kDa protein (Fig. 6, lane f). Immunoprecipitation of in vitro translation reactions showed that this protein is recognized by the anti-CG-4-β-galactosidase antiserum and comigrates with the capsid protein immunoprecipitated from pulse-labeled infected cell lysates (Fig. 6, lanes d and e). This protein was absent in immunoprecipitates from mock-infected cells (Fig. 6, lane c).

Precipitated proteins from both in vivo and in vitro synthesis appeared as 42-kDa doublets in SDS-polyacrylamide gels (Fig. 6, lanes d and e). The fact that both in vivo and in vitro-derived proteins migrated as doublets suggests that the doublet is not a result of posttranslational modification unless proteolysis is involved. A more likely possibility is that protein synthesis could be initiated at an internal methionine codon like the herpes simplex virus thymidine kinase mRNA (24). There are three methionine codons within the first 40 bp of the major ORF.

**DISCUSSION**

Through the use of an ORF selection-expression vector and immunological techniques, we identified the gene encoding the major capsid protein from AcMNPV. A molecular size of 39 kDa calculated from the predicted amino acid sequence agrees with the reported sizes of the capsid protein that range from 37 to 42 kDa (29, 34, 36–38). We demonstrated that we located the major capsid protein gene by (i) Western blot analysis of proteins from purified empty capsids, (ii) immunoprecipitation of pulse-labeled infected cell pellets, and (iii) analysis of products from in vitro translation hybrid-selected mRNA. We propose to call this gene vp39 for virus structural protein 39.

We classify the capsid gene as a late gene based on the temporal appearance of transcripts and the effects of inhibitors on capsid gene transcription. Late viral transcripts are distinguished from early transcripts by the absence of aphidicolin, a DNA synthesis inhibitor, and cycloheximide, a protein synthesis inhibitor. The reduction of capsid transcripts in the presence of cycloheximide indicates that virus-induced proteins are required for capsid gene transcription. This is consistent with a cascade model for regulation of the AcMNPV infection (10). Possible AcMNPV proteins that may be required for late transcription include positive or negative regulatory factors and a virus-specific RNA polymerase (12).

There are at least three initiation sites for capsid gene transcription as indicated by both nuclease protection and primer extension analyses. Like other late and very late baculovirus transcripts (17, 19, 47), the two major capsid gene transcripts initiate at the central A of a highly conserved ATAA sequence located within an A+T-rich region of the genome. It is interesting, however, that the most distal

---

Lanes A, C, G, and T contain the sequencing reactions. Sizes of molecular markers (lane MW) and the 5′ end-labeled probe (lane P) are shown on the right (the position of the probe is marked with an asterisk). The transcriptional start sites were determined by comparing the protected fragments and the extension products with the sequencing ladder. The sequence around each transcriptional start site is printed to the left of the sequencing ladder; arrows denote the 5′ initiation site. The positions of the 5′ ends are indicated by solid arrows in the diagram in panel A. The sizes of the major protected fragments and extension products are shown on the far left adjacent to the sequence. (D) Long exposure of a gel analyzing primer extension products and a sequencing ladder from the 15-base primer as described for panel C. The sequencing reactions are in lanes A, C, G, and T. Primer extensions from total RNA isolated from mock-infected cells and AcMNPV-infected cells at 6 and 12 h p.i. are in lanes M, 6, and 12, respectively. The sizes (in nucleotides) of molecular markers (lanes MW) are shown to the right of the panel. The sequences around the start sites are indicated to the left of the panel as described for panel C.

![FIG. 6. Hybrid selection, in vitro translation, and immunoprecipitation of the capsid protein. Capsid RNA was selected from total virus RNA isolated 12 h p.i. with a strand-specific DNA clone of the capsid ORF immobilized on nitrocellulose and translated in an RNA-dependent rabbit reticulocyte translation system. The radio-labeled translation products from hybrid-selected RNA (lane HS) and a control with no RNA (lane NR) are shown. Pulse-labeled proteins from mock-infected cells (lane M) and infected cells at 12 h p.i. (lane 12) are shown. Anti-CG-4-β-galactosidase antiserum was used to immunoprecipitate proteins from pulse-labeled lysates of mock-infected and infected cells at 12 h p.i. and from [35S]methionine-labeled in vitro translation products (lanes c to e). Molecular weights are shown at the left (K, 104). The 42-kDa capsid protein is indicated by an arrow, and the endogenous globin translation product (endog) is noted.](http://jvi.asm.org/content/63/1/205)
and least abundant of the three capsid gene transcripts initiates within a GTAAG sequence. Analysis of the polyhedrin promoter region by linker scan mutational analysis demonstrated that the ATAAG sequence at the polyhedrin transcriptional start site which overlaps an octanucleotide site, TAAGTA, is the major determinant of high-level polyhedrin gene expression (32). These studies also revealed that a mutation in which the ATAAG was changed to GTAAG still allowed strong expression, although it was reduced approximately fourfold. The presence of more than one of these GTAAG motifs in the capsid promoter is likely to be significant. Multiple initiation sites may allow the capsid gene to compete more effectively for RNA polymerase, allowing the message to be transcribed at high levels during BV synthesis (10 to 18 h p.i.) preceding occlusion.

The capsid gene is transcribed into at least three RNA species based on the existence of three different 5' initiation sites. A major transcriptional stop signal for the capsid transcript that terminates closest to the end of the capsid ORF, at approximately nucleotide 1560, is poorly recognized, and the majority of transcripts proceed through a downstream early ORF (Thiem and Miller, submitted), thus generating a bicistronic transcript. As in other regions of the AcMNPV genome that have been examined, there is little space between potential coding regions. In this region, the ATG of the downstream ORF is only two nucleotides from the translational termination signal (TAA) of the capsid gene. Readthrough may be important in the regulation of the downstream ORF.

Both late genes, such as the capsid protein and the basic 6.9-kDa core protein (47) genes, and occlusion-specific genes, such as the polyhedrin (17) and p10 (19) genes, contain the conserved 5' initiation sequence GTAAG. To determine whether there were any other features in the promoter regions of these genes that might account for differences in the temporal regulation of late and very late gene expression, we compared the sequences upstream from the ATG for any similarities. The upstream sequences of both the 6.9-kDa protein and the capsid protein genes contained two octamers, ATTAGGAA and ATTGCAAG, that were not observed in the occlusion-specific genes. The sequence ATTAGGAA was located at nucleotide positions −120 in the capsid sequence (Fig. 7A) and −143 in the core sequence (Fig. 7B) with respect to the initiation ATG. The sequence ATTAGCAAG was located at nucleotide position −200 (Fig. 7A) in the capsid and −209 (Fig. 7B) in the core. The placement of these sequences with respect to the ATAG transcriptional start sites may be of significance. The octamer ATTAGCAAG (double underline in Fig. 7) is approximately the same distance, 140 bp in the capsid and 148 bp in the 6.9-kDa core protein gene, from the ATAAG proximal to the initial methionine in each gene. Although the spacing of the octamers and the ATAG transcriptional initiation sites are not identical, the linear arrangement of the octamers and the ATAG sequences, ATTAGCAAG-ATTA GAA-ATAAG, is conserved between the two genes. The presence of these two octamer sequences leads to the intriguing hypothesis that protein factors could interact specifically at these sites and act either to initiate transcription at earlier times than polyhedrin or p10 or to repress transcription at very late times p.i. This hypothesis can be tested by linker scan mutational analysis.

The region of homology between the rotavirus capsid protein (4) and the AcMNPV capsid protein spanned 80 amino acids, 23% of the AcMNPV capsid protein. The
region with the most striking amino acid homology. 34 amino acids, is shown below:

\[
\text{KERG} \text{K}\text{R1LIPSATNYQDY} \text{FNLSMNMQA} \text{EQLIH} \text{FF} \quad (\text{capsid, amino acids 86 to 120})
\]

\[
\text{KLAG} \text{K}\text{R1NFDNUSIE} \text{1EWWLQ} \text{NRQ} \text{R7TGFVH} \quad (\text{human Vp rotavirus VP6})
\]

Identical amino acids are indicated by lines, and conservative changes are indicated by solid circles. The homology between the predicted amino acid sequence of the AcMNPV capsid protein and rotavirus capsid protein is not extensive, but it may reflect a common capsid structural motif.

ACKNOWLEDGMENTS

We thank David O’Reilly for many helpful discussions throughout the course of this work and preparation of the manuscript. We thank Karen Adams for technical assistance.

This work was supported in part by Public Health Service grant IR37 AI 23719 from the National Institute of Allergy and Infectious Diseases.

ADDITIONAL INFORMATION

A thymidine residue was inadvertently omitted from the sequence between nucleotides 232 and 233 in Fig. 3 and between nucleotides 133 and 132 in Fig. 7A. The polypeptide sequence of p39 from the Orgyia pseudotsugata nuclear polyhedrosis virus (OpMNPV) (Blissard et al., Virolology 168:354–362, 1989) is 60% identical to that of the AcMNPV major capsid protein. The nucleotide sequence between the distal ATAAAAG transcriptional start site and the initial methionine codon of the AcMNPV major capsid protein gene and the OpMNPV p39 gene are 73% identical with the greatest homology (81%) in the region between the two ATAAAAG motifs.

LITERATURE CITED


BACULOVIRUS CAPSID PROTEIN GENE 2017


