JOURNAL OF VIROLOGY, Feb. 1992, p. 971–982
0022-538X/92/020971-12$02.00/0

Vol. 66, No. 2

Identification and Expression of *rpo19*, a Vaccinia Virus Gene Encoding a 19-Kilodalton DNA-Dependent RNA Polymerase Subunit

BYUNG-YOON AHN, JOHANNES ROSEL,† NELSON B. COLE, AND BERNARD MOSS*

Laboratory of Viral Diseases, National Institute of Allergy and Infectious Diseases, Bethesda, Maryland 20892

Received 28 June 1991/Accepted 28 October 1991

The vaccinia virus DNA-dependent RNA polymerase subunit gene *rpo19* was identified, and its expression was examined at RNA and protein levels. Antibody to the multisubunit RNA polymerase purified from virions reacted with a polypeptide with an apparent M_r of 21,000 that was synthesized in reticulocyte lysates programmed with (i) mRNA from infected cells that was isolated by hybridization to DNA subclones of the viral genomic *Hin*dIII A fragment and (ii) mRNA made in vitro by transcription of the viral open reading frame A6R. Polyclonal antiserum, raised to a recombinant protein product of the A6R open reading frame which could encode an 18,996-Da protein with an acidic N terminus, reacted with M_r-21,000 and -22,000 polypeptides that cosedimented with purified RNA polymerase. Internal sequencing of the two polypeptides confirmed that both were encoded by A6R, and the gene was named *rpo19* to indicate the predicted molecular mass of the polypeptide in kilodaltons. Immunoblotting and metabolic labeling of infected cell proteins indicated that synthesis of the M_r-21,000 polypeptide started early and continued throughout virus infection, whereas the M_r-22,000 form appeared late following DNA replication. RNA analyses suggested that the *rpo19* mRNA was expressed from a dual early/late promoter and that the protein-coding region of the mRNA was directly preceded by a short 5′ poly(A) leader, apparently initiated within the TAAATG motif at the beginning of the open reading frame.

The encoding and virion packaging of a multisubunit DNA-dependent RNA polymerase are distinguishing features of vaccinia virus and other members of the *Poxviridae* that are in part responsible for the unique ability of these viruses to propagate within the cytoplasm of host cells (reviewed in references 27 and 28). Genes for 6 of the 8 to 10 subunits of the RNA polymerase have been identified and named *rpo147*, *rpo132*, *rpo35*, *rpo30*, *rpo22*, and *rpo18* according to the sizes of their predicted translation products in kilodaltons (2, 3, 5, 6, 11, 12, 31, 33). The two large subunit genes *rpo147* and *rpo132* are homologous to the corresponding-size subunit genes of prokaryotic and eukaryotic RNA polymerases (6, 11, 31). In addition, the sequence of the gene for one of the small subunits, *rpo30*, has significant similarities to that of the eukaryotic transcription elongation factor SII (2). By contrast, no homologs of the vaccinia virus RNA polymerase subunits encoded by *rpo35*, *rpo22*, and *rpo18* have been identified.

Knowledge of the complete subunit structure of the vaccinia virus RNA polymerase is required for genetic and functional studies and to determine whether modified forms of the enzyme regulate early, intermediate, and late gene expression. As a further step toward this goal, we describe the identification and expression of *rpo19*, a viral gene encoding another small subunit of the vaccinia virus RNA polymerase.

MATERIALS AND METHODS

Cells and virus. Vaccinia virus (strain WR) was propagated in HeLa S3 cells suspended in a minimum essential spinner medium supplemented with 5% horse serum (Quality Biologicals Inc.). Virions were purified by two successive sucrose gradient sedimentations as described previously (23). For in vivo labeling experiments, HeLa cell monolayers were grown in a minimum essential medium with 10% fetal calf serum.

Purification of RNA polymerase and protein sequencing. Vaccinia virus RNA polymerase was extracted from purified virions with deoxycholate (7, 13) and purified by chromatography on successive columns of DEAE-cellulose (DE52; Whatman), heparin agarose (BRL-Life Technologies, Inc.), single-stranded DNA agarose (BRL-Life Technologies), and phosphocellulose (P11; Whatman) and by glycerol gradient sedimentation. The enzyme activity was assayed by using a bacteriophage M13mp18 single-stranded DNA template as described previously (13). Approximately 100 μg of pure enzyme was obtained from 5×10^{13} virion particles.

The RNA polymerase used for protein sequencing was purified alternatively on columns of DEAE-cellulose, phosphocellulose, and Bio-Gel A1.5m (Bio-Rad). Approximately 100 μg of the purified enzyme was denatured by sodium dodecyl sulfate (SDS) and applied to a preparative 5 to 15% linear gradient polyacrylamide gel; the resolved subunit proteins were electrophoretically transferred to a nitrocellulose membrane (Schleicher & Schuell). Proteins were stained with 0.1% ponceau S (Sigma), and membrane strips containing appropriate bands were sent to William Lane at the Harvard Microchemistry Laboratory, where they were treated in situ with trypsin. The resultant peptides were separated by high-pressure liquid chromatography (HPLC) on a C_{18} column, and some were then analyzed on a gas-phase protein sequencer (ABI model 470A) connected to a 120A on-line phenylthiohydantoin-amino acid analyzer as described previously (1).

RNA hybridization and cell-free translation. Total cytoplasmic RNA was extracted from vaccinia virus-infected

* Corresponding author.
† Present address: CIBA-Geigy, Basel, Switzerland.

HeLa S3 cells and purified by CsCl centrifugation as described previously (14). Approximately 100 μg of RNA, prepared at 4 h after infection of cells in the presence of cycloheximide (100 μg/ml), was hybridized to 1-cm^2 pieces of nitrocellulose filter (BA85; Schleicher & Schuell) to which 20 to 25 μg of cloned denatured plasmid DNA was immobilized. Filters were washed stringently, and the bound RNA was eluted as described previously (24). In some experiments, RNA was synthesized in vitro from target DNA cloned in the Bluescript vector (Stratagene), using T7 RNA polymerase and the cap dinucleotide m^{7}G(5')ppp(5')G (29).

Hybrid-selected in vivo- or in vitro-synthesized RNAs were translated in the presence of [^{35}S]methionine (Amersham) in micrococcal nuclease-treated rabbit reticulocyte lysates as recommended by the supplier (Promega Biotec). A portion of the translated products was precleared by successive 2-h incubations with preimmune serum and protein A agarose beads (BRL-Life Technologies) at 4°C in a buffer containing 10 mM Tris-HCl (pH 7.4), 150 mM NaCl, 2 mM EDTA, 1% Nonidet P-40 detergent, and 0.05% aprotinin. After a 5-min low-speed centrifugation, the precleared supernatant was further incubated separately with preimmune serum or antiserum to vaccinia virus RNA polymerase for longer than 3 h. The reaction mixtures were finally incubated with protein A agarose beads, and the proteins that bound to the beads were analyzed by SDS-polyacrylamide gel electrophoresis (PAGE).

Cloning and DNA sequencing. The cosmid pHC79Hind IIIA, containing the 42-kbp vaccinia virus genomic *Hin*dIII A segment (25), was digested with *Hin*dIII and *Kpn*I, and the leftmost 13-kbp fragment was cloned in pUC19 to give pHAKpn1. The latter plasmid was the source of *Bam*HI subfragments used to construct pHABam 4 through 7 (Fig. 1). Cosmid or plasmid DNA, prepared by alkaline lysis of bacteria, was used for hybridization of viral mRNA.

For DNA sequencing, each *Bam*HI fragment was recloned into the phagemid vector Bluescript (Stratagene) in both orientations. Unidirectional deletions of inserts were made by exonuclease III digestion followed by treatment with nuclease S1 using the Pharmacia nested deletion kit. Single-stranded DNA of each deletion clone, prepared from 2-ml cultures of bacteria grown overnight in the presence of helper phage K-9, was sequenced by the dideoxynucleotide chain termination method using Sequenase (United State Biochemical) and [α-^{35}S]dATP (Amersham). The sequences at the junctions of subclones were confirmed by sequencing plasmid pHAKpn1 with appropriate synthetic oligonucleotide primers. Assembly, analysis, and translation of DNA sequences were performed by using the Staden program (39). Protein homology searches were performed with the FASTA program (32).

DNA containing specific open reading frames (ORFs) was cloned by polymerase chain reaction (PCR; Perkin Elmer-Cetus) of pHAKpn1, using pairs of synthetic oligonucleotide primers. Amplified DNA was inserted either into the Bluescript vector for in vitro translation studies or into the pET3c vector for expression of the protein in *Escherichia coli* (see below).

***E. coli* expression and antiserum preparation.** The full-length A6R protein-coding sequence was amplified by PCR and cloned into the bacteriophage T7 expression vector pET3c (36). The recombinant plasmids were introduced into *E. coli* BL21(DE3), a lysogenic strain harboring the T7 RNA polymerase gene under the control of the isopropylthiogalactoside-inducible promoter *lacUV5*. A protein with an apparent M_r of 21,000 was expressed from bacteria transformed with the recombinant plasmid. The identity of the expressed protein was confirmed by immunoblot analysis of the bacterial lysate, using antibody to the RNA polymerase and a ^{125}I-labeled protein A (data not shown). The highly expressed protein was separated by preparative SDS-PAGE, and protein bands were visualized by immersing the gel in a 0.3 M ice-cold KCl solution. Strips of gel containing the M_r-21,000 protein were cut out, and the recombinant protein was extracted by soaking the gel fragments in water overnight. The extracted protein was dialyzed and concentrated by lyophilization. Approximately 100 μg of recombinant protein was injected into New Zealand White rabbits in Freund's complete adjuvant, and the rabbits were given two

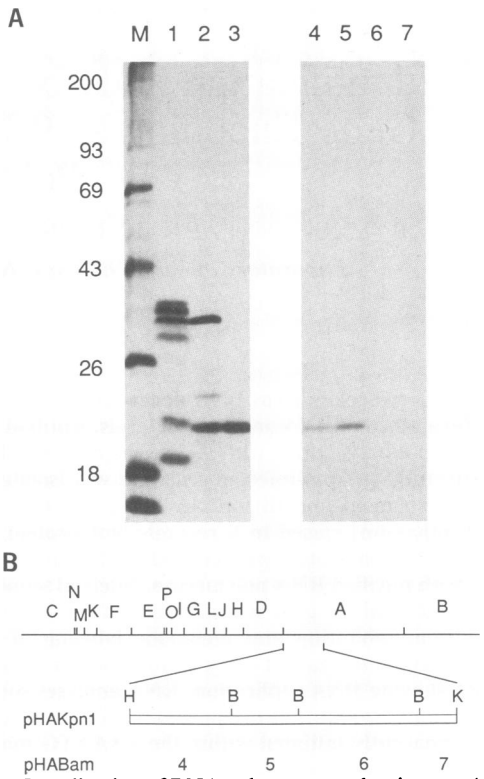


FIG. 1. Localization of RNA polymerase subunit gene. (A) Total early RNA, isolated from HeLa S3 cells at 4 h after vaccinia virus infection in the presence of cycloheximide (100 μg/ml), was hybridized to genomic DNA clones immobilized to nitrocellulose filters. The hybrid-selected mRNA was translated in rabbit reticulocyte lysates in the presence of [^{35}S]methionine, and the protein products were immunoprecipitated with antiserum to vaccinia virus RNA polymerase. An autoradiograph of the labeled proteins separated by SDS-PAGE on a 10% gel is shown. Lanes: M, ^{14}C-labeled protein markers, with their molecular masses in kilodaltons shown on the left; 1, proteins translated from 10 μg of total early RNA prior to hybridization; 2 through 7, translated products of RNA hybridized to a cosmid containing the entire 42-kbp *Hin*dIII A segment (lane 2), 13-kbp leftmost *Hin*dIII-*Kpn*I fragment cloned in pUC19 (lane 3), *Bam*HI subclones of the latter DNA, pHABam 4 through 7 (lanes 4 through 7, respectively). (B) *Hin*dIII restriction map of the entire vaccinia virus genome (upper line) and *Bam*HI restriction map of the leftmost 13-kbp *Hin*dIII-*Kpn*I subfragment (lower line) of the *Hin*dIII A segment. Numbers below the line represent each *Bam*HI subclone used for hybridization of RNA. B, *Bam*HI sites located at nucleotides 4138 (see Fig. 2), 6,761, and 11,763, numbering from the left-end *Hin*dIII site of the *Hin*dIII A fragment; K, *Kpn*I site at nucleotide 13110.

boosters of 50 μg of antigen. Anti-A6R sera were obtained 5 weeks after the first injection and thereafter.

In vivo labeling of viral proteins. Monolayer cultures of HeLa cells were infected with 20 PFU of virus per cell. For pulse-labeling, medium was removed at the indicated time and cells were incubated for 30 min in a methionine-free medium supplemented with 10% dialyzed fetal calf serum and 100 μCi of [^{35}S]methionine per ml. Where indicated, cytosine arabinoside (araC; 40 μg/ml) was included in the medium during infection. Cells were then washed with cold phosphate-buffered saline and lysed by freezing-thawing twice in RIPA buffer (10 mM Tris-HCl [pH 8], 140 mM NaCl, 1% bovine serum albumin, 1 mM phenylmethylsulfonyl fluoride, 0.1% Triton X-100, 0.1% sodium deoxycholate, 0.5% Nonidet P-40, 0.1% SDS). Cell lysates were clarified by centrifugation for 5 min in a microcentrifuge, and portions of the supernatants (equivalent of 2×10^5 cells) were used for immunoprecipitation in 500 μl of RIPA buffer. Samples were precleared by successive incubation with 20 μl of preimmune serum and 40 μl of protein A agarose beads (50% [vol/vol] suspension in phosphate-buffered saline) for 1 h at 4°C with constant rocking. The beads were removed by centrifugation, and the precleared samples were incubated with 20 μl of anti-A6R serum for 3 to 16 h and finally with 40 μl of protein A agarose beads for an additional 1 h. The beads were washed three times with buffer containing 50 mM Tris-HCl (pH 8), 300 mM NaCl, and 0.1% Triton X-100 and once with the latter buffer containing 0.1% SDS and 0.1% deoxycholate. The washed beads were boiled in the gel loading buffer, and the labeled proteins were analyzed by SDS-PAGE.

RNA analyses. Northern (RNA) blot (37), nuclease S1 digestion (3), and primer extension (4) techniques were used to characterize the *rpo19* mRNA as described previously. Total RNA extracted by the RNAzol method (Cinna BiotecX) from infected HeLa S3 cells was used for the Northern blot and nuclease S1 analyses. For the primer extension analysis, the extracted RNA was further treated with phenol-chloroform and precipitated with ethanol. The ^{32}P-labeled probe (extending from nucleotides 4260 to 4850) used in the Northern blot hybridization was made by asymmetric PCR of plasmid pHABam 5 with pairs of synthetic oligonucleotide primers (100:1 molar excess of the complementary strand) and [α-^{32}P]dCTP for internal labeling. The 5′-end-labeled probe, used in the nuclease S1 analysis of the 5′ ends of mRNAs, was prepared by 40 cycles of runoff PCR of plasmid pHAKpn1 (linearized by *Xba*I at nucleotide 3547 located 713 bp upstream of the ATG initiation codon of the *rpo19* gene) with a 5′-end-labeled (at nucleotide 4440) primer hybridized to the internal region (nucleotides 4415 to 4440) of the *rpo19* gene. The same primer was used for the primer extension analysis of the RNA 5′ end. T4 polynucleotide kinase and [γ-^{32}P]ATP were used to label the primer. The poly(dT)-containing probe, used to characterize the 5′ poly(A) leader, was prepared by two sequential PCRs as described previously (3).

The probe for the 3′-end mapping by nuclease S1 digestion was prepared by filling in the 3′-recessed end generated by *Spe*I digestion (at nucleotide 4357) of pHABam 5 with the Klenow form of DNA polymerase and [α-^{32}P]dCTP, followed by *Bgl*II digestion (at nucleotide 6110). Labeled probes were purified by electrophoresis on a 7 M urea–4% polyacrylamide gel before use.

Nucleotide sequence accession number. The 3,150-bp sequence from the *Hin*dIII A genomic fragment (Fig. 2) has been entered into GenBank under accession number M76473.

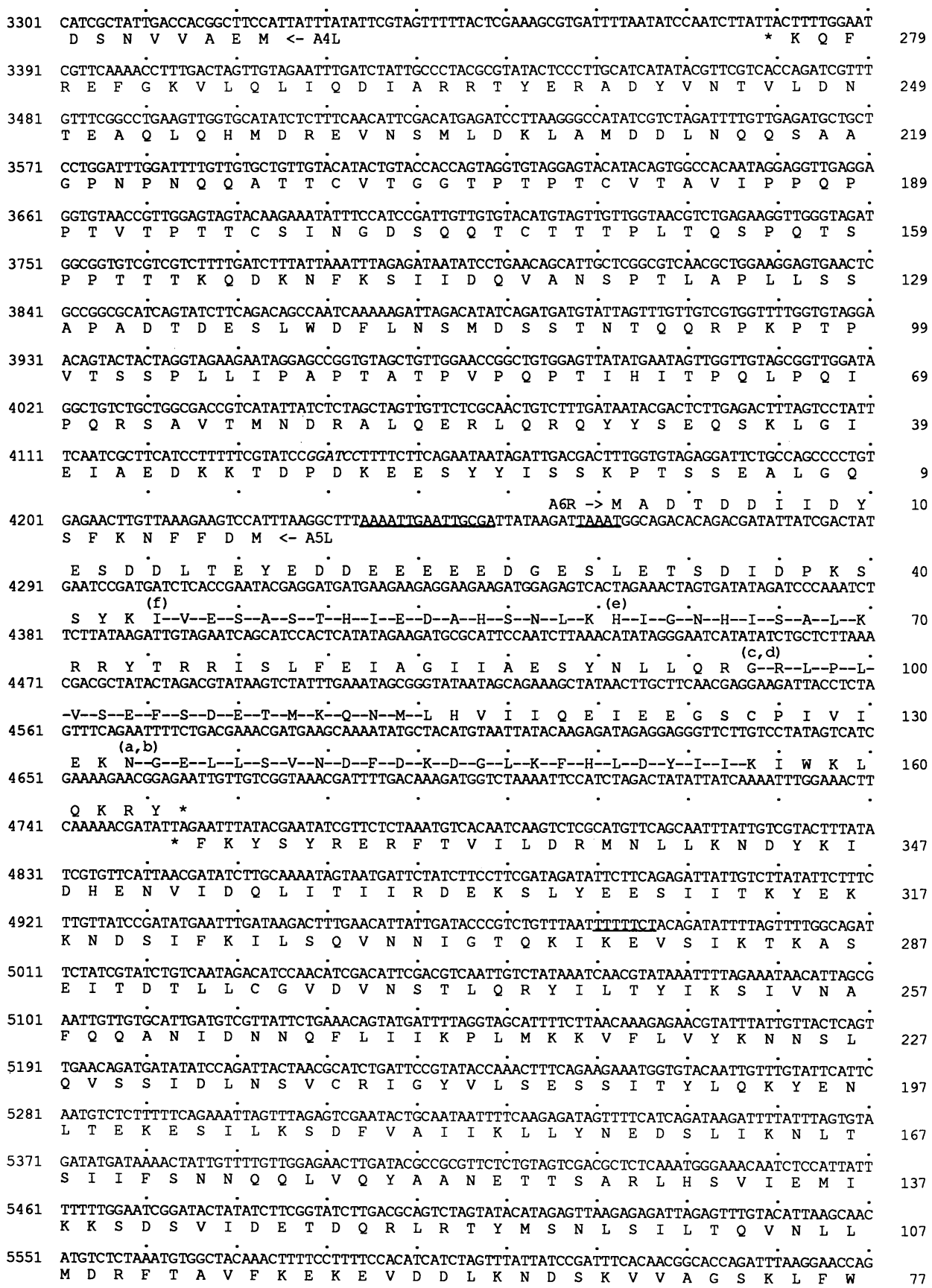
RESULTS

Localization of an RNA polymerase subunit gene encoding an M_r-21,000 polypeptide. It was previously demonstrated that mRNA obtained from vaccinia virus-infected cells directed a cell-free translation system to produce 8 to 10 polypeptides that reacted with antibody raised to the RNA polymerase purified from vaccinia virions (25). The approximate locations of the RNA polymerase genes were determined by selecting the mRNA by hybridization to cloned DNA fragments prior to translation in the micrococcal nuclease-treated reticulocyte system and immunoprecipitation. In this manner, evidence was obtained that several polypeptides estimated by SDS-PAGE to range from 130 to 17 kDa were encoded within the 42-kbp *Hin*dIII A genomic segment which had been cloned in a cosmid vector (25). Further analyses led to the precise localization of the genes encoding 132-kDa (6) and 35-kDa (5) subunits within the *Hin*dIII A fragment. In the present study, mapping of the gene encoding an M_r-21,000 RNA polymerase subunit was carried out.

As shown in Fig. 1A, several [^{35}S]methionine-labeled proteins were immunoprecipitated from reticulocyte extracts programmed with either total RNA (lane 1) or RNA hybridized to the entire *Hin*dIII A fragment (lane 2). The two large subunits (147 and 132 kDa) are usually expressed poorly in reticulocyte extracts and are not seen at all in this experiment. Only one major translation product, estimated to be about 21 kDa in size, was obtained with RNA that had hybridized to pHAKpn1 DNA containing the 13-kbp leftmost *Hin*dIII-*Kpn*I fragment (lane 3). Further mapping of this putative RNA polymerase gene was performed by using the *Bam*HI subclones (pHABam 4 to 7) of pHAKpn1 DNA. Only mRNAs that hybridized to pHABam 4 and 5 expressed an immunoreactive M_r-21,000 protein, suggesting that the gene may reside near the *Bam*HI sites present at the junction of the two clones.

Nucleotide sequencing and fine mapping. To identify the RNA polymerase gene, we needed the nucleotide sequence of the vaccinia virus DNA present in pHABam 4 and 5, of which only part had been reported at the time (34, 41). We therefore determined about 800 bp of sequence (proximal to the *Bam*HI site at nucleotide 4138) of pHABam 4 and the contiguous 2,300-bp sequence of pHABam 5 (Fig. 2). Examination of the nucleotide sequence revealed three complete ORFs, designated A5L, A6R, and A7L, respectively. These three ORFs are flanked by two ORFs identified previously: A4L, encoding a major structural protein P4b (34), and A8L, encoding a subunit of the vaccinia virus early transcription factor (18). Predicted molecular masses of the three proteins are 31 kDa for A5L, 19 kDa for A6R, and 43 kDa for A7L. None of the ORFs showed any significant amino acid sequence homology with entries in the National Biomedical Research Foundation protein data base (release 23) when analyzed by the FASTA program.

To further localize the RNA polymerase gene, we isolated viral RNA that hybridized to the *Bam*HI-*Xba*I DNA fragment (nucleotides 4138 to 4710; Fig. 2) derived from the pHABam 5 and demonstrated that it directed the synthesis of an M_r-21,000 polypeptide that was immunoprecipitated with the antiserum raised to the RNA polymerase (data not shown). This result narrowed the candidate genes to A5L and A6R. Because of its size, we considered ORF A6R to be a good candidate for the gene encoding an RNA polymerase subunit that was estimated to be 21 kDa by SDS-PAGE. However, the protein encoded by ORF A5L might have


3301 CATCGCTATTGACCACGGCTTCCATTATTTATATTCGTAGTTTTTACTCGAAAGCGTGATTTTAATATCCAATCTTATTACTTTTGGAAT
D S N V V A E M <- A4L * K Q F 279
3391 CGTTCAAAACCTTTGACTAGTTGTAGAATTTGATCTATTGCCCTACGCGTATACTCCCTTGCATCATATACGTTCGTCACCAGATCGTTT
R E F G K V L Q L I Q D I A R R T Y E R A D Y V N T V L D N 249
3481 GTTTCGGCCTGAAGTTGGTGCATATCTCTTTCAACATTCGACATGAGATCCTTAAGGGCCATATCGTCTAGATTTTGTTGAGATGCTGCT
T E A Q L Q H M D R E V N S M L D K L A M D D L N Q Q S A A 219
3571 CCTGGATTTGGATTTTGTTGTGCTGTTGTACATACTGTACCACCAGTAGGTGTAGGAGTACATACAGTGGCCACAATAGGAGGTTGAGGA
G P N P N Q Q A T T C V T G G T P T P T C V T A V I P P Q P 189
3661 GGTGTAACCGTTGGAGTAGTACAAGAAATATTTCCATCCGATTGTTGTGTACATGTAGTTGTTGGTAACGTCTGAGAAGGTTGGGTAGAT
P T V T P T T C S I N G D S Q Q T C T T T P L T Q S P Q T S 159
3751 GGCGGTGTCGTCGTCTTTTGATCTTTATTAAATTTAGAGATAATATCCTGAACAGCATTGCTCGGCGTCAACGCTGGAAGGAGTGAACTC
P P T T T K Q D K N F K S I I D Q V A N S P T L A P L L S S 129
3841 GCCGGCGCATCAGTATCTTCAGACAGCCAATCAAAAAGATTAGACATATCAGATGATGTATTAGTTTGTTGTCGTGGTTTTGGTGTAGGA
A P A D T D E S L W D F L N S M D S S T N T Q Q R P K P T P 99
3931 ACAGTACTACTAGGTAGAAGAATAGGAGCCGGTGTAGCTGTTGGAACCGGCTGTGGAGTTATATGAATAGTTGGTTGTAGCGGTTGGATA
V T S S P L L I P A P T A T P V P Q P T I H I T P Q L P Q I 69
4021 GGCTGTCTGCTGGCGACCGTCATATTATCTCTAGCTAGTTGTTCTCGCAACTGTCTTTGATAATACGACTCTTGAGACTTTAGTCCTATT
P Q R S A V T M N D R A L Q E R L Q R Q Y Y S E Q S K L G I 39
4111 TCAATCGCTTCATCCTTTTTCGTATCCGGATCCTTTTCTTCAGAATAATAGATTGACGACTTTGGTGTAGAGGATTCTGCCAGCCCCTGT
E I A E D K K T D P D K E E S Y Y I S S K P T S S E A L G Q 9
A6R -> M A D T D D I I D Y 10
4201 GAGAACTTGTTAAAGAAGTCCATTTAAGGCTTTAAAATTGAATTGCGATTATAAGATTAAATGGCAGACACAGACGATATTATCGACTAT
S F K N F F D M <- A5L
E S D D L T E Y E D D E E E E E D G E S L E T S D I D P K S 40
4291 GAATCCGATGATCTCACCGAATACGAGGATGATGAAGAAGAGGAAGAAGATGGAGAGTCACTAGAAACTAGTGATATAGATCCCAAATCT
(f) (e)
S Y K I--V--E--S--A--S--T--H--I--E--D--A--H--S--N--L--K H--I--G--N--H--I--S--A--L--K 70
4381 TCTTATAAGATTGTAGAATCAGCATCCACTCATATAGAAGATGCGCATTCCAATCTTAAACATATAGGGAATCATATATCTGCTCTTAAA
(c,d)
R R Y T R R I S L F E I A G I I A E S Y N L L Q R G--R--L--P--L- 100
4471 CGACGCTATACTAGACGTATAAGTCTATTTGAAATAGCGGGTATAATAGCAGAAAGCTATAACTTGCTTCAACGAGGAAGATTACCTCTA
-V--S--E--F--S--D--E--T--M--K--Q--N--M--L H V I I Q E I E E G S C P I V I 130
4561 GTTTCAGAATTTTCTGACGAAACGATGAAGCAAAATATGCTACATGTAATTATACAAGAGATAGAGGAGGGTTCTTGTCCTATAGTCATC
(a,b)
E K N--G--E--L--L--S--V--N--D--F--D--K--D--G--L--K--F--H--L--D--Y--I--I--K I W K L 160
4651 GAAAAGAACGGAGAATTGTTGTCGGTAAACGATTTTGACAAAGATGGTCTAAAATTCCATCTAGACTATATTATCAAAATTTGGAAACTT
Q K R Y *
4741 CAAAAACGATATTAGAATTTATACGAATATCGTTCTCTAAATGTCACAATCAAGTCTCGCATGTTCAGCAATTTATTGTCGTACTTTATA
* F K Y S Y R E R F T V I L D R M N L L K N D Y K I 347
4831 TCGTGTTCATTAACGATATCTTGCAAAATAGTAATGATTCTATCTTCCTTCGATAGATATTCTTCAGAGATTATTGTCTTATATTCTTTC
D H E N V I D Q L I T I I R D E K S L Y E E S I I T K Y E K 317
4921 TTGTTATCCGATATGAATTTGATAAGACTTTGAACATTATTGATACCCGTCTGTTTAATTTTTTCTACAGATATTTTAGTTTTGGCAGAT
K N D S I F K I L S Q V N N I G T Q K I K E V S I K T K A S 287
5011 TCTATCGTATCTGTCAATAGACATCCAACATCGACATTCGACGTCAATTGTCTATAAATCAACGTATAAATTTTAGAAATAACATTAGCG
E I T D T L L C G V D V N S T L Q R Y I L T Y I K S I V N A 257
5101 AATTGTTGTGCATTGATGTCGTTATTCTGAAACAGTATGATTTTAGGTAGCATTTTCTTAACAAAGAGAACGTATTTATTGTTACTCAGT
F Q Q A N I D N N Q F L I I K P L M K K V F L V Y K N N S L 227
5191 TGAACAGATGATATATCCAGATTACTAACGCATCTGATTCCGTATACCAAACTTTCAGAAGAAATGGTGTACAATTGTTTGTATTCATTC
Q V S S I D L N S V C R I G Y V L S E S S I T Y L Q K Y E N 197
5281 AATGTCTCTTTTTCAGAAATTAGTTTAGAGTCGAATACTGCAATAATTTTCAAGAGATAGTTTTCATCAGATAAGATTTTATTTAGTGTA
L T E K E S I L K S D F V A I I K L L Y N E D S L I K N L T 167
5371 GATATGATAAAACTATTGTTTTGTTGGAGAACTTGATACGCCGCGTTCTCTGTAGTCGACGCTCTCAAATGGGAAACAATCTCCATTATT
S I I F S N N Q Q L V Q Y A A N E T T S A R L H S V I E M I 137
5461 TTTTTGGAATCGGATACTATATCTTCGGTATCTTGACGCAGTCTAGTATACATAGAGTTAAGAGAGATTAGAGTTTGTACATTAAGCAAC
K K S D S V I D E T D Q R L R T Y M S N L S I L T Q V N L L 107
5551 ATGTCTCTAAATGTGGCTACAAACTTTTCCTTTTCCACATCATCTAGTTTATTATCCGATTTCACAACGGCACCAGATTTAAGGAACCAG
M D R F T A V F K E K E V D D L K N D S K V V A G S K L F W 77

```
5641  AATAAAAAACTCTGATAACTACAATATTTCATCATAGTTACGATTTTATCATCTTCTATAGTTGGTGTAATAGCGCATACCTTTTTCTCC
      F  L  F  S  Q  Y  S  C  Y  K  M  M  T  V  I  K  D  D  E  I  T  P  T  I  A  C  V  K  K  E     47

5731  AAGACTGGAACCAACGTCATAAAAATGTTTAAATCAAAATCCATATCAACATCTGATGCGCTAAGACCAGTCTCGCGTTCAAGATTATCT
      L  V  P  V  L  T  M  F  I  N  L  D  F  D  M  D  V  D  S  A  S  L  G  T  E  R  E  L  N  D     17

5821  TTACTAATGGTGACGAACTCATCATATAGAACTCTAAGTTTGTCCATTATTTATTTACAGATTTAGTTGTTTAATTTATTTGTGCTCTTC
      K  S  I  T  V  F  E  D  Y  L  V  R  L  K  D  M  <- A7L                 *  N  I  Q  A  R  G   705

5911  CAGAGTTGGGATAGTATTTTTCTAACGTCGGTATTATATTATTAGGATCTACGTTCATATGTATCATAATATTAATCATCCACGTTTTGA
        S  N  P  Y  Y  K  E  L  T  P  I  I  N  N  P  D  V  N  M  H  I  M  I  N  I  M  W  T  K  I   675

6001  TAAATCTATCTTTAGCTTCTGAAATAACGTATTTAAACAAAGGAGAAAAATATTTAGCTACGGCATCAGACGCAATAACATTTTTTGTAA
        F  R  D  K  A  E  S  I  V  Y  K  F  L  P  S  F  Y  K  A  V  A  D  S  A  I  V  N  K  T  F   645

6091  ATGTAACGTATTTAGACGACAGATCTTCGTTAAAAAGTTTTCCATCTATGTAGAATCCATCAGTTGTTAACACCATTCCCGCGTCAGATT
        T  V  Y  K  S  S  L  D  E  N  F  L  K  G  D  I  Y  F  G  D  T  T  L  V  M  G  A  D  S  Q   615

6181  GAATAGGAGTTTGAATAGTTTGTTTTGGAAATAGATCCTTCAATAACTTATAGTTGGGTGGGAAAAAATCGATTTTATCACTAGACTCTT
        I  P  T  Q  I  T  Q  K  P  F  L  D  K  L  L  K  Y  N  P  P  F  F  D  I  K  D  S  S  E  K   585

6271  TCTTTTTTACTATCATTACCTCATGAACTATTTCTTGAATGAGTATATGTATTTTCTTTCCTATATCGGACGCGTTCATTGGAAAATATA
        K  K  V  I  M  V  E  H  V  I  E  Q  I  L  I  H  I  K  K  G  I  D  S  A  N  M  P  F  Y  V   555

6361  CCATGTCGTTAACTATAAGAATATTTTTATCCTCGTTTACAAACTGAATAATATCAGATGTAGTTCGTAAACGAACTATATCATCACCAG
        M  D  N  V  I  L  I  N  K  D  E  N  V  F  Q  I  I  D  S  T  T  R  L  R  V  I  D  D  G  A   525
```

FIG. 2. Partial nucleotide sequence of the vaccinia virus (strain WR) *Hin*dIII A fragment. Numbering on the left starts from the left-end *Hin*dIII site (designated nucleotide 1) in the same orientation as that of the virus genome shown in Fig. 1B. Sequence of the top strand from nucleotides 3301 to 6450 is shown. Notation of the ORFs is according to Rosel et al. (35). Arrows indicate directions of each ORF. Predicted amino acid sequences are shown in single-letter code above the DNA sequence for rightward ORFs and below for leftward ORFs. Numbers on the right represent the amino acid residue at the right end of each line for each ORF. Only parts of the A4L and A8L ORF sequences are shown. The sequences from nucleotides 3301 to 3552 and from nucleotides 6398 to 6450 overlap sequences published previously (34, 41). The *Bam*HI site at nucleotide 4138 is shown in italic, and the three regulatory DNA sequences (discussed in the text) are underlined. Amino acid sequences shown with dashed lines are derived from N termini of six tryptic peptides (see Fig. 5).

migrated anomalously fast because of its unusual amino acid content (11% proline and 13% threonine, compared with the average values of 5 and 6%, respectively, for protein sequences in GenBank). In addition, ORF A5L spans the *Bam*HI site (Fig. 2), consistent with the selection of some mRNA encoding the M_r-21,000 polypeptide by hybridization to pHABam 4 as well to pHABam 5 (Fig. 1). To resolve these alternative possibilities, the DNAs for ORFs A5L and A6R were individually PCR amplified and cloned into a bacteriophage T7 expression vector (Bluescript), and the genes were transcribed in vitro with T7 RNA polymerase. The in vitro-synthesized RNAs were then translated in reticulocyte lysates in the presence of [^{35}S]methionine, and the labeled proteins were subsequently immunoprecipitated with RNA polymerase antibody. A truncated bacterial *lacZ* gene template served as a control for the specificity of translation and antibody binding. Translation of the RNA made from the A6R template produced an M_r-21,000 polypeptide which reacted with antiserum to the viral RNA polymerase but not with the preimmune serum (Fig. 3), whereas the A5L RNA produced an M_r-35,000 polypeptide that did not react with polymerase antiserum. Thus, these data were consistent with the encoding of the RNA polymerase subunit by the A6R ORF.

Subsequent to our determination of the sequence (Fig. 2) for the WR strain of vaccinia virus, the genomic sequence of the Copenhagen strain of vaccinia virus was reported (19). Comparison of the corresponding regions of the two vaccinia virus strains indicated only 20 differences in 3,150 nucleotides. Fifteen of the twenty different nucleotides were A:G or C:T transitions, and two were C:A and A:T transversions. The most extreme difference appeared as a three-nucleotide insertion in the Copenhagen strain (at nucleotide 5606) which caused one amino acid insertion in the A7L ORF. The predicted amino acid sequences of the corresponding WR and Copenhagen ORFs showed 98.2, 100, and 99.2% identity for A5L, A6R, and A7L, respectively.

Cosedimentation of the A6R gene product with virion RNA

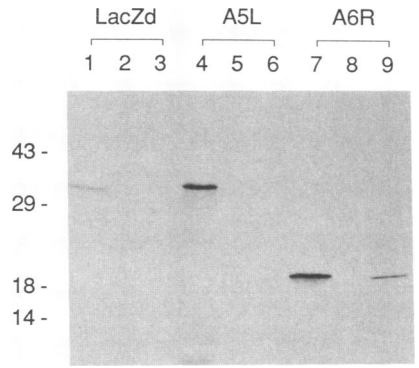


FIG. 3. Translation and immunoprecipitation of the in vitro-synthesized RNA. Bluescript plasmids containing the vaccinia virus ORFs or part of the bacterial *lacZ* gene sequences were linearized with restriction enzymes at sites downstream of the protein-coding region and transcribed by T7 RNA polymerase. RNA (1 μg) was translated in a reticulocyte lysate in the presence of [^{35}S]methionine, and the total labeled proteins (lanes 1, 4, and 7) or the proteins bound to preimmune serum (lanes 2, 5, and 8) or to antiserum raised to purified RNA polymerase (lanes 3, 6, and 9) were analyzed by SDS-PAGE on a 15% gel. As a control, RNA was transcribed from a Bluescript vector containing the *E. coli lacZ* gene linearized by cleavage at an internal *Cla*I site and translated (lanes 1 to 3). An autoradiograph is shown. The sizes of protein markers in kilodaltons are shown on the left.

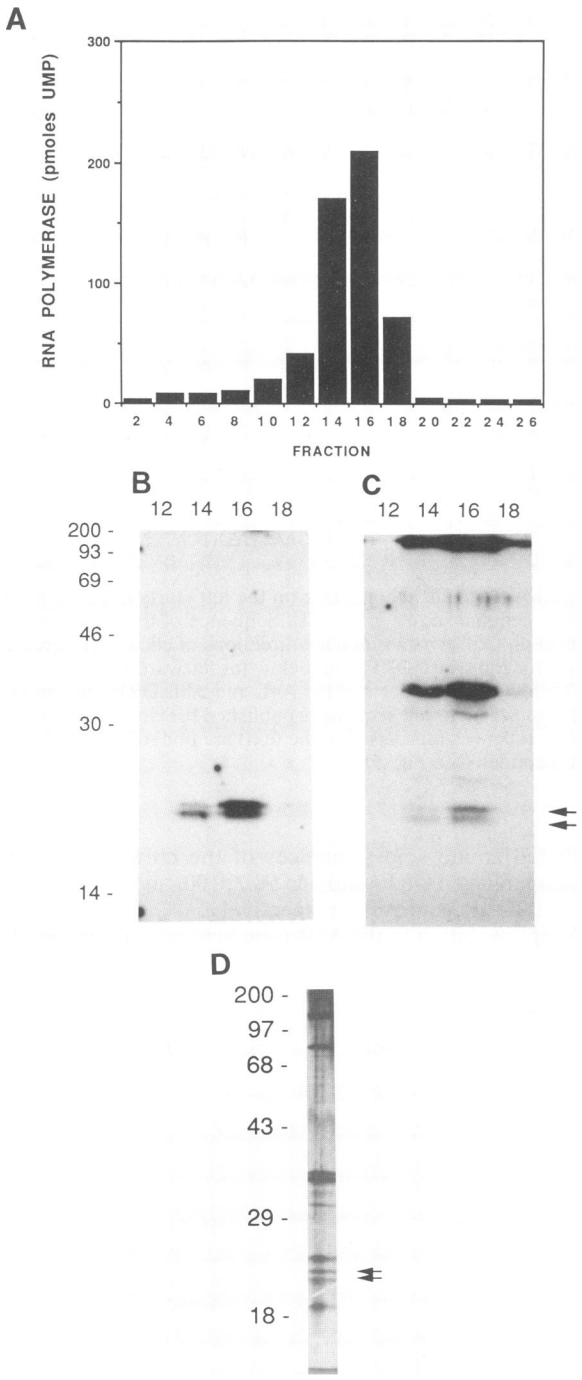


FIG. 4. Cosedimentation of the *rpo19* subunits with vaccinia virion RNA polymerase upon glycerol gradient centrifugation. Purified RNA polymerase (10 μg) was applied to a 15 to 35% linear glycerol gradient (11 ml in volume) and sedimented at 4°C at 41,000 rpm for 16 h (Beckman rotor SW41). Fractions of 0.4 ml were collected from the bottom of the tube and analyzed. (A) RNA polymerase activity of even-numbered fractions. Number of units in 5 μl of each fraction is indicated. A unit is equivalent to 1 pmol of [α-^{32}P]UMP incorporated under the conditions used (13). The fraction numbers are in an increasing order from the bottom to the top of the tube. The peak sedimentation positions of three protein markers that were run in a parallel gradient were as follows: fraction 12, thyroglobulin (M_r 690,000); fraction 20, bovine serum albumin (M_r 66,000); and fraction 24, carbonic anhydrase (M_r 29,000). (B and C) Immunoblot analyses. The proteins present in the individual glycerol gradient fractions were resolved by SDS-PAGE on a 12% gel and transferred to a nitrocellulose membrane. The membrane was incubated with antibody to the recombinant A6R protein (B) or with antibody to the whole RNA polymerase (C) and further incubated with ^{125}I-labeled staphylococcal protein A. Autoradiographs of gel lanes containing RNA polymerase peak fractions are shown. Positions of ^{14}C-labeled protein markers are shown in kilodaltons on the left. (D) Silver-stained SDS–10% polyacrylamide gel containing purified RNA polymerase. Arrows on the right indicate the two *rpo19* subunit polypeptides.

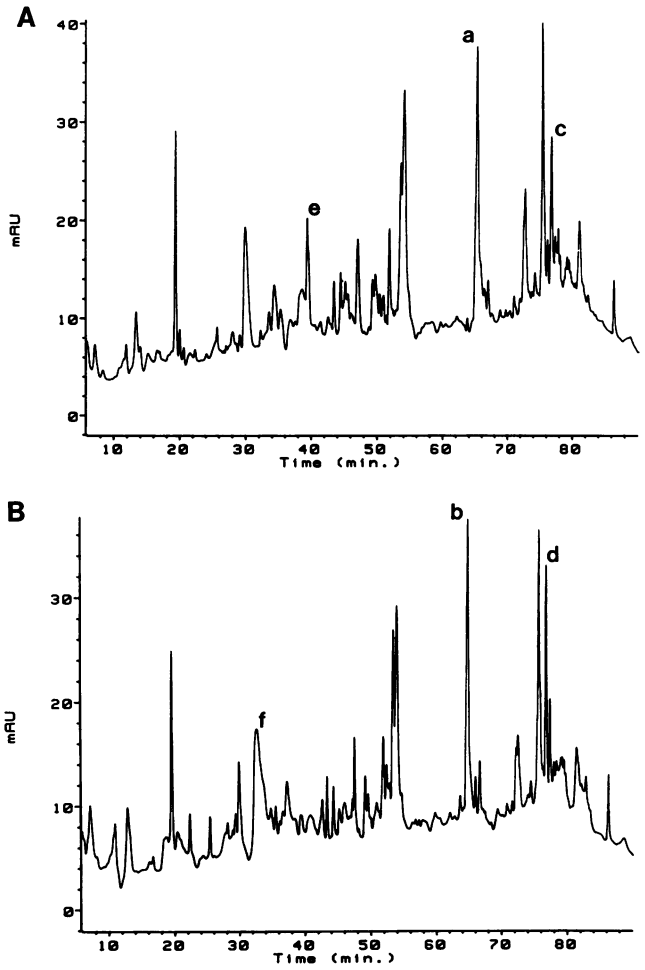


FIG. 5. Chromatograms of tryptic digests of the M_r-22,000 (A) and M_r-21,000 (B) polypeptides. In situ digestions of membranes containing approximately 200 pmol of each protein were separated on a C_{18} HPLC column. A_{210} (milliabsorbancy units [mAU]) of the effluent fractions is shown. N-terminal amino acid sequences of the six peptides (peaks a through f) were determined (see text).

polymerase. At this point, our identification of the A6R gene as an RNA polymerase subunit gene relied solely on the size of the in vitro-translated protein and its reactivity with antiserum to the purified multisubunit RNA polymerase. We wished to characterize the protein made in vivo and confirm its physical association with the RNA polymerase. To accomplish this, we prepared a polyclonal antibody to the A6R protein that we expressed in *E. coli*. We then used this

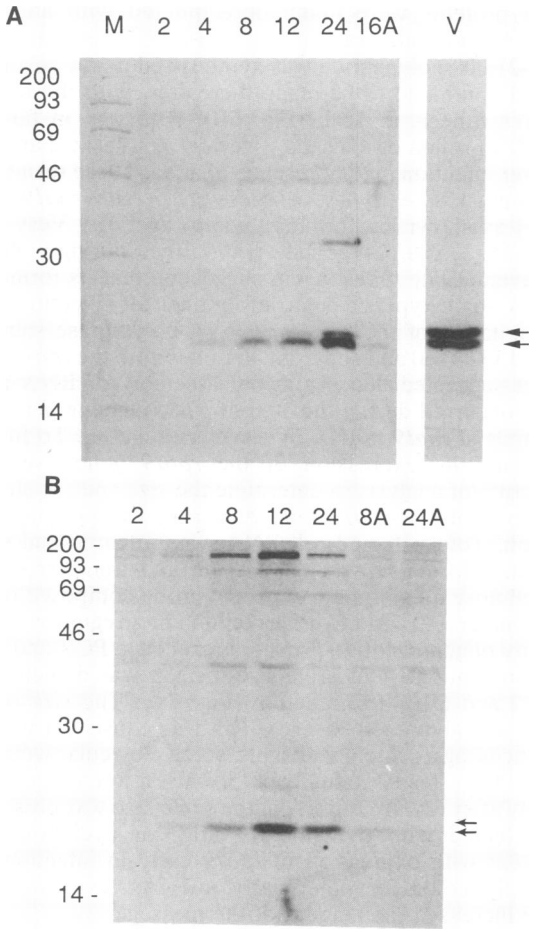


FIG. 6. Time course of *rpo19* protein synthesis. (A) Immunoblot analysis. Cells were harvested at the indicated times after infection, and cytoplasmic proteins were extracted as described previously (43). Aliquots of extract were separated on an SDS–12% polyacrylamide gel. Proteins were transferred to a nitrocellulose membrane and incubated successively with anti-A6R serum and ^{125}I-protein A. Lanes: M, ^{14}C-labeled protein markers, with sizes in kilodaltons shown on the left; 2 through 24, hours after infection; 16A, 16 h after infection in the presence of araC; V, approximately 5×10^{10} purified virion particles directly lysed in gel loading buffer. (B) In vivo pulse-labeling of the Rpo19 protein. Monolayer culture of HeLa cells were incubated for 30 min with [^{35}S]methionine starting at the indicated hour after infection. Lanes: 2 through 24, hours after infection; 8A and 24A, infection in the presence of araC for 8 and 24 h, respectively. Cells were lysed and the labeled proteins were analyzed by immunoprecipitation with anti-A6R serum. An autoradiograph of an SDS–12% polyacrylamide gel is shown. On the left are molecular masses (in kilodaltons) of ^{14}C-labeled protein markers. Arrows on the right indicate the two *rpo19* subunit polypeptides.

antiserum for an immunoblot analysis of fractions from a glycerol density gradient sedimentation of chromatographically purified vaccinia virus RNA polymerase. The autoradiograph indicated that the antiserum reacted with polypeptides of M_r 21,000 and 22,000 (Fig. 4B) that cosedimented with the peak of the RNA polymerase activity (Fig. 4A). Individual fractions, however, showed differences in the ratios of the two polypeptides suggesting heterogeneity in the subunit composition of RNA polymerase molecules. Polypeptides of M_r 21,000 and 22,000 were also seen along with several other subunit proteins on a replica immunoblot probed with antibody to the whole RNA polymerase (Fig. 4C) as well as on a silver-stained SDS-polyacrylamide gel of purified RNA polymerase (Fig. 4D). Analysis of glycerol gradient fractions of total soluble virion extracts prior to chromatographic purification, with the A6R-specific and the whole RNA polymerase antibodies, indicated that all of the packaged M_r-21,000 and -22,000 polypeptides were associated with the RNA polymerase (data not shown).

Protein sequence analysis. Since the M_r-21,000 and M_r-22,000 polypeptides associated with purified RNA polymerase reacted with antiserum that was raised to a gel-purified protein expressed from a cloned ORF, it seemed likely that they were products of the same gene. A less likely possibility, however, was that one of the two polypeptides cross-reacted with the antiserum to the A6R protein. Therefore, we wanted to confirm the identities of the two proteins by peptide sequencing. Purified RNA polymerase was resolved by electrophoresis on a preparative 5 to 15% linear gradient polyacrylamide–SDS gel, and the resolved subunit proteins were transferred to a nitrocellulose membrane. After staining, strips containing the M_r-21,000 and M_r-22,000 bands were cut out. Proteins on the strips were digested in situ with trypsin, and the resultant peptides were separated by microbore reverse-phase HPLC on a C_{18} column. Very similar chromatographic patterns were obtained from the two proteins, and most of their major tryptic peptides were eluted from the column at identical times (Fig. 5). Among several peaks that showed identical retention times between the two proteins, we chose peaks a and c from the M_r-22,000 protein and peaks b and d from the M_r-21,000 protein for Edman degradation. The sequence NGELLSVNDFDKDGLKF HLDYIIK, which matches a segment of the A6R ORF (Fig. 2, amino acid sequence connected by dashes), was obtained from both peaks a and b. Another sequence, GRLPLVS EFSDETMKQNML, that also matches part of the A6R ORF sequence was derived from peaks c and d. These results, combined with the enzyme purification and immunochemical evidence, led us to conclude that both RNA polymerase-associated polypeptides are encoded by the same gene, which we named *rpo19* according to the predicted molecular weight (M_r 18,996) of the protein encoded. In an attempt to understand the size difference between the two proteins, we chose for microsequencing two additional tryptic peptides (peak e from the M_r-22,000 protein and peak f from the M_r-21,000 protein) because they seemed somewhat anomalous in either their retention times or relative abundance. Sequences HIGNHISALK (peak e) and IVES ASTHIEDAHSNLK (peak f) were obtained, results again consistent with the *rpo19* sequence yet providing no information on possible modification of the proteins.

In vivo synthesis of the RNA polymerase subunit. To investigate the in vivo synthesis of the *rpo19* subunit, proteins were extracted from infected cells at different time after virus inoculation and analyzed by immunoblotting with the anti-A6R serum (Fig. 6A). The M_r-21,000 polypeptide was detected at 4 h after infection and increased in amount up to 24 h. By contrast, the M_r-22,000 polypeptide was prominent only late in infection and even at those times was less abundant than the more rapidly migrating one. An additional M_r-35,000 polypeptide was detected at late times. The intensity of this cross-reactive band, however, varied with the blocking agents used for the immunoblotting. When cells were infected in the presence of araC, an inhibitor of viral DNA replication, the amount of the M_r-21,000 polypeptide at 16 h was comparable to that at 4 h in a normal

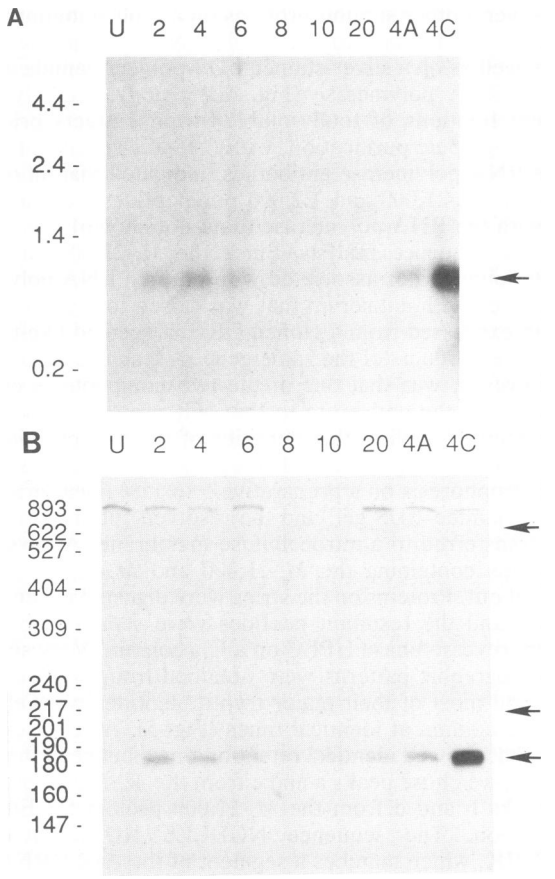


FIG. 7. Analysis of the *rpo19* transcript. (A) Northern blot analysis. Total RNA (20 μg) was isolated from cells at various times after infection and resolved on a formaldehyde–1% agarose gel. Lanes: U, uninfected cell; 2 through 20, hours after infection; 4A, 4 h after infection in the presence of araC; 4C, 4 h after infection in the presence of cycloheximide. The resolved RNAs were transferred to a nitrocellulose membrane and hybridized with a ^{32}P-labeled complementary DNA probe (10^8 dpm/μg of DNA). The arrow indicates the 0.75-kb discrete-size *rpo19* transcript. Sizes of marker RNAs (BRL-Life Technologies) in kilobases are shown on the left. (B) Determination of the 5′ ends of RNA by nuclease S1 analysis. RNA (20 μg) was hybridized with a 5′-end-labeled complementary probe and digested with nuclease S1. The digestion products were ethanol precipitated and analyzed on a 7 M urea–5% polyacrylamide gel. Lane designations are the same as for panel A. Shown on the left are the sizes (in bases) of *Msp*I-digested pBR322 DNA markers, endlabeled with ^{32}P (the size of the 893-base full-length probe is indicated at the top). Arrows on the right indicate the three nuclease-digested bands.

infection and the M_r-22,000 polypeptide was not detected at all. These results suggested that the more rapidly migrating polypeptide was synthesized early and continued to be made throughout the infection, whereas the more slowly migrating one formed at late times. Nearly equal amounts of the two proteins were present in mature virions (Fig. 6A, lane V; see also Fig. 4B, C, and D), suggesting either that some modification is coupled with virus packaging or that RNA polymerase containing the M_r-22,000 polypeptide is preferentially packaged.

To further investigate the time of synthesis of the two polypeptides, infected cells were pulse-labeled for 30 min with [^{35}S]methionine at various times after infection and the labeled proteins were immunoprecipitated with anti-A6R serum. An autoradiogram of the SDS-PAGE confirmed that the M_r-21,000 polypeptide was synthesized at 4 h postinfection and continued to be made thereafter, with the highest rate of synthesis at 12 h (Fig. 6B). Although in far less amount, a protein of the same size was synthesized at 8 and 24 h after infection in the presence of araC. These results are consistent with early/late expression of the *rpo19* gene product as shown by the immunoblot analysis. Very little labeled M_r-22,000 polypeptide was formed during a 30-min pulse even at late times, again suggesting that its formation involves some type of posttranslational modification. The coprecipitation of the two large RNA polymerase subunits (147 and 132 kDa) and some smaller ones with the *rpo19* gene product suggested that significant assembly of RNA polymerase occurred during the 30-min pulse-labeling period.

Analysis of *rpo19* mRNA by Northern blotting. To further investigate the expression of the *rpo19* gene, we used Northern blot analysis to determine the size and the steady-state level of the *rpo19* mRNA during various stages of infection. Total infected-cell RNA was prepared and separated on a formaldehyde-agarose gel and transferred to a nitrocellulose membrane. A specific probe composed of the complemetary strand of the protein-coding region was prepared by oligonucleotide-primed asymmetric PCR and used for hybridization. A discrete-size mRNA was detected between 2 and 6 h after infection (Fig. 7A). The size of the mRNA was about 750 bases, which was consistent with an ORF encoding a protein with a predicted molecular weight of 19,000. The steady-state level of this mRNA decreased abruptly after 6 h. An mRNA of the same size was present in cells infected with the inhibitor araC or cycloheximide, consistent with expression of *rpo19* early in infection. As commonly noted for other early mRNAs, cycloheximide greatly increased the amount of the message.

Neither the 750-nucleotide band or another of discrete size was detected by Northern blot analysis using RNA obtained later than 6 h after infection. From 4 h on, however, an RNA smear ranging up to 4 kb in size was noted. The absence of the smear with uninfected-cell RNA or with infected-cell RNA made early or in the presence of araC or cycloheximide suggests that the probe did hybridize to specific late RNAs containing at least part of the *rpo19* coding region. Translation of these heterogeneous RNAs late in infection would account for the continuous synthesis of the *rpo19* protein described above. Northern blot smears have been found for other late mRNAs and are believed to result mainly from heterogenous sites of transcriptional termination.

Analysis of the 5′ end of *rpo19* mRNA. The 5′ end of the *rpo19* transcript was determined by two methods: nuclease S1 analysis and primer extension. For nuclease S1 analysis, a single-stranded DNA probe complementary to the sequence from −713 to +180 (+1 at the ATG start codon) was end labeled at nucleotide +180. A major band of protected probe, approximately 180 bases in length, was obtained with RNA isolated in the presence of cycloheximide and araC (Fig. 7B). In the absence of drugs, the band was predominant with RNA harvested at 2, 4, and 6 h after infection but was detectable at all later time points as well. Two additional faint bands of 660 and 220 bases were also seen with late RNA; these will be discussed below. The size of the 180-base protected probe corresponds to a transcript with a 5′ end at or a few bases upstream of the translation initiation codon of the A6R ORF within the sequence TAAATG (Fig. 2). In this regard, we noted that the sequence AAAAT TGAATTGCGA, located 12 to 27 bases upstream of the

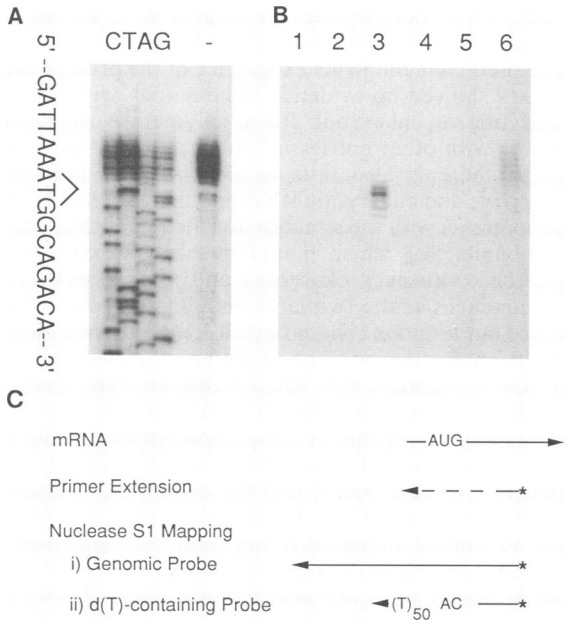


FIG. 8. Analysis of the 5′ poly(A) leader of the *rpo19* early transcript. (A) Primer extension. Early RNA, prepared from cells at 4 h after infection in the presence of cycloheximide, was hybridized with a 5′-end-labeled primer, and the primer was elongated by avian myeloblastosis virus reverse transcriptase in the presence of ddG (lane C), ddA (lane T), ddT (lane A), or ddC (lane G) or without any dideoxynucleotide (lane −). Reactions were separated on a 5% polyacrylamide sequencing gel, and an autoradiograph was taken. The DNA sequence around the ATG codon of the *rpo19* gene is shown on the left. (B) Nuclease S1 protection analysis. Early RNA was hybridized with either a 5′-end-labeled complementary genomic probe (lane 3) or a poly(dT)-containing probe (lane 6) and treated with nuclease S1 (see panel C). The digested material was analyzed on a 7 M urea–5% polyacrylamide sequencing gel. As controls, labeled probes prior to the nuclease digestion (lanes 1 and 4) and probes digested after hybridization with 20 μg of yeast tRNA (lanes 2 and 5) were loaded on the same gel. In lanes 1 and 4, the 893- and 230-base full-length probes, respectively, are not seen because the upper part of the autoradiograph was cut off. (C) Diagram of the primer and probes. Asterisks and arrowheads indicate labeled 5′ ends and polarity of strands, respectively. The 5′ end of the primer corresponds to nucleotide 4440 (180 bases downstream of the ATG codon). The complementary genomic probe contained sequences from −713 to +180 (numbering starts at the ATG codon). The poly(dT)-containing probe contains 50 thymidylate residues followed by completementary genomic sequences from +1 to +180.

ATG codon, resembled the consensus (AAAAATGAAA AAAAA) of viral early promoters (15). It was previously demonstrated that transcription initiates 7 to 15 bases downstream of this early promoter consensus sequence. TAAAT is an essential element of late promoters (8, 16, 20), consistent with the continued synthesis of RNA starting at this site.

The faint 660-base band, detected by nuclease S1 analysis only with RNA isolated at late times (Fig. 7B), corresponded to the start of an RNA 480 bases upstream of the A6R ORF (near nucleotide 3780 in Fig. 2). A TAAAT motif occurs at this location, but it lies within the A5L ORF. The significance of such an RNA start is not evident; although there are two successive ATG sequences downstream (nucleotide 3893 in Fig. 2), they are followed almost immediately by a stop codon. Translation initiation at the A6R ORF of such polycistronic RNA that starts 480 bases upstream, however, could account for the late synthesis of the Rpo19 protein.

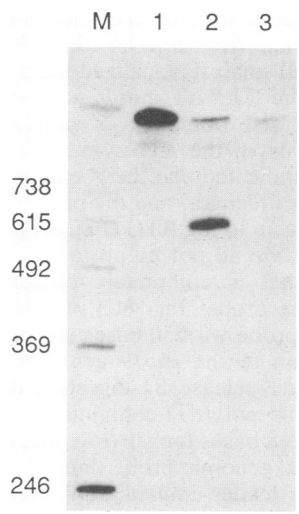


FIG. 9. Nuclease S1 analysis of the 3′ end of *rpo19* transcripts. Early RNA prepared at 4 h after infection in the presence of cycloheximide or late RNA prepared at 8 h after infection was hybridized with a probe that was 3′ end labeled at nucleotide 4357 and digested with nuclease S1. Nuclease-treated material was ethanol precipitated and analyzed by denaturing PAGE on a 4% gel. Lanes: M, DNA size markers (123-bp ladder; BRL-Life Technologies), with sizes in bases indicated on the left; 1, undigested probe of 1,754 bases (from the labeled *SpeI* site at nucleotide 4,357 to the *BglII* site at nucleotide 6110); 2, probe hybridized with early RNA; 3, probe hybridized with late RNA.

The faint 220-base nuclease S1 band was detected only with RNA harvested at late times (Fig. 7B). The size of this band could correspond to an RNA starting 40 bases upstream of the ATG codon (near nucleotide 4220 in Fig. 2). However, the 220-base band is probably an artifact resulting from RNA-RNA hybridization. A TAAAT motif is present on the complementary DNA strand, and nuclease S1 analysis with an appropriate probe indicated that a strong late RNA for the A5L ORF starts there (data not shown). Hybridization of an RNA that starts upstream of the A6R ORF to the complementary A5L transcript and to the labeled single-stranded DNA probe could lead to formation of a triple complex which would produce the observed 220-base labeled band upon nuclease digestion. In agreement with such an interpretation, the amount of the 220-base band decreased when a double-stranded DNA probe was used instead of a single-stranded one (data not shown). Evidently, the unlabeled strand of probe hybridized to the A5L transcript.

Early RNA start sites, upstream of the late RNA start site of the A5L ORF and complementary to the A6R transcript, also were detected (data not shown). Formation of networks by hybridization of A5L and A6R RNAs could lead to the selection of some mRNAs encoding the *rpo19* subunit by the DNA cloned in pHABam 4 (Fig. 1).

In summary, the RNA 5′-end analysis is most compatible with the presence of an overlapping dual early/late promoter resulting in transcription initiation within the TAAATG sequence before and after DNA replication. However, additional longer RNAs also may contribute to late expression of the *rpo19* gene.

Evidence for a short 5′ poly(A) leader on *rpo19* mRNA. The TAAATG motif, commonly found at the start of most late genes (20, 35) and some early ones (3, 22), includes both the RNA start site and the translation initiation codon. The motif

has been associated with heterogeneous-length 5′ poly(A) leaders of vaccinia virus mRNA (3, 4, 9, 16, 22, 30, 38). Since nuclease S1 analysis localized the 5′ end of the *rpo19* mRNA within the TAAAT motif, we considered that this transcript might also contain a 5′ poly(A) leader. Primer extension analysis of the RNA, made in the presence of cycloheximide, indicated that the 5′ end of the *rpo19* mRNA is heterogeneous in length, and the precise sequences could be identified only up to the ATG (Fig. 8A and C). Since such a poly(A) leader would not be protected from nuclease S1 with a probe that is completely complementary to the genome (genomic probe; Fig. 8C), we designed a complementary strand probe with 50 bases of poly(dT) attached to the coding region of the *rpo19* gene [poly(dT)-containing probe; Fig. 8C]. Nuclease S1 digestion of the early RNA hybridized to this poly(dT)-containing probe yielded protected bands 1 to 8 bases longer (Fig. 8B, lane 6) than those obtained with the genomic probe (lane 3), indicating a 4- to 11-base poly(A) leader immediately preceding the AUG codon.

Analysis of the 3′ ends of the *rpo19* mRNA. A single-stranded DNA probe, end labeled at nucleotide 4357 (Fig. 2), was used to determine the 3′ ends of *rpo19* transcripts by nuclease S1 analysis. Discrete 3′ ends were detected when early RNA was used (Fig. 9). The ends of the early RNAs were mapped 10 to 20 bp downstream of the sequence TTTTTCT (at nucleotide 4980; Fig. 2). The sequence motif TTTTTNT, in which N can be any nucleotide, was characterized as a general transcription termination signal for vaccinia virus early RNAs (44). The locations of the 5′ and 3′ ends of the early transcript, 730 nucleotides apart as determined by nuclease S1 analyses, are consistent with the 750-nucleotide size of the RNA determined by Northern blotting. In contrast to the result obtained with early mRNA, no bands indicative of discrete 3′ ends were detected with late RNA (Fig. 9). Extreme heterogeneity of 3′ ends would explain the smears seen in Northern blot analysis of late RNAs with use of the *rpo19* probe.

DISCUSSION

The gene *rpo19* was identified and shown to encode a small subunit of the vaccinia virus DNA-dependent RNA polymerase. Our initial localization of the gene depended on (i) selection of viral mRNAs that hybridized to cloned segments of the vaccinia virus genome, (ii) translation of these mRNAs in vitro, (iii) immunoaffinity purification of the translation products with antiserum to the RNA polymerase, and (iv) determination of the size of the antibody-bound polypeptides by SDS-PAGE. Sequencing of the positive DNA fragments revealed candidate ORFs, which were then cloned and transcribed in vitro. The translation product of the mRNA derived from the A6R ORF bound to the RNA polymerase antiserum and migrated as a single band with an apparent M_r of 21,000, only slightly greater than the molecular weight of 18,996 predicted from the sequence. To confirm the identification of *rpo19*, a recombinant M_r-21,000 protein was made in *E. coli* and used to prepare antiserum. The antiserum reacted with an M_r-21,000 protein of purified RNA polymerase as well as with one that had a slightly slower electrophoretic mobility. Both polypeptides were demonstrated to be products of the A6R ORF by tryptic peptide analysis and amino acid sequencing. At this time, we have not determined the structural basis for the differences in electrophoretic mobilities of the two polypeptides. Thus, of the 8 to 10 polypeptides associated with vaccinia virus RNA polymerase, two are encoded by *rpo19*.

The predicted amino acid sequence of the protein encoded by *rpo19* showed no evidence of homology with the known prokaryotic or eukaryotic RNA polymerase small subunit genes or with other entries in biological data bases. In this respect, there is also little or no similarity between the prokaryotic and eukaryotic RNA polymerase small subunits. This contrasts with the situation for the large RNA polymerase subunits, for which homology clearly exists between those of poxviruses, prokaryotes, and eukaryotes (6, 11, 31). The sequences of the two large subunits may be conserved to carry out functions shared by all RNA polymerases, such as RNA chain elongation and binding of nucleotides, DNA, and RNA, whereas the small subunits may have more specialized roles accounting for their diverse primary sequences among organisms. One apparent exception to this rule is the similarity between the small vaccinia virus RNA polymerase subunit *rpo30* and the eukaryotic transcription elongation factor SII (2). In addition, other subunits may share functional domains if not primary sequences. For example, near the N-terminal segment of the *rpo19* subunit, 21 of 37 amino acids are acidic. Similar stretches of acidic residues are present in some eukaryotic RNA polymerase small subunits (26).

The previously identified RNA polymerase subunits are synthesized throughout infection (2, 6, 11, 31) and therefore may participate in the expression of all three regulatory classes of vaccinia virus genes. The phenotype of some RNA polymerase subunit temperature-sensitive mutants indicates their involvement in viral late gene expression (17, 21, 40). The antibody raised to the recombinant product of the A6R ORF allowed us to demonstrate that synthesis of the *rpo19* subunit also occurred before and after DNA replication. Continued expression of *rpo19* is accomplished at least in part by an overlapping early/late promoter. There was an apparent discrepancy, however, between the amount of mRNA and protein. Nuclease S1 analysis demonstrated that the specifically initiated mRNAs were most abundant at 2 to 4 h after infection, whereas pulse-labeling with [^{35}S]methionine indicated that the protein was made at the highest rate at 12 h. We considered several possibilities to explain this difference: (i) late transcripts may have been underestimated by nuclease S1 analysis because of the presence of complementary antisense RNAs at late times (reference 10 and references therein); (ii) late RNA start sites were found far upstream of the *rpo19* gene, and such RNAs might serve as polycistronic messages; and (iii) late mRNAs may be translated more efficiently than early ones. The size heterogeneity of vaccinia virus late mRNAs contributes to the difficulty in studying them. This heterogeneity was also true for late *rpo19* RNA, which appeared as a smear by Northern blotting and had no discrete 3′ ends. In contrast, early *rpo19* mRNA appeared as a distinct band on Northern blots, and the location of the 3′ end, as determined by nuclease S1 analysis, was consistent with termination downstream of the TTTTTNT signal (44). The 5′ ends of the *rpo19* mRNAs have a short poly(A) leader that apparently arises by RNA polymerase slippage when transcribing the complement of the TAAAT motif present at the start of some early genes (3, 22). Thus far, no specific function has been associated with the untranslated poly(A) leader sequence.

When crude lysates of virus particles were applied to glycerol gradients, Western immunoblotting of the fractions with anti-A6R serum indicated that the immunoreactive 21- and 22-kDa polypeptides sedimented exclusively with the

multisubunit RNA polymerase. Furthermore, the two subunits were present in similar amounts. We do not yet have accurate numbers for the stoichiometry of the RNA polymerase subunits, partly because of the chromatographic microheterogeniety of the enzyme. The M_r-21,000 polypeptides appeared to be the primary *rpo19* gene product, since it comigrated in SDS-PAGE with the product of the A6R ORF made in *E. coli* and in reticulocyte lysates and is also the major labeled form present after a 30-min pulse with [^{35}S]methionine at early and late times after infection. Western blot analysis indicates that the M_r-22,000 form appears late in infection but is always much less abundant than the M_r-21,000 form when the total cytoplasm is examined. Therefore, the substantial amounts of the M_r-22,000 form present in virions may result either from preferential packaging or from modification during or after packaging. In this regard, a protein kinase is present within the core of vaccinia virus virions (25a), and there is precedent for the phosphorylation of the 190- and 18-kDa subunits of yeast RNA polymerase II (25b). Interestingly, the serine residue at position 78 of the *rpo19* polypeptide is within an Arg-Arg-Xaa-Ser cyclic AMP-dependent protein kinase substrate recognition motif (13a). Further studies are needed to determine the nature of the M_r-22,000 polypeptide and to ascertain whether it has a regulatory role.

The genome of vaccinia virus contains about 200 genes, and these genes are expressed in a highly programmed fashion. One might have anticipated, therefore, that the vaccinia virus RNA polymerase would have a simpler subunit composition than does eukaryotic RNA polymerases. This is not the case, however. Yeast RNA polymerase II has recently been shown to contain subunits that are expressed from 10 different genes (42). The present study brings the number of vaccinia virus RNA polymerase genes to seven, and there are at least two more that remain to be identified. Thus, poxvirus and eukaryotic RNA polymerases are of similar complexity.

ACKNOWLEDGMENTS

We thank Norman Cooper for preparing cells and virus, Jerry Sisler for synthesizing oligonucleotides, Elaine V. Jones for providing the initial *rpo19* gene localization data, Bernard Amegadzie and James Keck for critical reading of the manuscript, and Paul Gershon and Rafael Blasco for assistance with the computer analysis of sequences.

REFERENCES

1. **Aebersold, R. H., J. Leavitt, R. A. Saavedra, L. E. Hood, and S. B. H. Kent.** 1987. Internal amino acid sequence analysis of proteins separated by one or two dimensional gel electrophoresis after in situ protease digestion on nitrocellulose. Proc. Natl. Acad. Sci. USA **84**:6970–6974.
2. **Ahn, B.-Y., P. D. Gershon, E. V. Jones, and B. Moss.** 1990. Identification of *rpo30*, a vaccinia virus RNA polymerase gene with structural similarity to an eucaryotic transcription elongation factor. Mol. Cell. Biol. **10**:5433–5441.
3. **Ahn, B.-Y., E. V. Jones, and B. Moss.** 1990. Identification of the vaccinia virus gene encoding an 18-kilodalton subunit of RNA polymerase and demonstration of a 5′ poly(A) leader on its early transcript. J. Virol. **64**:3019–3024.
4. **Ahn, B.-Y., and B. Moss.** 1989. Capped poly(A) leader of variable lengths at the 5′ ends of vaccinia virus late mRNAs. J. Virol. **63**:226–232.
5. **Amegadzie, B. Y., B.-Y. Ahn, and B. Moss.** 1991. Identification, sequence, and expression of the gene encoding a Mr 35,000 subunit of the vaccinia virus DNA-dependent RNA polymerase. J. Biol. Chem. **266**:13712–13718.
6. **Amegadzie, B. Y., M. H. Holmes, N. B. Cole, E. V. Jones, P. L. Earl, and B. Moss.** 1991. Identification, sequence, and expression of the gene encoding the second-largest subunit of the vaccinia virus DNA-dependent RNA polymerase. Virology **180**: 88–98.
7. **Baroudy, B. M., and B. Moss.** 1980. Purification and characterization of a DNA-dependent RNA polymerase from vaccinia virions. J. Biol. Chem. **255**:4372–4380.
8. **Bertholet, C., P. Stocco, E. Van-Meir, and R. Wittek.** 1986. Functional analysis of the 5′ flanking sequence of a vaccinia virus late gene. EMBO J. **5**:1951–1957.
9. **Bertholet, C., E. Van Meir, B. ten Heggeler-Bordier, and R. Wittek.** 1987. Vaccinia virus produces late mRNAs by discontinuous synthesis. Cell **50**:153–162.
10. **Boone, R. F., R. P. Parr, and B. Moss.** 1979. Intermolecular duplexes formed from polyadenylated vaccinia virus RNA. J. Virol. **30**:365–374.
11. **Broyles, S. S., and B. Moss.** 1986. Homology between RNA polymerases of poxviruses, prokaryotes, and eukaryotes: nucleotide sequence and transcriptional analysis of vaccinia virus genes encoding 147-kDa and 22-kDa subunits. Proc. Natl. Acad. Sci. USA **83**:3141–3145.
12. **Broyles, S. S., and M. J. Pennington.** 1990. Vaccinia virus gene encoding a 30-kilodalton subunit of the viral DNA-dependent RNA polymerase. J. Virol. **64**:5376–5382.
13. **Broyles, S. S., L. Yuen, S. Shuman, and B. Moss.** 1988. Purification of a factor required for transcription of vaccinia virus early genes. J. Biol. Chem. **263**:10754–10760.

13a. **Cherry, J. R., T. R. Johnson, C. Dollard, J. R. Shuster, and C. L. Denis.** 1989. Cyclic AMP-dependent protein kinase phosphorylates and inactivates the yeast transcriptional activator ADR1. Cell **56**:409–419.

14. **Cooper, J. A., and B. Moss.** 1979. In vitro translation of immediate early, early and late classes of RNA from vaccinia virus infected cells. Virology **96**:368–380.
15. **Davison, A. J., and B. Moss.** 1989. The structure of vaccinia virus early promoters. J. Mol. Biol. **210**:749–769.
16. **Davison, A. J., and B. Moss.** 1989. The structure of vaccinia virus late promoters. J. Mol. Biol. **210**:771–784.
17. **Ensinger, M. J.** 1987. Phenotypic characterization of temperature-sensitive mutants of vaccinia virus with mutations in a 125,000-M_r subunit of the virion-associated DNA-dependent RNA polymerase. J. Virol. **61**:1842–1850.
18. **Gershon, P. D., and B. Moss.** 1990. Early transcription factor subunits are encoded by vaccinia virus late genes. Proc. Natl. Acad. Sci. USA **87**:4401–4405.
19. **Goebel, S. J., G. P. Johnson, M. E. Perkus, S. W. Davis, J. P. Winslow, and E. Paoletti.** 1990. The complete DNA sequence of vaccinia virus. Virology **179**:247–266.
20. **Hanngi, M., W. Bannwarth, and H. G. Stunnenberg.** 1986. Conserved TAAAT motif in vaccinia virus late promoters: overlapping TATA box and site of transcription initiation. EMBO J. **5**:1071–1076.
21. **Hooda-Dhingra, U., C. L. Thompson, and R. C. Condit.** 1989. Detailed phenotypic characterization of five temperature-sensitive mutants in the 22- and 147-kilodalton subunits of vaccinia virus DNA-dependent RNA polymerase. J. Virol. **63**:714–729.
22. **Ink, B. S., and D. J. Pickup.** 1990. Vaccinia virus directs the synthesis of early mRNAs containing 5′ poly(A) sequences. Proc. Natl. Acad. Sci. USA **87**:1536–1540.
23. **Joklik, W. K.** 1962. The preparation and characterization of highly purified radioactivity labeled poxvirus. Biochim. Biophys. Acta **61**:290–301.
24. **Jones, E. V., and B. Moss.** 1984. Mapping of the vaccinia virus DNA polymerase gene by marker rescue and cell-free translation of selected mRNA. J. Virol. **49**:72–77.
25. **Jones, E. V., C. Puckett, and B. Moss.** 1987. DNA-dependent RNA polymerase subunits encoded within the vaccinia virus genome. J. Virol. **61**:1765–1771.

25a. **Kleiman, J. H., and B. Moss.** 1975. Purification of a protein kinase and two phosphate acceptor proteins from vaccinia virions. J. Biol. Chem. **250**:2420–2429.

25b. **Kolodziej, P. A., N. Woychik, S.-M. Liao, and R. A. Young.** 1990. RNA polymerase II subunit composition, stoichiometry,

and phosphorylation. Mol. Cell. Biol. **10**:1915–1920.
26. **Mosrin, C., M. Riva, M. Beltrame, E. Cassar, A. Sentenac, and P. Thuriaux.** 1990. The *RPC31* gene of *Saccharomyces cerevisiae* encodes a subunit of RNA polymerase C (III) with an acidic tail. Mol. Cell. Biol. **10**:4737–4743.
27. **Moss, B.** 1990. Regulation of vaccinia virus transcription. Annu. Rev. Biochem. **59**:661–688.
28. **Moss, B., B.-Y. Ahn, B. Amegadzie, P. D. Gershon, and J. G. Keck.** 1991. Cytoplasmic transcription system encoded by vaccinia virus. J. Biol. Chem. **266**:1355–1358.
29. **Nielsen, D. A., and D. J. Shapiro.** 1986. Preparation of capped RNA transcripts using T7 RNA polymerase. Nucleic Acids Res. **14**:5936.
30. **Patel, D. D., and D. J. Pickup.** 1987. Messenger RNAs of a strongly-expressed late gene of cowpox virus contains a 5′-terminal poly(A) leader. EMBO J. **6**:3787–3794.
31. **Patel, D. D., and D. J. Pickup.** 1989. The second-largest subunit of the poxvirus RNA polymerase is similar to the corresponding subunits of procaryotic and eucaryotic RNA polymerases. J. Virol. **63**:1076–1086.
32. **Pearson, W. R., and D. J. Lipman.** 1988. Improved tools for biological sequence comparison. Proc. Natl. Acad. Sci. USA **85**:2444–2448.
33. **Quick, S. D., and S. S. Broyles.** 1990. Vaccinia virus gene D7R encodes a 20,000-dalton subunit of the viral DNA-dependent RNA polymerase. Virology **179**:603–605.
34. **Rosel, J., and B. Moss.** 1985. Transcriptional and translational mapping and nucleotide sequence analysis of a vaccinia virus gene encoding the precursor of the major core polypeptide 4b. J. Virol. **56**:830–838.
35. **Rosel, J. L., P. L. Earl, J. P. Weir, and B. Moss.** 1986. Conserved TAAATG sequence at the transcriptional and translational initiation sites of vaccinia virus late genes deduced by structural and functional analysis of the *Hin*dIII H genome fragment. J. Virol. **60**:436–439.
36. **Rosenberg, A. H., B. N. Lade, D. Chui, S. Lin, J. J. Dunn, and W. Studier.** 1987. Vectors for selective expression of cloned DNAs by T7 RNA polymerase. Gene **56**:125–135.
37. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed., p. 7.43–7.52. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
38. **Schwer, B., P. Visca, J. C. Vos, and H. G. Stunnenberg.** 1987. Discontinuous transcription or RNA processing of vaccinia virus late messengers results in a 5′ poly(A) leader. Cell **50**:163–169.
39. **Staden, R.** 1982. Automation of the computer handling of gel reading data produced by the shotgun method of DNA sequencing. Nucleic Acids Res. **10**:4731–4751.
40. **Thompson, C. L., U. Hooda-Dhingra, and R. C. Condit.** 1989. Fine structure mapping of five temperature-sensitive mutants in the 22- and 147-kilodalton subunits of vaccinia virus DNA-dependent RNA polymerase. J. Virol. **63**:705–713.
41. **Van Meir, E., and R. Wittek.** 1988. Fine structure of the vaccinia virus gene encoding the precursor of the major core protein 4a. Arch. Virol. **102**:19–27.
42. **Woychik, N., and R. A. Young.** 1990. RNA polymerase II: subunit structure and function. Trends Biochem. Sci. **15**:347–351.
43. **Wright, C. F., and B. Moss.** 1987. In vitro synthesis of vaccinia virus late mRNA containing a 5′ poly (A) leader sequence. Proc. Natl. Acad. Sci. USA **84**:8883–8887.
44. **Yuen, L., and B. Moss.** 1987. Oligonucleotide sequence signaling transcriptional termination of vaccinia virus early genes. Proc. Natl. Acad. Sci. USA **84**:6417–6421.