Molecular Characterization of the Vaccinia Virus Hemagglutinin Gene

CHARLES K. BROWN, PETER C. TURNER, AND RICHARD W. MOYER*

Department of Immunology and Medical Microbiology, College of Medicine, Box J-266, JHMC, University of Florida, Gainesville, Florida 32610-0266

Received 3 December 1990/Accepted 22 March 1991

The vaccinia virus hemagglutinin (HA) is a glycoprotein found on the plasma membrane of infected cells and the envelope of extracellular virus. Two forms of HA (85 and 68 kDa) are detected by immunoblot analysis. Although hemagglutination activity is only readily detectable late in infection, the 85-kDa HA appears early and accumulates throughout infection, whereas the 68-kDa form appears only late in the cycle. Production of the 68-kDa HA but not the 85-kDa HA was inhibited by either cytosine arabinoside or rifampin. Analysis of HA gene expression reveals a complex pattern of expression. The HA gene is transcribed early to yield a 1.65-kb dicistronic early transcript, consisting of the 945-bp HA open reading frame (ORF) fused to a 453-bp downstream ORF. Transcription from this site initiates 7 bases upstream of the AUG initiating codon of the HA ORF. Due to the discrepancy between the calculated size of the HA protein (33 kDa) and that reported for the unglycosylated HA protein derived from *in vitro* translation (58 kDa), we placed an early transcription termination signal (TTTTTAT) directly downstream of the 945-bp HA ORF. This led to a reduction in size of the early HA mRNA to 1.2 kb, as expected, but had no effect on the formation of either the 85- or 68-kDa protein. Transcripts originating from the early promoter are found throughout the infection cycle. However, after DNA replication, transcription from a second, late promoter ensues. The transcriptional start site of the late promoter is within a consensus TAAATG sequence located 135 bases upstream of the transcriptional start site of the first promoter. The late transcriptional start site is also found within an upstream ORF.

Vaccinia virus is the prototypic member of the orthopoxviruses and replicates within the cytoplasm of the host cell. The 185-kb genome consists of a linear double-stranded DNA with covalently linked "hairpin" termini and has a coding potential for greater than 150 proteins (3, 9, 25). Early genes are expressed before the onset of DNA replication, and the mRNAs are transcribed, capped, and 3' polyadenylated by virally encoded enzymes packaged within the virion (2, 22, 43). Early gene transcripts, which are generally monocistronic, are terminated downstream of a 3' transcriptional termination sequence (TTTTNT) (15). Cycloheximide, a potent inhibitor of protein synthesis, is commonly used to cause arrest of the infectious cycle, so that only early mRNAs are synthesized. Late gene expression begins at the onset of DNA replication and requires the previous expression of the early proteins. Progression into the late phase of the infectious cycle can be blocked by inhibitors of DNA synthesis, such as cytosine arabinoside (CAR) or hydroxyurea. Late transcripts differ from early transcripts in that in general they do not have a discrete 3' transcription termination signal and are therefore heterogeneous in size and lack discrete 3' ends (6, 18). Besides containing the usual 5' cap and 3' poly(A) tail, late transcripts also contain a 5' poly(A) tract that precedes the encoded transcript (1, 32). The initiation of late transcription usually occurs within the conserved hexanucleotide, TAAATPu, and it is thought that the 5' polyadenylation results from slippage of the RNA polymerase during the transcription initiation process (8, 31). Recently, a third, intermediate-early class of mRNA has been described, whose transcription is dependent on DNA replication but that, unlike late transcripts, does not depend on continuing protein synthesis (41).

After the synthesis of late proteins, morphogenesis leads to mature virion production. Morphogenesis and the proteolytic processing of virus products can be interrupted by the addition of rifampin (21, 26), leading to the accumulation of precursor proteins.

The vaccinia virus hemagglutinin (HA) is a glycoprotein found on both the cytoplasmic membrane of infected cells and the envelope of extracellular virus (26, 27). The HA gene is nonessential, in that viruses not expressing a functional HA product are viable in cell culture. The HA gene, like many other nonessential genes, has been used as a site for insertion of foreign DNA for construction of vaccinia virus expression vectors (34, 39). The HA gene has been considered a prototype for a strongly expressed nonstructural late gene in vaccinia virus. HA activity, as measured by the hemagglutination of chicken erythrocytes, is first detected at about 6 h postinfection, a time somewhat later than that seen for most other late genes (usually about 4 h postinfection) (30). However, our work indicates that the HA gene product appears early in the infectious cycle and is expressed from two separate promoters, one functioning early and a second active only late in infection.

MATERIALS AND METHODS

Cells and viruses. Wild-type vaccinia virus (IHD-J strain) and rabbitpox virus (Utrecht strain) were obtained from the American Type Culture Collection. HeLa and CV-1 cells were maintained as monolayers on Eagle minimal essential medium (GIBCO F-11) supplemented with 10% fetal bovine serum, 2 mM glutamine, 100 U of penicillin per ml, 100 µg of streptomycin per ml, and 0.1 mg of sodium pyruvate per ml.

Antisera. Monospecific anti-HA rabbit antiserum was generously provided by Sam Dales and has been described previously by Shida and Dales (37).

* Corresponding author.

Preparation of cell membranes. Plasma membranes were isolated from infected and uninfected cells by the two-phase polymer method of Brunette and Till (5) as modified by Weintraub and Dales (44). Plasma membranes were prepared for immunoblots as described below.

Preparation of infected and uninfected cell extracts for immunoblots. Confluent monolayers of HeLa cells (4×10^7 cells per 150-mm dish) were infected at a multiplicity of infection (MOI) of 10 PFU of purified virus per cell. The inhibitors CAR (50 $\mu\text{g/ml}$) and rifampin (100 $\mu\text{g/ml}$), used where indicated, were present throughout the infection. Cells were harvested at the appropriate time postinfection by scraping into medium and then centrifuging at $800 \times g$ for 5 min at 4°C . The cell pellets or purified plasma membranes were suspended in 0.5 ml of phosphate-buffered saline (PBS; 0.01 M sodium phosphate and 0.15 M NaCl [pH 7.2]) that contained 5 mM phenylmethylsulfonyl fluoride and 100 K units of aprotinin per ml. The pellet was briefly sonicated before the addition of $2\times$ sample buffer (60 mM Tris-HCl [pH 6.8], 4% [wt/vol] sodium dodecyl sulfate [SDS], 200 mM dithiothreitol, 40% glycerol, 0.12% bromophenol blue). Uninfected cell extracts were prepared in a similar manner.

Immunoblot analysis of infected and uninfected cell extracts. Infected and uninfected cellular extracts or plasma membranes prepared as described above were incubated at 100°C for 2 min and held at 37°C before electrophoresis. The samples (50 μg of protein per well, or as indicated) were resolved on 7.5 or 10% SDS-polyacrylamide gels as described by Moyer and Graves (23). Electrophoresis was for 14 h at room temperature in Tris-glycine buffer (50 mM Tris base, 0.38 M glycine, 0.1% SDS [pH 8.4]) at 50 V. The proteins were transferred to 0.1- μm -pore-size nitrocellulose paper by the method of Towbin et al. (40). Electrophoretic transfer was carried out at 250 mA for 3 h at 4°C in transfer buffer (25 mM Tris-HCl, 192 mM glycine, 20% methanol [pH 8.6]). After transfer was complete, the blots were blocked with a solution of 5% bovine serum albumin and 10% fetal bovine serum in PBS for 90 min at room temperature with agitation. The immunoblots were then washed in Tris-buffered saline (TBS; 10 mM Tris-HCl, 150 mM NaCl [pH 8.0]) and then incubated with the primary antibody overnight at room temperature with moderate agitation. The primary antibody was removed, and the blots were washed three times in TBS for 15 min per wash at room temperature, with moderate agitation. The blots were then incubated with the secondary antibody (goat anti-rabbit affinity-purified immunoglobulin G conjugated to alkaline phosphatase [Cooper/Cappel]) diluted 1:2,000 in TBS. The blots were incubated for 90 min at room temperature, with agitation. The secondary antibody was removed, and the blots were then washed three times with TBS at room temperature for 15 min each wash. The blots were then developed using a freshly prepared solution of Nitro Blue Tetrazolium (50 mg/ml in 70% dimethylformamide in water) and 5-bromo-4-chloro-3-indolyl phosphate (50 mg/ml in 100% dimethylformamide), diluted 1:15 and 1:30, respectively, in 100 mM Tris (pH 9.5)–100 mM NaCl–5 mM MgCl_2 . Finally, the blots were transferred to a solution of 100 mM Tris-HCl (pH 7.0)–10 mM EDTA to stop the development and allowed to air dry.

Virus and DNA purification. Vaccinia virus IHD-J was purified from infected cells by the method of Moyer and Rothe (24). DNA was prepared from purified virus suspended at 2×10^{11} PFU/ml in 10 mM Tris-HCl (pH 8.0)–100 mM NaCl–1% SDS–5 mM EDTA. Proteinase K (Boehringer Mannheim Biochemicals) was added to 250 $\mu\text{g/ml}$, and the suspension was incubated for 1 h at 37°C before extraction

with an equal volume of phenol-chloroform (1:1) three times and chloroform once. DNA was precipitated from the aqueous phase with ethanol and dissolved in 10 mM Tris-HCl (pH 8.0)–1 mM EDTA.

Vaccinia virus RNA. HeLa cells were infected at MOI of 10. Early viral RNA was prepared from cells infected for 4 h in the presence of cycloheximide (100 $\mu\text{g/ml}$). Late viral RNA was prepared from cells harvested at 18 h postinfection in the absence of inhibitors.

RNA was isolated from cells as described by Cox (7). Cells were washed twice with PBS and then lysed with 6 ml of 6 M guanidinium isothiocyanate in 0.1 M sodium acetate (pH 5.0) per 150-mm dish. The lysate from each dish was passed through a 20-gauge needle twice and through a 22-gauge needle once and then layered onto a 2-ml cushion of 5.7 M CsCl and centrifuged at 36,000 rpm overnight at 16°C in a Beckman SW41 rotor. The RNA pellet was dissolved in 10 mM Tris-HCl (pH 7.4)–5 mM EDTA–1% SDS, recovered by ethanol precipitation, and stored at -20°C .

Northern (RNA) blot analyses. Samples (10 μg) of total infected cellular RNA were denatured in 2.2 M formaldehyde–50% deionized formamide–0.5 $\times$ MOPS (morpholinepropanesulfonic acid) buffer (20 mM MOPS [pH 7.0], 0.5 mM EDTA, 5 mM sodium acetate) at 70°C for 2 min and then subjected to electrophoresis through a 1.4% agarose gel in MOPS buffer containing 2.2 M formaldehyde. RNAs were transferred onto nitrocellulose membranes by capillary blotting. The membranes were baked at 80°C for 1 h and prehybridized for 2 h at 42°C in 50% deionized formamide–5 $\times$ Denhardt solution (100 $\times$ Denhardt solution is 2% [wt/vol] each Ficoll, polyvinylpyrrolidone, and bovine serum albumin [Pentax fraction V])–5 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS–50 mM sodium phosphate (pH 6.5)–100 μg of calf thymus DNA per ml. Hybridizations with DNA and RNA probes were performed overnight at 42 and 55°C , respectively, in 50% deionized formamide–5 $\times$ SSC–20 mM sodium phosphate (pH 6.5)–1 $\times$ Denhardt solution–100 μg of calf thymus DNA per ml–5% dextran sulfate. Filters were washed twice in 2 $\times$ SSC–0.1% SDS and twice in 0.1 $\times$ SSC–0.1% SDS at room temperature for 15 min each before drying and exposing to X-ray film.

^{32}P -labeled HA-specific DNA probes were prepared by oligonucleotide extension (11). Random oligohexanucleotides were annealed to 100 ng of DNA to be labeled, and 2 U of DNA polymerase (large Klenow fragment from *Escherichia coli*, New England BioLabs) was added. Incubation was at room temperature for 1 h in the presence of 100 μM dATP, 100 μM dGTP, 100 μM dTTP, and 50 μCi of [α - ^{32}P]dCTP (800 Ci/mmol) in a final volume of 25 μl .

^{32}P -labeled single-stranded RNA hybridization probes were prepared as described by Melton et al. (20). DNA fragments containing coding sequences for the HA, p16, or p17 open reading frame (ORF), were cloned into the multiple cloning site of pBluescript KS(+) (Stratagene), and transcripts were synthesized in vitro from the T7 or T3 promoter. Transcription was at room temperature for 1 h in 40 mM Tris-HCl (pH 8.0)–8 mM MgCl_2 –2 mM spermidine–50 mM NaCl–30 mM dithiothreitol–0.4 mM ATP–0.4 mM GTP–0.4 mM CTP–2.4 μM [γ - ^{32}P]UTP (800 Ci/mmol) with 1 μg of plasmid DNA template and 10 U of T7 or T3 RNA polymerase (Stratagene) in a final volume of 25 μl .

Primer extension analysis and RNA sequencing. The oligonucleotide primer was 5' end labeled with ^{32}P by the method of Weaver and Weissman (42). Labeled primer was then annealed to 35 μg of total cellular RNA from vaccinia virus

IHD-J-infected cells as described by Lane et al. (16) and extended at 42°C for 1 h in a 15- μ l reaction containing 50 mM deoxynucleoside triphosphates, 75 mM KCl, 50 mM Tris-HCl (pH 8.3), 10 mM MgCl₂, 10 mM dithiothreitol, and 400 U of avian myeloblastosis virus reverse transcriptase (Life Sciences) per ml. Products were resolved on 8% polyacrylamide gels containing 7 M urea (19) and visualized by autoradiography.

RNA sequencing was performed as for primer extension, except that an appropriate dideoxynucleoside triphosphate was added to each of four separate reactions. The deoxynucleotide triphosphate/dideoxynucleotide triphosphate ratios in each termination reaction were 5.5, 1.4, 1.4, and 2.7 for T, A, G, and C reactions, respectively.

Construction of recombinant viruses. Recombinant vaccinia virus containing the *E. coli* xanthine-guanine phosphoribosyl transferase (*gpt*) gene was constructed by a method modified from that of Mackett et al. (17). The recombinant virus was selected by using mycophenolic acid resistance as described by Boyle et al. (4) and Falkner et al. (10). CV-1 cells (5×10^6 cells) in 60-mm dishes were infected with vaccinia virus IHD-J at an MOI of 0.05. At 2 h postinfection, the medium was removed and the cells were transfected with a mixture of 10 μ g each of calcium phosphate-precipitated plasmid DNA and vaccinia virus genomic DNA. After 30 min of incubation at room temperature, 5 ml of culture medium was added, and the cells were incubated at 37°C. After 3 h of incubation, the culture medium was removed and replaced with fresh medium containing 2.5 μ g of mycophenolic acid per ml, 15 μ g of hypoxanthine per ml, and 250 μ g of xanthine per ml; the culture was then incubated at 37°C for an additional 48 h. Confluent monolayers of CV-1 that had been pretreated with mycophenolic acid (2.5 μ g/ml) for 12 h were infected with virus from the transfections. After incubation for 2 h, infected monolayers were overlaid with *gpt* selection medium (GIBCO F-11, 5% fetal bovine serum, 2.5 μ g of mycophenolic acid per ml, 15 μ g of hypoxanthine per ml, and 250 μ g of xanthine per ml) containing 1% purified agar (Oxoid). After 2 days of incubation at 37°C, the cells were stained with neutral red (GIBCO) to visualize plaques.

Plasmid DNA clones. pHGN-1 is pUC18 containing a 1.8-kb *Sall*-*Hind*III fragment including the HA gene derived from the right end of the *Hind*III A fragment of vaccinia virus IHD-J (3a, 34). Plasmid pTK61-*gpt* contains the *E. coli gpt* gene fused to the vaccinia virus early-late P_{7.5} promoter as a 2.1-kb *Eco*RI fragment (*eco-gpt*) (10).

Plasmid pHA535 contains a 535-bp *Hpa*II DNA fragment from pHGN-1. This insert fragment spans from position 622 of the HA coding sequence to 212 bp downstream of the HA translation stop codon (34). The 535-bp insert was subcloned into *Xba*I-restricted pBluescript KS(+), after both ends of the vector and the insert were flushed with the *E. coli* DNA polymerase I large fragment and dNTPs, and then blunt end ligated with T4 DNA ligase (New England BioLabs). pHA535-*gpt* was constructed by inserting the 2.1-kb *eco-gpt* DNA fragment into the unique *Xba*I restriction site of pHA535 by blunt end ligation.

RESULTS

Expression of the vaccinia virus IHD-J HA in infected cells. In agreement with earlier studies (14), we have found that the appearance of the HA protein as measured by the ability of infected cell extracts to agglutinate chicken erythrocytes is first detectable about 6 to 7 h postinfection (data not shown). These observations are consistent with the sugges-

FIG. 1. Immunoblot analysis of kinetics of HA synthesis. HeLa cells were harvested after infection with vaccinia virus IHD-J, and 35- μ g protein samples were subjected to electrophoresis on SDS-polyacrylamide gels and transferred onto nitrocellulose membranes. The membrane was then probed with rabbit anti-HA polyclonal antiserum as described in Materials and Methods. Lanes: U, uninfected HeLa cell extract; 2 through 18, infected cell extracts isolated at the time points after infection (hours) as indicated; O, HeLa cell extract in which vaccinia virus IHD-J was added at a calculated MOI of 10 to an uninfected harvested cell pellet immediately before lysis and solubilization for analysis; V, like lane O, except with an MOI of 100. The samples in lanes O and V were prepared to insure that the HA protein observed at early times after infection represented de novo synthesis and was not a spurious contaminant of the infecting virus. Molecular masses in kilodaltons are listed on the left.

tion that the HA protein is derived from a late gene. However, in the presence of cytosine arabinoside, an inhibitor which blocks expression of late genes, reduced but significant levels of HA activity could be detected (data not shown).

Through use of anti-HA polyclonal sera, we examined protein blots derived from cell extracts isolated at various times postinfection (Fig. 1). We detected two immunoreactive protein species of 85 and 68 kDa. The 85-kDa protein can be detected as early as 2 h postinfection, whereas the 68-kDa protein is not observed before 5 h postinfection (Fig. 1) even if gels are overloaded. Control lanes in which virus and cells were mixed and immediately lysed before electrophoresis were negative for the 85- and 68-kDa proteins, even with 10-fold more virus than was used for the kinetic analysis (Fig. 1, lanes O and V). This demonstrated that the early appearance of the 85-kDa protein was not derived from HA contamination of the infecting virus.

The detection of a stable 68-kDa protein by the anti-HA serum was unexpected, since the only previously reported mature form of the HA has been a protein of 85 kDa (27, 36). To show that the two proteins are derived from the same gene, we insertionally inactivated the HA gene, which eliminates both the 85- and 68-kDa proteins (data not shown).

The rapid appearance of the 85-kDa HA protein within 2 h postinfection is consistent with expression of an early protein (Fig. 1), whereas the kinetics of appearance of the 68-kDa HA protein are more like those of a late protein. We then tested the effect of CAR, an inhibitor of viral DNA and late protein synthesis, on the formation and incorporation into membranes of the 85- and 68-kDa forms of HA. The data suggest that both forms of HA appear in membranes during

FIG. 2. Effect of CAR and rifampin on the synthesis of HA polypeptides in infected HeLa cells: (A) The proteins of cell membranes derived from infected cells at 18 h postinfection were separated on SDS-polyacrylamide gels, transferred to nitrocellulose membranes, and probed with anti-HA serum. Lanes: U, membranes derived from uninfected HeLa cells; A and B, membranes derived from vaccinia virus IHD-J-infected HeLa cells in the absence and presence, respectively, of 50 μ g of CAR per ml; C and D, samples derived from HeLa cells infected with rabbitpox virus (a phenotypically HA negative orthopoxvirus) in the absence and presence, respectively, of 50 μ g of CAR per ml. Each lane contains material purified from 10^7 cells. (B) Immunoblot of total protein isolated from cells 18 h postinfection as described in Materials and Methods. HeLa cells infected in the absence of inhibitors (N; 10 μ g of total protein), 50 μ g of CAR per ml (CAR; 70 μ g of total protein), or 100 μ g of rifampin per ml (RIF; 10 μ g of total cell protein) are shown. The different amounts of total cellular proteins loaded per lane have been adjusted so that similar amounts of HA protein may be visualized in each sample lane. The molecular masses in kilodaltons are indicated on the left in both panels.

a normal infection in the absence of inhibitors (Fig. 2A, lane A), but that the presence of CAR selectively blocks the appearance of the 68-kDa form of HA (Fig. 2A, lane B). In cell extracts of HeLa cells infected with rabbitpox virus, an orthopoxvirus which fails to express hemagglutination activity, neither form of HA is detectable (Fig. 2A, lanes C and D). The lack of any immunoreactive proteins in rabbitpox virus-infected cells is again consistent with the 85- and 68-kDa proteins originating from the HA gene.

In an attempt to explore further the relationship between the two forms of HA, we also examined the effect of CAR and rifampin on the accumulation of the two HA proteins in total cell extracts. Synthesis of the 68-kDa form of HA is selectively inhibited by either CAR or rifampin (Fig. 2B). This experiment eliminates the possibility that CAR blocks the insertion of the 68-kDa HA into cell membranes, since there is no 68-kDa form detectable even in total extracts of infected-cell protein.

Analysis of HA gene transcripts by Northern blots. The HA gene is located within a 1.8-kb *HindIII-SalI* fragment derived from the *HindIII* A fragment. Based on an analysis of ORFs, the HA coding sequence consists of 945 bp (34). We have found by sequencing to either side of the putative HA coding sequence an upstream ORF of 429 bp capable of encoding a 16-kDa protein (p16 ORF) and a downstream ORF of 453

bases capable of encoding a 17-kDa protein (p17 ORF) (4a). The C terminus of the upstream p16 ORF is separated from the N terminus of the putative HA ORF by a 49-bp intercistronic region, and the N terminus of the downstream p17 ORF is separated from the C terminus of the HA ORF by 144 bp (Fig. 3A). Double digestion of the *HindIII-SalI* fragment with *AccI* and *XbaI* releases a 904-bp fragment, derived entirely from within the putative HA ORF (Fig. 3A), which we have used as a probe for Northern blots (Fig. 3B). A discrete 1.65-kb transcript is detected in cells infected in the presence of cycloheximide (Fig. 3B, lane 4C) as well as early in a normal infection (lanes 2 and 3), particularly when gels are overexposed (data not shown). We have compared the level of the 1.65-kb early HA transcript relative to that of the viral thymidine kinase, a known early gene (Fig. 3C). Densitometric evaluation suggests the 1.65-kb transcript is present at approximately 20% of the level observed for the thymidine kinase mRNA. Later in a normal infection, HA mRNA becomes heterogeneous in size, consistent with the production of 3' polydisperse transcripts late in infection (Fig. 3B, lanes 4 through 18). These results tentatively suggest that an early transcript encompasses the HA coding region, which is consistent with the early appearance of a protein recognized by the anti-HA serum. However, the 1.65-kb early transcript may be derived from the DNA strand complementary to the HA coding strand. Even allowing for the addition of 3' poly(A), the 945-bp coding sequence of HA is much smaller than the 1.65 kb observed for the early transcript. The HA early mRNA may therefore be dicistronic and fused to the upstream or downstream ORF. We have addressed these questions with the use of strand-specific RNA probes. Labeled RNA transcripts within the HA, complementary to the HA, and from the adjacent upstream and downstream coding regions were prepared from cloned *RsaI* fragments in pBluescript KS(+) by using T7 or T3 promoters (Fig. 4A).

No hybridization occurs to either early or late mRNA from probe A, which was designed to measure activity from the strand complementary to that which encodes the HA protein (Fig. 4B, panel A). Sense probes from both the downstream (p17) and HA ORFs hybridize to a 1.65-kb early mRNA (Fig. 4B, panels B and C), whereas probe D, measuring transcription from the sense strand of the upstream p16 ORF, hybridizes to a larger 1.75-kb mRNA (Fig. 4B, panel D). Analysis of the sequence 2 bp downstream of the HA ORF shows the presence of the heptanucleotide TTTT TGA, an apparently nonfunctional variant of the early termination consensus signal (TTTTTNT). There is, however, a TTTTAT signal 42 bp downstream of the p17 ORF (4a), which should be functional. The simplest interpretation of the data of Fig. 4B is that the early transcription of the HA gene does not terminate immediately downstream of the HA ORF (945 bp) but continues through the downstream ORF of 453 bp and then terminates.

Expression of HA from a truncated transcript. Shida and Matsumoto reported that translation in vitro of HA mRNA isolated from vaccinia virus IHD-J-infected cells gave rise to a protein of 58 kDa, as estimated from mobility in SDS-polyacrylamide gels (38). This result is surprising because the in vitro product should be devoid of any modifying glycosylations to add to the apparent molecular weight. A protein of this size would require an mRNA of about 1.6 kb, far larger than the 945 bp needed to encode the ORF for the HA protein alone. Sequence analysis shows that the HA ORF is in the same reading frame as the downstream p17 ORF but separated by 144 bases of intervening sequence.

FIG. 3. (A) Map of the vaccinia virus IHD-J region containing the HA gene. A 1.8-kb *SalI*-*HindIII* fragment from the *HindIII* A fragment of vaccinia virus IHD-J is shown. A 904-bp probe derived by digestion with *AccI* and *XbaI* is indicated. Solid boxes indicate locations of the ORFs. Open boxes indicate noncoding intergenic regions of 49 bp for the p16-HA intergenic region and 144 bp for the HA-p17 intergenic region. The N-terminal portion of p16 ORF and the C terminus of the p17 ORF are located outside the 1.8-kb *SalI*-*HindIII* fragment. (B) Kinetics of HA transcription in vaccinia virus IHD-J-infected cells. RNA was isolated from uninfected and infected cells, resolved on formaldehyde-agarose gels, and transferred onto nitrocellulose as described in Materials and Methods. The blot was then probed with an HA-specific 904-bp *XbaI*-*AccI* DNA fragment (A). Each lane contains 10 μ g of total RNA. Lanes: U, RNA from uninfected cells; 4C, RNA from cells 4 h after infection maintained in the presence of 100 μ g of cycloheximide per ml (early viral message); 2 through 18, RNA isolated from a productive infection at the indicated times (hours). (C) Relative levels of early HA and thymidine kinase transcripts. A duplicate sample of the RNA from cells infected in the presence of cycloheximide (B) was probed with a mixture of radiolabeled thymidine kinase and HA probes of the same length and specific activity to estimate relative amounts of the two transcripts. Hybridizations were as described above.

Within this intervening sequence, there are five translation stop codons found in the same reading frame. Therefore, cotranslation of the HA and p17 ORFs, which theoretically could generate a 58-kDa protein, is not likely, unless the intervening sequence between the two ORFs is spliced or edited so as to bypass all five potential termination codons.

We have addressed the possibility of cotranslation of the HA and p17 ORFs by engineering an early transcription termination signal just downstream of the HA ORF, so that during transcription an mRNA containing sequences solely from the HA ORF can be obtained. This has been achieved by taking advantage of an *XbaI* site just downstream of the HA ORF to insert the *eco-gpt* fragment (the 2.1-kb *EcoRI* fragment of pTK61-gpt). At one end of this inserted fragment is an early transcription termination (TTTTTAT) signal, which through the cloning process is placed just downstream of the HA coding sequence (Fig. 5A), between the HA and p17 ORFs. The recombinant virus (IHD-J::HA535-gpt) should synthesize an artificially truncated HA transcript containing only sequences from the HA ORF (Fig. 5A). If

the HA product is actually a fusion protein of the HA and p17 ORFs, then this mutant should produce either no HA or a truncated HA protein.

The recombinant vaccinia virus IHD-J::HA535-gpt synthesizes a 1.2-kb early HA transcript that is truncated compared with that of the wild-type virus (Fig. 5B, lanes E), as expected. This suggests that the wild-type 1.65-kb transcript encompasses both the HA ORF and the downstream p17 ORF. An analysis of the HA protein synthesized by this mutant (Fig. 6) shows that the HA protein is indistinguishable from that found during wild-type vaccinia virus IHD-J infection; i.e., both the 68- and 85-kDa proteins are synthesized in the absence of any inhibitors, whereas in the presence of CAR the synthesis of the 68-kDa HA is blocked and that of the 85-kDa HA is expressed at a reduced level but to the same extent as for wild-type virus in the presence of CAR (Fig. 6). We conclude that the HA and p17 ORFs, while normally cotranscribed in wild-type virus, are not cotranslated as a fusion protein.

5'-end analysis of HA gene transcripts. Our data suggest

FIG. 4. Strand-specific RNA probes. (A) Structure of probes. *Rsa*I sites within the HA region are shown (R). The specific *Rsa*I fragments to be used for probes for the coding sequences of the upstream (p16), HA, or downstream (p17) ORFs were cloned into the *Eco*RV site of pBluescript KS(+). Single-stranded RNA transcripts were obtained as follows: probe A was from the T7 promoter of the *Sal*I-restricted pHA-A, probe B was from the T3 promoter of *Hpa*I-restricted pHA-B, probe C was from the T3 promoter of *Eco*RI-restricted pHA-C, and probe D was from the T3 promoter of *Eco*RI-restricted pHA-D. (B) Northern blots: hybridization of 10 μ g of total RNA per lane to probes A, B, C, and D as indicated. Lanes: U, total RNA isolated from uninfected HeLa cells; 4C, RNA isolated 4 h postinfection from cells infected in the presence of 100 μ g of cycloheximide per ml (early RNA); 10, RNA isolated from infected cells without inhibitors at 10 h postinfection (late RNA).

that HA expression occurs throughout the entire infectious cycle and that the HA product is not solely a late protein as previously thought. We analyzed the kinetics and location of the HA promoter(s) through primer extension. A 30-mer oligonucleotide, hybridizing to nucleotides 67 to 38 of the HA ORF, was synthesized and used to map the 5' end(s) of the HA transcript(s). Our results suggest an unusually complex mode of transcription. A single transcription start site is

detected for early viral transcripts (Fig. 7A, lane E), whereas this site is joined by a second site later in infection that is approximately 150 bp upstream of the early site (Fig. 7A, lane L). Transcripts beginning at the early start site are also detected at late times during infection. The second start site is referred to as a late promoter because it is detected only during the late stage of viral infection. This experiment suggests that the HA gene is regulated by two promoters, an

FIG. 5. (A) HA region of vaccinia virus IHD-J and recombinant virus IHD-J::HA535-gpt. The ORFs for HA and p17 are shown, together with the predicted size of the HA early transcripts based on the position of the transcription termination sequence (TTTTTAT) in wild-type and recombinant virus. (B) Comparison of HA-specific transcripts synthesized by vaccinia virus IHD-J and recombinant IHD-J::HA535-gpt by using Northern blots of RNA from vaccinia virus IHD-J- and IHD-J::HA535-gpt-infected HeLa cells probed with HA-specific RNA probe C (Fig. 4A). Lanes: E, RNA examined 4 h postinfection from cells infected in the presence of 100 μ g of cycloheximide per ml (early viral RNA); L, RNA isolated 12 h postinfection from cells infected in the absence of inhibitors (late viral RNA).

FIG. 6. Immunoblots of HA proteins in vaccinia virus IHD-J- and IHD-J::HA535-gpt-infected cells probed with anti-HA serum. Lanes: 1, proteins synthesized 18 h postinfection without inhibitors; 2, proteins isolated 18 h postinfection in the presence of 50 μ g of CAR per ml.

early promoter and a late promoter approximately 150 bp further upstream.

We have sequenced the 5' ends of the HA transcripts to map precisely the transcriptional start sites. The results indicate that the early transcriptional start site is found at the -7 position relative to the translational AUG start site, whereas that of the late transcriptional start site is 135 bases upstream at position -142 (Fig. 7B). The 5' poly(A) sequences typically present in late viral transcripts are apparent.

The sequence around the putative HA ORF AUG translation initiation site is shown in Fig. 8. The transcriptional start site for the late message is found within the AAA of the sequence TAAATG (boxed sequence, Fig. 8), which is thought to function as a transcriptional start site for other late genes (31). The HA translational start site (Fig. 8) is 142 bp downstream of the late transcriptional start site and out of frame with the upstream ATG adjacent to the late transcriptional start site. Within the 142-base leader region of the late HA transcript, two small ORFs are found upstream of the putative HA translation initiation codon (underlined sequences, Fig. 8). These minicistrons could potentially encode peptides of 4 and 3 amino acids, respectively. The start site for the early promoter is within 10 bp of the putative HA translational start site (Fig. 8).

DISCUSSION

We have detected two forms (85 and 68 kDa) of the HA protein that are produced during vaccinia virus IHD-J infection (Fig. 1). Although an HA-related protein of 68 kDa has been described previously, earlier reports suggested that the 68-kDa HA is a transient, partially glycosylated, intermediate form of the protein readily detectable only in certain mutants or when infections were carried out in the presence of inhibitors of N glycosylation such as tunicamycin (35, 36, 38). Evidence that both the 68- and 85-kDa proteins described here are derived from the HA gene is provided by the cross-reactivity of both proteins with anti-HA sera (Fig. 1 and 2), the disappearance of both proteins when the HA coding sequence is interrupted by insertional inactivation

FIG. 7. (A) Primer extension analysis of early and late viral messages. RNA from vaccinia virus IHD-J-infected HeLa cells was isolated and primer extended with RM07, a 30-mer oligonucleotide (5'-ATGCTCTGAGGAAAAGGTGTAGCGTATACTA-3') complementary to the region of positions 67 to 38 of the HA ORF (position 1 is the adenine residue of the translation start site, ATG). Lanes: U, RNA from uninfected HeLa cells; E, RNA isolated from HeLa cells 4 h after infection with vaccinia virus IHD-J in the presence of 100 μ g of cycloheximide per ml (early RNA); L, RNA isolated from cells infected in the absence of inhibitors 12 h postinfection (late RNA). (B) RNA sequencing of the 5' termini of the HA transcripts. RNA sequencing was performed with primer RM07 in the presence of dideoxynucleoside triphosphates as described in Materials and Methods. "Early" represents the sequence of RNA isolated from HeLa cells 4 h after infection with vaccinia virus IHD-J in the presence of 100 μ g of cycloheximide per ml. "Late" represents the sequence derived from RNA isolated from HeLa cells infected with vaccinia IHD-J at 12 h postinfection. Lanes: G, A, T, and C, positions of guanine, adenine, thymine, and cytosine, respectively; N, control reaction performed in the absence of any dideoxynucleoside triphosphates. A tract of 5' poly (A) in the late HA transcripts is indicated.

(unpublished results), and the absence of both proteins in rabbitpox virus-infected cells (Fig. 2A).

Standard biological assays for hemagglutinating activity suggest that the HA first appears late in infection at about 6 h, yet with immunoblots we can detect the 85-kDa form of the HA protein early in infection (Fig. 1). The simplest explanation for this apparent discrepancy is probably related to the relative insensitivity of the biological assay for HA activity in contrast to the very sensitive immunoblot assays. In the presence of rifampin, only the 85-kDa form of HA is found (Fig. 2B), yet hemagglutinating activity may be readily assayed in infected cells (14, 44), suggesting that this form rather than the 68-kDa protein is responsible for most of the observed HA activity. The failure to observe the 68-kDa

FIG. 8. Nucleotide sequence of the HA gene promoter region. The DNA sequence 263 bases upstream from the HA start codon and 87 bases downstream is shown. The predicted amino acid sequence of the HA polypeptide is indicated underneath the HA coding sequence. The late and early start sites are indicated by arrows. The boxed hexanucleotide (TAAATG) is a consensus sequence for vaccinia virus late promoters. Two minicistrons potentially encoding peptides of 4 and 3 amino acids are underlined.

form of HA in the presence of either CAR or rifampin (Fig. 2) suggests that the appearance of this protein may require late gene expression and rifampin-sensitive steps of virion assembly or protein maturation.

Our results with the recombinant virus IHD-J::HA535-gpt indicate that both the 85- and 68-kDa HA-related proteins are derived from the 945-bp HA ORF. The *in vitro* translation product of HA mRNA, with an apparent size of 58 kDa (38), presumably results from aberrant migration on SDS-polyacrylamide gels.

Although the HA gene was originally thought to be solely a late gene, we have shown that HA expression occurs throughout the viral cycle through the action of two promoters; i.e., an early promoter and a late promoter (Fig. 7). The early site of transcription initiation is at the -7 position (relative to the translational AUG start codon) of the HA ORF (Fig. 7B and 8) and apparently leads to synthesis of a 1.65-kb dicistronic HA transcript containing the HA ORF and the adjacent downstream p17 ORF as well (Fig. 4B). Despite the fact that the HA and p17 ORFs are probably cotranscribed, engineering of a transcription signal directly downstream of the HA ORF, resulting in monocistronic HA transcripts, has no effect on HA expression (Fig. 6). Transcripts starting at the early promoter are found early and late in infection (Fig. 7A), suggesting either that early transcripts are unusually stable or that the early promoter is active throughout the infection. The second transcriptional start site is found 135 bp upstream of the early transcriptional start site at position -142 relative to the translation initiating AUG and only becomes active after the initiation of DNA replication.

Late transcriptional start sites are usually found very near the start codon for translation. In fact, the TAAATG sequence of late transcriptional start sites often overlaps the translational start codon (31). In the case of HA the late transcription start site, although within a TAAATG motif, is 142 bases upstream of the initiating AUG of the HA ORF. This 142 bases of untranslated sequence, together with at least 40 bases of 5' poly(A), creates a 5'-untranslated RNA leader of almost 200 bases. Within this 5'-untranslated sequence, two small ORFs are found (Fig. 8). These minicistrons have been noted in both viral and eukaryotic transcripts (12, 13, 29). The presence of minicistrons is thought

to downregulate gene expression at the level of translation by decreasing the accessibility of the ATG of the downstream main ORF, thus resulting in a decreased rate of protein synthesis (28, 33). Regulation of gene expression in poxviruses is commonly believed to be primarily transcriptional, although regulation at the translational level has been proposed (32). The minicistrons in the late HA transcripts of vaccinia virus could provide an additional mechanism for the translational regulation of gene expression in poxviruses, a possibility which we are in the process of investigating.

ACKNOWLEDGMENTS

We thank S. Dales for the anti-HA rabbit antiserum, B. Moss for plasmid pTK61-gpt, and D. Bloom for plasmid pHGN-1.

This study was supported by Public Health Service grant AI 15722 from the National Institutes of Health.

REFERENCES

- Ahn, B. Y., and B. Moss. 1989. Capped poly(A) leaders of variable lengths at the 5' ends of vaccinia virus late mRNAs. *J. Virol.* **63**:226-232.
- 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.
- Berns, K. I., and C. Silverman. 1970. Natural occurrence of cross-linked vaccinia virus deoxyribonucleic acid. *J. Virol.* **5**:299-304.
- Bloom, D. C., and R. W. Moyer. Unpublished data.
- Boyle, D. B., and B. E. Coupar. 1988. A dominant selectable marker for the construction of recombinant poxviruses. *Gene* **65**:123-128.
- Brown, C. K., and R. W. Moyer. Virus Genes, in press.
- Brunette, D. M., and J. E. Till. 1971. A rapid method for the isolation of L-cell surface membranes using an aqueous two-phase polymer system. *J. Membr. Biol.* **5**:215-224.
- Cooper, J. A., R. Wittek, and B. Moss. 1981. Extension of the transcriptional and translational map of the left end of the vaccinia virus genome to 21 kilobase pairs. *J. Virol.* **39**:733-745.
- Cox, R. A. 1968. The use of guanidium chloride in the isolation of nucleic acids. *Methods Enzymol.* **12B**:120-129.
- de Magistris, L., and H. G. Stunnenberg. 1988. Cis-acting sequences affecting the length of the poly(A) head of vaccinia virus late transcripts. *Nucleic Acids Res.* **16**:3141-3156.
- Essani, K., and S. Dales. 1979. Biogenesis of vaccinia: evidence for more than 100 polypeptides in the virion. *Virology* **95**:385-394.
- Falkner, F. G., and B. Moss. 1988. *Escherichia coli gpt* gene provides dominant selection for vaccinia virus open reading frame expression vectors. *J. Virol.* **62**:1849-1854.
- Feinberg, A. P., and B. Vogelstein. 1983. A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. *Anal. Biochem.* **132**:6-13.
- Hackett, P. B., R. B. Petersen, C. H. Hensel, F. Albericio, S. I. Gunderson, A. C. Palmenberg, and G. Barany. 1986. Synthesis *in vitro* of a seven amino acid peptide encoded in the leader RNA of Rous sarcoma virus. *J. Mol. Biol.* **190**:45-57.
- Hughes, S., K. Mellstrom, E. Kosik, F. Tamanoi, and J. Brugge. 1984. Mutation of a termination codon affects *src* initiation. *Mol. Cell. Biol.* **4**:1734-1746.
- Ichihashi, Y., and S. Dales. 1971. Biogenesis of poxviruses: interrelationship between hemagglutinin production and polykaryocytosis. *Virology* **46**:533-543.
- Ink, B. S., and D. J. Pickup. 1989. Transcription of a poxvirus early gene is regulated both by a short promoter element and by a transcriptional termination signal controlling transcriptional interference. *J. Virol.* **63**:4632-4644.
- Lane, D. J., B. Pace, G. J. Olsen, D. A. Stahl, M. L. Sogin, and N. R. Pace. 1985. Rapid determination of 16S ribosomal RNA sequence for phylogenetic analyses. *Proc. Natl. Acad. Sci. USA* **82**:6955-6959.
- Mackett, M., G. L. Smith, and B. Moss. 1985. The construction

- and characterization of vaccinia virus recombinants expressing foreign genes, p. 191–211. In D. M. Glover (ed.), DNA cloning: a practical approach. IRL Press, Oxford.
18. Mahr, A., and B. E. Roberts. 1984. Arrangement of late RNAs transcribed from a 7.1-kilobase *EcoRI* vaccinia virus DNA fragment. *J. Virol.* **49**:510–520.
 19. Maxam, A., and W. Gilbert. 1977. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA* **74**:560–564.
 20. Melton, D. A., P. A. Kreig, M. R. Rebagliati, T. Maniatis, K. Kinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**:7035–7056.
 21. Moss, B., and E. N. Rosenblum. 1973. Letter: Protein cleavage and poxvirus morphogenesis: tryptic peptide analysis of core precursors accumulated by blocking assembly with rifampicin. *J. Mol. Biol.* **81**:267–269.
 22. Moss, B., E. N. Rosenblum, and E. Paoletti. 1973. Polyadenylate polymerase from vaccinia virions. *Nature (London) New Biol.* **245**:59–63.
 23. Moyer, R. W., and R. L. Graves. 1982. The white pock mutants of rabbit poxvirus. IV. The late white pock (mu) host range (hr) mutants of rabbit poxvirus are blocked in morphogenesis. *Virology* **119**:332–346.
 24. Moyer, R. W., and C. T. Rothe. 1980. The white pock mutants of rabbit poxvirus. I. Spontaneous host range mutants contain deletions. *Virology* **102**:119–132.
 25. Oie, M., and Y. Ichihashi. 1981. Characterization of vaccinia polypeptides. *Virology* **113**:263–276.
 26. Payne, L. G. 1979. Identification of the vaccinia hemagglutinin polypeptide from a cell system yielding large amounts of extracellular enveloped virus. *J. Virol.* **31**:147–155.
 27. Payne, L. G., and K. Kristensson. 1985. Extracellular release of enveloped vaccinia virus from mouse nasal epithelial cells *in vivo*. *J. Gen. Virol.* **66**:643–646.
 28. Peabody, D. S., S. Subramani, and P. Berg. 1986. Effect of upstream reading frames on translation efficiency in simian virus 40 recombinants. *Mol. Cell. Biol.* **6**:2704–2711.
 29. Pelletier, J. M., E. Flynn, G. Kaplan, V. Racaniello, and N. Sonenberg. 1988. Mutational analysis of upstream AUG codons of poliovirus RNA. *J. Virol.* **62**:4486–4492.
 30. Pennington, T. H. 1974. Vaccinia virus polypeptide synthesis: sequential appearance and stability of pre- and post-replicative polypeptides. *J. Gen. Virol.* **25**:433–444.
 31. 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 *HindIII* H genome fragment. *J. Virol.* **60**:436–449.
 32. 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.
 33. Sedman, S. A., P. J. Good, and J. E. Mertz. 1989. Leader-encoded open reading frames modulate both the absolute and relative rates of synthesis of the virion proteins of simian virus 40. *J. Virol.* **63**:3884–3893.
 34. Shida, H. 1986. Variants of vaccinia virus hemagglutinin altered in intracellular transport. *Mol. Cell. Biol.* **6**:3734–3745.
 35. Shida, H. 1986. Nucleotide sequence of the vaccinia virus hemagglutinin gene. *Virology* **150**:451–462.
 36. Shida, H., and S. Dales. 1981. Biogenesis of vaccinia: carbohydrate of the hemagglutinin molecules. *Virology* **111**:56–72.
 37. Shida, H., and S. Dales. 1982. Biogenesis of vaccinia: molecular basis for the hemagglutination-negative phenotype of the IHD-W strain. *Virology* **117**:219–237.
 38. Shida, H., and S. Matsumoto. 1983. Analysis of the hemagglutinin glycoprotein from mutants of vaccinia virus that accumulates on the nuclear envelope. *Cell* **33**:423–434.
 39. Shida, H., T. Tochikura, T. Sato, T. Konno, K. Hirayoshi, M. Seki, Y. Ito, M. Hatanaka, Y. Hinuma, M. Sugimoto, F. Takahashi-Nishimaki, T. Maruyama, K. Miki, K. Suzuki, M. Morita, H. Sashiyama, and M. Hayami. 1987. Effect of the recombinant vaccinia viruses that express HTLV-I envelope gene on HTLV-I infection. *EMBO J.* **6**:3379–3384.
 40. Towbin, H., T. Staehelin, and J. Gordon. 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedure and some applications. *Proc. Natl. Acad. Sci. USA* **76**:4350–4354.
 41. Vos, J. C., and H. G. Stunnenberg. 1988. Derepression of a novel class of vaccinia virus genes upon DNA replication. *EMBO J.* **7**:3487–3492.
 42. Weaver, R. F., and C. Weissman. 1979. Mapping of RNA by a modification of the Berk-Sharp procedure: 15S beta-globin mRNA precursor and mature 10S beta-globin mRNA have identical map coordinates. *Nucleic Acids Res.* **7**:1175–1193.
 43. Wei, C. M., and B. Moss. 1974. Methylation of newly synthesized viral messenger RNA by an enzyme in vaccinia virus. *Proc. Natl. Acad. Sci. USA* **71**:3014–3018.
 44. Weintraub, S., and S. Dales. 1974. Biogenesis of poxviruses: genetically controlled modifications of structural and functional components of the plasma membrane. *Virology* **60**:96–127.