JOURNAL OF VIROLOGY, June 1988, p. 1889–1897
0022-538X/88/061889-09$02.00/0

Vol. 62, No. 6

Sequence and Transcriptional Analysis of the Vaccinia Virus *Hin*dIII I Fragment

JACKY F. C. SCHMITT AND HENDRIK G. STUNNENBERG*

European Molecular Biology Laboratory, Meyerhofstrasse 1, P.O. Box 10.2209, 6900 Heidelberg, Federal Republic of Germany

Received 30 November 1987/Accepted 26 February 1988

The complete sequence of the vaccinia virus *Hin*dIII I fragment, which is composed of 6,498 base pairs, encodes six complete and two incomplete open reading frames (ORFs). Computer analysis revealed an amino acid sequence homology between ORF I 4 and the large subunit of the ribonucleotide reductase complex. The two small polypeptides derived from ORFs I 2 and I 5, with molecular weights of 8,500 and 8,700, respectively, have a very high hydrophobic amino acid sequence composition. S1 analysis revealed that ORF I 4 is expressed at early stages of infection, ORFs I 1, I 2, I 5, and I 7 are expressed in the late phase of infection, and ORF I 3 is constitutively expressed. Screening a vaccinia virus genomic library revealed a large vaccinia virus insert overlapping the *Hin*dIII I and O fragments which contains a previously undetected *Hin*dIII P fragment of approximately 300 base pairs. S1 analysis revealed an early (O1) and a late (O2) start site of transcription initiation located within the *Hin*dIII O fragment.

Transcription during the life cycle of the cytoplasmic vaccinia virus is generally divided into two distinct phases. The genes which are expressed immediately after the penetration of the virus and before the second uncoating are generally referred to as early genes. The transcription of these early genes is shut off with the start of viral DNA replication, and the late genes are switched on (for a review, see reference 17). Whether a given gene is actively transcribed before or after replication is routinely determined using primary (arabinose C or hydroxyurea) or secondary (cycloheximide) inhibitors of DNA replication (17).

Recent studies have revealed that the regulation of gene expression in vaccinia virus is not merely separated into two temporally distinct phases, but also the mode of transcription within these phases appears to be different (3, 23). Early transcripts are initiated 10 to 50 base pairs (bp) upstream of their AUG start codon, and the transcripts are terminated at a discrete site downstream of the coding sequences (30). The sequence T_5NT is an essential *cis*-acting element in this process, and the actual termination occurs 50 to 70 bases downstream of this element (20). Late transcripts appear to be discontinuously synthesized with a noncontiguously encoded capped poly(A) head at the 5′ end (23). The junction between the A head and the coding body of the messenger is located within the conserved TAAATG motif, which is late specific (7, 21). Deletions and substitutions within this motif abolish the transcription (7). A further characteristic of late transcripts is the absence of discrete 3′ termini, and early termination signals are not recognized as such in the late phase of infection (12, 29). Until more information becomes available, it remains an open question whether all late genes are synthesized in the same discontinuous manner.

The present study was undertaken to learn more about the genomic organization and to identify early as well as late genes, study their temporal regulated expression, and compare their regulatory sequences.

* Corresponding author.

MATERIALS AND METHODS

Virus and cells. Vaccinia virus (strain WR) was grown in HeLa cell suspension cultures that were maintained in Eagle medium containing 5% fetal calf serum.

Sequencing. The *Hin*dIII I fragment of vaccinia virus was isolated after *Hin*dIII restriction endonuclease digestion of purified viral genomic DNA and cloned into pUC9 (26). Subfragments were isolated by gel electrophoresis onto DEAE membranes (NA 45) according to the manufacturer (Schleicher and Schuell) and subcloned into bacteriophage m13 derivatives (16). Sequencing was performed as described (22).

RNA analysis. RNA from vaccinia virus-infected HeLa cells (multiplicity of infection of 5) was extracted at 3, 6, 9, and 16 h after infection and from infected cells incubated for 6 h in the presence of 100 μg of cycloheximide per ml. Purification of RNA was done as described (14). A 5-μg sample of total RNA was hybridized at 42°C to asymmetrically end-labeled DNA probes as described (14). Single-stranded DNA was digested after overnight hybridization with 50 U of S1 nuclease for 1 h at room temperature. The protected fragments were resolved on denaturing polyacrylamide gels.

Computer analysis. DNA sequence data were managed and analyzed by using the programs of the University of Wisconsin genetics computer group (6). Protein data base searching was done by using the FASTP program written by Lipman and Pearson (11) for the National Biomedical Research Foundation protein library.

RESULTS

Sequencing. We have sequenced the *Hin*dIII I fragment of vaccinia virus by the dideoxynucleotide chain termination method (22). We have chosen the strategy of subcloning restriction fragments into phage m13 derivatives (26), and both strands of the vaccinia DNA were sequenced. The entire sequence of the *Hin*dIII I fragment comprises 6,498 base pairs (Fig. 1).

Analysis of ORFs. The nucleotide sequence was translated in each of the six possible frames to locate open reading

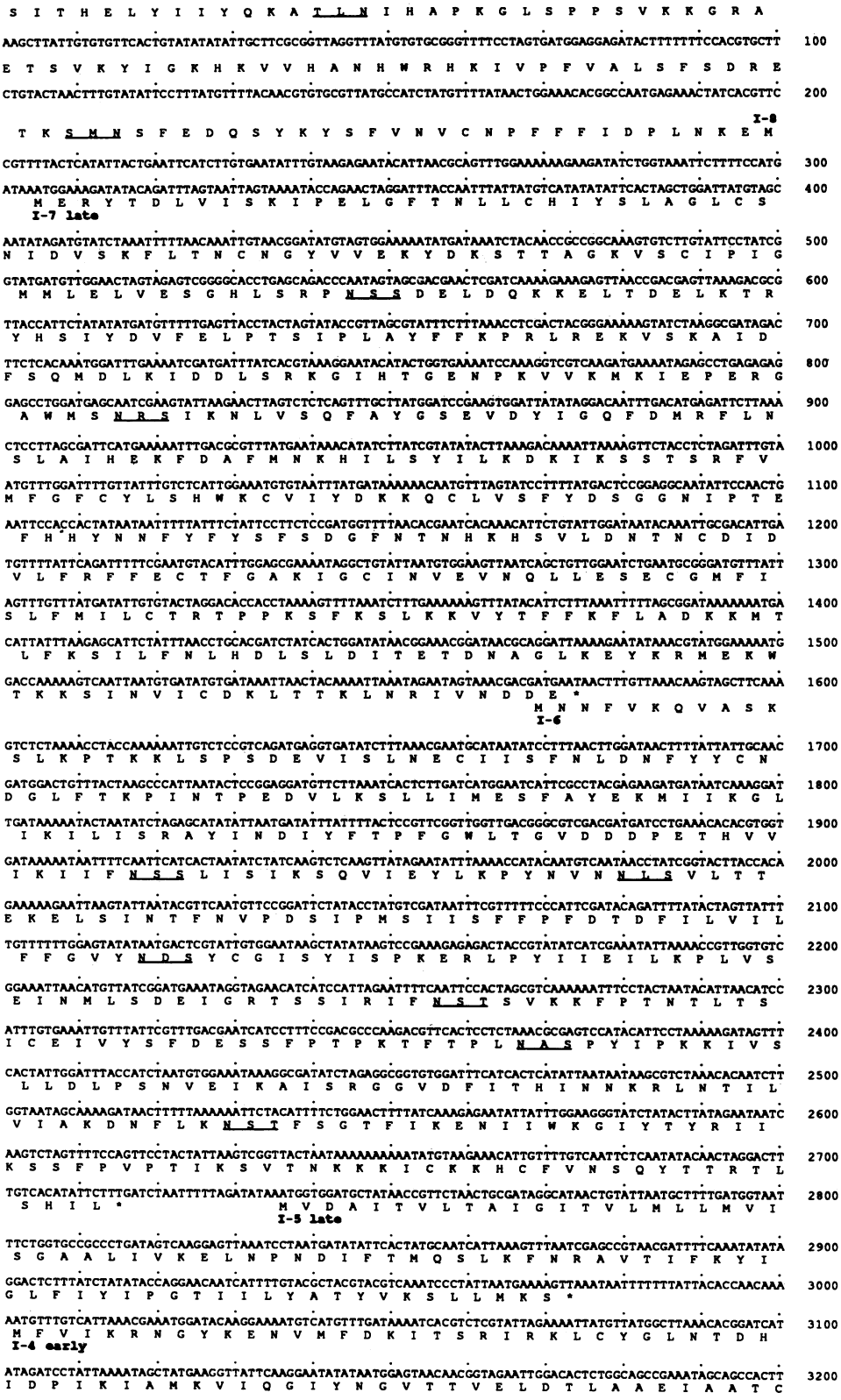

FIG. 1. Nucleotide and amino acid sequences. Numbering starts on the right side of the *Hin*dIII I fragment in the opposite orientation from that in the genome. Potential asparagine-linked N-glycosylation sites are underlined in the protein sequences.

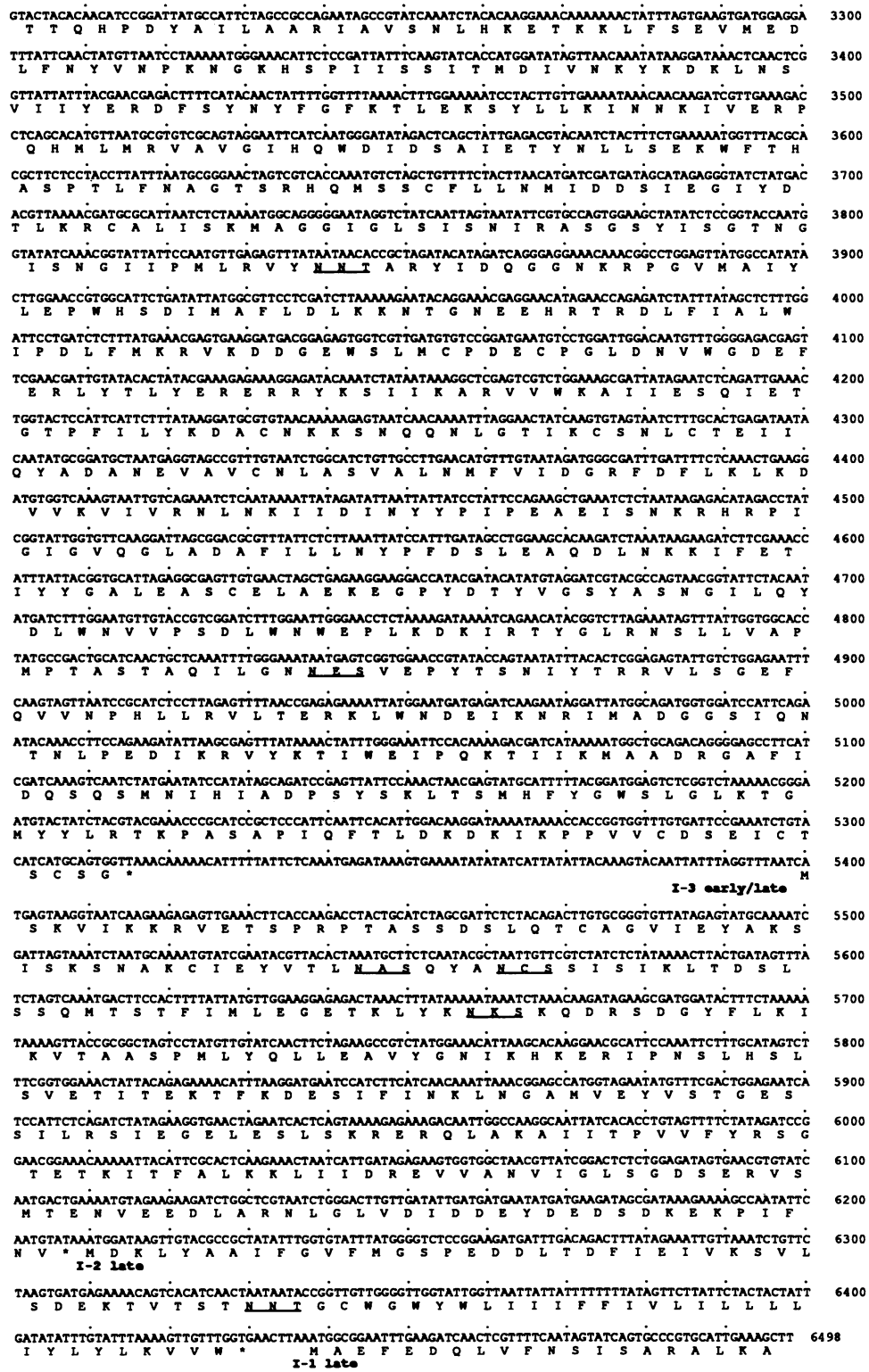

FIG. 1—*Continued.*

frames (ORFs). Six complete major ORFs are located in the *Hin*dIII I fragment, and two putative ORFs are only partly contained within the vaccinia virus *Hin*dIII I fragment. The positions of the ORFs and the respective sizes of the polypeptides are shown in Fig. 2. The numbering of the ORFs was according to Rosel et al. (21), using the letter (I) designating the *Hin*dIII fragment in which they originated and then numbering successively from left to right. The

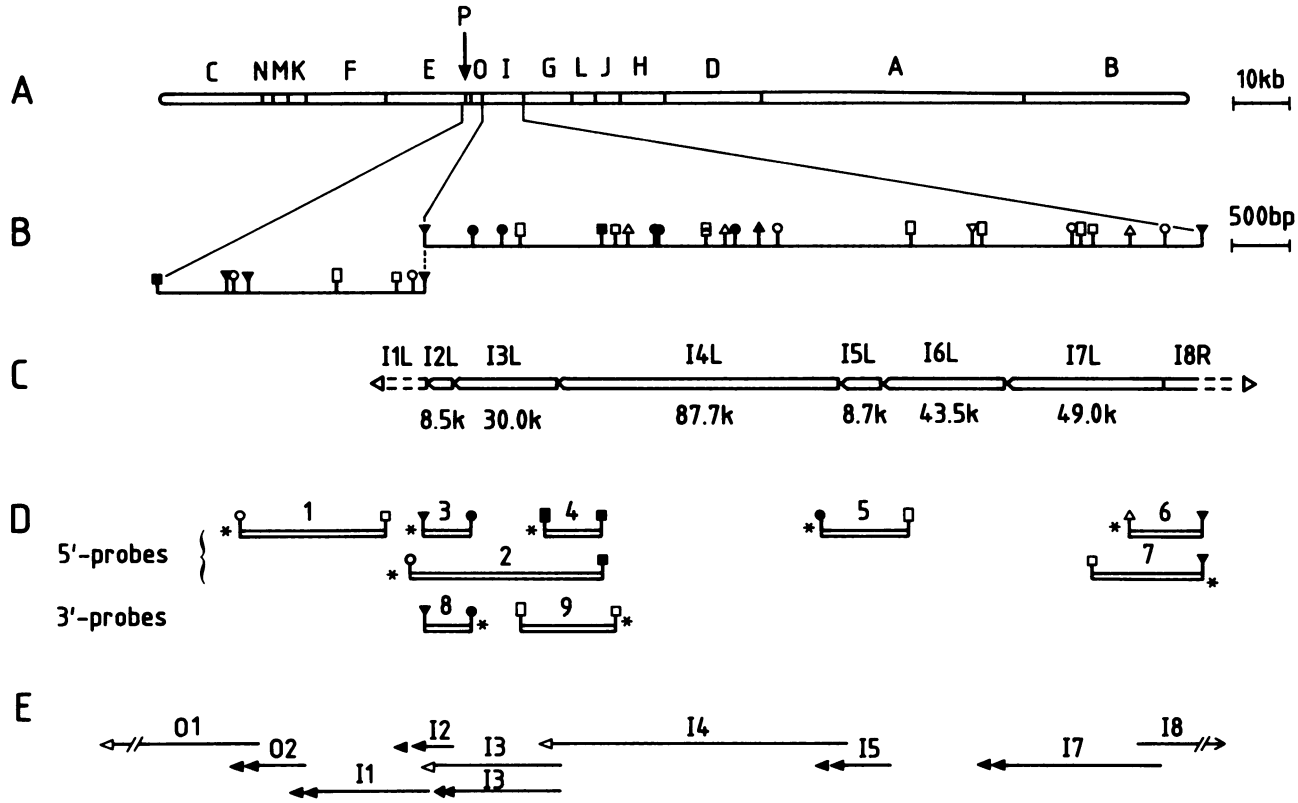

FIG. 2. Location of the ORFs and transcripts. (A) *Hin*dIII map of the vaccinia genome: the fragments are designated A to P according to their size. (B) The upper line represents a detailed restriction map of the sequenced *Hin*dIII I fragment. The lower line identifies *Hin*dIII-O, -P, and part of -E. (C) Positions of the ORFs and molecular weights of the putative polypeptides. (D) DNA fragments used as probes for S1 mapping. The 5' probes were labeled with T4 polynucleotide kinase, and the 3' probes were labeled with *E. coli* DNA polymerase large fragment. (E) RNA transcripts identified in *Hin*dIII I and O fragments. Open arrowheads indicate early transcripts up to the site of transcription termination, and closed, double arrowheads indicate late specific readthrough transcripts. Symbols indicate the position of the restriction sites: *Acc*I ( ), *Bam*HI ( ), *Bgl*II ( ), *Cla*I ( ), *Eco*RI ( ), *Hin*dIII ( ), *Hpa*II ( ), *Kpn*I ( ), *Mbo*I ( ), *Pst*I ( ), *Sal*I ( ), *Xba*I ( ), and *Xho*I ( ).

reading frame direction is indicated by adding an L (left) or R (right) after the numbering. All complete ORFs are located on one strand of the DNA having the direction of transcription from right to left. The only exception is the putative ORF I 8R, which is located at the right-hand side of the *Hin*dIII fragment and probably continues into *Hin*dIII G with the direction of transcription from left to right.

Computer analyses permitted the determination of the amino acid composition of putative polypeptides derived from the ORFs (Table 1). Furthermore, we performed hydropathicity plots of the polypeptides according to the method of Hopp and Woods (8). Two polypeptides (ORFs I 2 L and I 5L, of 8.5 and 8.7 kilodaltons [kDa], respectively) showed a high degree of hydrophobic amino acids (>50%). The hydropathicity plots indicate that both short polypeptides have two large N- and C-terminal hydrophobic domains separated by small hydrophilic domains (Fig. 3). Both hydrophobic regions can function as membrane-spanning segments, indicating a possible localization in the membrane envelope. The other polypeptides do not have striking amino acid compositions. All ORFs with the exception of I 5L contain potential asparagine-linked N-glycosylation sites; the respective amino acid sequences are underlined in Fig. 1.

TABLE 1. Amino acid composition

ORF	Amino acids					
	Total	% Acidic	% Basic	% Aromatic	% Hydrophobic	Mol wt
I2	73	11	5	18	57	8,500
I3	269	14	13	13	32	30,000
I4	772	10	12	10	36	87,700
I5	79	4	8	10	54	8,700
I6	383	9	11	11	41	43,500
I7	424	12	13	12	37	49,000

We also compared the amino acid sequences of the proteins with previous entries in the National Biomedical Research Foundation protein library. The 87.7-kDa polypeptide (ORF I 4L) shows a relatively high degree of homology to the ribonucleoside diphosphate reductase of *Escherichia coli* (4) (24.6% in 589 amino acids) and to the 140-kDa ribonucleotide reductase of Epstein-Barr virus (EBV) (1) (29.7% in 202 amino acids). The respective dot plots (Fig. 4) reveal that the central part of the vaccinia virus protein (from amino acids 300 to 600) has a high degree of similarity with the *E. coli* polypeptide. The N termini have only a limited similarity, whereas the C termini are not homologous at all. The homology with the EBV protein is restricted to the central and C-terminal part of the vaccinia virus polypeptide.

Physical map of lambda EMBL vv43. An *Mbo*I genomic library of the vaccinia virus genome, cloned into the lambda

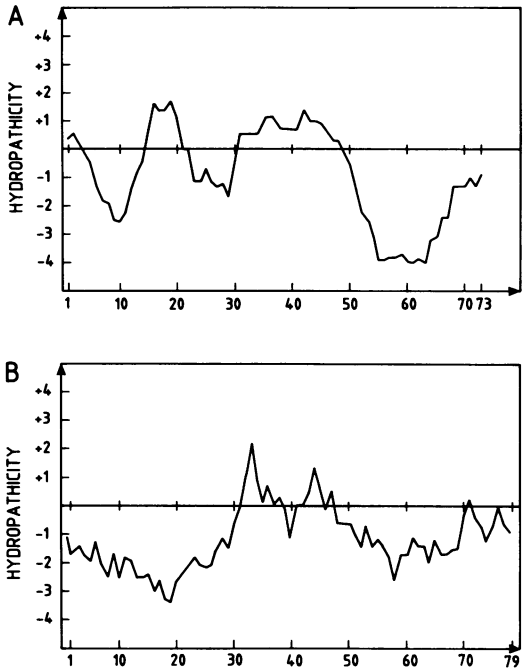

FIG. 3. Hydropathicity plots of polypeptides derived from ORF I 2 (A) and ORF I 5 (B). Plots were made according to the method of Hopp and Woods (8). The *x* axis indicates the amino acid position, and the *y* axis indicates degree of hydropathy.

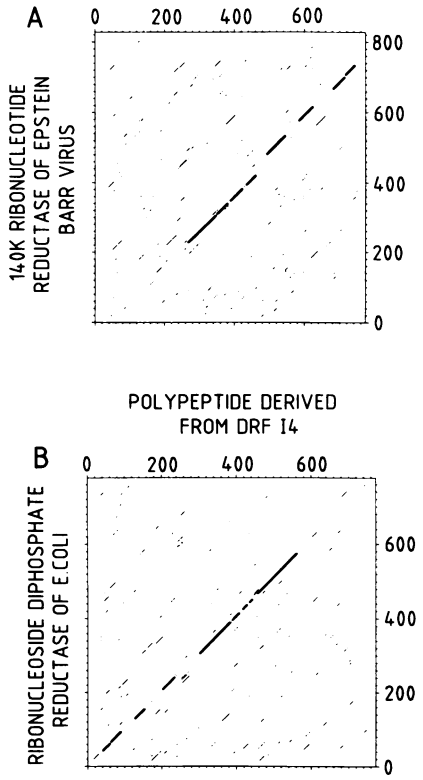

FIG. 4. Dot plots. Dot-plot comparison was performed by the method of Maizel and Lenk (13) between the amino acid sequence of ORF I 4 with (A) the 140-kDa ribonucleotide reductase of EBV and (B) the 87.5-kDa ribonucleoside diphosphate reductase of *E. coli*. The *x* axis indicates amino acid position in the polypeptide from the National Biomedical Research Foundation data bank, and the *y* axis indicates amino acid position in the polypeptide derived from ORF I 4. The window size was 30 and stringency was 13.

phage EMBL-4, was screened with nick-translated probes of the *Hin*dIII I and O fragments. Positive clones were isolated, and the inserts were analyzed by restriction digests and Southern blotting. The clone vv43 contains a vaccinia virus genomic fragment spanning the complete *Hin*dIII I and O fragments and additional sequences of the adjacent *Hin*dIII E and G fragments. A small *Hin*dIII fragment of approximately 300 bp was contained within the genomic insert which is located between the E and O fragments. This small fragment had escaped previous detection by restriction analysis with purified vaccinia virus DNA because of its small size. We will refer to this fragment as *Hin*dIII-P (Fig. 2), according to the nomenclature suggested by DeFilippes (5). Different restriction fragments were subcloned and used for S1 mapping of RNA transcripts.

Nuclease S1 mapping of mRNA 5′ ends. The RNAs transcribed from the *Hin*dIII I fragment were analyzed by nuclease S1 mapping. The structures of the probes are indicated in Fig. 2. A precise mapping of RNA start sites was achieved by coelectrophoresis of the sequence ladder derived from the labeled S1 fragments by using the chemical cleavage procedure (15) (only shown in some cases). A shift in electrophoretic mobility of 1.5 bases, due to the presence of the 3′ phosphate as a result of the chemical cleavage reaction, was taken into account. The electrophoretic analysis of the nuclease S1-protected fragments, obtained with RNAs prepared at different times after infection, is shown in Fig. 5. ORFs I 1L, I 2L, I 5L, and I 7L are expressed late in infection, and the RNA start sites were mapped immediately upstream of the AUG within a TAAAT motif (Fig. 5B, C, E, and F, respectively). S1-protected fragments were not generated if early RNA (3 h) or RNA from cycloheximide-treated cells was used, confirming the late phenotype of the genes. ORF I 4L is expressed in the early phase of infection; the start of transcription initiation was mapped 19 bp upstream of its AUG. The start of transcription initiation maps within a stretch of T residues, which was not expected on the basis of cap analysis studies showing that the penultimate nucleotide is a purine (27). The region surrounding the start site is highly AT rich, and therefore it cannot be excluded that the S1-protected fragment is generated by nibbling of the nuclease S1. I 4L transcripts were detectable up to 9 h after infection, and a strong signal was obtained if RNA derived from cycloheximide-treated cells was used (Fig. 5E). ORF I 3L is expressed early as well as late in infection. The different RNA start sites are separated by approximately 33 nucleotides; the early start is proximal to the ORF (27 bp upstream of the AUG), and the late start is located further upstream (60 bp from the AUG) and maps within a CAAAT motif.

Reproducible S1 mapping of potential transcripts from I 6L was not possible. This might be due to a high level of readthrough transcripts initiating at I 7L which mask the transcripts initiated at the putative I 6 promoter. Several S1-protected fragments were generated with late RNA, suggesting the presence of multiple late RNA start sites located 300, 330, 370, and 540 bp from the *Hin*dIII restriction site. These start sites do not coincide, however, with a TAAAT motif as anticipated for late genes. Furthermore, the start site immediately upstream of the putative ATG of the ORF I 8 (position 298) coincided with the 5′ end of the I 7 RNA, which is transcribed on the opposite strand. Some of

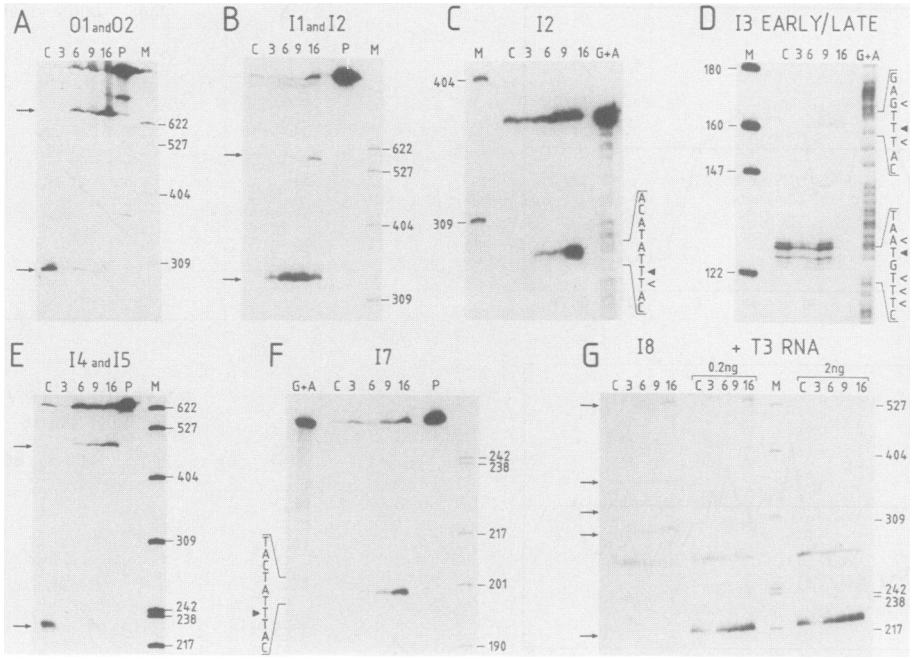

FIG. 5. S1 mapping of mRNA 5′ ends. Asymmetrically labeled S1 probes were hybridized to RNA extracted from HeLa cells at different times after infection. Mapping of transcripts from the different ORFs (indicates above the gels) are shown. The different lanes represent S1 mappings using RNA extracted from cells infected for 3, 6, 9, or 16 h or in the presence of cycloheximide (C); lanes P correspond to the input S1 probe, and lanes M contain ^{32}P-labeled *Hpa*II-digested pBR322 size marker. On gels C, D, and F, the lane marked G+A is a Maxam and Gilbert (15) sequence reaction on the S1 probe. The corresponding sequence is indicated next to the gel; major start sites are indicated by closed arrowheads, and minor start sites are shown by open arrowheads. In vitro RNA, synthesized from a Bluescribe vector using phage T3 RNA polymerase, was added to the S1 mapping procedure as indicated.

these multiple late start sites might be generated by a heteroduplex formation between the complementary RNAs, transcribed in the opposite direction from ORFs I 7 and I 8, and the S1 probe. This possibility was tested by the synthesis of artificial RNA transcripts in vitro, using a Bluescribe vector (Vector Cloning Systems), and addition of these in vitro RNAs to the S1 mapping (schematically indicated in Fig. 6). This artificial transcript was complementary to I 8 RNA up to the *Eco*RI site at position 219 (Fig. 5G). A new S1-protected fragment of 220 nucleotides (nt) was generated which corresponded to an artificially introduced S1 "start site" mapping at the *Eco*RI site. This fragment corresponds to a protection of the DNA probe from *Hin*dIII up to the 5′ end of the added in vitro transcripts. Furthermore, the S1-protected fragments of 300, 330, 370, and 540 nt in length, generated in the absence of the in vitro-synthesized RNA, gradually disappeared upon addition of the T3 polymerase-derived synthetic RNA. This experiment indicated that S1-protected fragments can be generated by heteroduplex formation between complementary RNA transcripts which might not correspond to the start site of transcription initiation.

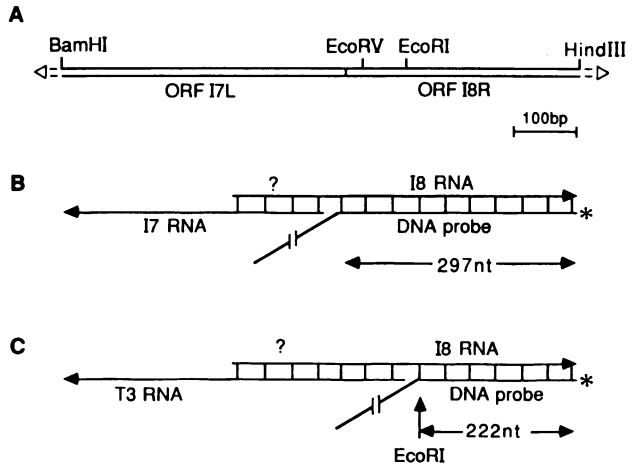

FIG. 6. Schematic presentation of S1 mapping in the presence of complementary T3 RNA. (A) Positions of ORFs I 7 and I 8. Postulated heteroduplex formation occurring during S1 mapping of I 8 transcripts in the absence (B) and presence of T3 transcript (C).

The results of the 5′ S1 mapping are summarized in Fig. 7; the solid triangles indicate the major S1 signals and the carets indicate the minor protections.

S1 mapping of the 3′ ends. The 3′ ends of the transcripts were analyzed as indicated in the legend of Fig. 2. Discrete 3′ termini could not be detected with transcripts from ORFs I 1, I 2, I 3-late, I 5, and I 7, which is in agreement with their late character. Discrete 3′ termini could be detected with transcripts originating from ORFs I 3-early and I 4 (Fig. 8). Remarkable is the site of transcription termination of transcripts from ORF I 3-early, which is located approximately 190 nt downstream of the translation stop codon and includes most of the coding sequences of the late ORF I 2. I 4 transcripts are terminated approximately 48 nt downstream of the stop codon. The described termination signal T_5NT is present upstream of the 3′ ends of the I 3-early and I 4 transcripts. The results of the 3′ S1 mapping are summarized

in Fig. 7; the solid triangles indicate the major S1 signals, and the carets indicate the minor protections.

DISCUSSION

The entire sequence of the *Hin*dIII-I fragment was determined, and the nucleotide and predicted amino acid compositions were analyzed using computer programs (Table 1 and Fig. 3). Six complete ORFs were revealed that are preceded by an AUG start codon. Two potential but incomplete ORFs are located at the extremities of the fragment. The N-terminal sequences of ORFs I 1L and I 8R are encoded within the *Hin*dIII fragment at the left- and right-hand sides. ORF I 8R extends into the *Hin*dIII G fragment for an additional 2 kilobases (Z. Fathi and R. Condit, personal communication).

The genomic organization in the *Hin*dIII I fragment is similar to that observed in other regions of the vaccinia virus genome. ORFs I 1 to I 7 are transcribed in the same direction, whereas I 8 is transcribed in the opposite direction. It has been reported that large segments of the vaccinia virus genome are transcribed in one direction, e.g., the 5.1-kilobase fragment spanning the *Hin*dIII-D–A junction (28), whereas, e.g., the central part of the genome is transcribed in both directions (19, 21). Remarkable is the high density of protein coding sequences: in general, the ORFs are separated by a few base pairs of spacer sequence (e.g., between I 1-I 2 and I 5-I 6) whereas other ORFs even partially overlap (e.g., I 6-I 7). A similar organization has been reported for other regions of the genome (18, 19). Large noncoding regions have not been observed in the vaccinia virus genome. It is therefore likely that ORF I 6 is transcribed into RNA despite the fact that the transcripts could not be detected by S1 mapping.

Late regulatory elements:

	-27	-17	-7	+1	+4
I1L	GTATTTAAAA	GTTGTTTGGT	GAAC**TAAATG**		**G**CGGAA
I2L	AAAGAAAAGC	CAATATTCAA	TGTA**TAAATG**		GATAAG
I3L	GTTAAACAAA	AACATTTTTA	TTCTC**AAATG**		AGATAA
I5L	ATTCTTTGAT	CTAATTTTTA	GATA**TAAATG**		GTGGAT
I7L	TATCTGGTAA	ATTCTTTTCC	ATGA**TAAATG**		GAAAGA
	Consensus		TAAATG		G

Early promoter sequences:

	-30	-20	-10	+1
I3L	AGATAAAGTG	AAAATATATA	TCATTATATT	ACAAAGTACA
			^	▲ ^^^
	+10	+20		
	ATTATTTAGG	TTTA**A**TC**ATG**	**A**	
	-30	-20	-10	+1
I4L	AATCCCTATT	AATGAAAAGT	TAAATAATTT	TTTTATTACA
			^	^^
	+10	+20		
	CCAACA**AAAA**	**TGT**		

Early termination sequences:

I3L	**TTTTTAT**AGT	TCTTATTCTA	CTACTATTGA	TATATTTGTAT
			^▲^	
I4L	**TTTTTAT**TCT	CAAATGAGAT	AAAGTGAAAA	TATATCATTAT
				^▲^

FIG. 7. Regulatory sequences. The nucleotide sequences surrounding the site of transcription initiation and termination are indicated. The symbols indicate the positions of the major (▲) and minor (^) S1 signals in the sequence. Conserved sequence elements are underlined. Boldface letters indicate the positions −3 and +4 with respect to the initial ATG start codon.

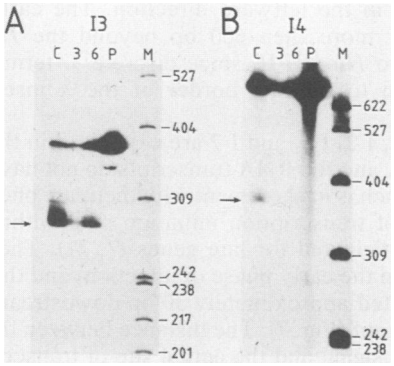

FIG. 8. 3′ S1 mapping of early mRNAs. S1 mapping of transcripts from ORF I 3 (A) and ORF I 4 (B). The S1 probes are indicated in Fig. 2D. The different lanes are S1 mappings using RNA extracted from cells at 3, 6, 9, and 16 h postinfection and at 7 h postinfection in the presence of cycloheximide (C); lanes P correspond to the input S1 probe, and lanes M contain ^{32}P-labeled *Hpa*II-digested pBR322 size marker.

The calculated molecular weights of the different polypeptides (Table 1) are in agreement with those detected by Belle Isle and collaborators using hybrid selection and subsequent in vitro translation of the selected mRNAs (2). They detected two early polypeptides of 80 and 32 kDa which probably correspond to I 4 and I 3-early. Two late polypeptides of 46 kDa (I 6 or I 7 or both) and 33 kDa (I 3) were found, as well as a polypeptide of 35 kDa (I 1), which could be selected using *Hin*dIII-E, -O, and -I, respectively. The two small polypeptides encoded by ORFs I 2 and I 5 are very rich in hydrophobic amino acids (>50%). Both hydrophobic regions can function as membrane-spanning segments, indicating a possible localization in the membrane envelope (Fig. 3). Amino acid homology search revealed a similarity between the polypeptide encoded by ORF I 4 and the large subunit of the 140-kDa ribonucleotide reductase of EBV (1) and the 87.5-kDa ribonucleoside diphosphate reductase of *E. coli* (4). The putative small subunit of the reductase complex was identified also on the basis of amino acid sequence homology with the small subunit of the reductase of EBV. The vaccinia virus gene appears to be located within the *Hin*dIII F fragment (24) and not in *Hin*dIII-H, as previously reported on the basis of sequence comparison (21). The two subunits of the reductase complex are genomically linked in *E. coli* and EBV, but they do not seem to be linked in the vaccinia virus genome.

Restriction mapping of the insert of the λ clone vv43 containing an *Mbo*I fragment revealed the presence of a previously undetected small *Hin*dIII fragment which was designated *Hin*dIII-P. The transcripts derived from the different ORFs within the *Hin*dIII I fragment and from the region covered by the λ clone vv43 were analyzed by S1 mapping. An early (O1) and a late (O2) RNA start site could be detected within the *Hin*dIII O fragment; both genes are

transcribed in the leftward direction. The early gene (O1) extends for more then 600 bp beyond the *Hin*dIII-P sequences into *Hin*dIII-E, since discrete 3′ termini were not obtained up to the left border of the λ insert (data not shown).

ORFs I 1, I 2, I 5, and I 7 are expressed in the late phase of infection, and the RNA transcripts do not have discrete 3′ termini, which is in agreement with their late phenotype. The start sites of transcription initiation map within a TAAAT motif as anticipated for late genes (7, 21). The I 4 gene is expressed in the early phase of infection, and the transcripts are terminated approximately 30 bp downstream of a T_5NT sequence motif (Fig. 7). The distance between this *cis*-acting termination signal and the actual site of transcription termination is significantly shorter in the case of I 3 early and I 4 transcripts (16 and 29 nt, respectively) than observed for the vaccinia virus growth factor gene (50 to 70 nt) (20). ORF I 3 appears to be expressed in the early as well as the late phase of infection; the late start site is located 33 bp upstream of the early start site. The level of transcription in the late phase of infection from the upstream promoter is significantly lower than that from the downstream early promoter. This low level of late transcripts might be due to the presence of a CAAAT motif at the site of transcription initiation in the I 3 gene instead of the conserved TAAAT motif of late genes (7, 21). The 7.5K gene (25) and the ORF 5 gene within the *Hin*dIII D fragment (28) are also constitutively expressed from tandemly arranged promoters. These genes, however, have a higher rate of transcription in the late phase of infection.

RNA transcripts from ORF I 6 could not be detected by S1 analysis. Different DNA probes were used which were either end labeled or homogeneously labeled obtained by primer extension of single-stranded m13 DNA containing I 6 RNA-like sequences. A possible explanation could be that I 6 and I 7 are not two independent ORFs as suggested by sequencing of different independent subclones. The stop codon between ORFs I 6 and I 7 (position 1574) could be confirmed in in vitro transcription-translation experiments (data not shown). Finally, translation frameshift experiments performed as described for the *gag-pol* fusion in Rous sarcoma virus by Jacks and Varmus (9) were not indicative for a frameshift phenomenon (data not shown). A feasible explanation seems to be that I 6 is expressed at a very low level and that the detection of its transcripts by nuclease S1 experiments is complicated by a high level of readthrough transcripts initiated at the I 7 gene.

Other difficulties were encountered in the S1 analysis of I 8 transcripts which might be due to the arrangement of the I 7 and I 8 genes. The S1 mapping data are indicative for the presence of multiple start sites in the late phase of infection which, however, do not coincide with late specific TAAAT motifs (Fig. 1 and 5G). S1 mapping in the presence of in vitro RNA which was partially complementary to I 8 RNA sequences resulted in the introduction of an artificial S1-protected fragment and simultaneous disappearance of the multiple start sites. This experiment clearly indicated that some of the multiple start sites might be generated because of the presence of RNA-RNA heteroduplexes between I 7 and I 8 transcripts (Fig. 5G and 6). Alternatively, it cannot be excluded that I 8 belongs to a different class of late genes which lack the TAAAT motif. Further experiments are necessary to identify the start site of transcription initiation.

The sequences surrounding the ATG that starts each ORF are according to the Kozak rules: an A or G residue is preferred at the −3 position, and a G residue is most frequently found at position +4 with respect to the AUG start codon used for translation (10). ORFs I 1, I 2, I 5, and I 7 would have an unfavored T residue at position −3, encoded by the DNA. The late 11K and 4b mRNAs, however, were shown to be discontinuously synthesized, obtaining a poly(A) stretch at the 5′ end (3, 23). The junction between the nontranslated poly(A) leader RNA and the protein coding sequences is located within the conserved TAAAT motif. As a consequence of this process, an A residue is found at the −3 position in the mRNA. It remains to be determined whether this phenomenon applies to all late genes. Primer extension experiments, cDNA cloning, and analyses of RNA transcripts in in vitro cell-free transcription systems are in progress to answer this question.

ACKNOWLEDGMENTS

We thank G. Chinalli (CEINGE Naples, Italy) for his generous gift of the vaccinia virus genomic library, and Z. Fathi and R. Condit (Buffalo, N.Y.) for the sequence information. We also thank our colleagues for discussions and critical reading of the manuscript.

LITERATURE CITED

1. **Baer, R., A. T. Bankier, M. D. Biggin, P. L. Deiniger, P. J. Farrell, T. J. Gibson, G. Hatfull, G. S. Hudson, S. C. Satchwell, P. S. Tuffnell, and B. G. Barrell.** 1984. DNA sequence and expression of the B95-8 Epstein-Barr virus genome. Nature (London) **310**:207–211.
2. **Belle Isle, H., S. Venkatesan, and B. Moss.** 1981. Cell-free translation of early and late mRNAs selected by hybridization to cloned DNA fragments derived from the left 14 million to 72 million daltons of the vaccinia virus genome. Virology **112**:306–317.
3. **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.
4. **Carlson, J., J. A. Fuchs, and J. Messing.** 1984. Primary structure of the Escherichia coli ribonucleotide diphosphate reductase operon. Proc. Natl. Acad. Sci. USA **81**:4294–4297.
5. **DeFilippes, F. M.** 1982. Restriction enzyme mapping of vaccinia virus DNA. J. Virol. **43**:136–149.
6. **Devereux, J., P. Haeberli, and O. Smithies.** 1984. A comprehensive set of sequence analysis programs for the VAX. Nucleic Acids Res. **12**:387–395.
7. **Hänggi, 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.
8. **Hopp, T. P., and K. R. Woods.** 1981. Prediction of antigenic determinants. Proc. Natl. Acad. Sci. USA **78**:3821–3828.
9. **Jacks, T., and H. E. Varmus.** 1985. Expression of the Rous sarcoma virus pol gene by ribosomal frameshifting. Science **320**:1237–1242.
10. **Kozak, M.** 1984. Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
11. **Lipman, D. J., and W. R. Pearson.** 1985. Rapid and sensitive protein similarity searches. Science **227**:1435–1441.
12. **Mahr, A., and B. Roberts.** 1984. Arrangement of late mRNAs transcribed from a 7.1-kilobase EcoRI vaccinia virus DNA fragment. J. Virol. **49**:510–520.
13. **Maizel, J. V., Jr., and R. P. Lenk.** 1981. Enhanced graphic matrix analysis of nucleic acid and protein sequences. Proc. Natl. Acad. Sci. USA **78**:7665–7669.
14. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
15. **Maxam, A. M., and W. Gilbert.** 1977. A new method for sequencing DNA. Proc. Natl. Acad. Sci. USA **74**:560–564.
16. **Messing, J., and J. Vieira.** 1982. A new pair of M13 vectors for selecting either DNA strand of double-digest restriction fragments. Gene **19**:269–276.

17. **Moss, B.** 1985. Replication of poxviruses, p. 685–703. *In* B. N. Fields (ed.), Virology. Raven Press, New York.
18. **Niles, E. G., R. C. Condit, P. Caro, K. Davidson, L. Matusick, and J. Seto.** 1986. Nucleotide sequence and genetic map of the 16-kb vaccinia virus HindIII D fragment. Virology **153**:96–112.
19. **Pluciennicczak, A., E. Schroeder, G. Zettlmeisl, and R. E. Streeck.** 1985. Nucleotide sequence of a cluster of early and late genes in a conserved segment of the vaccinia virus genome. Nucleic Acids Res. **13**:985–998.
20. **Rohrmann, G., L. Yuen, and B. Moss.** 1986. Transcription of vaccinia virus early genes by enzymes isolated from vaccinia virions terminates downstream of a regulatory sequence. Cell **46**:1029–1035.
21. **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**:436–449.
22. **Sanger, F., S. Nicklen, and A. R. Coulson.** 1977. DNA sequencing with chain termination inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
23. **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.
24. **Slabaugh, M., R. D. Roseman, and C. Mathews.** 1988. Vaccinia virus-encoded ribonucleotide reductase: sequence conservation of the gene for the small subunit and its amplification in hydroxyurea-resistant mutants. J. Virol. **62**:519–527.
25. **Venkatesan, S., B. J. Baroudy, and B. Moss.** 1981. Distinctive nucleotide sequences adjacent to multiple initiation and termination sites of an early vaccinia virus gene. Cell **25**:805–813.
26. **Vieira, J., and J. Messing.** 1982. The pUC plasmids, an M13 mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. Gene **19**:259–268.
27. **Wei, C. M., and B. Moss.** 1975. Methylated nucleotides block 5′-terminus of vaccinia virus messenger RNA. Proc. Natl. Acad. Sci. USA **72**:318–322.
28. **Weinrich, S. L., and D. E. Hruby.** 1987. Noncoordinate regulation of a vaccinia virus late gene cluster. J. Virol. **61**:639–645.
29. **Weir, J. P., and B. Moss.** 1984. Regulation of expression and nucleotide sequence of a late vaccinia virus gene. J. Virol. **51**:662–669.
30. **Yuen, L., and B. Moss.** 1986. Multiple 3′ ends of mRNA encoding vaccinia virus growth factor occur within a series of repeated sequences downstream of T clusters. J. Virol. **60**:320–323.