JOURNAL OF VIROLOGY, June 1987, p. 1901–1908
0022-538X/87/061901-08$02.00/0

Vol. 61, No. 6

Sequence and Structural Organization of Murine Cytomegalovirus Immediate-Early Gene 1

GÜNTHER M. KEIL, ANGELIKA EBELING-KEIL, AND ULRICH H. KOSZINOWSKI*
Federal Research Centre for Virus Diseases of Animals, D-7400 Tübingen, Federal Republic of Germany

Received 10 December 1986/Accepted 3 March 1987

In murine cytomegalovirus, abundant immediate-early transcription originates from 0.769 to 0.815 map units of the genome. This region contains the immediate-early gene (gene *ie*I) which encodes pp89, a phosphoprotein active in transcriptional regulation. In this paper we report on the precise location, structural organization, and sequence of gene *ie*I. The predominant *ie*I transcript, a 2.75-kilobase mRNA, is generated by splicing and composed of four exons. The precise termini of the 2.75-kilobase mRNA and the positions of the exons were determined by nuclease digestion experiments with either 5′ or 3′ end-labeled DNA fragments or in vitro transcribed cRNA probes. Exons of 300, 111, 191, and 1,703 nucleotides are separated by introns of 825, 95, and 122 nucleotides. The first AUG is located in the second exon of 111 nucleotides, and a single open reading frame of 1,785 nucleotides predicts a protein of 595 amino acids with a calculated molecular weight of 66,713. The N-terminal region of the protein contains sequences similar to a consensus sequence of histone 2B proteins. The regulatory function of pp89 and the role of this protein as an immunodominant antigen are discussed in relation to the amino acid sequence.

As with other herpesviruses, expression of murine cytomegalovirus (MCMV) genes in infected cells is temporally controlled and regulated in a cascade fashion (7). In the absence of protein synthesis, predominant transcription occurs from three immediate-early (IE) transcription units designated ie1, ie2, and ie3 (10). IE proteins are the first proteins synthesized after infection or reactivation from latency. At least one herpesviral IE protein is required for synthesis of early and late mRNAs. The dominant IE protein of MCMV could be identified as a phosphoprotein (12) with an apparent molecular weight of 89,000 (pp89). Expression of pp89 encoded by gene *ie*I contained in transcription unit 1 is under control of a complex enhancer sequence (2). Functional analysis revealed that, after transient expression or stable transfer of ie1 into fibroblasts, pp89 activates heterologous promoters (14). Transactivation as a function of herpesviral IE proteins has also been demonstrated for ICP4, the major IE protein of herpes simplex virus type 1, and the major IE proteins of human cytomegalovirus (HCMV), pseudorabies virus, and varicella-zoster virus (4). Analysis of the different regulatory genes and comparison of the different proteins should help us to understand the mechanism by which this activation is achieved.

Other studies focused on the property of MCMV proteins as target structures for the protective immune response of the host (30, 31, 33). Viral antigens of the IE phase were detected which are recognized by cytolytic T lymphocytes (CTL) (31). Cells with specificity for MCMV IE antigens constitute the major proportion of MCMV-specific CTL (31). In subsequent studies we could demonstrate that (i) IE antigen presentation for CTL throughout the replication cycle correlates with the synthesis of pp89 (29) and (ii) cells that express pp89 after IE gene transfection are lysed by cloned MCMV IE antigen-specific CTL (15, 28).

A viral protein active in gene regulation should contain signals for nuclear transport, whereas viral antigen recognition by CTL requires the presence of the antigen at the plasma membrane. It was of interest to learn whether analysis of the sequence of the gene would explain how these two requirements are fulfilled. In this paper we present the physical organization, as well as the deduced amino acid sequence and secondary structure, of pp89 and the complete nucleotide sequence of *ie*I. The relation of the *ie*I gene to other known herpesviral IE genes and membrane expression of the protein are discussed.

* Corresponding author.

MATERIALS AND METHODS

Virus and cell culture. MCMV (Mouse salivary gland virus, strain Smith, ATCC VR-194) was propagated on BALB/c mouse embryo fibroblasts as previously described (3).

Isolation of IE RNA. Mouse embryo fibroblasts were infected with MCMV at a multiplicity of 20 to 40 PFU per cell in the presence of cycloheximide (50 μg/ml) (11). Preparation of whole-cell RNA and selection of poly(A)$^+$ RNA by oligo(dT)-cellulose chromatography were carried out as described previously (11).

Cloning of DNA. Recombinant plasmids were prepared by published procedures (20). BAL 31 treatment and generation of blunt ends with the Klenow polymerase were as previously described (20).

Analysis of RNA. IE RNA (5 μg) was either (i) hybridized to 0.03 to 0.1 pmol of end-labeled MCMV DNA fragments and then digested with S1 nuclease (2) or (ii) hybridized to 2×10^5 cpm of uniformly labeled cRNA, transcribed by the SP6 RNA polymerase, and then digested with RNase A and RNase T1 as previously described (23). Nuclease-resistant fragments were size fractionated on denaturing sequencing gels.

Labeling, strand separation, and sequence analysis of DNA. End labeling of cloned DNA fragments and preparation of uniquely labeled fragments by strand separation or secondary restriction enzyme cleavage were performed as previously described (20). DNA sequence analysis was done by the method of Maxam and Gilbert (21), except that the A+G modification reaction was 3 min at 22°C in 70% formic acid.

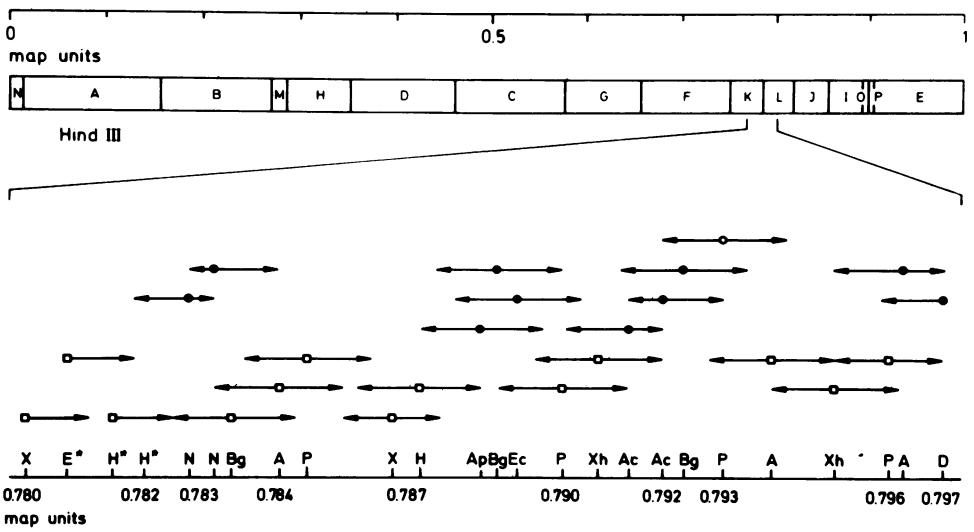

FIG. 1. *Hin*dIII cleavage map of the MCMV genome and sequencing strategy for *ie*I. The region between map units 0.780 and 0.797 is expanded. Restriction enzyme cleaveage sites used for DNA sequencing are shown. DNA fragments labeled at the 5′ or 3′ end were sequenced as indicated by the arrows. Arrows connected by open squares represent fragments analyzed with both 5′ and 3′ labeled ends. Open circles indicate fragments analyzed after 3′ labeling, and closed circles indicate fragments analyzed after 5′ labeling. Abbreviations: A, *Asp*718; Ac, *Acc*I; Ap, *Apa*I; Bg, *Blg*II; D, *Dra*I; E, *Eco*RI; Ec, *Eco*RV; H, *Hin*dIII; N, *Nco*I; P, *Pst*I; X, *Xba*I; Xh, *Xho*I. The *Eco*RI and *Hin*dIII sites marked with asterisks were introduced by insertion of synthetic linkers after BAL 31 digestion from the *Xba*I site (E*) at map unit 0.780 or the *Nco*I site (H*) at map unit 0.783.

RESULTS

Nucleotide sequence of MCMV *ie*I. The 5′ cap site of the 2.75-kb mRNA was located at a position 280 nucleotides (n) upstream of the *Xho*I site at map unit 0.795 (2) (Fig. 1). RNase protection experiments with cRNA transcribed in vitro by SP6 polymerase from the *Pst*I site (0.796 map units) to the *Dra*I cleavage site (0.797 map units) confirmed the start position at 50 n upstream of the *Pst*I site (0.796 map units). Transcription of *ie*I terminates to the right of map position 0.781 (10), and because cloned DNA from map units 0.780 to 0.782 did not select *ie*I-specific mRNA (unpublished data), the 3′ end of the 2.75-kb *ie*I transcript is probably close to map unit 0.782. Therefore, the entire sequence between map units 0.797 and 0.780 was obtained by repeated determinations by the method of Maxam and Gilbert (21). The physical map for the restriction cleavage sites used for 3′ and 5′ end labeling of DNA fragments of *ie*I is shown in Fig. 1. Figure 2 shows the nucleotide sequence of *ie*I, starting with the 5′ cap site of the 2.75-kb mRNA close to map unit 0.796 and terminating 34 n downstream of the polyadenylation signal AATAAA located at map unit 0.782.

Structural analysis of the *ie*I gene. Structural analysis of the *ie*I gene was performed by nuclease digestion experiments with either 5′ or 3′ end-labeled DNA fragments or uniformly labeled cRNA transcribed in vitro by SP6 RNA polymerase. The experimental strategy and results are shown in Fig. 3. Hybridization of IE RNA to labeled cRNA transcribed from map unit 0.769 (*Bam*HI) to map unit 0.797 (*Dra*I) (panel C, lane 6) revealed splicing. Protected fragments of 111, 191, 300, and about 1,700 n were detected. This finding was confirmed with smaller probes which were used to locate the exons precisely. A *Hin*dIII (0.787 map units)-*Pst*I (0.790 map units) DNA fragment labeled at the 5′ end (panel A1) revealed a protected sequence of 521 n. A cRNA transcribed from map units 0.787 (*Hin*dIII) to 0.797 (*Dra*I) (panel C, lane 5) showed the same 521-n fragment and also the three smaller fragments of 300, 191, and 111 n. A cRNA transcribed from map unit 0.787 (*Hin*dIII), which terminated already at map unit 0.794 (*Sac*I) (panel C, lane 4), did not show the 300-n fragment, whereas the other fragments (521, 191, and 111 n) remained detectable. This suggested that the 300-n fragment represents the first exon of the 2.75-kilobase (kb) IE mRNA and the large fourth exon starts 521 n upstream of the single *Hin*dIII site in *ie*I, whereas the two exons of 111 and 191 n should be located in between.

Hybridization of cRNA transcribed from map unit 0.787 (*Hin*dIII) to map unit 0.791 (*Xho*I) (panel C, lane 1) resulted in the 521-n fragment seen before and only one further fragment of 162 n, and it was concluded that the protected fragment of 162 n represented a part of the 191-n exon. Therefore, the 191-n fragment should represent the third exon, which begins at a position of 29 n upstream of the *Xho*I site close to map unit 0.791 and terminates 162 n downstream.

The location of the 111-n second exon was mapped by S1 analysis with the small 160-n *Acc*I restriction fragment located between map units 0.791 and 0.792 (panel A2). The fragment was labeled at both ends, and after separation each strand was used for sequencing and S1 analysis. After S1 digestion, protected fragments of 98 and 96 n were seen. cRNAs that terminated at the right *Acc*I cleavage site (panel C, lane 2) failed to detect the 109- to 111-n fragment. The 191-n exon appeared to be heterogeneous in some determinations (panel C, lanes 2 and 5); the size heterogeneity of the second 111-n exon, however, was always found when cRNAs containing the sequences of this exon were used for hybridization (panel C, lanes 4 to 6). A fragment size heterogeneity could be a technical artifact due to imprecise cleavage. However, this size heterogeneity could also occur in vivo since a T-rich region (nucleotides 1112 to 1123, Fig. 2) is followed by a repeated splice acceptor site (AGAGAG) sequence. This exon could therefore start after the first or second AG dinucleotide. As shown below, this differential

GTCGCAGTCT TCGGTCTGAC CACCGTAGAA CGCAGAGCTC CTCGCTGCAG TATCCTGCCC AAGAAAGAGC GACCTTCCCC GGCTACTGAA GCGCATCGGA CCCGCTCGGT TCTATTGGGA 120

CCACGCCCAT CCGCCATATT GGAACAGCCG CCATATTGGT AACGCCGCCA TCTTGATTGG GACCACCGCC CCTATTGGGA CCACCGCCCA ATTGGGACCA CCGCCCACCC AGCGACCCAA 240

GCCCAGAGGC TACGTCACCA AGACCCCCGG CTGCCCAGCG CCCCCTCGAG GACACACAAT GTAAGTAGAC CAAGGCCTTT TATTGGCTGC GGTGTTAAGG CAAGCTGCCT GTCTATCCCT 360

ATCTATCTAC TATCCACCTG GCTCTCTAGC TCCATACTGC GTACGGTGTT TGTTAGCTTA GGGTGATAGC AAGTAGAACG CTGGTGATGG TTGCTCTCTA TCAGTTTAGC GCTGGTGATG 480

GTGGCCTCCT TCTCCATTTC TTATATATAT ATGCTGGTGA TGGTCCCTCT TCCATGGCCG CGGTGTGCTT CGGTGCCGGG TACCGTAACG TGGTACAGAA AAAGGCCTGA CACGTACGCG 600

TCTCTCCCTA GGATGCTATG TGTCTGTAGC ACCGGCGGAG CTTGCGAGCT CGATGTCCAA GATAGAGTCT CCATGACTGC TCAGATCGTC TCCTGTGTTC ACACACCAGA TATTACAAGC 720

GACGGCCGCC ATGATAAGTC CAAAAATGAG CAGAATACAG AATATTGTAC TCAGAGGCTG CAGCCATCTT GGTGGCCTGC TGTGGTGCTC TTTTCCCGTG GAGGACTCCG TCAGTTCCTG 840

CGTGTTGTTG CTCGTACTGC GAAGATCTCT CTTGCCCCGT CCTGAAAACC TAGGCCCTGA CAGAAAAAAG GGAGAGTTAG GAGAGGGCAC AGAGGATTCT GTCTGTGTCA AGGGGGCTTC 960

CATGTTGTTT CTTTACCTGA GATCTCTTCC ACCAGTTGTT TTTTATCAGC AACTGACAGG TAGTATATCA AAATGAGTGC ACTGATGATT ATTAGCACTG TATACTTTAG ATATCTTCTG 1080

GTCTCTGTGG ACATCTGTTG ATGATAAAAA ATTATATTTT TTTAGAGAGA TGGAGCCCGC CGCACCCAGT TGCAACATGA TCATGATCGC CGATCAGGCA TCCGTCAACG CCCACGGCCG 1200

TCACTTGGAT GAGAACCGTG TCTACCCATC AGACAAGTAA GTTTCTGTTT CCATTTCCTC TCTTGTGACA CATTTCATGA CTCTGCATTT ACATAACCCC TGACTCATGG TCTTCCCACT 1320

CTTGTTCTCT CAGGGTGCCA GCTCATGTCG CCAACAAGAT CCTCGAGTCT GGAACCGAAA CCGTCCGCTG TGACCTGACT CTGGAGGACA TGTTGGGAGA CTATGAGTAT GATGATCCCA 1440

CTGAGGAAGA GAAGATCCTC ATGGACCGCA TCGCTGACCA CGTGGGGAAT GATGACAGCG ACATGGCCAT CAAGCATGCT GCAGGTGAGT GTCTCAGTTA TTCACATCAT GACTCTTCTA 1560

ACTATCTTAT CTCCTCTACC TGTCTCTGTC TTTCATATTC ACCCACACAG AACACTTGAG TTATACAAAC CACTCTTATA TTCCAGTGAG GAGCGTTCTG TTGTCCTGTA AGATTGCACA 1680

CCTCATGATA AAACAGAACT ATCAATCAGC CATCAACTCT GCTACCAACA TACTCTGTCA GCTAGCCAAT GATATCTTCG AGCGCATCGA AAGACAACGC AAGATGATAT ACGGCTGTTT 1800

CAGATCTGAG TTTGATAATG TCCAGCTAGG GAGGCTCATG TATGACATGT ACCCACACTT CATGCCCACT AATCTAGGGC CCTCAGAAAA GAGAGTCTGG ATGAGCTACG TAGGTGAGGC 1920

CATAGTGGCA GCCACCAACA TTGACCACGC ACTAGATGAG CGTGCCGCAT GGGCCAAGAC TGATTGTTCA TTGCCTGGGG AGTTTAAGCC TGAGTTGTGT GTCCTGGTTG GGGCCATCAG 2040

GAGGCTTCAT GATCCACCCT GTTACACCAA GCCTTTCCTG GATGCCAAGT CTCAGCTGGC AGTATGGCAG CAGATGAAGG CTATAGAGAG TGAGAGCGTA TCCACACATG TGGTAGTGGT 2160

GGAAGCTTTG AAGCTCAGGG AGAATCTGGC TAAGGCAGTG CAGGAAACGA TTGCCTATGA GAGACACCAA TATCACAGGG TGTGTCAGAT GATGTGTAAT AACATGAAAG ACCATCTAGA 2280

GACCACCTGT ATGCTGGCCA GAGGCAGAAC CTTGGCTACC TTAGCTGATC TGAGGAGCAC CAGATACAAT CTAGCACTTT TTCTTCTGAG TGAGATGCAC ATCTTTGATA GCTTCACCAT 2400

GCCTAGGATC AGAGGTGCTA TGAAACAAGC AAGGTGTATG AGTTATGTGG AGAGGACCAT CTCTCTGGCC AAGTTCAGAG AGCTAGCTGA TAGAGTCCAT AACAGGAGTG CACCATCACC 2520

CCAGGGAGTG ATAGAAGAAC AGCAGCAGGC AGGGGAGGAA GAACAGCAAC AGCAGCAGGA GATAGAGTAT GATCCTGAGA TGCCCCCTCT GGAAAGAGAG GAAGAGCAGG AGGATGAGCA 2640

GGTAGAGGAG GAGCCCCCTG CAGATGAGGA GGAAGGAGGA GCTGTTGGTG GTGTCACTCA AGAGGAACCT GCTGGTGAAG CCACTGAGGA AGCTGAAGAA GATGAGTCCC AGCCTGGACC 2760

AAGTGACAAC CAGGTGGTAC CTGAGTCTTC AGAGACACCA ACACCTGCTG AGGATGAGGA GACCCAAAGT GCTGATGAAG GAGAATCTCA GGAGCTGGAA GGCTCACAGC AACTCATCCT 2880

ATCCAGACCT GCTGCCCCTC TGACTGACAG CGAAACTGAT TCTGATTCTG AAGATGATGA TGAGGTGACC CGCATCCCAG TGGGATTCAG TCTCATGACA TCTCCTGTCC TGCAACCTAC 3000

GACCAGATCT GCAACAGCGG CAGCTTCTTC TGGCACTGCT CCTCGCCCAG CTCTCAAGAG GCAGTACGCC ATGGTGCATA CCAGATCCAA GTCTTCTGAG AATCAACAGC AGCCCAAGAA 3120

GAAGAGCAAG AAGTGAATCC ACAAGCCCAG CTCCCCACCC ATCACCATGG ACAATGTTTT TTTACTAACA CTTGGACAAT GATGGATACT TTTTTACTAA CACTTGGACA ATGATGATGA 3240

TACACTCCTC ACCTGCCCAC TTAGACACAA TTACTAACAC CACACCCCCT CTTTTATTTC TCTGTACTTA ATGTTTTCTG AATAAAGTGA TCCATGACCA CATTAAAGAC TTCTTTAACT 3360

FIG. 2. Nucleotide sequence of *ie*I. The sequence shown starts with the 5′ cap site (+1) of the 2.75-kb mRNA (2) and extends 34 n downstream of the AATAAA polyadenylation signal (boxed). The exon sequences are underlined, and the open reading frame, translated into the pp89, is indicated by a heavy line. The arrow shows the 3′ end of the 2.75-kb mRNA.

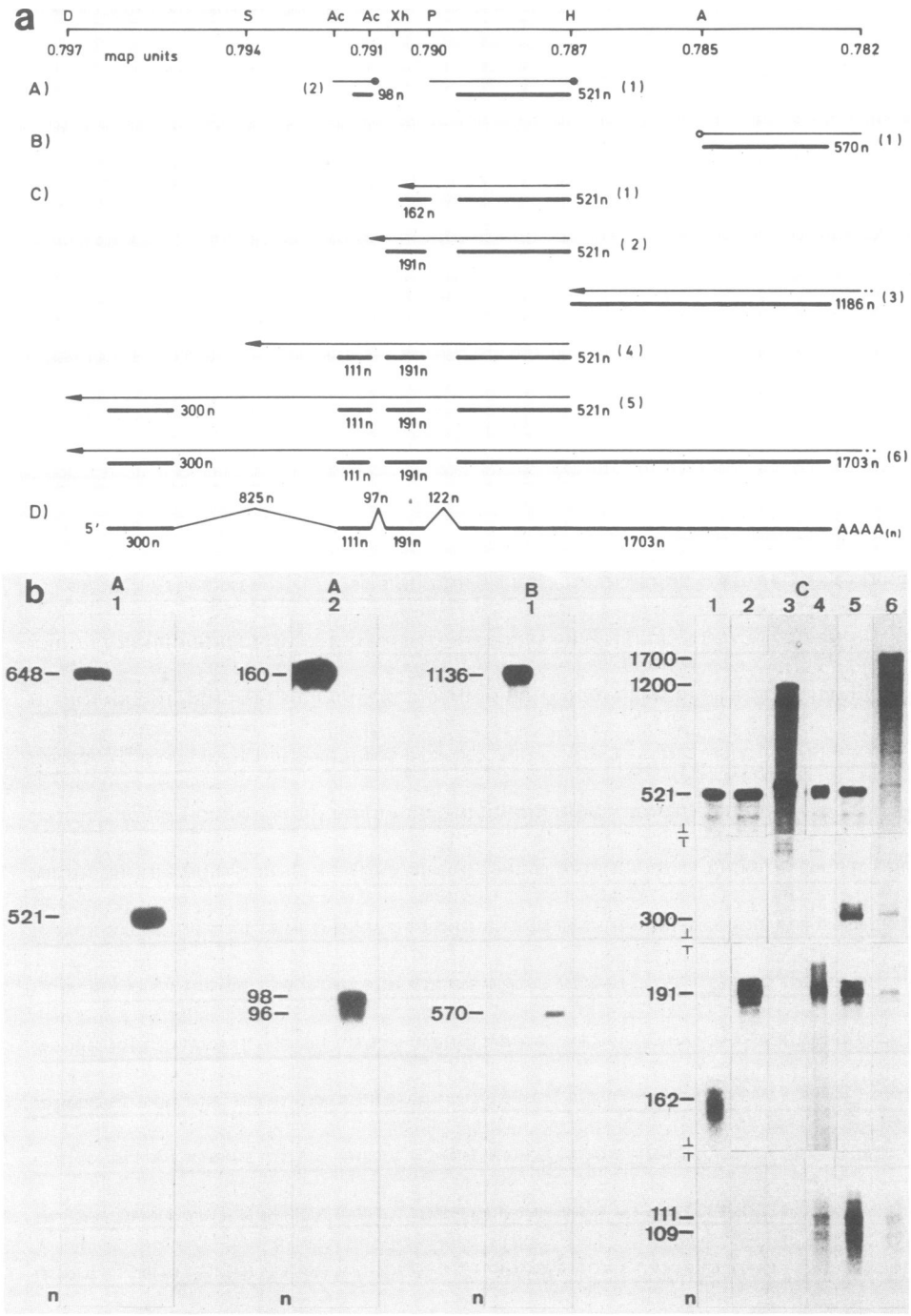

FIG. 3. Structural analysis of *ie*I RNA. The experimental protocol for mapping of the 5′ and 3′ ends, as well as the locations of exons, is depicted in panel a, and experimental data are shown in panel b. To present all relevant information in one figure, the autoradiographs of some sequencing gels were cropped at identical positions and refitted as indicated (panel C, lanes 1 to 6). (a) Experimental protocol. IE RNA, hybridized to (i) 5′-labeled (A, closed circle) or 3′-labeled (B, open circle) DNA fragments or (ii) uniformly labeled cRNA (C, arrows indicate the direction and termination of cRNA transcription) were incubated with S1 nuclease (A and B) or RNases A and T1 (C). The cRNAs used in C, lines 3 and 6, started at map units 0.780 and 0.769, respectively. Solid bars show the locations of protected fragments, and the number of nucleotides is indicated. The results are summarized in D. The abbreviations used are the same as in Fig. 1. S, *Sac*I. (b) Autoradiographs of protected fragments separated by electrophoresis in sequencing gels. Sequencing gels (60 cm) of 8%, 6%, or 4% polyacrylamide were used. The sizes of labeled fragments before S1 nuclease treatment in A and B were taken from the sequence. When sequencing ladders served as size markers, the size of the protected fragments was determined by using $+1n$ correction (6). Additional size markers were (i) from commercial suppliers and (ii) pBR322 DNA cleaved with *Hpa*II.

splicing would have no effect on the amino acid sequence of the protein. In conclusion, the second exon of 111 n spans from 96 n upstream to 15 n downstream of the *Acc*I site (map unit 0.792).

The 521-n fragment detected by cRNAs that started at map unit 0.787 (*Hin*dIII) was not found with cRNA that started at map unit 0.769 (panel C, lane 6), which revealed instead a protected fragment of about 1,700 n. cRNA transcribed from map unit 0.780 (*Xba*I) to map unit 0.787 (*Hin*dIII) (panel C, lane 3) produced a protected fragment of about 1,200 n. Taken together, these results suggested that the fourth exon begins 521 n upstream and terminates about 1,200 n downstream of the *Hin*dIII site at map unit 0.787. An additional protected fragment of 530 n (panel C, lane 3) probably represents products of transcription unit ie3 (10). The analysis of this region is under study. To locate the 3′ end more precisely, S1 analysis was performed with a DNA fragment labeled at the 3′ end after cleavage with *Asp*718 at map unit 0.785 and then digested with *Xba*I, which cleaved at map unit 0.780. Hybridization of this fragment to IE RNA followed by digestion with nuclease S1 resulted in a protected fragment of 570 nucleotides (panel B1). This result placed the 3′ end of the fourth exon 23 n downstream of the AATAAA RNA polyadenylation consensus sequence (27). There was no indication of an additional exon within the major IE mRNA. Thus, the 2.75-kb IE mRNA coding for pp89 contains four exons of 300, 111, 191, and 1,703 n with a sum of 2.35 kb. The sizes of the introns are 825, 97, and 122 n (Fig. 3D). The splice signals (Fig. 2) follow the AG-GT rule for splice junctions with the presence of C- or T-rich regions upstream of the AG dinucleotide splice acceptor sequence (25, 35) and the GT$^{A}_{G}$AGT hexanucleotide at the splice donor sites (25, 35). The initiation and polyadenylation signals are also consistent with the respective eucaryotic transcription signals. In summary, *ie*I is a viral gene which uses the normal eucaryotic transcription signals for expression.

Deduced amino acid composition and predicted secondary structure of IE pp89. The sequence of the major IE mRNA, consisting of the four exons, contains an open reading frame of 1,785 n that starts with an AUG located within the second exon and terminates with a UGA. Differential splicing at the 5′ end of the second exon upstream of the first AUG therefore has no effect on the amino acid sequence of the protein. This open reading frame codes for a polypeptide of 595 amino acids (Fig. 4). It consists of 105 (17.6%) acidic and 60 (10.1%) basic amino acids. The COOH terminus is basic, followed by an acidic region, in which 51 (38%) of 136 amino acids are 38 Glu and 13 Asp. A similar acidic region is found at the COOH terminus of the HCMV major IE protein (37), where within 51 residues 22 are Glu and 2 are Asp. A second acidic region is found between amino acids 51 and 89 near the NH_2 terminus, where within 39 amino acids 6 are Glu and 8 are Asp.

Comparisons of the *ie*I nucleotide and amino acid sequences with known eucaryotic viral and cellular sequences were carried out. No significant matches with viral proteins were found except a nuclear location signal found in polyomavirus and adenovirus (17, 18) located at the COOH terminus of pp89. The major IE proteins of HCMV and pp89 show similarities only in a Glu-rich part of the amino acid sequence. A total of 23 amino acids in the amino-terminal portion of pp89 were found to be identical with amino acids of the histone H2B consensus sequence. Identical amino acids at homologous positions were found in a region of 63 amino acids located in the amino-terminal part of both proteins.

Based on the amino acid sequence, the secondary structure of pp89 was predicted as described by Chou and Fasman (1) (Fig. 5). Based on this analysis, we have eliminated the introns and have recently been able to express the open reading frame in a vaccinia virus (H. Volkmer et al., manuscript in preparation).

DISCUSSION

We have shown previously that gene *ie*I of MCMV is located between map units 0.781 and 0.796 (10, 11). The 5′ end of the 2.75-kb mRNA transcribed from *ie*I was mapped downstream of a long enhancer-promoter element and initiates with a 304-n nontranslated leader sequence 23 n downstream of the TATA box at the sequence GTCGCA (2). The first ATG is located at the 5′ end of the second 111-n exon, which is followed by the third 191-n exon. The major constituent of the mRNA is contributed by the 3′ terminal exon of 1,703 n. It terminates 23 n downstream of the consensus sequence for polyadenylation, AATAAA (27). This location of *ie*I is in full agreement with data published before (2, 10, 14). The splice donor and acceptor sequences follow the consensus sequences of other eucaryotic and viral mRNAs. The GT$^{A}_{G}$AGT sequence was present at all three splice donor sites. No additional splice donor consensus sequence is present within *ie*I. A T- or C-rich region was always found 5′ to the splice acceptor sites (25, 35).

The amino acid sequence of pp89 was deduced from the mRNA sequence. The first AUG in the mRNA appears at n 305, a second is at n 332, and a third is at n 338. All three AUGs could represent the codon for the first amino acid of pp89 because they contain the consensus sequence for translation initiation, A/G NNATG (16). Since translation of most eucaryotic proteins starts at the first AUG, it is likely that the AUG at n 305 represents the codon for the NH_2 terminal methionine of pp89. A single open reading frame extends from n 305 to n 2,090, where a UGA stop codon is located. This 1,785-n open reading frame is translated into a protein of 595 amino acids with a calculated molecular weight of 66,713. The distribution of charged amino acids and the phosphorylation of pp89 of MCMV are most probably responsible for overestimation of the molecular weight of pp89 after separation in denaturing sodium dodecyl sulfate-polyacrylamide gels. It should be noted that molecular weights of the HCMV major IE protein (37), ICP4 (Vmw 175) (22), and IE110 of HSV-1 (26) have also been overestimated. In all three IE proteins the apparent molecular weight is, as with pp89, about 1.3-fold higher than the calculated molecular weight.

pp89 is a nuclear protein and should therefore contain nuclear location signals. A nuclear location signal has been defined for the simian virus 40 large-T protein and consists of a stretch of predominantly basic amino acids (Pro-Lys-Lys-Lys-Arg-Lys-Val) located in the amino-terminal part of the protein (18, 34). This sequence appears to represent a prototypic nuclear location signal also found in other viral proteins (36). pp89 contains a similar sequence (Pro-Lys-Lys-Lys-Ser-Lys-Lys) located at the carboxy terminus. The possibility that such a sequence functions as a nuclear location signal is derived from deletion studies of a similar sequence present in the carboxy termini of the E1A proteins of adenovirus 5, which resulted in a reduced rate of nuclear accumulation (17). We therefore suggest that this sequence represents the nuclear location signal of pp89.

In contrast to the gene coding for ICP4 of herpes simplex virus, which is transcribed into an unspliced mRNA (22), the

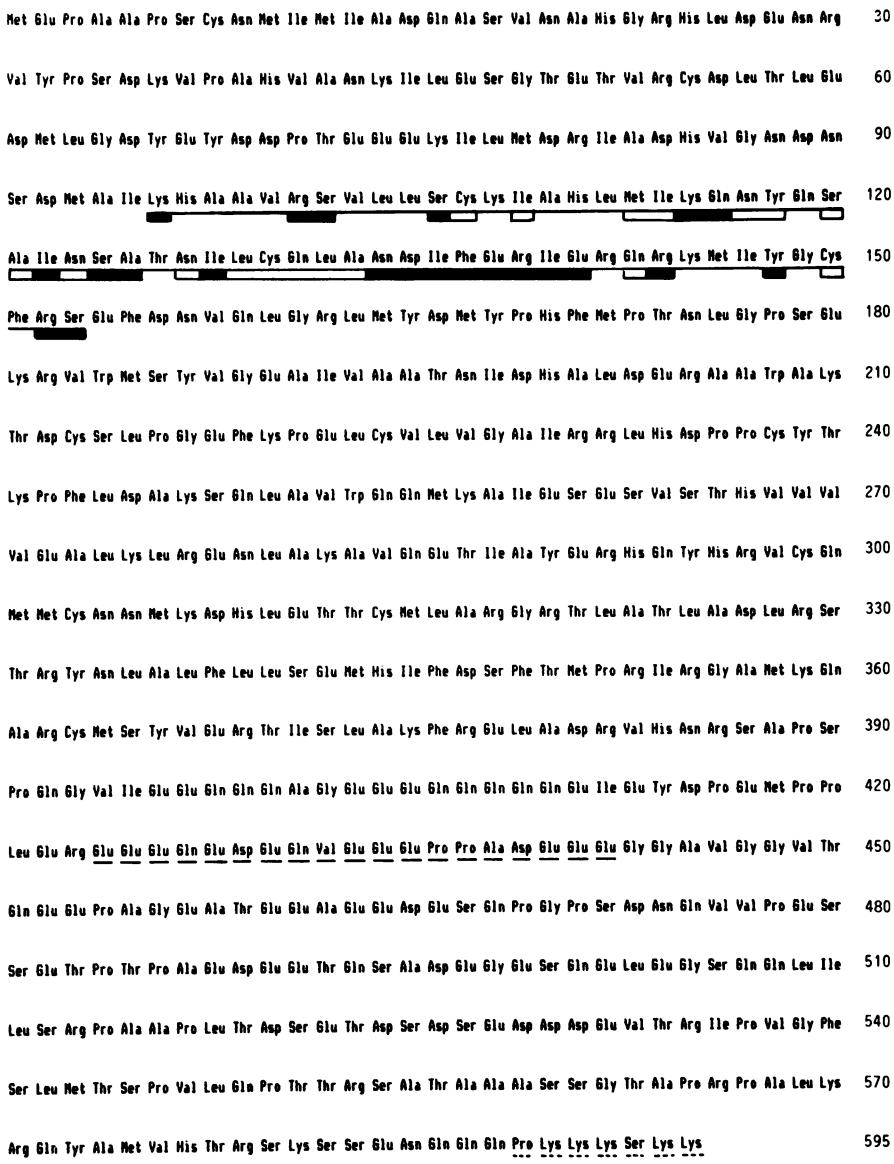

FIG. 4. Deduced amino acid sequence of IE pp89. Broken line, homology to HCMV major IE protein; dotted line, homology to simian virus 40 large T antigen; solid line, region homologous to histone H2B. Solid bars indicate identical amino acids at the same position in both proteins, and open bars indicate conservative amino acid exchanges (comparison with histone H2B) that do not affect the polarity or charge of the protein.

major IE gene in both MCMV and HCMV (37) contains four exons. In both viruses the first exon is nontranslated, the second exon starts with the first ATG, and the fourth exon is the major component of the mRNA. In addition, the functions of both gene products seem to be very similar (4, 14). Remarkably, there was no significant homology between the nucleotide sequences of these genes. Also, after comparing the protein sequences, only one region with homology extending over more than 8 amino acids was found in a Glu-rich part of both IE proteins, where amino acids 424 to 434 of pp89 and 426 to 437 of the major IE protein of HCMV share 8 of 12 amino acids.

A notable similarity was found between amino acids 27 to 87 of the histone H2B consensus sequence (39) and amino acids 96 to 153 of pp89. In this region, 23 (37%) of 63 amino acids were found to be identical and located at a corresponding position in the amino acid sequence. The codons for these 23 amino acids display 83% identity, whereas only 48% homologous nucleotides would be required to encode these amino acids. The similarity within this region increases to 62% when conservative amino acid exchanges are considered as similar. The H2B proteins of different species maintain a high degree of sequence conservation in certain regions of the polypeptide, including amino acids 67 to 73 (Asn-Asp-Ile-Phe-Glu-Arg-Ile), which are represented in an identical sequence in pp89 (amino acids 134 to 140). The region between residues 37 and 114 of H2B is important for the interaction of H2B with histone H2A (8, 9). It could be speculated that this region in pp89 plays a role in an interaction of IE pp89 with cellular chromatin.

After transfection of the DNA sequences which encode pp89, the lymphocyte-detected IE antigen (LYDIEA) (32)

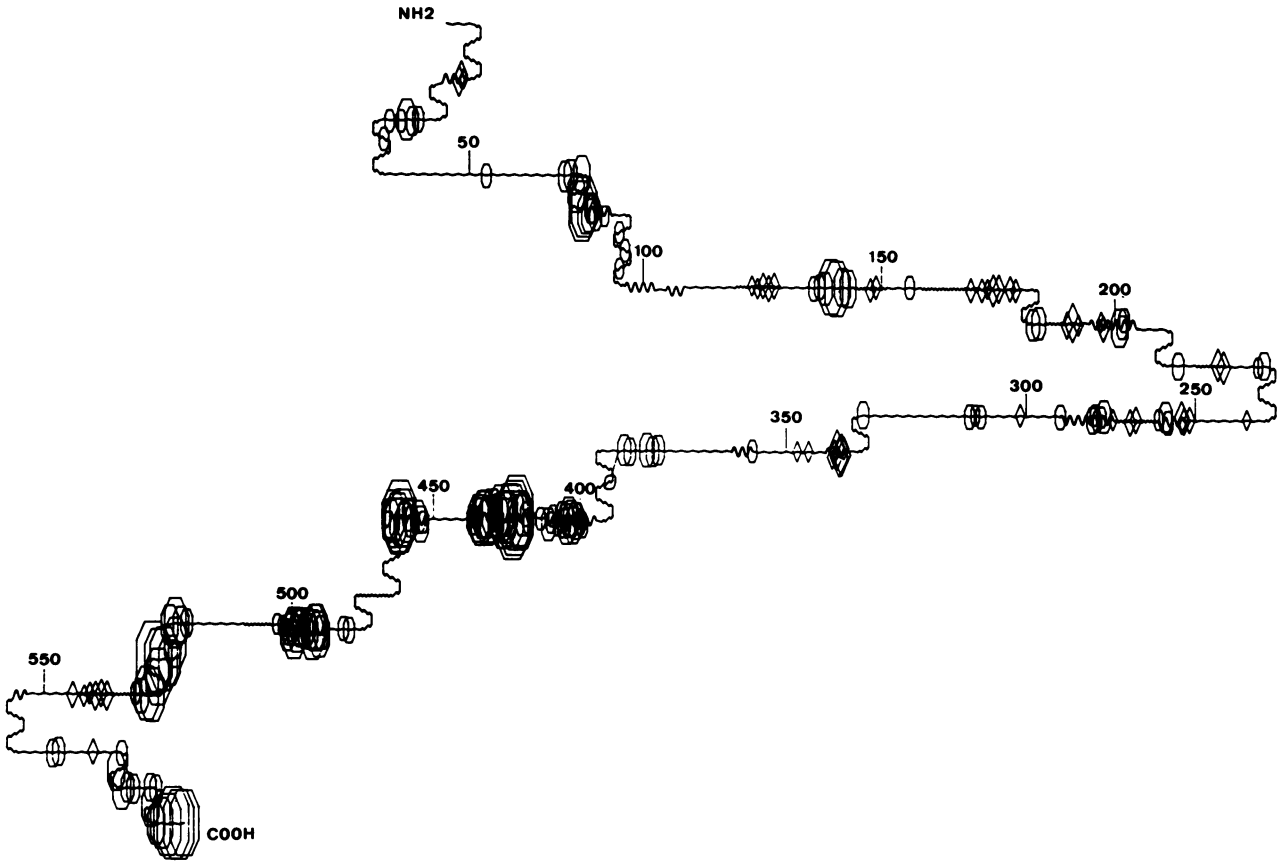

FIG. 5. Secondary structure of IE pp89. The method of Chou and Fasman (1) was used. Octagons indicate regions of hydrophilicity of >0.7, and rhombic symbols stand for regions of hydrophobicity of <−0.7.

was found to be expressed (15). The only other herpesviral protein reported to play a role in the host response to herpesviral infection is LYDMA, an antigen detected by specific T lymphocytes on Epstein-Barr virus-transformed cells. It is thought that LYDMA is a part of the latent-infection membrane protein. Analysis of the amino acid sequence of the latent-infection membrane protein predicted repeated hydrophobic regions separated by hydrophilic segments, an arrangement indicative of membrane insertion (5, 19). The predicted amino acid sequence of pp89, however, does not contain such regions, and there is no evidence for membrane-spanning domains. Thus, pp89 is structurally different from the Epstein-Barr virus latent-infection membrane protein, and pp89 probably requires processing for expression at the cell membrane. The absence of structural characteristics that indicate cell surface expression does not contradict the immunological findings. Nonglycosylated internal viral proteins can reach the cell membrane by alternative routes of processing and degradation (24). The known features of pp89 appear similar to some properties of a simian virus 40 nonstructural protein, the large-T antigen. Similar to pp89, simian virus 40 large-T antigen is a phosphoprotein involved in regulation of simian virus 40 genes. The majority of simian virus 40 large-T antigen in the cell is concentrated in the nucleus, but a small amount of the protein is present on the cell surface, probably after acylation (13). Surface-expressed simian virus 40 large-T antigen serves as a target structure for CTL (38). Work is in progress to study the surface expression of pp89.

ACKNOWLEDGMENTS

This study was supported by grant Ko 571/8 from the Deutsche Forschungsgemeinschaft and by grant PTB 8614 from the Bundesministerium für Forschung und Technologie.

We thank A. Klug for comments, H. Niller and W. Bodemer for technical help, S. Modrow for secondary-structure analysis, B. Mächtig, G. Röhrs, and M. Votteler for technical assistance, and S. Grau for manuscript preparation.

LITERATURE CITED

1. **Chou, P. Y., and G. D. Fasman.** 1974. Prediction of protein conformation. Biochemistry **13**:222–245.
2. **Dorsch-Häsler, K., G. M. Keil, F. Weber, M. Jasin, W. Schaffner, and U. H. Koszinowski.** 1985. A long and complex enhancer activates transcription of the gene coding for the highly abundant immediate early mRNA in murine cytomegalovirus. Proc. Natl. Acad. Sci. USA **82**:8325–8329.
3. **Ebeling, A., G. M. Keil, E. Knust, and U. H. Koszinowski.** 1983. Molecular cloning and physical mapping of murine cytomegalovirus DNA. J. Virol. **47**:421–433.
4. **Everett, R. D.** 1984. Trans activation of transcription of herpes virus products: requirement for two HSV-1 immediate-early polypeptides for maximum activity. EMBO J. **3**:3135–3141.
5. **Fennewald, S., V. van Santen, and E. Kieff.** 1984. Nucleotide sequence of an mRNA transcribed in latent growth-transforming virus infection indicates that it may encode a membrane protein. J. Virol. **51**:411–419.
6. **Green, M. R., and R. G. Roeder.** 1980. Definition of a novel promoter for the major adenovirus-associated virus mRNA. Cell **22**:231–242.
7. **Honess, R. W., and B. Roizman.** 1974. Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthe-

sis of at least three groups of viral proteins. J. Virol. **14**:8–19.
8. **Isenberg, I.** 1979. Histones. Annu. Rev. Biochem. **48**:159–191.
9. **Kedes, L. H.** 1979. Histone genes and histone messengers. Annu. Rev. Biochem. **48**:837–870.
10. **Keil, G. M., A. Ebeling-Keil, and U. H. Koszinowski.** 1987. Immediate-early genes of murine cytomegalovirus: location, transcripts, and translation products. J. Virol. **61**:526–533.
11. **Keil, G. M., A. Ebeling-Keil, and U. H. Koszinowski.** 1984. Temporal regulation of murine cytomegalovirus transcription and mapping of viral RNA synthesized at immediate early times after infection. J. Virol. **50**:784–795.
12. **Keil, G. M., M. R. Fibi, and U. H. Koszinowski.** 1985. Characterization of the major immediate-early polypeptides encoded by murine cytomegalovirus. J. Virol. **54**:422–428.
13. **Klockmann, U., and W. Deppert.** 1983. Acylated simian virus 40 large T antigen: a new subclass associated with a detergent-resistant limina of the plasma membrane. EMBO J. **2**:1151–1157.
14. **Koszinowski, U. H., G. M. Keil, H. Volkmer, M. R. Fibi, A. Ebeling-Keil, and K. Münch.** 1986. The 89,000-M_r murine cytomegalovirus immediate-early protein activates gene transcription. J. Virol. **58**:59–66.
15. **Koszinowski, U. H., M. J. Reddehase, G. M. Keil, and J. Schickedanz.** 1987. Host immune response to cytomegalovirus: products of transfected viral immediate-early genes are recognized by cloned cytolytic T lymphocytes. J. Virol **61**:2054–2058.
16. **Kozak, M.** 1983. Comparison of initiation of protein synthesis in procaryotes, eucaryotes, and organelles. Microbiol. Rev. **47**:1–45.
17. **Krippl, B., B. Ferguson, N. Jones, M. Rosenberg, and H. Westphal.** 1985. Mapping of functional domains in adenovirus E1A proteins. Proc. Natl. Acad. Sci. USA **81**:6988–6992.
18. **Lanford, R. E., and J. S. Butel.** 1984. Construction and characterization of an SV40 mutant defective in nuclear transport of T antigen. Cell **37**:801–813.
19. **Liebowitz, D., D. Wang, and E. Kieff.** 1986. Orientation and patching of the latent infection membrane protein encoded by Epstein-Barr virus. J. Virol. **58**:233–237.
20. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
21. **Maxam, A. M., and W. Gilbert.** 1980. Sequencing with base specific chemical cleavages. Methods Enzymol. **65**:499–560.
22. **McGeoch, D. J., A. Dolan, S. Donald, and D. H. K. Brauer.** 1986. Complete DNA sequencing of the short repeat region in the genome of herpes simplex virus type 1. Nucleic Acids Res. **14**:1727–1745.
23. **Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, 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.
24. **Mills, K. H. G.** 1986. Processing of viral antigens and presentation to class II restricted T cells. Immunol. Today **7**:260–263.
25. **Mount, S. M.** 1982. A catalog of splice junction sequences. Nucleic Acids Res. **10**:459–472.
26. **Perry, L. J., F. R. Rixon, R. D. Everett, M. C. Frame, and D. J. McGeoch.** 1986. Characterization of the IE110 gene of herpes simplex virus type 1. J. Gen. Virol. **67**:2365–2380.
27. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ Non-coding region sequences in eucaryotic messenger RNA. Nature (London) **263**:211–214.
28. **Reddehase, M. J., H.-J. Bühring, and U. H. Koszinowski.** 1986. Cloned long-term cytolytic T-lymphocyte line with specificity for an immediate-early membrane antigen of murine cytomegalovirus. J. Virol. **57**:408–412.
29. **Reddehase, M. J., M. R. Fibi, G. M. Keil, and U. H. Koszinowski.** 1986. Late-phase expression of a murine cytomegalovirus immediate-early antigen recognized by cytolytic T lymphocytes. J. Virol. **60**:1125–1129.
30. **Reddehase, M. J., G. M. Keil, and U. H. Koszinowski.** 1984. The cytolytic T lymphocyte response to the murine cytomegalovirus. I. Distinct maturation stages of cytolytic T lymphocytes constitute the cellular immune response during acute infection of mice with the murine cytomegalovirus. J. Immunol. **132**:482–489.
31. **Reddehase, M. J., G. M. Keil, and U. H. Koszinowski.** 1984. The cytolytic T lymphocyte response to the murine cytomegalovirus. II. Detection of virus replication stage-specific antigens by separate populations of in vivo active cytolytic T lymphocyte precursors. Eur. J. Immunol. **14**:56–61.
32. **Reddehase, M. J., and U. H. Koszinowski.** 1984. Significance of herpesvirus immediate early gene expression in cellular immunity to cytomegalovirus infection. Nature (London) **312**:369–371.
33. **Reddehase, M. J., F. Weiland, K. Münch, S. Jonjic, A. Lüske, and U. H. Koszinowski.** 1985. Interstitial murine cytomegalovirus pneumonia after irradiation: characterization of cells that limit viral replication during established infection of the lungs. J. Virol. **55**:264–273.
34. **Richardson, W. D., B. L. Roberts, and A. E. Smith.** 1986. Nuclear location signals in polyoma virus large-T. Cell **44**:77–85.
35. **Sharp, P. A.** 1981. Speculations on RNA splicing. Cell **23**:643–646.
36. **Smith, A. E., D. Kalderon, B. L. Roberts, W. D. Colledge, M. Edge, P. Gillett, A. F. Markham, E. Paucha, and W. D. Richardson.** 1985. The nuclear location signal. Proc. R. Soc. Lond. B Biol. Sci. **226**:43–58.
37. **Stenberg, R. M., D. R. Thomsen, and M. F. Stinski.** 1984. Structural analysis of the major immediate early gene of human cytomegalovirus. J. Virol. **49**:190–199.
38. **Tevethia, S. S., D. C. Flyer, and R. Tjian.** 1980. Biology of simian virus 40 (SV40) transplantation antigen (TrAg). Virology **107**:13–23.
39. **Wells, D. E.** 1986. Compilation analysis of histones and histone genes. Nucleic Acids Res. **14**:119–149.