

# Promoter-regulatory region of the major immediate early gene of human cytomegalovirus

(RNA polymerase II/transcriptional control elements)

DARRELL R. THOMSEN, RICHARD M. STENBERG, WILLIAM F. GOINS, AND MARK F. STINSKI

Department of Microbiology, School of Medicine, University of Iowa, Iowa City, IA 52242

Communicated by Bernard Roizman, September 19, 1983

**ABSTRACT** The DNA templates containing immediate early (IE) genes of human cytomegalovirus (CMV) were transcribed *in vitro* by using a HeLa cell extract. When IE region 1, 2, and 3 were used, transcription was detected qualitatively only from IE region 1. Transcription was detected with DNA representing IE region 2 when the IE region 1 promoter was not present. DNA sequence analysis of the upstream regulatory region of IE region 1 detected two distinct repeats of 19 and 18 nucleotides, both being repeated four times. A putative cruciform structure could form through the surrounding sequences with each 18-nucleotide repeat being located in the unpaired region. The potential secondary structure and the repeat sequences in the regulatory region of IE region 1 are presumably related to the high level of transcription of this IE gene.

Human cytomegalovirus (CMV), a member of the herpesvirus classification group, has a large double-stranded DNA genome of 240 kilobases (kb). The viral genome consists of a long and short unique region flanked by different repeat sequences that are inverted relative to each other. Four genome arrangements, resulting from the possible combination of inversions of the two sections of the genome, are present in DNA preparations in approximately equal amounts (1-7).

At immediate early (IE) times after infection—i.e., in the absence of *de novo* viral protein synthesis, 88% or more of the viral RNA originates from a region in the long unique component of the viral genome (6, 8, 9) between 0.660 and 0.751 map units for the Towne strain (8, 10). One or more of the IE viral genes presumably codes for a viral regulatory protein that stimulates transcription from other regions of the viral genome.

Based on the high steady-state levels of viral mRNA and the abundance of its translation product in the infected cell, the IE gene between 0.739 and 0.751 map units is highly expressed and has been designated IE gene 1 or the major IE gene (11, 12). Adjacent IE genes from 0.732 to 0.739 (region 2) and from 0.709 to 0.728 (region 3) map units are expressed at relatively low levels and, consequently, are considered minor IE genes (12). Transcription under IE conditions is also detectable from another adjacent region of approximately 0.660-0.685 map units (6, 8), but we have failed to translate *in vitro* hybrid-selected RNA encoded by this region; consequently, the expression of this region requires further investigation.

Because CMV IE gene expression is dominated *in vivo* by the expression of a single gene, we were interested in determining the properties of the promoter-regulatory region and whether or not region 1 was highly transcribed *in vitro* relative to regions 2 and 3. The DNA sequence upstream of IE region 1 of CMV may constitute the earliest point at which

expression of the viral genome is regulated at the level of transcription.

## MATERIALS AND METHODS

**Genetic Map and Recombinant Plasmids.** Physical maps of the entire CMV genome were developed by LaFemina and Hayward (5). The cloning, purification, and characterization of recombinant plasmids containing insertions of CMV DNA have been described (13). Recombinant plasmid pCB42 and pSmaF are gifts from R. LaFemina and P. Weil, respectively. A physical map of the *Xba* I fragment E and the recombinant plasmids representing this region have been described by Stinski *et al.* (12). Restriction endonucleases were obtained from Bethesda Research Laboratories or New England Biolabs. The conditions were as described by the supplier. After digestion, the DNA was extracted twice with phenol/chloroform, 1:1 (vol/vol), and twice with chloroform, precipitated twice with ethanol, and resuspended in 10 mM Tris-HCl, pH 7.9/1 mM EDTA.

**Preparation of HeLa Cell Extracts.** HeLa cells were obtained from W. C. Summers. Spinner cultures were grown to a density of  $4-5 \times 10^5$  cells per ml. *In vitro* transcription extracts were prepared by the method of Manley *et al.* (14).

***In Vitro* Transcription and RNA Fractionation.** *In vitro* transcription was as described by Manley *et al.* (14). Recombinant plasmids cut with various restriction enzymes to generate linear templates were at a concentration of 100  $\mu$ g per ml. Some reactions contained  $\alpha$ -amanitin (1  $\mu$ g/ml; Sigma) to inhibit RNA polymerase II activity. The  $^{32}$ P-labeled RNA was subjected to electrophoresis in 1.5% agarose gels containing 10 mM methylmercury (II) hydroxide as described by Bailey and Davidson (15). Molecular weight standards were 23S (3.3 kb) and 16S (1.7 kb) *Escherichia coli* rRNA (16), 28S (5.3 kb) and 18S (2.0 kb) human cell rRNA (17), and approximately 0.160 kb tRNA. To visualize the RNA, the slab gels were stained in a solution containing 0.5 M ammonium acetate, 0.005 M 2-mercaptoethanol, and 1  $\mu$ g of ethidium bromide per ml. The gels were dried and exposed to Kodak X-Omat AR film. RNA sizes were interpolated from a standard curve.

**DNA Sequence Analysis.** Recombinant plasmid pXEP 22 containing the 5' end of the major IE RNA (18) and its promoter-regulatory region (12) were digested with the appropriate restriction endonucleases, fractionated by electrophoresis in agarose or acrylamide gels, and eluted electrophoretically. The methods used for labeling DNA *in vitro* and for sequence determination by the chemical modification and degradation procedure of Maxam and Gilbert (19) have been described (18).

**Estimation of Secondary Structure.** The free energies for the base-paired regions in the putative cruciform structures were calculated by the method of Tinoco *et al.* (20).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. § 1734 solely to indicate this fact.

Abbreviations: IE, immediate early; CMV, cytomegalovirus; kb, kilobase(s).

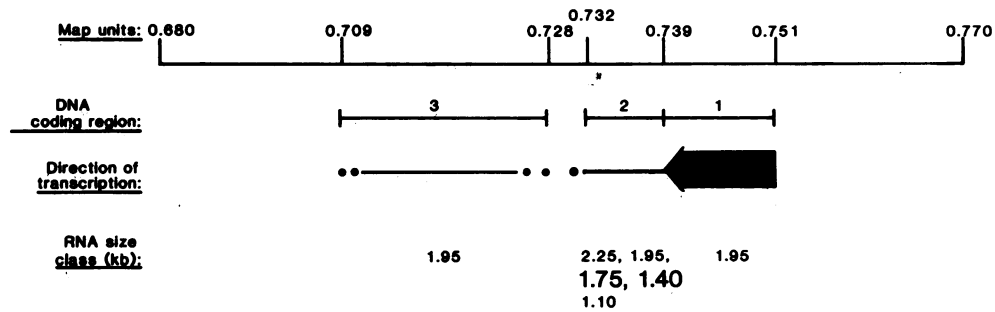


FIG. 1. Summary of the IE RNAs coded within the *Xba* I fragment E DNA region. The map units of coding regions 1, 2, and 3 depict the limits of the probes used to detect viral RNAs. The direction of transcription is indicated for coding region 1. The direction of transcription in region 2 requires further investigation to determine which direction predominates at IE and early times after infection. The thickness of the bar represents the relative abundance of the IE RNAs originating from the various coding regions. The size classes of the viral RNAs *in vivo* are indicated in kb. The data for the above is taken from Stinski *et al.* (12).

## RESULTS

### *In Vitro* Transcription Using DNA Templates for IE Genes.

At least three promoters between 0.709 and 0.751 map units influence IE transcription after infection with CMV (12). One IE viral gene (IE region 1) is highly expressed, whereas the other (IE regions 2 and 3) are expressed at relatively low levels based on steady-state levels of mRNA in the cytoplasm (12). These viral genes are also referred to as the major and minor IE genes. Fig. 1 summarizes the map location

of these viral genes, direction of transcription for IE region 1, and the RNA size classes originating from the various regions as described (12). We previously had designated the transcription in IE region 2 from left to right based on 3' cDNA hybridizations. However, recent evidence obtained by one of us (unpublished data) does not support this interpretation. The direction of IE transcription in this region requires further investigation.

*In vitro* transcription of these IE DNA templates was analyzed to obtain a general map location of the promoters and

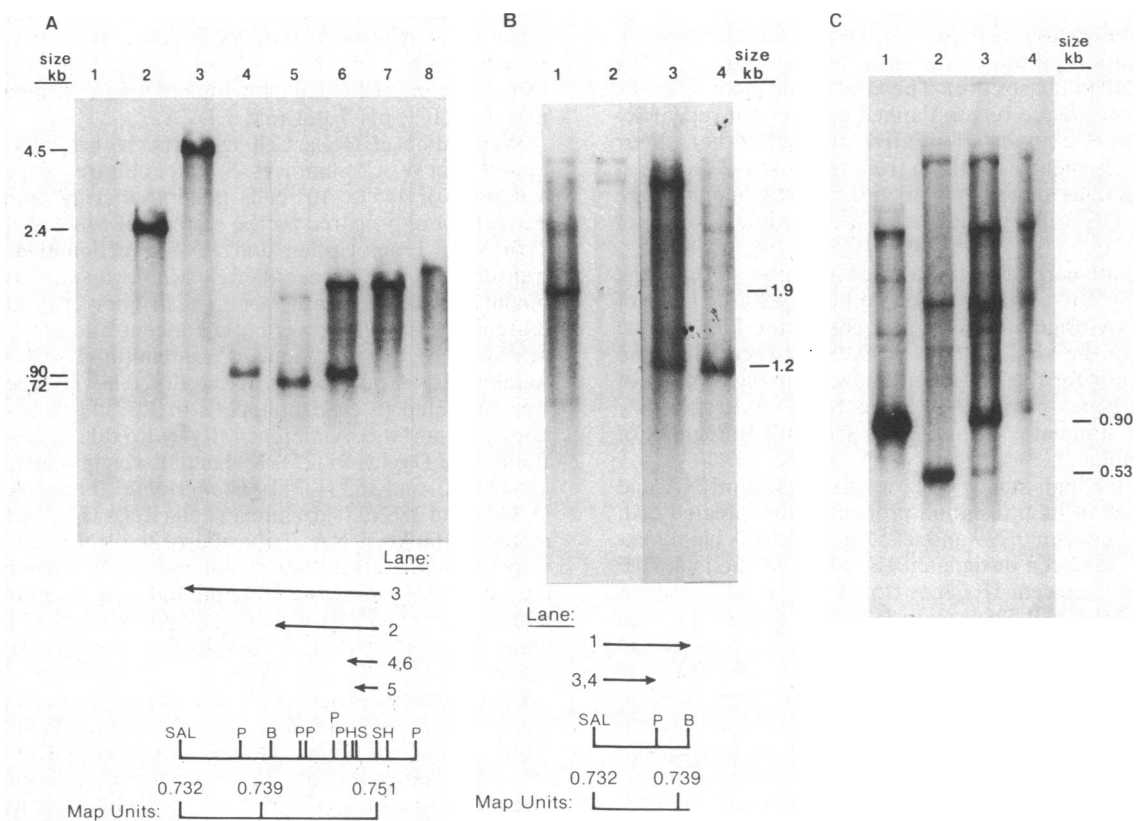


FIG. 2. Autoradiogram of *in vitro* transcripts with DNA templates for the IE genes. RNA was synthesized in standard reactions with various DNA templates, extracted, denatured, and fractionated by electrophoresis in denaturing 1.5% agarose gels containing methylmercury(II) hydroxide as described. (A) DNA templates (100  $\mu$ g/ml) for both major and minor IE genes. Lanes: 1, no added DNA; 2, *Bam*HI-cleaved pXbaIE; 3, *Sal* I-cleaved pXbaIE; 4, *Pst* I-cleaved pXbaIE; 5, *Hinc*II-cleaved pXEP22; 6, *Pst* I-cleaved pXEP22; 7, *Pst* I-cleaved pXEP22 with  $\alpha$ -amanitin (1  $\mu$ g/ml); 8, *Sac* I-cleaved pXEP22. (B) DNA templates (100  $\mu$ g/ml) for IE region 2. Lanes: 1, *Bam*HI-cleaved pCB42; 2, *Bam*HI-cleaved pCB42 with  $\alpha$ -amanitin (1  $\mu$ g/ml); 3, *Pst* I-cleaved pCB42; 4, *Bam*HI/*Pst* I-cleaved pCB 42. (C) DNA templates for IE region 1 and the major late adenovirus promoter. Lanes: 1, *Pst* I-cleaved pXEP22 (100  $\mu$ g/ml); 2, *Sma* I-cleaved pSmaF (100  $\mu$ g/ml); 3, *Pst* I-cleaved pXEP22 (50  $\mu$ g/ml) plus *Sma* I-cleaved pSmaF (50  $\mu$ g/ml); 4, no added DNA. The sizes of the RNAs are shown in kb. Restriction enzyme sites *Sal* I (Sal), *Pst* I (P), *Bam*HI (B), *Hinc*II (H), and *Sac* I (S) relative to region 1 and region 2 DNA coding regions as well as the direction of transcription on the prototype arrangement of the viral genome are designated.

to test if these viral promoters are recognized by RNA polymerase II. Three types of recombinant plasmids were used. Recombinant plasmid pXbaIE contained IE regions 1 (0.739–0.751 map units), 2 (0.732–0.739 map units), and 3 (0.709–0.728 map units) (Fig. 1). Recombinant plasmid pXEP22 contained the promoter for IE region 1, and pCB42 contained a promoter in region 2. However, it is not known whether this is an IE or early promoter.

When all three promoters were present on a single plasmid (pXbaIE), *in vitro* transcription was detected only from IE promoter region 1. Fig. 2A demonstrates that transcripts of 2.4, 4.5, and 0.90 kb were truncated at the *Bam*HI (lane 2), *Sal*I (lane 3), and *Pst*I (lane 4) sites, respectively. When the recombinant plasmid containing only IE promoter region 1 (pXEP22) was used, 0.72- and 0.90-kb transcripts (Fig. 2A, lanes 5 and 6) were truncated at the *Hinc*II and *Pst*I sites, respectively. However, digestion of this recombinant plasmid with *Sac*I eliminated detectable *in vitro* transcription (Fig. 2A, lane 8). *In vitro* transcription was also inhibited by treatment with  $\alpha$ -amanitin at 1  $\mu$ g/ml (Fig. 2A, lane 7). The band at the top represents the typical end-labeled or read-through product obtained with the HeLa cell lysate. Lanes 2, 3, and 4 required longer exposures to see the end-labeled bands. These data indicated that the major IE promoter was located left of the *Hinc*II site and right of the *Sac*I site and that transcription was by host cell RNA polymerase II.

To test for *in vitro* transcription from promoters in region 2, recombinant plasmid pCB42 was used for *in vitro* transcription. This plasmid contains a viral DNA insert extending approximately 5.2 kb left of IE DNA coding region 1 and represents the *Bam*HI fragment B within the *Xba*I fragment E (12) or the *Bam*HI fragment T for the *Bam*HI physical map of the viral genome (unpublished data). *In vitro* transcription with IE region 2 was possible when IE promoter region 1 was not present. Fig. 2B demonstrates that 1.9-kb (lane 1) and 1.2-kb transcripts (lanes 3 and 4) were truncated by the *Bam*HI and *Pst*I sites, respectively. *In vitro* transcription was inhibited by  $\alpha$ -amanitin at 1  $\mu$ g/ml (lane 2). Because of the complexity of RNAs in this region, it is presently not known whether the promoter located at approximately 0.732 map units is an IE or early promoter.

The above data suggested that region 1 IE promoter competed for RNA polymerase II and any other host cell proteins necessary for *in vitro* transcription to the point that activity of promoters in IE regions 2 and 3 were not detectable. To further evaluate this qualitative difference in DNA templates, an equal mixture of each (50  $\mu$ g/ml) recombinant plasmid containing IE promoter region 1 of CMV (pXEP22) and the late adenovirus promoter (pSmaF) was tested for *in vitro* transcription. Fig. 2C is an autoradiogram of the fractionated <sup>32</sup>P-labeled RNAs with the DNA template for IE region 1 of CMV (lane 1), the DNA template for the major late adenovirus promoter (lane 2), and a mixture of both templates (lane 3) for *in vitro* transcription. Even though the late adenovirus promoter was slightly greater in molar equivalents of DNA because of the smaller size of the DNA fragment, the major IE promoter of CMV permitted transcription approximately 2-fold greater than the late adenovirus promoter based on incorporation of [<sup>32</sup>P]GTP of newly synthesized RNA. This calculation was determined by isolating the 0.90-kb RNA made from the CMV DNA template and the 0.53-kb RNA made from the adenovirus DNA template and determining the amount of [<sup>32</sup>P]GTP associated with each RNA molecule and then dividing by the size of the RNA molecule. Therefore, with equal amounts of the two promoters, the level of synthesis of the transcripts were significantly different.

**DNA Sequence of the Major Promoter-Regulatory Region.** Fig. 3 shows the nucleotide sequence upstream of the initiation site of IE region 1. The sequences were confirmed by

```

(5') GCGCACC GCC -480
CAGCGACCCC CCCCCTTGA CGTCAATAGT GACGTATGTT CCCATAGTAA CGCCAATAGG -420
GACTTCCAT TGACGTCAAT GGGTGGAGTA TTTACGGTAA ACTGCCCACT TGGCAGTACA -360
TCAAGTGTAT CATATGCCAA GTCCGCCCCC TATTGACGTC AATGACGGTA AATGGCCCCG -300
CTAGCATTAT GCCCAGTACA TGACCTTAGC GGAGTTTCTC ACTTGGCAGT ACATCTACGT -240
ATTAGTCATC GCTATTACCA TGGTGATGCG GTTTTGGCAG TACACCAATG GGC GTGGATA -180
CGGGTTTGC TCACGGGGAT TTCCAAGTCT CCACCCCAAT GACGTCAATG GGAGTTTGT -120
TTGGCACCAA AATCAACGGG ACTTTCCAAA ATGTCGTAAT AACCCCGCCC CGTTGACGCA -60
AATGGCGCGT AGGCGGTGAC GGTGGGAGGT ATATATAGCA GAGCTCGTTT AGTGAACCGT +1 (3')
Cap
    
```

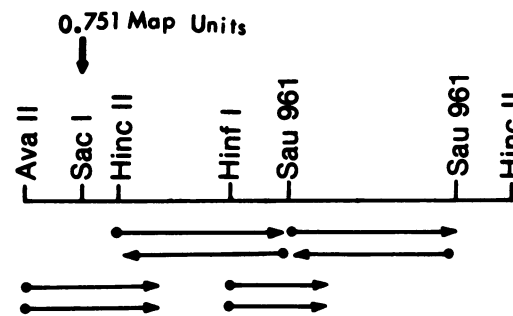


Fig. 3. (Upper) Nucleotide sequence for the promoter-regulatory region of the major IE gene. The sequences of IE region 1 promoter-regulatory regions were sequenced in both directions by the chemical method as described. The numbers above the sequences represent plus or minus nucleotides from the cap site. The transcription initiation site *in vivo* was determined by Stenberg *et al.* (18). The TATA and CAAT boxes are enclosed. Relevant restriction enzyme sites are underlined and designated. (Lower) The sequence assay strategy for the prototype arrangement of the Towne strain. ●, Termini labeled at either the 5' or 3' end; arrow, direction of sequence determination.

analysis of both complementary DNA strands. The initiation site is designated +1 and represents the *in vivo* cap site as determined by Stenberg *et al.* (18). The sequences reveal typical Hogness-Goldberg boxes and "CAAT" boxes (21, 22) at the predicted distance and in the expected orientations for eukaryotic promoter regions. Relevant restriction enzyme sites are underlined and designated. In the IE promoter-regulatory region, a *Sac*I site is located slightly downstream of the "TATA" box (Fig. 3). This explains why *in vitro* transcription of IE region 1 is eliminated by digestion of the DNA template with *Sac*I. A *Hinc*II site is located upstream of the CAAT box and, consequently, *in vitro* transcription with this DNA template was possible, but the amount of transcription was reduced approximately half relative to DNA templates containing the upstream regulatory sequences (see Fig. 2). The locations of the 19- and the 18-nucleotide repeat sequences are illustrated in Figs. 3 and 4. The 19-nucleotide repeat that overlaps into a 18-nucleotide repeat between -397 and -415 was not designated.

Both the 19- and the 18-nucleotide sequence are repeated four times with a 83–95% fidelity. The 19-nucleotide repeat sequence characteristically has a CAAT box-like sequence. One of these is located approximately 60 nucleotides from the cap site. A central sequence was highly conserved within the 18-nucleotide repeat and is underlined (Fig. 4). A 16-nucleotide repeat with the consensus <sup>5</sup>C-T-T-G-G-C-A-G-T-A-C-A-T-C-A-A-3' is also repeated four times with a 63–100% fidelity but is not designated.





in 100 to 468 nucleotides upstream of the major IE gene are palindromic sequences and repeat sequences. Whether or not these repeat sequences are associated with cruciform structures in the DNA molecule is hypothetical. Nevertheless, we proposed that these sequences and their surrounding dyad symmetry play a role in the relative amount of gene expression.

Our sincere thanks go to Pamela Witte for expert assistance. We thank Mark Urbanowski for advice in sequence assay and C. Martin Stoltzfus for a critical review of this manuscript. This investigation was supported by Public Health Service Grant AI13526 from the National Institute of Allergy and Infectious Diseases and by Grant 1-697 from the National Foundation March of Dimes. M.F.S. is the recipient of Public Health Service Career Development Award AI100373 from the National Institute of Allergy and Infectious Diseases. R.M.S. is the recipient of a fellowship from the Ladies Auxiliary of the Veterans of Foreign Wars.

1. Kilpatrick, B. A. & Huang, E. S. (1977) *J. Virol.* **24**, 261-276.
2. Geelen, J. L. M. C., Walig, C., Wertheim, P. & Van der Noordaa, J. (1978) *J. Virol.* **26**, 813-816.
3. DeMarchi, J. M., Blankship, M. L., Brown, G. D. & Kaplan, A. S. (1978) *Virology* **89**, 643-646.
4. Weststrate, M. W., Geelen, J. L. M. C. & Van der Noordaa, J. (1980) *J. Gen. Virol.* **49**, 1-22.
5. LaFemina, R. L. & Hayward, G. S. (1980) in *Animal Virus Genetics*, eds. Fields, B. N. & Jaenish, R. (Academic, New York), pp. 39-55.
6. DeMarchi, J. M. (1981) *Virology* **114**, 23-28.
7. Spector, D. H., Hock, L. & Tamashiro, J. C. (1982) *J. Virol.* **42**, 558-582.
8. Wathen, M. W. & Stinski, M. F. (1982) *J. Virol.* **41**, 462-477.
9. McDonough, S. H. & Spector, D. H. (1983) *Virology* **125**, 31-46.
10. Wathen, M. W., Thomsen, D. R. & Stinski, M. F. (1981) *J. Virol.* **38**, 446-459.
11. Stinski, M. F., Thomsen, D. R. & Rodriguez, J. E. (1982) *J. Gen. Virol.* **60**, 261-270.
12. Stinski, M. F., Thomsen, D. R., Stenberg, R. M. & Goldstein, L. C. (1983) *J. Virol.* **46**, 1-14.
13. Thomsen, D. R. & Stinski, M. F. (1981) *Gene* **16**, 207-216.
14. Manley, J. L., Fire, A., Cano, A., Sharp, P. A. & Gelfand, M. L. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 3855-3859.
15. Bailey, J. M. & Davidson, N. (1976) *Anal. Biochem.* **70**, 75-85.
16. Bishop, D. H. L., Claybrook, J. R. & Spiegelman, S. (1967) *J. Mol. Biol.* **26**, 373-387.
17. Anderson, K. P., Costa, R. H., Holland, L. E. & Wagner, E. K. (1980) *J. Virol.* **34**, 9-27.
18. Stenberg, R. M., Thomsen, D. R. & Stinski, M. F. (1983) *J. Virol.* **49**, 190-199.
19. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499-560.
20. Tinco, I., Borer, P., Dengler, B., Levine, M. D., Uhlenbeck, O. C., Crothers, D. M. & Gralla, J. (1973) *Nature (London)* **246**, 40-41.
21. Chambon, P. & Breathnach, R. (1981) *Annu. Rev. Biochem.* **50**, 349-383.
22. Liebhaber, S. A., Goossens, M. J. & Wai Kan, Y. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7054-7058.
23. Saragosti, S., Cereghini, S. & Yaniv, M. (1982) *J. Mol. Biol.* **160**, 133-146.
24. Shakhov, A., Nedospasov, S. A. & Georgiev, G. P. (1982) *Nucleic Acids Res.* **10**, 3951-3965.
25. Clark, S. P. & Mak, T. W. (1982) *Nucleic Acids Res.* **10**, 3315-3330.
26. Mackem, S. & Roizman, B. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 4917-4921.
27. Mackem, S. & Roizman, B. (1982) *J. Virol.* **44**, 939-949.
28. Mackem, S. & Roizman, B. (1982) *J. Virol.* **43**, 1015-1023.
29. Post, L. E., Mackem, S. & Roizman, B. (1981) *Cell* **24**, 555-565.
30. Post, L. E., Norrild, B., Simpson, T. & Roizman, B. (1982) *Mol. Cell Biol.* **2**, 233-240.