VIROLOGY 191, 355–367 (1992)

Characterization of the Murine Cytomegalovirus Genes Encoding the Major DNA Binding Protein and the ICP18.5 Homolog

MARTIN MESSERLE,* GÜNTHER M. KEIL,† KARIN SCHNEIDER,* AND ULRICH H. KOSZINOWSKI*,[1]

*Department of Virology, Institute for Microbiology, University of Ulm, 7900 Ulm, Germany; and †Federal Research Centre for Virus Diseases of Animals, 7400 Tübingen, Germany

Received May 27, 1992; accepted July 13, 1992

In several herpesviruses the genes for the major DNA binding protein (MDBP), a putative assembly protein, the glycoprotein B (gB), and the viral DNA polymerase (pol) collocate. In murine cytomegalovirus (MCMV), two members of this gene block, pol (Elliott, Clark, Jaquish, and Spector, 1991, *Virology* 185, 169–186) and gB (Rapp, Messerle, Bühler, Tannheimer, Keil, and Koszinowski, 1992, *J. Virol.*, 66, 4399–4406) have been characterized. Here the two other MCMV genes are characterized, the gene encoding the MDBP and the ICP18.5 homolog encoding a putative assembly protein. Like in human cytomegalovirus (HCMV) the genes order is pol, gB, ICP18.5, and MDBP. The 4.2-kb MDBP mRNA is expressed first in the early phase, whereas the 3.0-kb ICP18.5 mRNA is a late transcript. The open reading frame of the MDBP gene has the capacity of encoding a protein of 1191 amino acids with a predicted molecular mass of 131.7 kDa. The MCMV ICP18.5 ORF is translated into a polypeptide of 798 amino acids with a calculated molecular mass of 89.1 kDa. Comparison of the amino acid sequences of the predicted proteins of MCMV with the respective proteins of HCMV, Epstein–Barr virus (EBV), and herpes simplex virus type-1 (HSV-1) reveals a striking homology ranging from 72% (HCMV), 50% (EBV), to 45% (HSV-1) for the MDBP sequence and from 74% (HCMV), 51% (EBV), to 49% (HSV-1) for the ICP18.5 sequence. These results establish the close relationship of the two cytomegaloviruses, and underline the usefulness of the murine model for studies on the biology of the CMV infection.

INTRODUCTION

Cytomegaloviruses (CMV) are highly species-specific herpesviruses that infect many animals. Human cytomegalovirus (HCMV) is an ubiquitous human pathogen. Although primary infection in the immunocompetent host is often not accompanied by an apparent disease, a state of livelong persistence is established (Ho, 1982). Severe immunosuppression or immunodeficiency can cause reactivation of the HCMV infection from the latent state. Therefore, HCMV is a major threat for organ allograft recipients and HIV patients (Pass, 1991). Apparently, host defense mechanisms rather than viral properties define the different manifestations of CMV disease. Therefore, the development of an efficacious vaccine or treatment against HCMV needs a thorough analysis of the viral pathogenesis and of the immune response against CMV antigens. Due to the species specificity of HCMV the experimental analysis of virus–host interactions is limited. Thus, the availability of animal models are important in defining the key elements of CMV control.

A series of similarities have been reported between murine (MCMV) and human CMV (Ho, 1982). Both MCMV and HCMV lead to the infection of the spleen, salivary and adrenal glands, and lungs of their respective hosts. Pneumonia is the major cause of mortality following MCMV or HCMV infection. Therefore, the infection of mice with MCMV has been used as a model system to study various parameters of CMV infections including the reactivation following immunosuppressive treatment (Jordan *et al.*, 1977) and the cellular immune response against CMV antigens (Koszinowski *et al.*, 1990).

Despite of the obvious similarities of the two viruses in biology and pathogenesis, it is not known whether the similarity is also reflected at the molecular level. However, a molecular approach to virus control will only be feasible when the factors contributing to viral pathogenesis and antiviral immunity are defined at the genetic level. Whereas the complete sequence of the HCMV genome has been determined (Chee *et al.*, 1990), only a few MCMV genes and proteins have been analyzed so far (Keil *et al.*, 1987; Bühler *et al.*, 1990; Elliott *et al.*, 1991; Loh *et al.*, 1991; Messerle *et al.*, 1992; Rapp *et al.*, 1992). In order to study the homology of the two CMV genomes we have determined the nucleic acid sequence of a region of the MCMV genome, which contains a series of genes conserved throughout all herpesviruses.

Here we report the nucleic acid sequence and the transcriptional patterns of two MCMV genes encoding

The nucleotide sequence data reported in this paper have been submitted to the GenBank nucleotide sequence database and have been assigned Accession Number X67021.

[1] To whom reprint requests should be addressed.

proteins with a high degree of amino acid sequence homology to the major DNA binding protein and a putative assembly protein of HCMV and herpes simplex virus (HSV).

MATERIALS AND METHODS

Virus and cell culture

MCMV (mouse salivary gland virus strain Smith, ATCC VR-194) was propagated on BALB/c mouse embryonal fibroblasts (MEF) as described previously (Keil *et al.*, 1984).

Cloning procedures and sequence analysis

*Bam*HI subclones of the MCMV *Hin*dIII D fragment were generated and these were further subcloned as *Eco*RI or *Pst*I fragments. The nucleic acid sequence was determined by a modified version of the dideoxy sequencing method (Tabor and Richardson, 1987), using nuclease *Bal*31-generated subclones of the *Eco*RI and *Pst*I fragments. When necessary deaza-dGTP was used in the sequencing reactions instead of dGTP to avoid compressions. Sequences were either determined for both strands or several overlapping subclones were used to confirm the sequence. Appropriate primers were synthesized and used to define the overlapping sequences between the *Eco*RI and *Pst*I subclones, respectively. Overlaps between the *Bam*HI fragments were determined by sequence analysis of polymerase chain reaction amplified fragments. The Genetics Computer Group software package version 7.0 from April 1991 (Devereux *et al.*, 1984) was used for the analysis of the nucleic acid sequences and for the deduction of the amino acid sequences. Comparison of the nucleotide sequences of the identified open reading frames and of the deduced amino acid sequences was performed to the entries of the Genbank database release 69 from September 1991.

Isolation of RNA

MEF were MCMV infected with 20 PFU per cell by using the technique of centrifugal enhancement of infectivity at 800 *g* for 30 min. At different time points postinfection whole-cell RNA was prepared from MCMV-infected cells following established procedures (Chirgwin *et al.*, 1979).

Northern (RNA) blot hybridization and nuclease S1 protection experiments

RNA samples were size fractionated by gel electrophoresis and transferred to nitrocellulose filters as described previously (Keil *et al.*, 1984). DNA probes used for hybridization were labeled with [α-^{32}P]dCTP by using a multiprime DNA labeling system (Boehringer-Mannheim, Germany). For estimation of RNA sizes, *Escherichia coli* 16 S and 23 S rRNAs and 18 S and 28 S mouse rRNAs were used as size markers. Whole-cell RNA was hybridized to 5′ or 3′ end-labeled DNA fragments and digested with nuclease S1 as described previously (Keil *et al.*, 1987). Nuclease-resistant fragments were size fractionated on denaturing sequencing gels.

3′ end determination

The 3′ end of the MCMV ICP18.5 mRNA was determined according to the protocol of Frohman *et al.* (1988) using the oligonucleotides (dT)17-R1-R0 (5′-AAGGATCCGTCGACATCGATAATACGACTCACTA-TAGGGATTTTTTTTTTTTTTTT-3′), R0 (5′-AAG-GATCCGTCGACATC-3′), and End1 (5′-GCAA-GCTTATGTCAGGTCGAGAG-3′). The primer End1 corresponds to positions 6278 to 6295 in Fig. 2, immediately upstream of the *Sma*I site at map unit 0.375 (Fig. 4).

RESULTS

Nucleotide sequence of the MCMV MDBP and ICP18.5 genes

A conserved gene block containing the genes for the major DNA-binding protein (MDBP), a putative assembly protein, the glycoprotein B (gB), and the viral DNA polymerase (pol) has been described for all herpesviruses (Kouzarides *et al.*, 1987). We assumed that the analysis of this gene block could help to define the degree of relatedness between murine and human CMV. The pol and gB genes of MCMV have been sequenced (Elliott *et al.*, 1991; Keil, unpublished data; Rapp *et al.*, 1992), and the location of gB has been mapped to the left end of the MCMV *Hin*dIII D fragment and pol overlaps the border between the *Hin*dIII D and H fragments (Fig. 1). Assuming the same gene order as in HCMV, we expected that the MCMV genes encoding the MDBP and the putative assembly protein, which we call MCMV ICP18.5 gene in analogy to the gene in the HSV-1 genome (Pellett *et al.*, 1986), to be located upstream of the glycoprotein B gene. Consequently, we analyzed the nucleic acid sequence upstream of the glycoprotein B gene (Fig. 1).

The nucleotide sequence from the *Xho*I site at map unit 0.401 up to the *Eco*RI site at map unit 0.371 is shown in Fig. 2. Two large open reading frames (ORF) of 3573 and 2394 nucleotides (nt) extending from positions 256 to 3828 and from positions 4038 to 6431 were identified. A third ORF starting at position 6334 encodes the MCMV glycoprotein B (Rapp *et al.*, 1992).

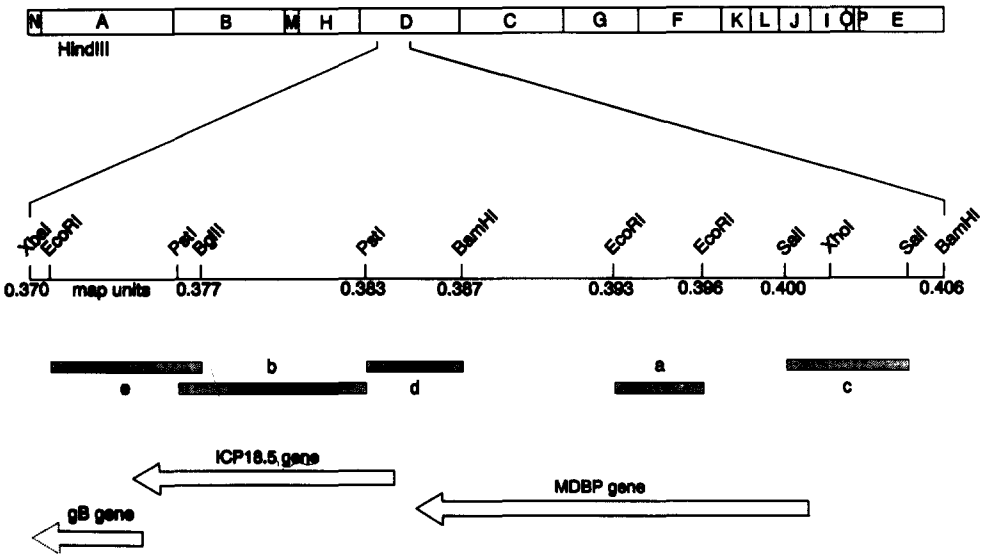

FIG. 1. Localization of the MCMV genes encoding the ICP18.5 homolog and the major DNA binding protein (MDBP). (Top) *Hin*dIII cleavage map of the MCMV strain Smith genome. (Middle, enlarged) Positions (in map units) of restriction enzyme sites in the analyzed part of the *Hin*dIII D fragment. The hatched bars represent DNA probes used in the Northern blot analysis. (Bottom) Localization of the identified open reading frames (open bars).

The ORF of the MCMV ICP18.5 gene overlaps the ORF of the glycoprotein B gene by 32 codons. The nucleotide sequences of the MCMV ICP18.5 and MDBP ORFs showed 65 and 60% homology to the UL56 and UL57 ORFs of the HCMV genome (Chee *et al.*, 1990). Remarkably, the sequences upstream of the MDBP ORF (map units 0.401 to 0.406) and the 200 nt between the MDBP and ICP18.5 ORFs display almost no similarity to the corresponding sequences in the HCMV genome.

Expression kinetics of the MCMV MDBP and ICP18.5 genes

Northern (RNA) blot hybridizations were performed to determine the expression kinetics of the MDBP and ICP18.5 genes. Hybridization with a 821-bp *Eco*RI fragment (Fig. 1, probe a; map units 0.396 to 0.393) from the MDBP ORF resulted in a band of 4.2 kb (Fig. 3a). The MDBP mRNA appeared as early as 2 hr p.i. and remained present throughout the replication cycle (Fig. 3a, lanes 3 to 6).

The size of the ICP18.5 mRNA and the expression kinetics of the ICP18.5 gene was determined by using a 1734-bp *Pst*I fragment from the ICP18.5 ORF (Fig. 1, probe b; map units 0.383 to 0.376) as ^{32}P-labeled probe. A transcript of 3.0 kb was seen 24 hr p.i. (Fig. 3b). Thus, the MCMV ICP18.5 gene represents a late gene.

To define the approximate ends of the MDBP and ICP18.5 transcripts Northern blot hybridizations were performed using DNA probes which overlap the ends of the putative open reading frames (see Fig. 1, probes c to e). A 1034-bp *Sal*I fragment (Fig. 1, probe c; map units 0.4045 to 0.400) showed a weak hybridization with the 4.2-kb MDBP mRNA (Fig. 3c). This result suggested that the 5′ end of the MDBP mRNA is located within this *Sal*I fragment. Hybridization with a 856-bp *Bam*HI/*Pst*I fragment (Fig. 1, probe d; map units 0.387 to 0.383) revealed the MDBP transcript as well as the late ICP18.5 transcript (Fig. 3d, lanes 3 and 4). Therefore, the 3′ end of the MDBP mRNA and the 5′ end of the ICP18.5 transcript are located within the *Bam*HI/*Pst*I fragment. The 1389-bp *Bgl*II/*Eco*RI fragment (Fig. 1, probe e; map units 0.377 to 0.371) showed hybridization to the 3.0-kb ICP18.5 transcript, to a 3.3-kb transcript, and to additional bands of higher molecular mass. The 3.3-kb band represents the glycoprotein B mRNA and the larger transcripts also represent gB-

FIG. 2. Nucleotide sequence of the MCMV MDBP and ICP18.5 genes and deduced amino acid sequence of the encoded proteins. The nucleotide sequence is shown from the *Xho*I site at map unit 0.401 up to the *Eco*RI site at map unit 0.371. Open reading frames are indicated by capital letters and the deduced amino acid sequence is shown below in the one-letter code. 5′ cap sites and 3′ ends of the mRNAs are marked by stars and dots, respectively. Putative binding sites for the transcription factor SP1 are underlined, and potential TATA-boxes are underlined and marked by capital letters. Some restriction enzyme sites are shown for orientation.

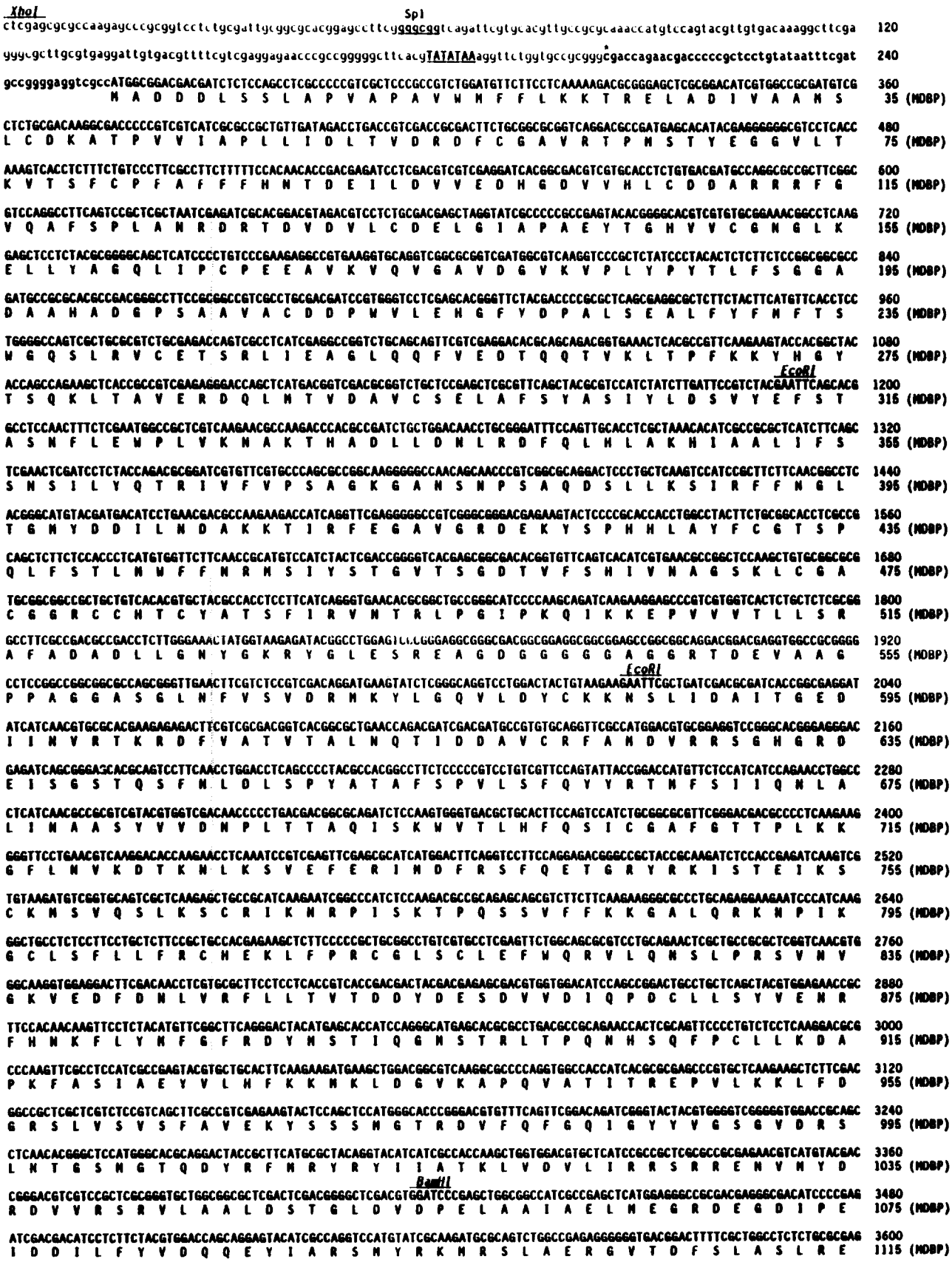

FIG. 2—*Continued*

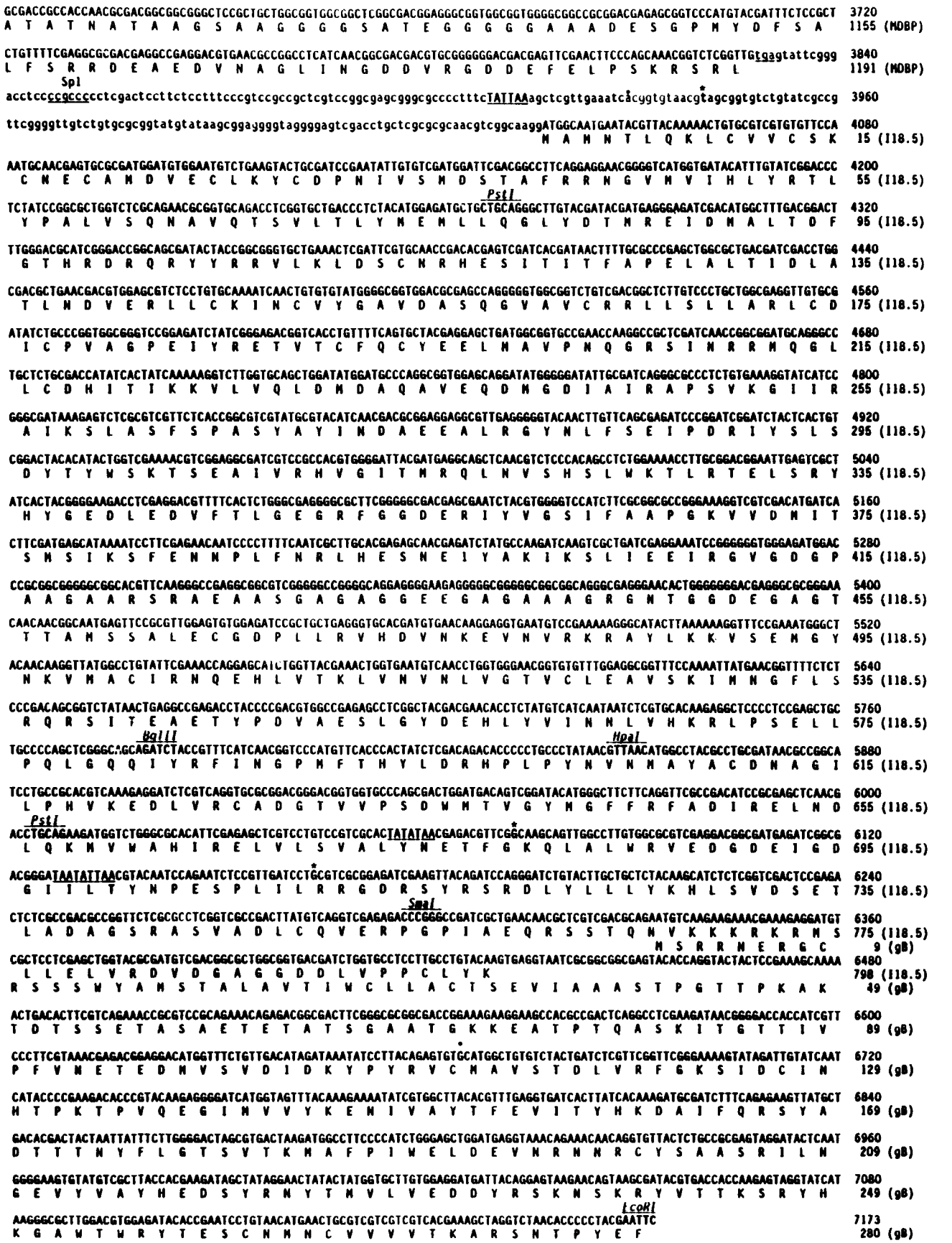

FIG. 2—*Continued*

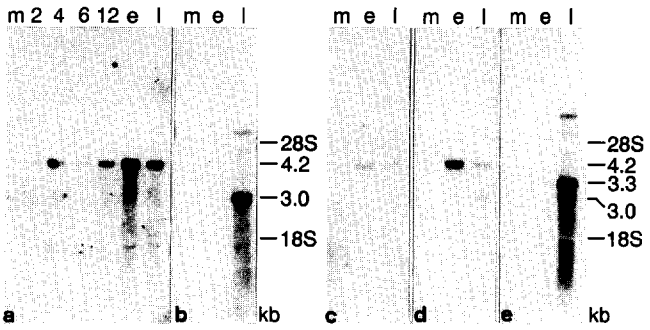

FIG. 3. Expression kinetics of the MDBP and ICP18.5 genes. Northern blot analysis with ^{32}P-labeled DNA probes (for location of the probes see Fig. 1) was performed to determine the expression kinetics of the genes. MEF were mock infected (m) or infected in the presence of PAA (e) or without inhibitors. Whole-cell RNA was isolated at 2, 4, 6, 12 (lanes 2 to 12), 20 (lanes e), and 24 (lanes l) hr p.i.

specific messages as they were seen with DNA probes located further downstream within the gB open reading frame (Rapp *et al.*, 1992).

Identification of the 5′ and 3′ ends of the MDBP and ICP18.5 transcripts

Nuclease S1 protection experiments were performed to locate the ends of the transcripts. The 5′ end of the MDBP transcript was mapped with a 414-nt *Xho*I/*Sal*I fragment (map units 0.401 to 0.400; Fig. 4A (1)), which was 5′ end-labeled at the *Sal*I site (Fig. 4B (1), lane 1). After hybridization of the probe to RNA isolated from cells in the early phase of infection and subsequent nuclease S1 digestion a protected fragment of 215 nt was seen (Fig. 4B (1), lane 3). The 5′ cap site of the MDBP mRNA is therefore located 215 nt upstream of the *Sal*I site (map units 0.400) at position 205 in Fig. 2. Thus, the MDBP mRNA has a short leader sequence of only 51 nt in front of the first AUG. A putative TATA box was found 20 nt upstream of the determined 5′ cap site (positions 180 to 185 in Fig. 2). The sequence GGGCGG at positions 56 to 61 represents a putative binding site for the transcription factor SP1. Nuclease S1 analysis with RNA from cells in the late phase of infection revealed the same protected fragment (Fig. 4B (1), lane 4). Thus, the same transcription start site is used early and late in infection.

The 3′ end of the MDBP mRNA was determined with a 856-nt *Bam*HI/*Pst*I fragment (map units 0.387 to 0.383; Fig. 4A (2)), 3′ end-labeled at the *Bam*HI site (Fig. 4 (2), lane 1). RNA from cells in the early and late phase of infection protected a fragment of 513 nt from nuclease S1 digestion (Fig. 4B (2), lanes 3 and 4). Thus, the 3′ end is located at position 3930 in the nucleic acid sequence (Fig. 2), 99 nt downstream of the stop codon of the MDBP ORF. The sequence ATTAAA (positions 3911 to 3916) 14 nt upstream of the determined 3′ end could represent a polyadenylation signal of the rarely used type (Birnstiel *et al.*, 1985). GT-rich sequences were found downstream of the 3′ end (positions 3932 to 3968). However, these sequences fit only weakly to the 3′ end consensus sequence YGTGTTYY (Birnstiel *et al.*, 1985). Considering the size of the polyA-tail, the calculated size of 3726 nt for the MDBP mRNA corresponds well to the observed size of 4.2 kb in the Northern blot hybridizations.

A 444-nt *Not*I/*Sau*3AI fragment (map units 0.386 to 0.384; Fig. 4A (3)), 5′ end-labeled at the *Sau*3AI site was used to map the 5′ end of the ICP18.5 transcript (Fig. 4B (3), lane 1). The nuclease S1 analysis using RNA from cells in the late phase of infection revealed a protected fragment of 188 nt (Fig. 4B (3), lane 3). Therefore, the 5′ cap site of the ICP18.5 mRNA is located 188 nt upstream of the *Sau*3AI site at position 3941 in Fig. 2. Thus, the 5′ end of the ICP18.5 mRNA maps only 10 nt downstream of the 3′ end of the MDBP mRNA. The length of the untranslated leader sequence of the ICP18.5 mRNA is 97 nt. The sequence TATTA at positions 3910 to 3014 (Fig. 2), 27 nt upstream of the 5′ cap site, could represent a TATA box. The sequence CCGCCC at positions 3846 to 3852 represents a putative binding site for the transcription factor SP1.

The sequence analysis revealed that the ICP18.5 and gB genes, and even the ORFs, overlap. Therefore, it was expected that the 3′ end of the ICP18.5 mRNA and the 5′ ends of the gB mRNAs are collinear. To determine the 3′ end of the ICP18.5 mRNA, DNA fragments were used, which were 3′ end-labeled upstream of the 5′ cap sites of the gB mRNAs. This experimental procedure should avoid misinterpretation resulting from hybridization and protection of the probes with the gB mRNA. A 1319-nt *Hpa*I/*Eco*RI (map units 0.3766 to 0.371; Fig. 4A (5)) and a 1389-nt *Bgl*II/*Eco*RI fragment (map units 0.377 to 0.371; Fig. 4A (4)), 3′ end-labeled at the *Hpa*I and at the *Bgl*II sites, respectively (Fig. 4B, (4) and (5), lanes 1), were used. Protected fragments of approximately 820 and 900 nt were identified following the nuclease S1 digestion (Fig. 4B, (4) and (5), lanes 3). Thus, the ICP18.5 mRNA contains a trailer sequence of approximately 250 nt downstream of the stop codon of the ICP18.5 ORF. To determine the 3′ end of the ICP18.5 mRNA more precisely, the ICP18.5 mRNA was reverse transcribed using the primer (dT)17-R1-R0, and the 3′ end of the cDNA was amplified using primers End1 and R0 according to the method of Frohman *et al.* (1988). A 300-bp fragment was identified following amplification and subcloned into pUC19. Sequence analysis located the 3′ end of the ICP18.5 mRNA to position 6665 in Fig. 2.

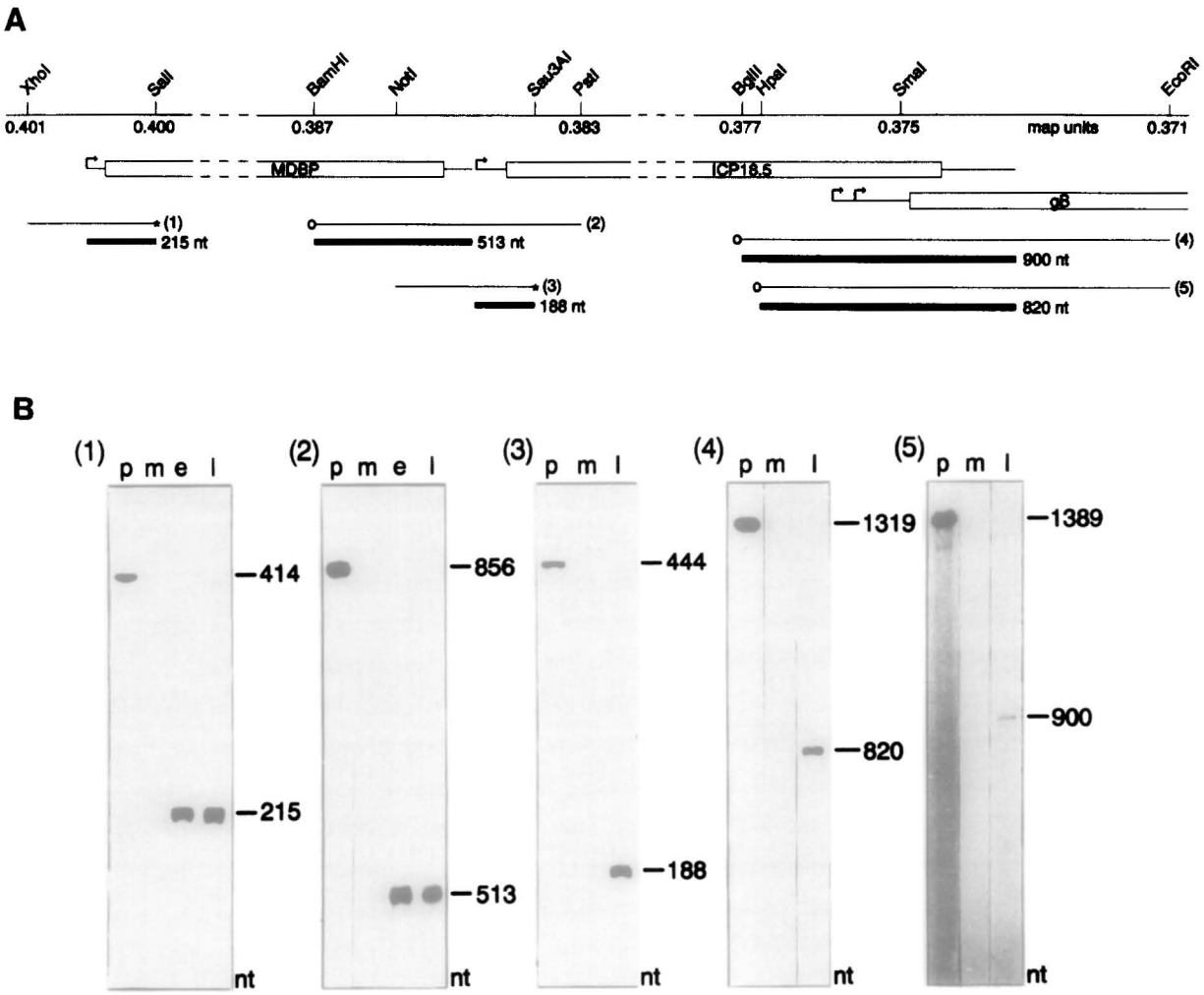

FIG. 4. Localization of the 5′ and 3′ ends of the MDBP and ICP18.5 mRNAs. (A) Experimental design; (B) experimental data. (A, top) Cleavage sites of restriction enzymes, which were used for preparation of the probes are indicated above the open reading frames (open bars) and the deduced structure of the mRNAs. Gaps were introduced (- - -) into the drawing to present all relevant information in one picture. (A, bottom) 5′ (—*) and 3′ end-labeled DNA fragments (O—) were hybridized to whole-cell RNA and then digested with nuclease S1. Solid bars and numbers indicate the location and length, respectively, of the protected fragments. (B) Autoradiographies of protected fragments separated by electrophoresis in 6% (experiments (1) and (3)) or 4% denaturing polyacrylamide gels (experiments (2), (4), and (5)). Lanes p show the labeled fragments before the nuclease treatment, and lanes m, e, and l show the nuclease-resistant fragments after hybridization either to RNA from mock-infected cells, or to early RNA isolated 20 hr p.i. from cells which were infected with MCMV in the presence of PAA or to late RNA isolated from MCMV-infected cells 24 hr p.i. Size markers were 5′ end-labeled *Hpa*II cleavage products of pBR322 and the 123-bp and 1-kb ladders (GIBCO BRL).

An A/T-rich sequence was found 14 nt upstream of the determined 3′ end (ATAAATAT; positions 6644 to 6651), which does not represent a typical polyadenylation signal, whereas the sequence TGTGTCT (positions 6672 to 6678) 6 nt downstream of the determined 3′ end shows a partial homology to the 3′ consensus sequence YGTGTTYY (Birnstiel *et al.*, 1985). The calculated size of 2725 nt for the ICP18.5 transcript is in accordance with the observed size of 3.0 kb for the ICP18.5 mRNA in the Northern blot experiments if the increase in size by polyadenylation is taken into account.

Amino acid sequence conservation between the MCMV MDBP and the homologous proteins of HCMV, EBV, and HSV-1

The amino acid sequence of the MCMV major DNA binding protein was deduced from the nucleic acid sequence of the identified open reading frame. The MDBP ORF is capable of encoding a polypeptide of 1191 amino acids (Fig. 2), and the calculated molecular mass of the MCMV MDB protein is 131.7 kDa.

The amino acid sequence of the MCMV MDB protein was compared by the Genetic Computer Group

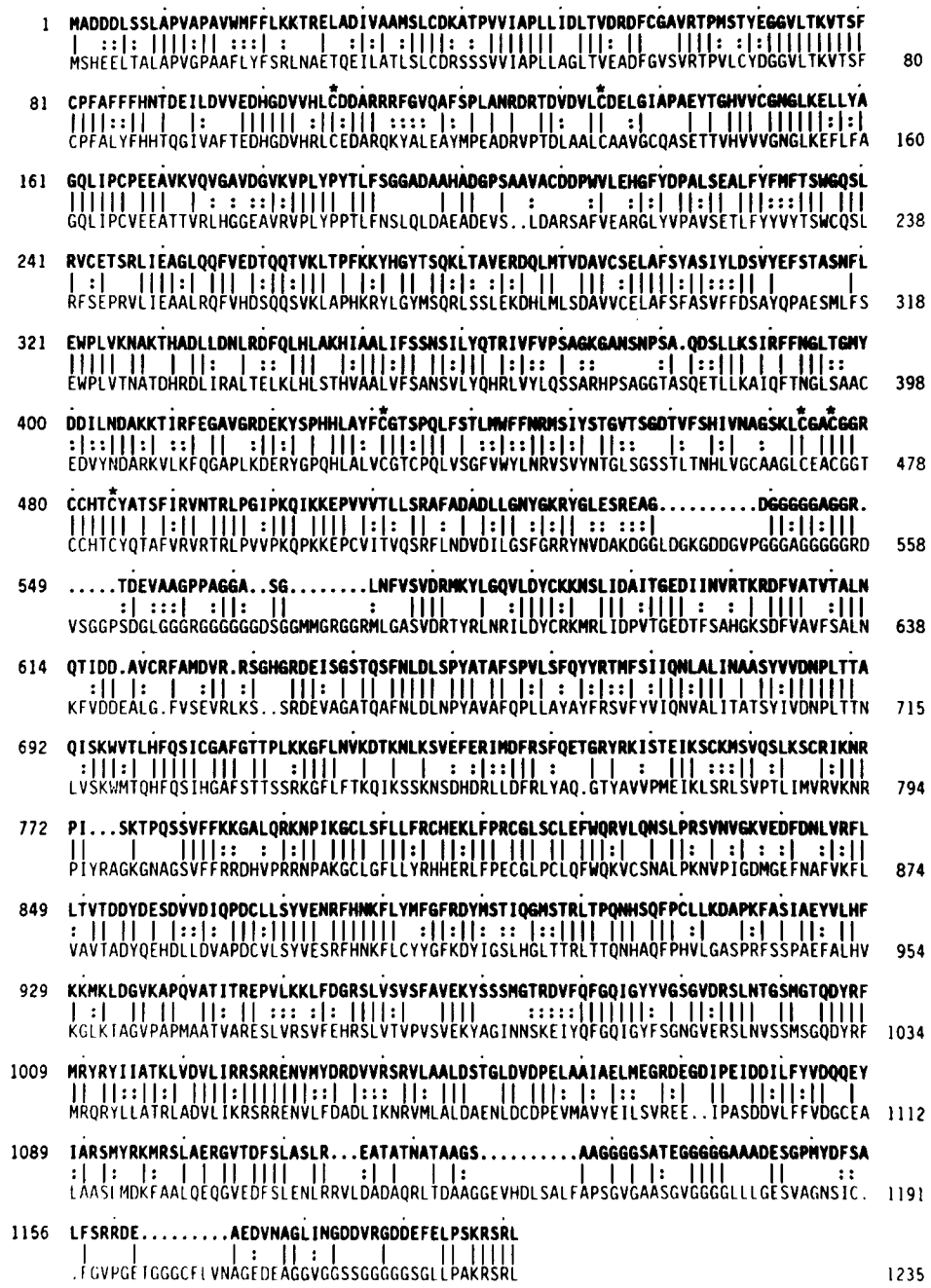

FIG. 5. Comparison of the amino acid sequences of the MCMV and HCMV MDB proteins. The amino acid sequences of the MCMV (upper line, bold faced) and HCMV MDB proteins (lower line) were aligned. Identical residues are connected by bars (|) and conservative exchanges are marked by dots (:). Gaps were introduced to maximize the alignment of identical amino acids. Cysteine residues at identical positions in the sequence of the corresponding HSV-1 MDB protein are labeled by stars.

software program BESTFIT to the sequence of the HCMV MDB protein (Fig. 5). In addition, dot matrix analyses were performed with the amino acid sequences of the corresponding proteins of HCMV, EBV, and HSV-1 (Fig. 7a–7c). Figure 5 shows the alignment of the MCMV and HCMV MDBP sequences. The amino acid sequences of the MCMV and HCMV MDB proteins show 72% similarity. 604 of the 1191 amino acids of the MCMV MDB protein (50.7%) are identical with the corresponding residues of the HCMV MDB protein, and 220 amino acids are replaced by conserved residues. Only the carboxy terminus is less stringent conserved (amino acids 1115 to 1191). Gaps in the alignment occur between amino acids 539 and 567. Remarkably, these gaps are not seen if the MCMV MDBP amino acid sequence is aligned to the HSV-1 ICP8 se-

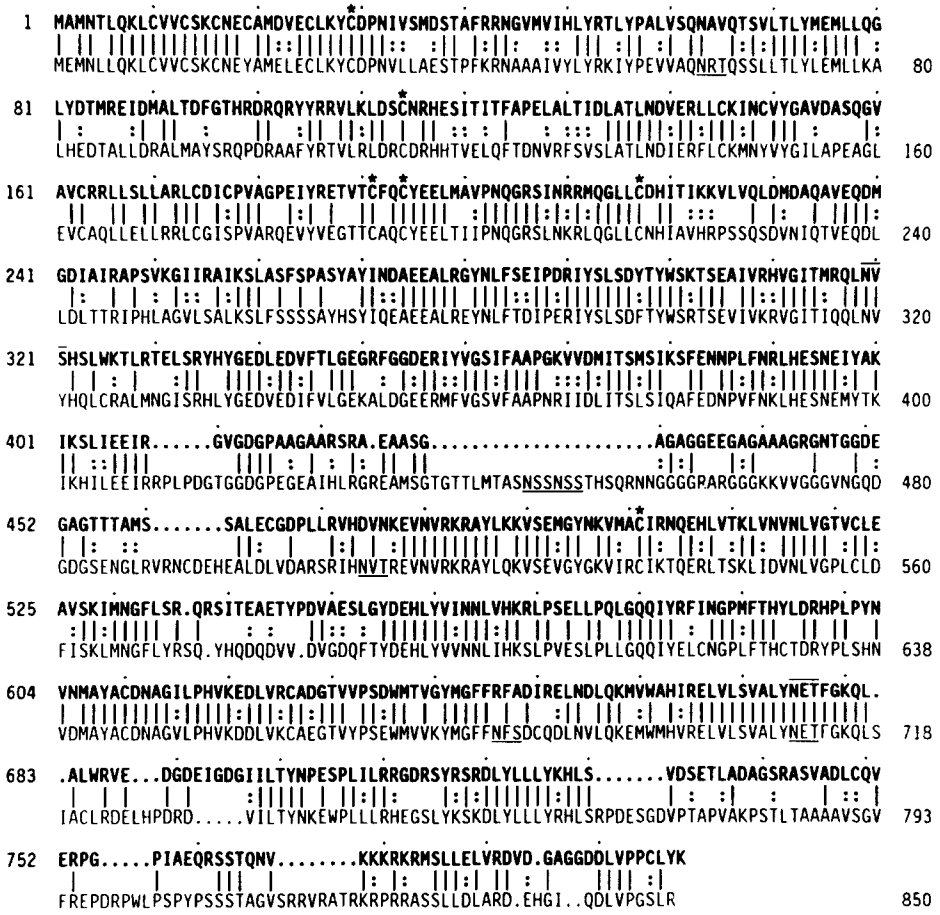

FIG. 6. Alignment of the amino acid sequences of the MCMV ICP18.5 and the HCMV UL56 proteins. Labeling is as described in Fig. 5. Putative N-glycosylation sites (NXS/T) are over- or underlined.

quence (not shown). Similar gaps have been observed previously, when the simian CMV and HCMV MDBP sequences were compared (Anders, 1990). Thus, we conclude that the HCMV MDB protein contains an insertion of 26 residues at this position compared to the MCMV and SCMV MDB proteins (Anders, 1990) and to HSV-1 ICP8. 16 of the 27 cysteine residues are located at identical positions in the MCMV and HCMV MDBP amino acid sequences.

The MCMV and EBV MDBP amino acid sequences show a significant degree of conservation represented by 50% identical and similar amino acids (Fig. 7b). The similarity between the sequences of the MCMV MDBP and HSV-1 ICP8 protein is 45% (Fig. 7c). Six cysteine residues in the MCMV MDBP are at identical positions as in the HSV-1 ICP8 sequence (amino acids 107, 134, 431, 473, 476, and 484), and the same number of cysteine residues is conserved to the EBV MDBP sequence (amino acids 107, 295, 473, 476, 484, and 705). Three of these conserved cysteine residues (positions 473, 476, and 484) are located within a motif which has been implicated with DNA binding and could perhaps fold into a zinc finger structure (Gao *et al.*, 1988).

Amino acid sequence conservation between the MCMV ICP18.5 protein and the homologous proteins of HCMV, EBV, and HSV-1

The MCMV ICP18.5 ORF is translated into a polypeptide of 798 amino acids (Fig. 2). The calculated molecular mass of the MCMV ICP18.5 protein is 89.1 kDa.

Analogous to the comparison of the MDB proteins, the amino acid sequence of the MCMV ICP18.5 protein was compared to the sequences of the HCMV UL56 protein (Figs. 6 and 7d), the EBV homolog (BALF2) (Fig. 7e), and the HSV-1 ICP18.5 protein (Fig. 7f). The amino acid sequences of the MCMV and HCMV ICP18.5 proteins display 74.4% similarity. Four hundred forty amino acids are identical (55%) and 133 residues are replaced by conserved residues (Fig. 6). There are again gaps in the alignment between amino acids 409 and 461, which result from additional residues in the HCMV UL56 sequence. Only the carboxy

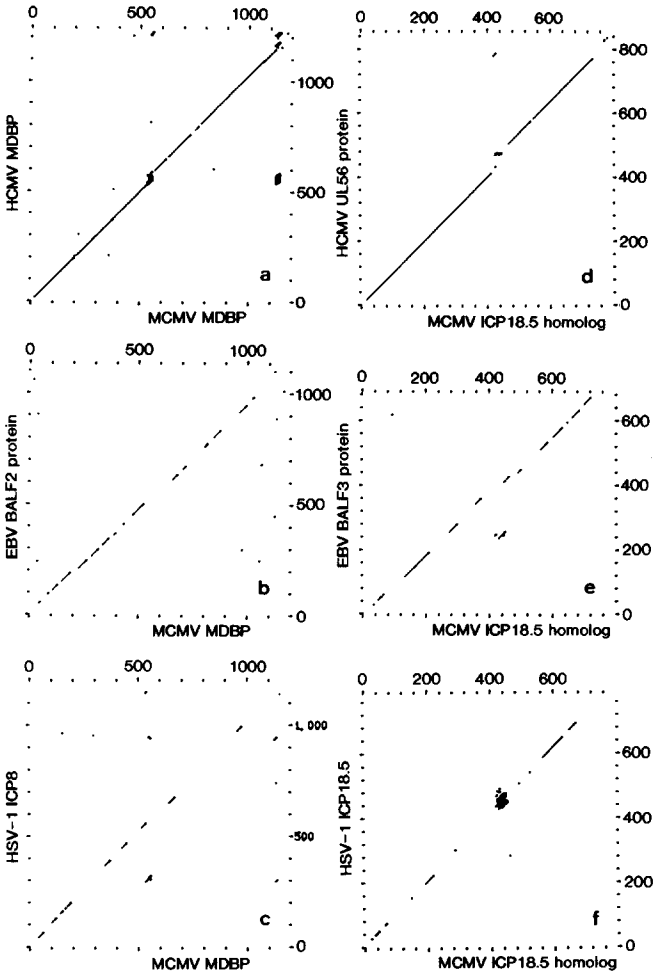

FIG. 7. Dot matrix comparison of the MCMV MDB protein (a,b,c) and the ICP18.5 homolog (d,e,f) to the homologous proteins of HCMV, EBV, and HSV-1. The following parameters were used: window, 30 aa; and stringency, 18 aa. The N termini of the sequences are in the bottom left corner in each graph.

terminal amino acids (the last 68 residues) are less conserved. According to the same criteria, the similarity between the amino acid sequences of the MCMV ICP18.5 protein and the EBV homolog is 51% (Fig. 7e), and the similarity to the HSV-1 ICP18.5 protein is 49% (Fig. 7f). Whereas the similarity between the MCMV ICP18.5 and the EBV BALF3 protein is equally distributed throughout the complete sequence (Fig. 7e), the comparison to the HSV-1 ICP18.5 sequence shows a higher conservation of carboxy-terminal sequences (amino acids 574 to 669; Fig. 7f). Sixteen of the 22 cysteine residues in the MCMV ICP18.5 sequence are at identical positions as in the HCMV UL56 sequence and 6 cysteine residues are conserved in comparison with the HSV-1 ICP18.5 (amino acids 29, 114, 191, 194, 217, and 501) and with the EBV BALF2 sequence (amino acids 29, 191, 194, 217, 501, and 610). Only two potential N-glycosylation sites (NXT/S) were identified in the MCMV ICP18.5 sequence (positions 319 to 321 and 675 to 677), whereas 6 and 3 glycosylation sites were found in the HCMV UL56 and HSV-1 ICP18.5 sequences, respectively. The sequence KKKRKR (amino acids 768 to 773) at the end of the MCMV ICP18.5 amino acid sequence represents a consensus nuclear targeting sequence.

DISCUSSION

In this communication we report on the nucleotide sequence of the MCMV genes encoding the major DNA binding protein and the ICP18.5 homolog. The expression kinetics of the genes was analyzed and the 5′ and 3′ ends of the mRNAs were determined. The homology of the MCMV MDB and ICP18.5 proteins to the respective proteins of other herpesviruses is shown.

The genes encoding the viral DNA polymerase, the glycoprotein B, a putative assembly protein and the major DNA binding protein constitute a gene block, which is conserved throughout all herpesviruses (Kouzarides *et al.*, 1987). In MCMV, only two genes of this block, the DNA polymerase and glycoprotein B genes, had been identified so far (Elliott *et al.*, 1991; Rapp *et al.*, 1992). We have now analyzed the region upstream of the gB gene and found the MCMV genes encoding the MDB protein and the ICP18.5 homolog. Thus, the organization of this gene block in MCMV exactly mirrors the situation in HCMV.

The MCMV MDBP gene encodes a protein of 1191 amino acids with a predicted molecular mass of 131.7 kDa. The HCMV strain AD169 MDB protein consists of 1235 amino acids (Chee *et al.*, 1990) and the apparent molecular masses of the MDB proteins of simian (strain Colburn) and human CMV have been determined to 129 and 140 kDa (Anders *et al.*, 1986; Kemble *et al.*, 1987). Thus, the predicted molecular mass of the MCMV MDB protein is in the same range as the observed molecular masses of the SCMV and HCMV MDB proteins.

The highest degree of similarity was observed between the amino acid sequences of the MCMV and HCMV MDB proteins. However, the similarity to the amino acid sequences of the HSV-1 and EBV MDB proteins is remarkable. This conservation points to the important function of the MDB proteins. For the HSV-1 ICP8 protein it has been shown that it is essential for the replication of the viral DNA (Wu *et al.*, 1988). It is assumed that the MDB proteins bind to and stabilize the single-stranded DNA immediately after unwinding of the double strand, thereby facilitating the entry of the viral DNA polymerase and the replication of the DNA (Wu *et al.*, 1988). The conservation between the MDB

proteins of so distantly related herpesviruses as HSV-1, EBV, and the CMVs shows that at least for some domains only subtle changes are allowed in order to preserve the function of the protein.

A 4.2-kb mRNA has been identified which originates from the MCMV MDBP gene and which has the capacity to encode the MDB protein. The expression kinetics of the MCMV MDBP gene is consistent with the function of the MDB protein. In HSV-1 it was shown that the ICP8 protein migrates to prereplicative sites prior to DNA replication (Gao and Knipe, 1989), and it was proposed that ICP8 serves as organizational protein for targeting other replication proteins to these sites (Quinlan *et al.,* 1984). For the HCMV and SCMV MDB proteins a subnuclear localization was observed which is reminiscent of the HSV-1 ICP8 protein (Anders *et al.,* 1987; Kemble *et al.,* 1987). Whether the MCMV MDB protein has a similar function and shows a similar subnuclear compartmentalization awaits the production of a specific antiserum and further analyses.

The size of the MDBP mRNA is in accordance with the determined 5′ and 3′ ends of this mRNA and is sufficient to encode a 1191 amino acid polypeptide. Transcripts of similar sizes (3.9 and 4.5 kb, respectively) were detected for the SCMV and HSV-1 MDBP genes (Anders and Gibson, 1988; Rafield and Knipe, 1984). However, one additional transcript of 10 kb was found in SCMV and HSV-1 (Anders and Gibson, 1988; Rafield and Knipe, 1984) and in HCMV only a 10- to 12-kb transcript arising from the MDBP gene could be detected (Kemble *et al.,* 1987). In MCMV only the 4.2-kb mRNA and no larger transcript was seen, even after long exposures of the Northern blot. Therefore, the transcriptional pattern of the MCMV MDBP gene is slightly different from that of the other CMV MDBP genes.

A putative TATA box was found immediately upstream of the determined 5′ cap site of the MDBP mRNA. The determined 3′ end of the MDBP mRNA and the 5′ cap site of the ICP18.5 mRNA are only separated by 10 nt. Thus it is obvious that the 3′ end of the MDBP gene and the promoter of the ICP18.5 gene overlap and it is difficult to dissect the termination signals for the MDBP mRNA from promoter consensus sequences of the ICP18.5 gene. For example, it remains open whether the sequence TATTAAA (positions 3910 to 3916) constitutes a TATA box for the ICP18.5 gene or it represents an atypical polyadenylation signal (Birnstiel *et al.,* 1985) for the MDBP mRNA. Putative binding sites for the transcription factor Sp1 were found 144 and 89 nt upstream of the 5′ cap sites of the MDBP and ICP18.5 mRNAs. This proximal location of Sp1 binding sites is typical for a series of cellular and viral promoters.

The amino acid sequences between the MCMV ICP18.5 protein and its homologs are as highly conserved as the sequences of the MDB proteins and of the glycoproteins B (Rapp *et al.,* 1992). The function of the HSV-1 ICP18.5 protein and its homologs in other herpesviruses was unknown for a long time and has not been completely unraveled until now. Previous analyses with temperature-sensitive HSV-1 mutants pointed to a function in the transport of glycoproteins to the membrane (see for example Pancake *et al.,* 1983). However, recent studies suggested a function of the protein in virus assembly (Addison *et al.,* 1990). This is consistent with a nuclear localization of the Pseudorabies virus ICP18.5 protein (Pederson and Enquist, 1991). The amino acid sequence of the MCMV ICP18.5 protein is not indicative for a particular function of the protein. No typical transmembrane region was found and the sequence contains only two potential N-glycosylation sites. Therefore, a transport of the protein to the cell membrane and glycosylation of the protein appears quite unlikely, whereas the putative nuclear transport signal of the MCMV ICP18.5 protein suggests a nuclear location of the protein.

An interesting feature of the MCMV ICP18.5 gene is the overlapping of the ICP18.5 and gB open reading frames. A similar overlap or at least a close spacing of the ICP18.5 and gB ORFs was also found in other herpesviruses (reviewed in Pederson and Enquist, 1991). Thus, the promoter of the gB gene is located within the coding region of the ICP18.5 gene. The 3′ end of the MCMV ICP18.5 mRNA was mapped within the ORF of the gB gene. This is different from the situation in HSV-1 where a 3′ collinearity of the ICP18.5 and gB transcripts was found and the usage of the same polyadenylation signals for both transcripts is likely (Holland *et al.,* 1984). It is an intriguing question how the transcriptional machinery terminates the MCMV ICP18.5 mRNA when at the same position a readthrough of the gB gene occurs. It is possible that the untypical polyadenylation and 3′ end consensus sequences of the MCMV ICP18.5 gene play a role in ICP18.5 mRNA termination and gB transcription.

The gene block encoding the MDBP, the ICP18.5 homolog, gB, and the DNA polymerase represent the most conserved region in the herpesviral genomes (Kouzarides *et al.,* 1987). All four genes have now been characterized in MCMV (Elliott *et al.,* 1991; Rapp *et al.,* 1992). In the two CMV genomes the spatial relationship of the genes is identical. The conservation between the amino acid sequences of the encoded MCMV and HCMV proteins is striking, and ranges from 64% (gB), 72% (MDBP), 73% (pol), to 74% (ICP18.5) similarity. The slightly higher similarity to the amino acid sequences of the homologous EBV proteins than

to the HSV-1 proteins (Elliott *et al.*, 1991; Rapp *et al.*, 1992) suggests a closer relationship of the cytomegaloviruses to the γ-herpesviruses than to the α-herpesviruses.

In conclusion, this report suggests the notion that MCMV and HCMV are not only similar in biology and pathogenesis but also exhibit extensive homology at the nucleic acid level of genes and in the amino acid sequences of proteins. The close relationship of the two cytomegaloviruses underlines the importance of the murine cytomegalovirus as an appropriate model for studying the biology of CMV infections.

ACKNOWLEDGMENTS

We thank M. Rapp for critical reading of the manuscript and A. Lüske and B. Reutter for preparation of the photographs. This investigation was supported by SFB 322 Grant C6 from the Deutsche Forschungsgemeinschaft.

REFERENCES

ADDISON, C., RIXON, F. J., and PRESTON, V. G. (1990). Herpes simplex virus type 1 UL28 gene product is important for the formation of mature capsids. *J. Gen. Virol.* **71,** 2377–2384.

ANDERS, D. G. (1990). Nucleotide sequence of a cytomegalovirus single-stranded DNA-binding protein gene: Comparison with alpha- and gammaherpesvirus counterparts reveals conserved segments. *J. Gen. Virol.* **71,** 2451–2456.

ANDERS, D. G., and GIBSON, W. (1988). Location, transcript analysis, and partial nucleotide sequence of the cytomegalovirus gene encoding an early DNA-binding protein with similarities to ICP8 of herpes simplex virus type 1. *J. Virol.* **62,** 1364–1372.

ANDERS, D. G., IRMIERE, A., and GIBSON, W. (1986). Identification and characterization of a major early cytomegalovirus DNA-binding protein. *J. Virol.* **58,** 253–262.

ANDERS, D. G., KIDD, J. R., and GIBSON, W. (1987). Immunological characterization of an early cytomegalovirus single-strand DNA-binding protein with similarities to the HSV major DNA-binding protein. *Virology* **161,** 579–588.

BIRNSTIEL, M. L., BUSSLINGER, M., and STRUB, K. (1985). Transcription termination and 3′ processing: The end is in site. *Cell* **41,** 349–359.

BÜHLER, B., KEIL, G. M., WEILAND, F., and KOSZINOWSKI, U. H. (1990). Characterization of the murine cytomegalovirus early transcription unit e1 that is induced by immediate–early proteins. *J. Virol.* **64,** 1907–1919.

CHEE, M. S., BANKIER, A. T., BECK, S., BOHNI, R., BROWN, C. M., CERNY, R., HORSNELL, T., HUTCHISON III, C. A., KOUZARIDES, T., MARTIGNETTI, J. A., PREDDIE, E., SATCHWELL, S. C., TOMLINSON, P., WESTON, K. M., and BARRELL, B. G. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD 169. *Curr. Top. Microbiol. Immunol.* **154,** 125–169.

CHIRGWIN, J. M., PRZYBALA, A. E., MACDONALD, R. J., and RUTTER, W. J. (1979). Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. *Biochemistry* **18,** 5294–5299.

DEVEREUX, J., HABERLI, P., and SMITHIES, O. (1984). A comprehensive set of sequence analysis programs for the Vax. *Nucleic Acids Res.* **12,** 387–395.

ELLIOTT, R., CLARK, C., JAQUISH, D., and SPECTOR, D. (1991). Transcription analysis and sequence of the putative murine cytomegalovirus DNA polymerase gene. *Virology* **185,** 169–186.

FROHMAN, M. A., DUSH, M. K., and MARTIN, G. R. (1988). Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* **85,** 8998–9002.

GAO, M., BOUCHEY, J., CURTIN, K., and KNIPE, D. M. (1988). Genetic identification of a portion of the herpes simplex virus ICP8 protein required for DNA-binding. *Virology* **163,** 319–329.

GAO, M., and KNIPE, D. M. (1989). Genetic evidence for multiple nuclear functions of the herpes simplex virus ICP8 DNA-binding protein. *J. Virol.* **63,** 5228–5267.

HO, M. (1982). "Cytomegalovirus: Biology and Infection." Plenum, New York.

HOLLAND, L. E., SANDRI-GOLDIN, R. M., GOLDIN, A. L., GLORIOSO, J. C., and LEVINE, M. (1984). Transcriptional and genetic analyses of the herpes simplex virus type 1 genome: Coordinates 0.29 to 0.45. *J. Virol.* **49,** 947–959.

JORDAN, M. C., SHANLEY, J. D., and STEVENS, J. G. (1977). Immunosuppression reactivates and disseminates latent murine cytomegalovirus. *J. Gen. Virol.* **37,** 419–425.

KEIL, G. M., EBELING-KEIL, A., and KOSZINOWSKI, U. H. (1984). Temporal regulation of murine cytomegalovirus transcription and mapping of viral RNA synthesized at immediate early times after infection. *J. Virol.* **50,** 784–795.

KEIL, G. M., EBELING-KEIL, A., and KOSZINOWSKI, U. H. (1987). Sequence and structural organization of murine cytomegalovirus immediate-early gene 1. *J. Virol.* **61,** 1901–1908.

KEMBLE, G. W., MCCORMICK, A. L., PEREIRA, L., and MOCARSKI, E. S. (1987). A cytomegalovirus protein with properties of herpes simplex virus ICP8: Partial purification of the polypeptide and map position of the gene. *J. Virol.* **61,** 3143–3151.

KOSZINOWSKI, U. H., DEL VAL, M., and REDDEHASE, M. J. (1990). Cellular and molecular basis of the protective immune response to cytomegalovirus infection. *Curr. Top. Microbiol. Immunol.* **154,** 189–220.

KOUZARIDES, T., BANKIER, A. T., SATCHWELL, S. C., WESTON, K., TOMLINSON, P., and BARRELL, B. G. (1987). Large-scale rearrangement of homologous regions in the genomes of HCMV and EBV. *Virology* **157,** 397–413.

LOH, L. C., BALACHANDRAN, N., and BRITT, W. J. (1991). Characterization of a membrane-associated phosphoprotein of murine cytomegalovirus (pp50) and its immunological cross-reactivity with a human cytomegalovirus protein. *Virology* **183,** 181–194.

MESSERLE, M., BÜHLER, B., KEIL, G. M., and KOSZINOWSKI, U. H. (1992). Structural organization, expression, and functional characterization of the murine cytomegalovirus immediate-early gene 3. *J. Virol.* **66,** 27–36.

PANCAKE, B. A., ASCHMAN, D. P., and SCHAFFER, P. A. (1983). Genetic and phenotypic analysis of herpes simplex type 1 mutants conditionally resistant to immune cytolysis. *J. Virol.* **47,** 568–585.

PASS, R. F. (1991). *In* "Progress in Cytomegalovirus Research" (M. P. Landini, Ed.), pp. 3–15. Elsevier Science, Amsterdam.

PEDERSON, N. E., and ENQUIST, L. W. (1991). Overexpression in bacteria and identification in infected cells of the pseudorabies virus protein homologous to herpes simplex virus type 1 ICP18.5. *J. Virol.* **65,** 3746–3758.

PELLETT, P. E., JENKINS, F. J., ACKERMANN, M., SARMIENTO, M., and ROIZMAN, B. (1986). Transcription initiation sites and nucleotide sequence of a herpes simplex virus 1 gene conserved in the Epstein–Barr virus genome and reported to affect the transport of viral glycoproteins. *J. Virol.* **60,** 1134–1140.

QUINLAN, M. P., CHEN, L. B., and KNIPE, D. M. (1984). The intranuclear location of a herpes simplex virus DNA-binding protein is determined by the status of viral DNA replication. *Cell* **36,** 857–868.

RAFIELD, L. F., and KNIPE, D. M. (1984). Characterization of the major mRNAs transcribed from the genes for glycoprotein B and DNA-

binding protein ICP8 of herpes simplex virus type 1. *J. Virol.* **49,** 960–969.

RAPP, M., MESSERLE, M., BÜHLER, B., TANNHEIMER, M., KEIL, G. M., and KOSZINOWSKI, U. H. (1992). Identification of the murine cytomegalovirus glycoprotein B gene and its expression in vaccinia virus. *J. Virol.*, **66,** 4399–4406.

TABOR, S., and RICHARDSON, C. C. (1987). DNA sequence analysis with a modified bacteriophage T7 DNA polymerase. *Proc. Natl. Acad. Sci. USA* **84,** 4767–4771.

WU, C. A., NELSON, N. J., MCGEOCH, D. J., and CHALLBERG, M. D. (1988). Identification of herpes simplex virus type 1 genes required for origin-dependent DNA synthesis. *J. Virol.* **62,** 435–443.