Vol. 203, No. 1, 1994 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1994 Pages 472-478

CLONING AND SEQUENCE ANALYSIS OF MURINE CYTOMEGALOVIRUS PROTEASE AND CAPSID ASSEMBLY PROTEIN GENES[1]

Jeannette M. Loutsch[φ,ξ], Nancy J. Galvin[φ], Martin L. Bryant[ξ] and Barry C. Holwerda[ξ,2]

[ξ]Searle Infectious Diseases Research, 700 Chesterfield Parkway N., St. Louis, MO 63198

[φ]Department of Pathology, St. Louis University Health Sciences Center, 3635 Vista Ave., St. Louis, MO 63104

Received July 13, 1994

The murine cytomegalovirus U_L80 open reading frame was cloned and the predicted amino acid sequence compared with those from other herpesviruses. The open reading frame encodes a fused protease-capsid assembly protein precursor and maintains conserved features including the active site serine, conserved regions CD1 through CD5, the release and maturation sites, and a potential internal cleavage site within the protease. However, the murine cytomegalovirus protease differs in comparison with the other proteases because it contains a unique 15-16 amino acid insertion between CD3 and CD1. The assembly protein sequences are relatively divergent, but they can be arranged into groups defined by herpesvirus subfamily, with each group possessing a conserved motif at its carboxyl terminus.

MCMV is a herpesvirus closely related to the human pathogen, HCMV [1]. All members of the herpesvirus family encode a unique serine protease in association with a capsid assembly protein [2, 3]; the associated ORF is designated U_L80 in HCMV, APNG in SCMV and U_L26 in HSV-1. Transcription from this ORF produces multiple 3' co-terminal mRNAs in virus-infected cells with the largest transcript encoding a fused precursor of the protease and capsid assembly protein and a smaller, more abundant, transcript encoding the assembly protein alone [4]. The protease is autocatalytically removed from the initial translation product at the "release" site and

[1]The nucleotide sequence reported in this paper has been submitted to the GenBank/EMBL Data Bank with accession number L34690.

[2]To whom correspondence should be addressed.

Abbreviations used:
MCMV: murine cytomegalovirus; HCMV: human cytomegalovirus; ORF: open reading frame; SDS: sodium dodecyl sulfate; nt: nucleotide; kb: kilobase pairs; SCMV: simian cytomegalovirus; HSV-1: herpes simplex virus type-1; EHV: equine herpesvirus; VZV: varicella-zoster virus; ILTV: infectious laryngotracheitis virus; EBV: Epstein-Barr virus; HVS: herpesvirus saimari.

0006-291X/94 $5.00

further cleaves the assembly protein at a single "maturation" site; both cleavage sites have the consensus sequence LxA/S or VxA/S ("/" denotes the scissile peptide bond and "x" denotes any amino acid) [3, 5]. There is considerable interest in these proteases because they are necessary for viral replication [6-8] making them attractive targets for antiviral drugs.

We have cloned and sequenced the MCMV U_L80 ORF. Alignment with other herpesviruses shows that the MCMV protease is highly conserved and that the assembly proteins can be grouped by herpesvirus subfamily.

EXPERIMENTAL PROCEDURES

Viral DNA Preparation MCMV (Smith strain) was obtained from American Type Culture Collection and propagated on murine cell line C127 in Earle's minimal essential media containing 5% fetal bovine serum at 37° in a 5% CO_2 atmosphere. Viral DNA was isolated from the media of viral cultures as described previously [9].

Cloning of MCMV U_L80 ORF Restriction endonuclease fragments of MCMV DNA were separated by agarose gel electrophoresis and transferred to nylon membranes (0.45 μm, PhotoGene, Gibco BRL) as described previously [10]. The membrane was prehybridized in 6x SSPE (1x SSPE is 10 mM sodium phosphate buffer, pH 7.4 containing 150 mM NaCl and 1 mM EDTA), 0.1% SDS, 1x Denhardt's solution (0.02% Ficoll, 0.02% polyvinylpyrrolidone, 0.02% bovine serum albumin), and 100 μg/ml calf thymus DNA (Sigma) for 4 h at 51° before addition of the radioactively-labeled DNA probe. Following hybridization for 8 h at 51° the membrane was washed with 2x SSPE containing 0.1% SDS for 30 min at 20° before exposure of X-ray film (Kodak X-Omat AR) at -80°. The probe was made by random priming [11] in the presence of [^{32}P]dCTP of a cloned DNA fragment encoding HCMV U_L80 protease region (HCMV AD169 nt 115198 to nt 115969) [12]. Several MCMV PstI fragments approximately 5 kb in size were recovered from an electrophoretic gel, ligated into the PstI site of pUC18 and the ligation mixture used to transform *Escherichia coli* strain DH5α [13]. Clones containing 5 kb PstI fragments were further screened by hybridization to the HCMV U_L80 probe as described above. Membranes were washed with 2x SSPE, 0.1% SDS for 20 minutes at 20°, 30°, 35°, 40°, 45°, and 50°C with exposure of X-ray film between washes.

DNA Sequence Analysis DNA sequence was determined using the dideoxynucleotide chain termination method [14] with *Thermus aquaticus* DNA polymerase [15] (PRISM Ready Reaction DyeDeoxy Terminator Cycle Sequencing Kit, Applied Biosystems) on an Applied Biosystems Model 373A DNA Sequencing System. DNA sequences were compiled using University of Wisconsin GCG fragment assembly program.

RESULTS AND DISCUSSION

We used the HCMV U_L80 protease sequence as a probe to identify the homologous MCMV U_L80 ORF. Under low stringency wash conditions the HCMV protease probe hybridized weakly to a single 5-kb fragment in PstI-digested MCMV DNA (Fig. 1B, lane 5), 18- and 5-kb fragments in BamHI digests (Fig 1B, lane 3) and a single >20 kb-fragment in HindIII digests (Fig. 1B, lane 4). No hybridization signals were detected in SalI digests (Fig. 1B, lane 6). This was probably due to cleavage of the MCMV U_L80 ORF into three fragments at two SalI sites (see Fig. 3) coupled with moderate DNA sequence homology (59%) with the HCMV protease probe. However, hybridization to the MCMV DNA fragments was specific because the probe bound only to the expected U_L80-containing fragment in HCMV digests (i.e. 6.2-kb BamHI; Fig.1B, lane 7) [12].

Several different 5-kb fragments were generated by digestion of MCMV DNA with PstI (Fig 1A, lane 5). After cloning into pUC18 [16], recombinants containing unique PstI inserts

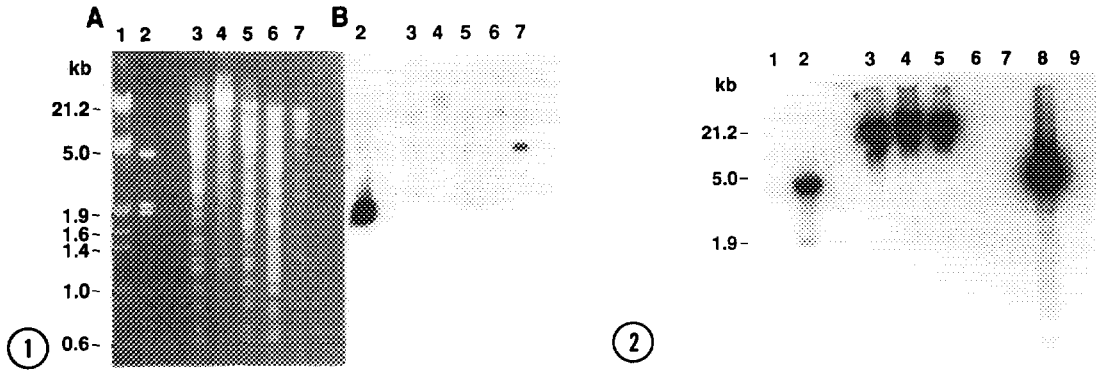


Fig. 1. Electrophoretic gel (**A**) and autoradiograph (**B**) of MCMV DNA hybridized with HCMV protease probe. The ethidium bromide-stained 0.8% agarose gel in panel (**A**) was used to make the blot in panel (**B**) hybridized with a radioactively-labeled HCMV protease probe. *Lane 1*, DNA size standards (bacteriophage lambda DNA digested with HindIII plus EcoRI); *lane 2*, PstI digest of pMON22203 (HCMV BamHI nt 112520 to SalI 117499 [1] containing the U_L80 open reading frame cloned in pUC18); MCMV DNA digested with: *lane 3*, BamHI; *lane 4*, HindIII; *lane 5*, PstI; and, *lane 6*, SalI; *lane 7*, HCMV DNA digested with BamHI. The sizes of the DNA lambda standards are indicated on the left.

Fig. 2. Autoradiograph of MCMV DNA probed with MCMV U_L80 probe. An 0.8% agarose electrophoretic gel was transferred to a membrane and subjected to hybridization with the MCMV U_L80-containing insert (PstI) of pJML28. The blot was washed for 1 h at 65° before exposing film for 14 h. *Lane 1*, pMON22203 PstI digest; MCMV DNA digested with: *lane 2*, PstI; *lane 3*, EcoRI; *lane 4*, HindIII; and *lane 5*, XbaI; PstI digests of independent 5-kb PstI MCMV fragments cloned into pUC18: *lane 6*, pJML23; *lane 7*, pJML24; *lane 8*, pJML28; and *lane 9*, pJML29.

were screened by hybridization to the HCMV protease probe (data not shown). A single clone, pJML28, was selected and further tested for hybridization to specific MCMV DNA fragments. The MCMV genome is collinear with HCMV [17] and, therefore, a probe containing MCMV U_L80 should bind to the mapped MCMV restriction fragments HindIII C, XbaI B, and EcoRI A each of which is greater than 20 kb in size [18, 19]. When used as a probe, the pJML28 PstI insert bound to single MCMV EcoRI, HindIII and XbaI fragments of appropriate sizes (>20 kb) as predicted for a probe containing MCMV U_L80 (Fig. 2, lanes 3, 4, and 5, respectively). The same probe hybridized to a 5-kb MCMV PstI fragment (Fig. 2, lane 2), but not to other independent PstI clones (Fig. 2, lanes 6,7, and 9). Based on these results pJML28 was chosen for DNA sequence analysis.

Our initial sequence data was obtained using a degenerate oligonucleotide primer, 5'-TATAAGCTTCCSCTNAAYRTCCAAYCAC-3', that hybridized to the conserved CD2 region of the protease [3, 20]. This primer generated DNA sequence encoding a polypeptide with the conserved active site serine in the CD3 motif in addition to the I-site [20]. Sequencing was completed for both strands using primers designed from accumulated sequence data and extended beyond the 3'- and 5'-ends of the ORF (Fig. 3). The MCMV U_L80 ORF encodes a 697 residue protease-assembly protein precursor with a number of conserved features including the active site Ser^{128} (Fig. 3, boxed) and the release and maturation cleavage sites at Ala^{269} and Ala^{655},

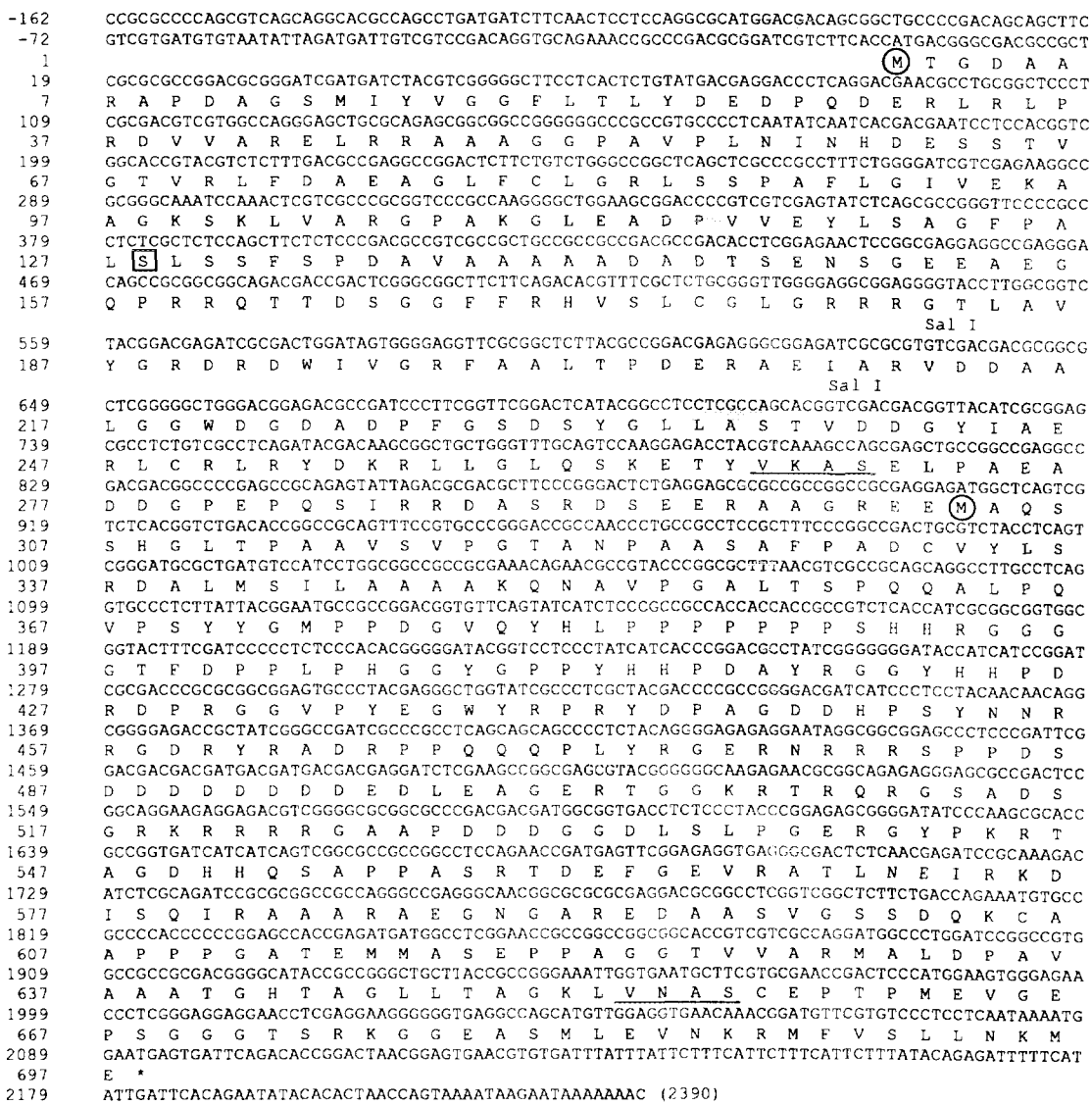


Fig. 3. Nucleotide sequence and deduced amino acid sequence of the MCMV U_L80 open reading frame. Nucleotides are numbered in the 5' to 3' direction beginning at the first base of the initiator ATG codon. The deduced amino acid sequence, numbered from the methionine initiating in the protease domain, is shown *below* the nucleotide sequence using *single-letter* code. The initiator methionines for the protease and capsid assembly protein, at *positions 1 and 303*, respectively, are enclosed in *circles* and the active site serine is enclosed in a *box*. The release cleavage site, *amino acids 267 to 270*, and the maturation cleavage site, *amino acids 653 to 656*, are *underlined*. The positions of SalI cleavage sites are indicated *above* the nucleotide sequence.

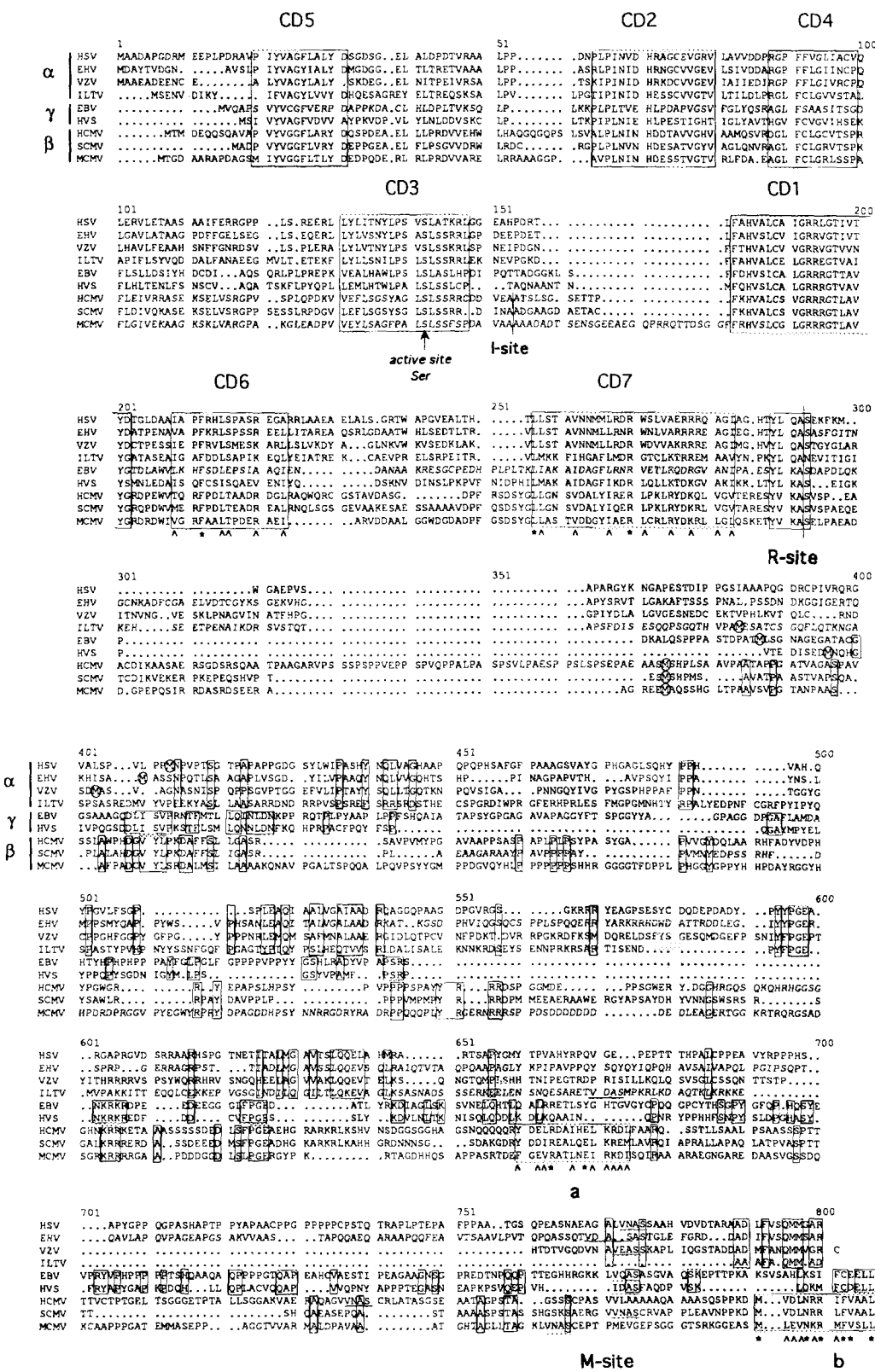
respectively (Fig. 3, underlined) sharing the consensus VxA/S. The respective initiating methionines for the protease and assembly proteins were assigned by alignment to the homologous ORFs of other herpesviruses (see Fig. 4) and compliance with the rules of Kozak [21].

Multiple sequence alignment with other herpesviruses indicates that the MCMV U_L80 ORF possesses many of the common features of the protease and assembly protein (Fig. 4). The protease domain, defined by the release cleavage site at Ala^{269}, maintains conserved domains, CD1 through CD5 [3, 20], and the release site (Fig. 4, "R-site") with additional conservation of the internal cleavage site found in other cytomegalovirus proteases (Fig. 4, "I-site") [20, 22-24]. The I-site has the sequence VEA/A for HCMV, INA/A for SCMV and putatively VAA/A for MCMV. In contrast to the other enzymes, MCMV protease has a unique 15-16 amino acid insertion following the I-site (Fig. 4, positions 166 to 181) that may constitute an extended loop in the enzyme. In addition to CD1 through CD5 [3, 20], we have designated two further regions, CD6 and CD7, that are conserved among herpesvirus proteases. CD6 (Fig. 4, positions 209 through 223) contains an invariant Phe plus 5 chemically conserved residues and CD7 (Fig. 4, positions 257 through 283) contains an invariant Leu and Arg residues plus 8 chemically similar residues.

The capsid assembly proteins show greater divergence among herpesviruses in comparison to the proteases (Fig. 4). This is seen by the reduced incidence of invariant amino acids and the greater number of gaps inserted to promote the alignment. The MCMV assembly protein has a predicted initiator Met at position 374 and a maturation site (Fig. 4, "M-site") 42 residues from the carboxyl terminus. Although the assembly proteins have diverged across the entire herpesvirus family, it is evident they are more closely related within subfamilies as denoted in Figure 4 for the respective alpha- (HSV, EHV, VZV, and ILTV), gamma- (EBV and HVS), and betaherpesviruses (HCMV, SCMV, and MCMV). Particularly strong conservation occurs within the betaherpesvirus assembly protein sequences in two regions designated "**a**" and "**b**" (Fig. 4). Region "**a**" contains conserved Arg and Glu residues plus 9 amino acids with similar chemical properties. Region "**b**", at the extreme carboxyl terminus, is more conserved with 9 invariant residues in addition to 6 with chemical similarity in the motif MxxxNxRxFVxxLNKxE. Conservation at the carboxyl terminus is also seen for the other herpesvirus subfamilies, but with different motifs. The alphaherpesvirus assembly proteins have the motif, ADxFxxQMMxAR, while the gammaherpesviruses have the

---

Fig. 4. Amino acid sequence alignment of herpesvirus protease and capsid assembly protein homologues. The deduced amino acid sequences of MCMV U_L80 and eight herpesvirus homologues were compared using the University of Wisconsin GCG program PileUp. Deduced amino acid sequences were obtained by translation of the respective nucleotide sequences retrieved from GenBank for the herpes simplex virus type-1 U_L26 ORF (*HSV*), equine herpesvirus ORF (*EHV*), varicella-zoster ORF 33 (*VZV*), infectious laryngotracheitis virus p80 ORF (*ILTV*), Epstein-Barr virus BVRF2 ORF (*EBV*), herpesvirus saimari ORF (*HVS*), human cytomegalovirus AD169 U_L80 ORF (*HCMV*), and simian cytomegalovirus APNG ORF (*SCMV*). Five previously noted conserved regions, *CD1* through *CD5* [2, 3],in addition to new regions designated here, *CD6* and *CD7*, are enclosed in *boxes* in the protease domain (*positions 1 through 292*). The internal cleave site, *I-site*, is indicated by a *line* between positions 153 and 154 and the release cleavage site, *R-site*, is enclosed in a *box* with a line through the scissile bonds between *positions 292 and 293*. The maturation cleavage sites, *M-site*, are underlined for each assembly protein homologue at the appropriate locations within *positions 737 to 776*. The putative maturation cleavage site for ILTV is underlined from *position 670 to 673*. The putative initiator methionines for the assembly proteins (*positions 374 to 813*) are *circled*. Identical residues for the gamma- (γ) and betaherpesviruses (β) and those with 3 identical out of 4 in the alphaherpesviruses (α) are enclosed in *boxes* to highlight the clustering of the assembly proteins within herpesvirus subfamilies. Identical residues are indicated by asterisks (*) and chemically similar residues by carets (^) within the conserved regions, *CD6* and *CD7*, of the proteases and "**a**" and "**b**" in the assembly proteins.


CD5
CD2
CD4
CD3
CD1
CD6
CD7
active site
Ser
I-site
R-site
M-site
a
b
α
γ
β
HSV
EHV
VZV
ILTV
EBV
HVS
HCMV
SCMV
MCMV

motif FCxELLNK (Fig. 4). The role of these conserved motifs is unknown, but their conservation suggests that it will tolerate little variation. Since these motifs are specific to herpesvirus subfamilies, it is possible they interact with other subfamily-specific components. This possibility is particularly intriguing since cleavage at the maturation site (Fig. 4, underlined), an event necessary for viral replication [6-8], separates this motif from the bulk of the assembly protein.

In summary, sequence analysis of the MCMV U_L80 open reading frame shows that, with the exception of a 15-16 residue insertion, the protease is conserved with other herpesviruses. However, the capsid assembly proteins are clustered into groups defined by herpesvirus subfamily with each possessing a conserved motif at the carboxyl terminus.

REFERENCES

1. Ho, M. (1991) in Cytomegalovirus Biology and Infection (2 ed.) pp. 327-353, Plenum Medical Book Company, New York.
2. Liu, F., and Roizman, B. (1991) J. Virol 65, 5149-5156.
3. Welch, A. R., Woods, A. S., McNally, L. M., Cotter, R. J., and Gibson, W. (1991) Proc. Natl. Acad. Sci. USA 88, 10792-10796.
4. Welch, A. R., McNally, L. M., and Gibson, W. (1991) J. Virol. 65, 4091-4100.
5. DiIanni, C. L., Drier, D. A., Deckman, I. C., McCann III, P. J., Liu, F., Roizman, B., Colonno, R. J., and Cordingley, M. G. (1993) J. Biol. Chem. 268, 2048-2051.
6. Preston, V. G., Coates, J. A. V., and Rixon, F. J. (1983) J. Virol. 45, 1056-1064.
7. Preston, V. G., Rixon, F. J., McDougall, I. M., McGregor, M., and Kobaisi, M. F. A. (1992) Virol. 186, 87-98.
8. Gao, M., Matusick-Kumar, L., Hurlburt, W., DiTusa, S. F., Newcomb, W. W., Brown, J. C., McCann III, P. J., Deckman, I., and Colonno, R. J. (1994) J. Virol. 68, 3702-3712.
9. D'Aquila, R. T. D., and Summers, W. C. (1987) J. Virol. 61, 1291-1295.
10. Southern, E. M. (1975) J. Mol. Biol. 98, 503-517.
11. Feinberg, A. P., and Vogelstein, B. (1983) Anal. Biochem. 132, 6-13.
12. 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., *et al.* (1990) in Current Topics in Microbiology and Immunology pp. 125-169, Springer-Verlag, Berlin/Heidelberg.
13. Maniatis, T., Fritsch, E. F., and Sambrook, J. (1982) in Book Molecular cloning: A Laboratory Manual (1 ed) 545 pages, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
14. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463- 5467.
15. Innis, M. A., Myambo, K. B., Gelfand, D. H., and Brow, M. A. D. (1988) Proc. Natl. Acad. Sci. USA 85, 9436-9440.
16. Yanisch-Perron, C., Vieira, J., and Messing, J. (1985) Gene 33, 103-119.
17. Rawlinson, W. D., Farrell, H. E., and Barrell, B. G. (1993) in XVIII International Herpesvirus Workshop. Pittsburgh, PA, abstract A-61.
18. Ebeling, A., Keil, G. M., Knust, E., and Koszinowski, U. H. (1983) J. Virol. 47, 421-433.
19. Mercer, J. A., Marks, J. R., and Spector, D. H. (1983) Virol. 29, 94-106.
20. Welch, A. R., McNally, L. M., Hall, M. R. T., and Gibson, W. (1993) J. Virol. 67, 7360-7372.
21. Kozak, M. (1987) Nucleic Acids Res. 15, 8125-8148.
22. Baum, E. Z., Bebernitz, G. A., Hulmes, J. D., Muzithras, V. P., Jones, T. R., and Gluzman, Y. (1993) J. Virol. 67, 497-506.
23. Holwerda, B. C., Wittwer, A. J., Duffin, K. L., Smith, C., Toth, M. V., Carr, L. S., Wiegand, R. C., and Bryant, M. L. (1994) submitted.
24. Jones, T. R., Sun, L., Bebernitz, G. A., Muzithras, V. P., Kim, H.-J., Johnston, S. H., and Baum, E. Z. (1994) J. Virol. 68, 3742-3752.