Novel virus-associated proteins encoded by UL112–113 of human cytomegalovirus

Shang-Kwei Wang,1 Cheng-Hui Hu,1 Miao-Chan Lu,1 Chang-Yih Duh,2 Pao-Chi Liao3 and Yu-Chang Tyan3

1Department of Microbiology, Institute of Medicine, College of Medicine, Kaohsiung Medical University, 100 Shih-Chuan 1st Road, Kaohsiung 80708, Taiwan, ROC.
2Asia-Pacific Ocean Research Center, National Sun Yat-sen University, 70 Lien-Hai Road, Kaohsiung 80424, Taiwan, ROC
3Department of Environmental and Occupational Health, College of Medicine, National Cheng Kung University, 138 Sheng-Li Road, Tainan 704, Taiwan, ROC

Evidence suggests that the products of the human cytomegalovirus (HCMV) UL112–113 genes are involved in viral DNA replication during lytic infection. A polyclonal antibody was raised against the UL112 open reading frame (ORF) to characterize its function in detail. Immunoblots utilizing the UL112 antibody identified seven distinct protein bands (p20, p26, p28, p34, p43, p50 and p84) expressed during the HCMV infectious cycle. After screening a cDNA library constructed from cells 72 h after infection with HCMV, only four different cDNA protein-producing constructs were obtained, and their ORFs corresponded to p34, p43, p50 and p84. The proteins p20, p26 and p28 were further shown to be selectively included within mature HCMV particles, virions, non-infectious enveloped particles and dense bodies. Immunoaffinity protein purification was used to prepare the samples for liquid chromatography coupled to tandem mass spectrometry. This analysis revealed that p20, p26 and p28 were derived from the UL112 ORF, most likely through post-translational proteolytic cleavage.

INTRODUCTION

Human cytomegalovirus (HCMV), a member of the Herpesviridae family, is a pathogen that is ubiquitously present in the human population. Most people who acquire HCMV display no signs or have mild symptoms and no long-term health consequences during its life-long latency in healthy individuals. Nevertheless, the virus causes severe congenital abnormalities in neonates and reactivation of the latent virus often complicates opportunistic infections in immunosuppressed patients (Britt, 2008; Mocarski et al., 2007).

Our research goal is to understand the regulation of HCMV lytic replication. The HCMV lytic replication cycle manifests itself in a temporal regulatory mode. Major immediate-early proteins (MIEs; IE1 and IE2) act as promiscuous activators that manipulate cellular gene expression and regulate early and late viral gene expression. Expressed as an early gene, UL112–113 is synergistically activated by the action of MIE proteins (Mocarski et al., 2007). Additionally, computational analyses indicate that UL112–113 may potentially be regulated by miR-UL112-1 (Grey et al., 2007) and miR-UL54-1 (Murphy et al., 2008).

Several lines of evidence suggest that the UL112–113 locus is implicated in HCMV orLyt-dependent replication in a cell culture system (Pari et al., 1993). When UL112–113 expression is blocked via a constitutively expressed anti-sense RNA, the efficiency of viral replication is severely hindered (Yamamoto et al., 1998). Again, this result is consistent with the findings that HCMV with a genomic deletion in UL112 or UL113 results in greatly reduced production of progeny (Dunn et al., 2003). At the early stages of infection, UL112–113 proteins appear in nuclear domain 10 (ND10) (Ahn et al., 1999; Penfold & Mocarski, 1997). UL112–113 proteins act as pre-replication factors that seed the replication complex by sequentially recruiting single-stranded DNA-binding protein (pUL57) and DNA polymerase processivity factor (pUL44) to ND10 (Park et al., 2006; Penfold & Mocarski, 1997). Furthermore, this observation is strengthened by the fact that UL112–113 interacts with UL44 (Iwayama et al., 1994; Park et al.,...
2006). Previous reports indicate that UL112–113 acts with other essential replication genes (UL36–38, UL122–123, IRS1 and TRS1) to upregulate the UL54 and UL44 promoters (Iskenderian et al., 1996; Kerry et al., 1996). A recent investigation further illustrates that UL112–113 is able to fully activate lytic replication of Kaposi’s sarcoma-associated herpesvirus (KSHV) (Wells et al., 2009).

UL112–113 has been shown to encode a class of cognate phosphoproteins (p34, p43, p50 and p84) that share 252 aa at the N terminus, while the variable C-termini are generated from alternative splicing (Iwayama et al., 1994; Staprans & Spector, 1986; Wright & Spector, 1989; Wright et al., 1988). In this study, we attempted to characterize the functions of the proteins encoded by UL112–113. By cloning UL112–113 and expressing its gene products, we found that this locus produces three additional, not previously described, proteins: p20, p26 and p28. Moreover, these novel proteins are selectively associated with mature HCMV particles, while the four associated known proteins are not.

In the present study, we reveal the context of these novel proteins (p20, p26 and p28). Direct proteomic evidence showed that novel virus-associated proteins are probably derived from proteolytic cleavage in a common region of the N terminus of UL112.

METHODS

Antibodies. The generation of polyclonal antibody to UL112 ORF is described in detail in the Supplementary Material (available in JGV Online). Murine monoclonal antibodies against pUL55 (gB) and pUL99 (tegument protein) were obtained from the Rumbaugh-Goodwin Institute for Cancer Research.

Virus production and cell culture. Human embryonic lung (HEL) cells (HEL299) and HCMV strain AD169 (ATCC VR-538) were obtained from the ATCC. Virus propagation and cell culture maintenance were conducted as described previously (Wang et al., 2000, 2004). For the time-course of viral infection, 1.5 × 10⁶ HEL cells were plated in 60 mm culture dishes. Twenty-four hours later, cells were infected with HCMV AD169 at an m.o.i. of 2–3 p.f.u. per cell. At the indicated time points, the infected cells were harvested by washing three times with PBS buffer. Cells were lysed in RIPA buffer. 50 mM Tris (pH 7.5), 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate and 1 mM EDTA) containing a complete protease inhibitor cocktail (Roche), and were incubated on ice for 30 min. Protein extracts were recovered from the supernatant after centrifuging the cell lysates at 20000 g at 4 °C for 30 min.

cDNA library construction, gene screening and protein production in a cell-free system. To construct the cDNA library, HEL cells were infected with HCMV at an m.o.i. of 2–3 p.f.u. per cell. After 3 days of infection, total RNA was prepared by following the acid-phenol-thiocyanate protocol. High purity polyadenylated RNA was obtained by passing the total RNA through an oligo(dT)-cellulose column three times. A unidirectional lambda ZAP express cDNA library was constructed according to the manufacturer’s guidelines (Stratagene). For the cDNA library screening, 1 × 10⁷ p.f.u. of primary recombinant phage were plated out in dishes before being immobilized in Nytran N² membranes. A 429 bp Pst–BamHI fragment (nt 161791–162215), which encoded aa 91–233 of UL112 ORF (Chee et al., 1999; Davison et al., 2003), was labelled with [³²P]dCTP by random priming. Phage colonies with isotope-labelled signals underwent either secondary or tertiary screening until a single colony was isolated for each. Co-inoculation with helper phage ExAssist (Stratagene) recovered the phagemid DNA from the phage culture. The DNA was then used for transformation and subsequent identification.

Protein products of the UL112–113 cDNA clones were subcloned behind the SV40 early gene promoter in the phHK-3 cloning vector. Eukaryotic expression constructs containing p34, p43, p50 and p84 ORFs were constructed and designated pECEp34, pECEp43, pECEp50 and pECEp84, respectively. The p34 ORF was PCR-amplified from the pCEp34 plasmid (described in Results) with a 5′ primer containing a Sac restriction site (5′-GATCGAGCTCCAGGAGGTTCGCT-3′; restriction sites are underlined) and a 3′ primer with a Sal restriction site (5′-ATGTCGACCAACACCGCTGCTGCTG-3′). The p34 PCR product and vector phHK-3 were both digested with SacI and SalI and ligated together. The resulting construct was designated pREP34. The p43 ORF sequence was amplified from the pCEp43 plasmid using a 5′ primer with a Sac site (5′-ATGTCGACCAACACCGCTGCTGCTG-3′) and a 3′ primer with an EcoRI site (5′-ATGTCGACCAACACCGCTGCTGCTG-3′). The p34 PCR product and vector pHK-3 were both digested with SacI and EcoRI and then ligated. The resulting construct was designated pREP43. To clone the p50 ORF, pCEp50 was deleted from the 3′ untranslated region (3′-UTR) by cutting with XhoI, and religated with the remaining sequences. The construct was designated pREP50. All the constructs with frameshift mutations described in this report were at the PstI site of exon1 (aa 91, nt 161791). To create the frameshift mutation, expression plasmids were cut with PstI, blunted with T4 polymerase and religated. These mutated constructs were denoted by adding an ‘m’ after the name of their wild-type counterparts, i.e. pREP50 m.

HCMV particle purification and fractionation of viral proteins. The mature viruses were purified by a modified rate-velocity glycerol-tartrate gradient centrifugation as described previously (Irmiere & Gibson, 1983; Talbot & Almeida, 1977; Wang et al., 2004). Virion particles were subjected to a second gradient centrifugation in order to ensure their purity. Virions were treated with detergent and the envelope and tegument/capsid of the purified virion particles were fractionated into a detergent-soluble supernatant and an insoluble pellet, according to previously described methods (Lehner et al., 1989; Wang et al., 2004; Yao & Courtney, 1992).

Immuonaffinity purification of proteins. The UL112 antibody was purified by ammonium sulphate precipitation. Initially, the unwanted fraction of the rabbit serum that contained the UL112 antibody was removed by the addition of saturated ammonium sulphate solution at a final concentration of 30%. After overnight incubation at 4 °C, the supernatant was collected by centrifugation at 3000 g at 4 °C for 30 min. Rabbit serum with UL112 antibody was then precipitated by the addition of saturated ammonium sulphate at a final concentration of 50%. The precipitated pellet enriched with UL112 antibody was harvested by centrifugation, redissolved and dialysed against antibody binding/washing buffer [50 mM sodium borate (pH 8.2)]. UL112 antibody (11 mg) was bound to 2 ml protein A-Sepharose beads (Pierce) by inverting the slurry for 30 min at room temperature, and the unbound antibody was washed away with antibody binding/washing buffer. In order to immobilize and covalently conjugate the antibody to the protein A-Sepharose, disuccinimidyl substrate was used to cross-link the antibodies to the beads. The remaining unreacted NHS–ester group of the disuccinimidyl substrate was
blocked in 0.1 M ethanolamine (pH 8.2). The antibody that was not cross-linked with protein A was eluted in wash buffer containing 0.1 M glycine (pH 2.5). Afterwards, the slurry was packed in a column and equilibrated in antibody binding/washing buffer.

To prepare the protein extracts, 120 15 cm dishes of HEL cells were infected with HCMV at an m.o.i. of 2–3 p.f.u. per cell. After 72 h infection with HCMV, HEL cells were washed three times with PBS and cell pellets were harvested by centrifugation at 1000 g at 4 °C for 20 min. To prepare the cell lysates, the pellets were thawed and resuspended in RIPA lysis buffer. To release the cellular proteins completely, the cell suspension was further homogenized with a dounce homogenizer. RNase and DNase were added at 100 µg ml⁻¹ to the cell suspension and the digestion was carried out at 25 °C for 1 h. Soluble proteins were collected from the supernatant of the cellular suspension after centrifugation at 64 000 g at 4 °C for 1 h. The protein extract was added to an equal volume of binding buffer [50 mM Tris (pH 7.5)], applied to the protein A-Sepharose column with immobilized UL112 antibody and allowed to incubate in the column at 4 °C with rocking for 2 h. Bound proteins were eluted in 1 ml aliquots from the column with eight vols elution buffer [0.1 M glycine (pH 2.7)]. Fractions were immediately neutralized by adding 50 µl 1 M Tris (pH 9.5). The eluate was dialysed against double distilled water for 24 h and lyophilized.

**Mass spectrometry.** Protein fractions purified from the immunofinity column were separated by SDS-PAGE. Western blot was carried out in parallel and the corresponding protein bands that were immunoreactive with UL112 antibody were excised from the Coomassie brilliant blue-stained gel. Each protein band was subjected to protein sequencing analysis using high performance liquid chromatography–tandem mass spectrometry (HPLC–MS/MS) followed by mapping mass spectral data to protein sequence databases (Tyan et al., 2005). In brief, sliced gel pieces were in-gel digested with trypsin (Promega) and the digested peptides were extracted from the gel pieces by incubation with 50 % (v/v) acetonitrile/5 % (v/v) formic acid. Dry peptides were redissolved in 20 µl 5 % (v/v) acetonitrile/0.1 % (v/v) formic acid and loaded into an HPLC–MS/MS analyser for protein identification.

Tryptin-digested peptides were first fractionated using a C18 microcapillary column (75 µm × 15 cm) with a reverse phase nano-HPLC apparatus (LC Packings) coupled with ionization tandem mass spectrometry (LCQ DECA XP Plus, ThermoFinnigan). The mobile phases for the peptide elution were buffer A (5 % acetonitrile/0.1 % formic acid) and buffer B (80 % acetonitrile/0.1 % formic acid). The protein sequences were matched against a virus protein database using the MASCOT software (Matrix Science; http://www.matrixscience.com/search_form_select.html). The amino acid modifications that were allowed were oxidation at methionine, carbamidomethylation at cysteine/lysine, pyroglutamate at asparatic acid/glutamic acid, and deamidation at asparagine/glutamine. Identification of a protein was considered to be significant according to the Mowse score (P<0.05).

**DNA and protein sequences.** DNA and degenerated protein sequence data were compiled and analysed with the Genetic Computational Groups software package. Complete nucleotide sequences of the full-length cDNAs for p34, p43, p50 and p84 were deposited in GenBank with the accession numbers U57433, U57434, U57431 and U57432, respectively.

**RESULTS**

**Expression of UL112 proteins during the HCMV replication cycle**

To characterize the protein products of UL112–113, an anti-UL112 antibody was generated (Supplementary Fig. S1, available in JGV Online) and used to investigate the expression kinetics of UL112-encoded proteins during the HCMV infectious period. Viral permissive HEL cells were infected with HCMV AD169 and protein extracts were harvested at 3, 8, 16, 24, 48 and 72 h post-infection (p.i.). Previously reported proteins with apparent molecular masses of 34, 43, 50 and 84 kDa were detected with the UL112 antibody and designated p34, p43, p50 and p84 (Iwayama et al., 1994; Wright et al., 1988). Throughout the entire course of infections, the polyclonal antiserum repeatedly reacted with three additional protein bands with apparent molecular masses of 20, 26 and 28 kDa. These have been tentatively designated p20, p26 and p28, and they represent a novel group of proteins. At 8 h p.i., the p20, p26, p28, p34, p43 and p50 proteins began to appear (Fig. 1, lane 3) and increase steadily throughout the experiment (Fig. 1, lanes 4–8). In contrast, p84 was usually detected at 72–96 h p.i. (Fig. 1, lane 8). The proteins with apparent molecular masses of 97 kDa, observed both in mock- and viral-infected cells, were non-specific cellular proteins that were cross-reactive with UL112 antiserum. Protein products that were consistently detected by the anti-UL112 antibody were evaluated for matches to cloning sequences and/or their amino acid sequences were verified by mass spectrometric analysis. Protein bands that failed to meet these criteria were designated unassigned proteins (e.g. see the 38 kDa protein in Figs 1, 2c and 4b).

**Fig. 1.** Western blot analysis of UL112-encoded proteins during the HCMV replication cycle. HEL cells were infected with an m.o.i. of 2–3 p.f.u. per cell and cell lysates harvested at the indicated times p.i. were resolved by SDS-PAGE. A Western blot of these samples was probed with anti-UL112 antibody. Lanes: 1, protein extracts from mock-infected cells; 2–7, protein extracts harvested at 3, 8, 16, 24, 48 and 72 h p.i., respectively. Sizes of the molecular mass markers are shown on the left. The proteins detected by the anti-UL112 antibody are named according to their apparent molecular masses as assessed by acrylamide gel migration, as indicated on the right. The unassigned protein band is marked with an asterisk (*).
Fig. 2. (a) A schematic depiction of the exon–intron layout of the full-length cDNA producing p34, p43, p50 and p84 encoded by the HCMV UL112–113. Restriction endonuclease sites in the UL112–113 gene region are shown. The corresponding nucleotides in the HCMV genomic sequence (annotated HCMV genome) are given. The PstI–BamHI (thick line) fragment was used as a probe for screening the cDNA library. The closed downward arrow denotes the cDNA polyadenylation site (nt 163864). Exons are boxed and hatched regions indicate the homologous amino acid sequences of exon1 and exon1a. The black regions are the identical amino acid sequences between exon2 and exon5. The nucleotide lengths of the exons and introns are numbered at the top. Closed diamonds (●) indicate the stop codon. (b) Protein products of UL112–113 full-length cDNA clones (pCEp34, pCEp43, pCEp50 and pCEp84) synthesized in a reticulocyte cell-free TNT system. Lanes: 1, pCEp34; 2, pCEp43; 3, pCEp50; 4, pCEp84. (c) In vivo transient eukaryotic expression of proteins from constructs encoding full-length cDNA regions of the UL112–113 genes and their respective frameshift mutations. Plasmids pECEp34, pECEp43, pECEp50 and pECEp84 were transfected into HEL cells (lanes 2–5). As controls, the cloning vector pHK-3 and frameshift mutants pECEp34 m, pECEp43 m, pECEp50 m and pECEp84 m were transiently expressed in parallel (lanes 1 and 6–9). The unassigned protein band is marked with an asterisk (*). (d) Transient expression of genes encoding UL112–113 ORF cDNAs but lacking the 3′-UTR, as well as their corresponding frameshift mutants. Plasmids pREp34, pREp43, pREp50 and pREp84 were transfected into HEL cells (lanes 2–5). As controls, the cloning vector pHK-3 and frameshift mutants, pREp34 m, pREp43 m, pREp50 m and pREp84 m were transiently expressed in parallel (lanes 1 and 6–9). The areas within the open boxes were overexposed and are shown in the lower panels. To detect transient expression of UL112–113 protein products, Western blot analyses were performed using the anti-UL112 antibody. Proteins recognized by the anti-UL112 antibody are marked.
Molecular cloning of cDNAs encoding putative ORFs for p34, p43, p50 and p84

In order to characterize the cDNAs encoded by the UL112–113 ORF, and to study their biological function in detail, an HCMV cDNA library was constructed. Poly(A)⁺ RNA was purified from HEL cell cultures 72 h after infection with HCMV, which is the time at which all the protein bands were observed (Fig. 1). After screening the cDNA library, 300 recombinant phages with positive hybridization signals were isolated as single colonies. We also noticed that both strands of UL112–113 were transcribed, as the poly(A)⁺ RNA Northern blot hybridized to both sense and antisense riboprobes (data not shown). To exclude the unwanted antisense cDNA, a reference single-stranded phage with sense directionality was used for a complementation test with the selected phages. In total, 28 cDNAs contained inserts with the same sense reading frame as the proposed early proteins. These all terminated at the same nucleotide (163864 of the HCMV AD169 annotated genomic sequence) (Fig. 2a). Four cDNA constructs representing each of the four different splicing patterns and each containing the first putative ATG codon (nt 161520) were chosen for further study (pCEP34, pCEP43, pCEP50 and pCEP84; Fig. 2a).

Sequencing confirmed that plasmids pCEP34, pCEP43, pCEP50 and pCEP84 contained inserts of 2418, 1953, 1848 and 2270 nt, respectively. The profiles of the introns and exons for each deduced protein are summarized in Fig. 2(a). The molecular masses of the protein products from pCEP34, pCEP43, pCEP50 and pCEP84 were determined using a cell-free TNT-coupled reticulocyte lysate system. The protein products from pCEP34, pCEP43, pCEP50 and pCEP84 had apparent molecular masses of 34, 43, 50 and 84 kDa (Fig. 2b and Supplementary Fig. S2). However, the ORFs of pCEP34, pCEP43, pCEP50 and pCEP84 had predicted molecular masses of 28.14, 36.95, 44.05 and 70.27 kDa. The discrepancy between the predicted and the apparent molecular masses implies post-translational protein modifications such as phosphorylation or glycosylation (Wright & Spector, 1989). We did not observe any other splicing pattern in the clones.

Protein products encoded by the UL112–113 cDNA constructs

From the nucleotide sequencing data, we determined that pCEP34 with the p34 ORF encodes all the potential splicing sites for the generation of p43, p50 and p84. In addition, pCEP84, which produces p84 in a cell-free system, contains the splicing sequences for the generation of p43 and p50. Since the splicing machinery is not present in reticulocyte lysates, confirmation of the gene products of the four cDNAs expressed in a eukaryotic system was required. From constructs utilizing the SV40 early gene promoter, each cDNA was transiently expressed in HEL cells. As shown in lane 2 of Fig. 2(c), the immunoblotting assay revealed that pCEP34 produced p34 as well as p20, p26, p28, p43 and p50; however, p84 was only minimally observed. Plasmid pECEp43 produced p43 as well as p20, p26 and p28 (Fig. 2c, lane 3). Plasmid pECEp50 produced p50 as well as p20, p26 and p28 (Fig. 2c, lane 4). Plasmid pECEp84 produced p84 as well as p20, p26, p28, p43 and p50 (Fig. 2c, lane 5). In summary, p20, p26 and p28 were concurrently synthesized during transient expression of the four full-length cDNA clones in HEL cells. Proteins produced from constructs containing frameshift mutations of each ORF (pECEp34 m, pECEp43 m, pECEp50 m and pECEp84 m) failed to produce any unique products that reacted with the anti-UL112 antibody (Fig. 2c, lanes 6–9).

In order to exclude the alternative possibility that these three proteins are gene products from rare splicing patterns, clones were constructed containing only the ORFs corresponding to p34, p43 and p50, and coordinate clones with frameshift mutations in the protein sequence were created at the PstI site as controls. Constructs expressing the single ORF or the frameshift mutation were transiently expressed in HEL cells and protein products were analysed by Western blotting. As shown in Fig. 2(d), the single ORF constructs pREP34, pREP43 and pREP50 produced not only the corresponding protein products but also p20, p26 and p28 (Fig. 2d, lanes 2–5 and lower panel). However, the frameshift mutation constructs pREP34 m, pREP43 m, pREP50 m and pECEp84 m did not produce any detectable proteins (Fig. 2d, lanes 6–9 and lower panel). These results indicate that the three proteins are likely derived from the UL112 ORF, because they were recognized by anti-UL112 antibody and also because they repeatedly and consistently appeared in HCMV-infected cells and in cells transfected with UL112–113 constructs.

The novel proteins p20, p26 and p28 are associated with mature HCMV particles

According to a previous proteomic analysis of HCMV virion particles, protein derived from the UL112 reading frame is found in the virion (Varnum et al., 2004). To further characterize the proteins included within mature virions, gradient-purified HCMV particles were analysed by Western blotting with the anti-UL112 antibody. Only three proteins, migrating at the positions of p20, p26 and p28, were recognized within the virions, non-infectious envelope particles (NIEP) and dense body particles (Fig. 3a). These three proteins were present in approximately equal abundance in the virions and NIEP, and in a lower abundance in the dense bodies (DB) (Fig. 3a). As a control for protein loading, the same blot was analysed by Western blotting using antibodies to pUL55 and pUL99 (Fig. 3b, c).

To investigate the localization of these proteins in the virion, highly purified virion particles were disrupted by detergent fractionation. Secondary gradient-purified virions were incubated with the detergents NP-40 and deoxycholate and then centrifuged to separate the detergent-soluble supernatant and insoluble pellet. In principle, many proteins of the envelope can be partitioned into the
soluble fraction and most proteins associated with the tegument/capsid remain in the insoluble pellet. Equal amounts of protein recovered from the virion-soluble and insoluble fractions were resolved by SDS-PAGE and subjected to Western blot analysis. The antibody to UL112 recognized p20, p26 and p28 in both the virion and insoluble tegument/capsid fractions, but not in the soluble envelope fraction (Fig. 3d). As controls, the antibodies recognizing the envelope protein pUL55 and tegument protein pUL99 were used in the Western blot experiment to confirm the purity of the fractionated proteins (Fig. 3e, f). These results suggest that virion-associated p20, p26 and p28 are unlikely to be non-specific proteins recognized by the anti-UL112 antibody.

**Proteomic confirmation that p20, p26 and p28 are protein products of UL112**

Immunoaffinity chromatography was used to purify the proteins recognized by the anti-UL112 antibody in order to obtain material suitable for mass spectrometric analysis and protein identification. Cell extracts containing proteins encoded by UL112–113 were prepared from HCMV-infected HEL cells at 72 h p.i. Total cell extracts, non-binding proteins and proteins selectively bound to the UL112 antibody-charged column were analysed by Coomassie brilliant blue staining and Western blotting (Fig. 4). After removing non-specific proteins, we could visualize the enrichment of purified protein bands corresponding to p20, p26, p28, p34 and p43 (Fig. 4a, lane 3). These protein bands were further analysed and confirmed by Western blotting (Fig. 4b). Protein bands corresponding to p20, p26 and p28 were excised from the Coomassie brilliant blue-stained gel and subjected to in-gel trypsin digestion and mass spectrometric analysis. Electroelution was performed simultaneously to recover each individual protein band from the gel. As shown in Fig. 4(c), the eluted protein samples were resolved by PAGE and made visible by silver staining. As an additional control, Western blotting was carried out to confirm that the excised protein bands were indeed p20, p26 and p28 (Fig. 4d). Concurrently, in-gel trypsin digestion, HPLC–MS/MS analysis and subsequent bioinformatics database mapping for protein identification were carried out. The identified amino acid sequences of p20, p26 and p28 were matched with the exon1 sequence of UL112. In summary, the calculated molecular masses for the maximal protein lengths were 18.28, 18.28 and 24.38 kDa for p20, p26 and p28, respectively (Fig. 4e). In addition, the unassigned 38 kDa protein (Figs 1, 2c and 4b) was also subjected to sequence analysis (Fig. 4b), which showed that the protein was derived from UL112 exon1 reading frame (Fig. 4e). However, we could not obtain a cDNA clone for the mRNA of this protein.

**DISCUSSION**

It is tentatively understood that the HCMV UL112–113 gene region governs HCMV replication (Grey et al., 2005; Pari et al., 1993; Penfold & Mocarski, 1997). A review of the literature reveals that the UL112-113 region encodes several gene products, including four phosphorylated proteins (p34, p43, p50 and p84), two putative ORFs (C-ORF18 and C-ORF19) (Murphy et al., 2003) produced from complementary sequence, as well as two microRNAs, miR-UL111a (Grey et al., 2005) and miR-UL112-1 (located at the UL114 region) (Grey et al., 2007).

In the present report, we provide evidence that authenticates the existence of novel gene products encoded by UL112–113. Initially, a polyclonal antibody was generated that recognized UL112. Western blot analyses showed that, seven protein bands consistently appeared during the
The identities of these proteins were verified by genetic cloning and/or protein purification followed by mass spectrometric analysis. In addition to the four previously described proteins (p34, p43, p50 and p84; Staprans & Spector, 1986; Wright & Spector, 1989; Wright et al., 1988), three novel proteins were detected: p20, p26 and p28 (Fig. 1). Subsequently, p20, p26 and p28 were also found to be expressed from constructs containing full-length cDNAs for p34, p43, p50 and p84 (Fig. 2c, d). Most likely, the expression of these three novel proteins was from derivatives of exon1 and occurred in the absence of other viral proteins (Fig. 2c, d).

For the first time, we have demonstrated that these three novel proteins, and not the previously identified UL112–113 proteins, are constituents of mature HCMV particles (Fig. 3a). Our findings not only agree with a previous report that a peptide of UL112 is present within virions (Varnum et al., 2004) but also validate the biological authenticity of these novel proteins. Specifically, these three proteins were present in the tegument/capsid fraction of virions (Fig. 3d). It is not surprising that only selective forms of viral proteins are assembled within the virion. For instance, only a specific phosphorylated form of UL69 is found in the virion (Winkler & Stamminger, 1996).

Subsequent investigations were undertaken to confirm the identity of these three novel proteins. p20, p26 and p28 were purified from HCMV-infected cells by immunoaffinity chromatography (Fig. 4a–d). In preliminary experiments, these proteins were not recognized in Western blot analyses by an antibody that specifically recognized the C terminus of p34, p43, p50 and p84 (S.-K. Wang, unpublished results). Thus, these proteins were neither
unmodified forms of p34 (calculated molecular mass 28141.86 Da) nor containing the unique C-terminal sequences of p34, p43, p50 or p84. Mass spectrometry analysis of their amino acid sequences demonstrated that p20, p26 and p28 originated from an ORF of UL112 exon1 (Fig. 4e). The coverage sequence of p28 reached a calculated molecular mass of 24 kDa. We also learned throughout the study that UL112–113-derived proteins were unstable, as several protein bands reacted with the UL112 antibody inconsistently. Taking these results together, we speculated that proteolysis of the mature UL112–113 protein was occurring.

Proteolytic modification of viral proteins plays pivotal roles in many steps of viral infection, including entry, replication, assembly and egress, as well as the induction of immune responses. In search of the mechanism responsible for this proteolysis, we performed computational analyses of the protein structure. A number of short half-life proteins exhibit intrinsically disordered/unstructured sequences that may act as signals to promote proteolytic cleavage (Dyson & Wright, 2005; Piwko & Jentsch, 2006). Consistent with this idea, we identified several disordered/unstructured regions at the C-termini of UL112–113-encoded proteins by using the computational program DisEMBL (http://dis.embl.de/) (Linding et al., 2003). In addition, we found PEST motifs embedded in UL112–113 protein sequences by using an ePESTfind program maintained by the EXPASy Proteomics Server (http://emboss.bioinformatics.nl/cgi-bin/emboss/epestdn). PEST sequences, which are regions rich in proline (P), glutamic acid (E), serine (S) and threonine (T), serve as a proteolytic signal conferring instability to many PEST-containing proteins (Belizario et al., 2008; Garay-Malpartida et al., 2005; Rechsteiner, 1990; Rechsteiner & Rogers, 1996; Rogers et al., 1986; Tompa et al., 2004). Most importantly, these analyses showed the PEST motifs within p34, p43, p50 and p84 to be highly significant. Recent studies have revealed that degradation of PEST-containing proteins is mediated by three molecular pathways: ubiquitin-proteasome system, the calpain proteases and the caspases (Belizario et al., 2008; Rechsteiner & Rogers, 1996). Therefore, these pathways may mediate the proteolysis of UL112–113-encoded proteins.

In order to gain insight into the functions of p34, p43, p50 and p84 during the HCMV infection cycle without the interference of degradation products that could complicate the interpretation of the functional analysis, it will be necessary to identify the specific proteases that mediate these cleavages. This information is required in order to eliminate signals that mediate degradation. To this end, site-directed mutagenesis experiments are currently underway in our laboratory.

ACKNOWLEDGEMENTS

This study was supported by research grants (NSC87-2314-B-037-067, NSC94-2311-B-037-004 and NSC95-2320-B-037-034) awarded to S.-K. W. from the National Science Council, Taiwan. The authors thank Dr John Sinclair of Cambridge University for providing the pHK-3 cloning vector and National Cheng-Kung University Proteomics Research Core Laboratory for assistance with mass spectrometry analysis.

REFERENCES


S.-K. Wang and others


