CHARACTERIZATION OF UL1, A MEMBER OF THE HUMAN CYTOMEGALOVIRUS RL11 GENE FAMILY

Medya Shikhagaie

Doctoral Thesis UPF-2011

DIRECTORS

Miguel López-Botet Arbona and Mar Albà Soler

Departament de Ciencies Experimentals i de la Salut

UNIVERSITAT POMPEU FABRA

To my family

THESIS ABSTRACT

We have investigated the previously uncharacterized human cytomegalovirus (HCMV) *UL1* open reading frame (ORF), a member of the lately evolving HCMV *RL11* gene family. *UL1* is HCMV specific; absence of *UL1* in chimpanzee CMV (CCMV) and sequence analysis studies suggests that *UL1* may have originated by duplication of an ancestor gene from the *RL11-TRL*-cluster (*TRL11, TRL12* and *TRL13*). Sequence similarity searches against human immunoglobulin (Ig) containing proteins revealed that HCMV UL1 protein shows significant similarity to the cellular carcinoembrionic antigen related (CEA) protein family N-terminal Ig domain, which is responsible for CEA ligand recognition. Northern blot analysis revealed that *UL1* is transcribed during the late phase of the HCMV replication cycle, both in fibroblast-adapted and endotheliotropic strains of HCMV. We characterized the protein encoded by an HA-tagged *UL1* in the AD169-derived HB5 background. *UL1* is expressed as a 224 amino acid type I transmembrane glycoprotein which becomes detectable not earlier than 48 h to 72 h post infection. In infected human fibroblasts, pUL1 co-localized at the cytoplasmic site of virion assembly and secondary envelopment together with TGN-46, a marker for the *trans*-Golgi network, and viral structural proteins including envelope glycoprotein B and tegument phosphoprotein pp28. Furthermore, analysis of highly purified AD169 UL1-HA epitope tagged virions revealed that UL1 glycoprotein is a novel constituent of the HCMV envelope. Importantly, deletion of *UL1* in HCMV TB40/E resulted in a reduced growth in a cell type-specific manner, suggesting that pUL1 may be implicated in regulating HCMV cell tropism.

RESUM DE LA TESI

En aquest treball hem investigat la pauta oberta de lectura de UL1 del Cytomegalovirus humà (HCMV), membre de la família gènica apareguda rcentment en l'evolució RL11 del HCMV. El gen UL1 es específic del HCMV; la seva absència en el CMV de ximpanzé (CCMV) i els anàlisis de seqüència suggereixen que UL1 es pot haver

originat per la duplicació d'un gen ancestral del grup RL11-TRL (TRL11, TRL12 i TRL13). La comparació per similitud de seqüència amb proteïnes humanes que contenen dominis immunoglobulina (Ig) va revelar que la proteïna UL1 del HCMV presenta una semblança significativa al domini Ig N-terminal de la família de proteïnes cel·lulars relacionades a l'antigen carcinoembriònic (CEA), que és la part de la proteïna responsable del reconeixement del lligand de CEA. Els anàlisis de Northern Blot ens mostren que UL1 es transcriu durant la fase tardana del cicle de replicació del HCMV, tan en soques de HCMV adaptades a fibroblasts com en soques amb tropisme per endoteli. Hem caracteritzat la proteïna UL1 modificada amb un epítop HA en la soca HB5, derivada de AD169. L'UL1 s'expressa com una glicoproteïna de tipus I de 224 aminoàcids que es pot detectar a les 48 i 72h post-infecció. En fibroblasts humans infectats, UL1 co-localitza al citoplasma, al lloc d'assemblatge del virió, amb TGN-46, marcador de la xarxa *trans*-Golgi, i amb proteïnes estructurals del virus, incloent la glicoproteïna B de l'envolta i la fosfoproteïna pp28 del tegument. A més a més, els anàlisis de virions AD169 purificats que contenen UL1-HA mostren que UL1 és un nou constituent de l'envolta del HCMV. És important remarcar que la delecció de UL1 en el context de la soca TB40/E del HCMV disminueix el creixement viral de manera selectiva en determinats tipus cel·lulars, suggerint que UL1 podria estar involucrat en la regulació del tropisme cel·lular del HCMV.

PREFACE

Human cytomegalovirus (HCMV) has the largest genome of any characterized human viruses and with this large protein coding capacity the virus has evolved a variety of mechanisms to interact with host cells and to interfere with the host immune response. The identification and characterization of these genes and their putative cellular homologues should improve our knowledge of the virus-host interaction and pathogenesis. Here we investigated the similarity of the HCMV-specific glycoprotein UL1 to the N-terminal IgV domain of members of the CEA family that are involved in homotypic and heterotypic adhesion mechanisms.

ACKNOWLEDGEMENTS

It is a true pleasure to thank all the people who made the completion of this thesis possible, everyone who contributed with their scientific, technical or moral support.

First and foremost, it is with immense gratitude that I acknowledge the expertise, support and guidance of my thesis directors Miguel López-Botet Arbona and Mar Albá Soler.

A special thank you to Ana Angulo for her positive attitude, for her generous and continuous help.

Special thanks also deserve Harmut Hengel for giving me the opportunity to visit his lab in Dusseldorf, a stay that have been crucial to making this project work.

I would like to thank my colleagues at the Immunology unit and the Evolutionary Genomics group: Gemma Heredia, Aura Muntasell, Giuliana Magri, Diogo Baia, Neus Romo, Eugenia Corrales-Aguilar, Andrea Vera, Jose Martinez, Noemi Marina, Andrea Saez, Rosa Berga, Giulia Lunazzi, Mari Carmen Ortells, Maria Buxade and Maria Alberdi for being a great group of people to work with and making everyday life in the lab enjoyable during four long years.

I would like to thank Pablo Engel, Cristina Lopez Rodriguez, Jose Aramburu Beltran, Eva Merce Maldonado, Vu Thuy Khanh Le, Carola Prato and Julia V. Blume for scientific advice. Many thank you to German Anders and Xavier Sanjuan for technical support.

Thanks to friends in Barcelona, for unforgettable memories and invaluable emotional support: Sarah Djebali, Joshua Weitzer, Deborah Sandoval, Francisco Camara, Hagen Tilgner, Gunesh Gundem, Regina Lopez, Michael Breen, Blanca Gumez, Julien Lagard, Julien Quelen, Carol Dunlop, Emre Guney, Natalia Gabrielli, Aisha Prigann, Andres Bartos, Miriam Giordano and Elena Suarez Robles.

I would like to thank my family: My parents Tahmine and Siamand.
My siblings Mandana, Roxana and Diako.
My thanks to Marin, Dunja, Milan, Zejlko, Jassa and Sakine.
They have unconditionally supported me, and loved me. *Zor spas bo awayka hamisha baweratan be mn bo, hamotanm zor xosh dawe.*
I am greatful to Jens for his patience, inspiration and love.

ABBREVIATIONS

aa: amino acid
BSA: bovine serum albumin
°C: degrees Celsius
DMEM: Dulbecco's modified eagle medium
DBs: dense bodies
E: early
EGFR: epidermal growth factor receptor
ELISA: enzyme-linked immunosorbent assay
FBS: fetal bovine serum
GAPDH: glyceraldehyde dehydrogenase
gp: glycoprotein
h: hour(s)
h p.i.: hours post infection
HCMV: human cytomegalovirus
HLA: human leucocyte antigens
IE: immediate early
IFN: interferon
IL: Interleukin
Ig: immunoglobulin
IP: immunoprecipitation
IRL: internal repeats long
IRS: internal repeats short
Kbp: kilobase pairs
kDA: Kilo Dalton
L: late
Min: minute(s)
mAb: Monoclonal antibody
MCMV: Murine Cytomegalovirus
MHC: Major Histocompatibility Complex
MOI: multiplicity of infection (pfu/cell)
NIEPs: non-infectious enveloped particles
NK: Natural killer
ORF: open reading frame(s)
PBS: phosphate buffered saline
PCR: polymerase chain reaction
Pfu: plaque forming units
pp: phosphoprotein
RT-qPCR: real-time quantitative polymerase chain reaction
RT: room temperature
SDS-PAGE: sodium dodecyl sulfate-polyacrylamide gel electrophoresis
TNF: Tumor necrosis factor
TRL: Terminal repeats long
TRS: Terminal repeats short
UL: unique long
US: unique short

TABLE OF CONTENTS

ABSTRACT……………………………………………………………………iii
PREFACE………………………………………………………………….....v
ACKNOWLEDGMENTS…………………………………………………...vi
ABBREVIATIONS……………………………………………………… ix

PART I INTRODUCTION AND AIMS

1. INTRODUCTION..1
 1.1 Transmission and Pathogenesis of HCMV.. 2
 1.2 Molecular Virology of HCMV 3
  1.2.1 Genome Organization and Genetic Content....................................... 3
  1.2.2 HCMV Strains and BAC Mutagenesis 4
 1.3 HCMV Virion.. 7
  1.3.1 The Capsid 7
  1.3.2 The Tegument.. 8
  1.3.3 The Envelope... 8
  1.3.4 Other components related to the virion 9
 1.4 HCMV replication cycle.. 10
  1.4.1 Entry into the host cell 12
  1.4.2 Maturation and Egress.. 13
 1.5 Immune responses to HCMV... 14
  1.5.1 Innate and Adaptive Immunity 14
  1.5.2 Immunoregulatory mechanisms... 16
 1.6 Primate CMV RL11 gene family... 19

2. AIMS ……………………………………………………………….25

PART II MATERIAL AND METHODS

3. MATERIAL AND METHODS...29
 3.1 Sequence analysis and evolutionary studies.. 29
 3.2 Cell lines 30
 3.3 Antibodies 30
 3.4 Preparation of eukaryotic expression constructs and DNA transfection 31
 3.5 Expression and production of UL1-Fc fusion proteins............................ 33
 3.6 Viruses 34
 3.7 Construction and generation of recombinant viruses 35

3.8 UL1 RNA analysis 40
3.9 Western blot analyses and N-glycosidase F treatment of HCMV infected cells 41
3.10 Purification of HCMV particles and analysis of the virion protease-sensitivity 41
3.11 Immunoprecipitation of UL1 protein... 42
3.12 Immunofluorescence microscopy 42
3.13 Viral growth kinetics.. 43
3.14 Measurement of IE1 mRNA levels... 43

PART III RESULTS

4. THE UL1 PROTEIN IS A HCMV-SPECIFIC MEMBER OF THE RL11 FAMILY 47
4.1 Rapidly evolving and highly variable RL11 family 47
4.2 Members of RL11 family show sequence similarity to several cellular proteins 50
4.3 UL1 shows significant similarity to CEA proteins.................................... 51
4.4 The UL1 protein is a HCMV-specific member of the RL11 family 54
4.5 Sequence similarity between UL1 and other HCMV RL11 family members 57

5. RECOMBINANT EXPRESSION OF HCMV UL1 59
5.1 UL1 sequence fused to HA and Myc tags.. 60
5.2 Expression and production of UL1-Fc fusion proteins............................ 62

6. UL1 PROTEIN HCMV INFECTED CELLS AND VIRIONS...........................65
6.1 UL1 is transcribed with a late gene kinetics 65
6.2 Expression of the UL1 protein in HCMV infected cells 68
6.3 Subcellular localization of pUL1 in HCMV infected cells 70
6.4 Co-localization of pUL1 with tegument and envelope HCMV proteins 74
6.5 pUL1 is a component of the HCMV virion envelope 75
6.7 Analysis of UL1 mutant viruses... 77
6.7.1 Deletion of *UL1* results in HCMV growth defects in epithelial cells 77
6.7.2 Entry of TB40/E *UL1* deletion viruses 78

PART IV DISCUSSION

7. DISCUSSION......83

PART V CONCLUSIONS

8. CONCLUSIONS......93

ANNEX 1

Article submitted for publication: Shikhagaie *et al.,* 2011. *The HCMV-specific UL1 gene encodes a late phase glycoprotein incorporated in the virion envelope.*

REFERENCES......101

PART I INTRODUCTION & AIMS

1. INTRODUCTION

Members of the *Herpesviridae* family are classified on the basis of biological criteria such as differences in genome content and organization, host tropism, site of latency and growth kinetics. They are grouped into the following subfamilies *α-herpesviruses*, *β-herpesviruses* and *γ-herpesviruses.* These subfamilies comprise a diverse group of viruses and are estimated to have arisen 180-220 million years ago (McGoech *et al.*, 1995), before the major mammal radiation. The speciations within sublineages took place in the last 80 million years, probably with a major component of co-speciation with host lineages. Herpesviruses have characteristic virion morphology with an icosahedral capsid structure composed of 162 capsomeres reaching a diameter of ~100-110 nm. The capsid contains a linear double-stranded DNA genome, ranging in size from about 125 to 240 Kbp and in nucleotide composition ranging from 32 to 75% G+C content (Reviewed in Mocarski *et al.,* 2007).

The *α-herpesviruses* have the smallest genomes, *γ-herpersviruses* have an intermediate size and the *β-herpesviruses* are the largest. In spite of the high levels of sequence variability, gene content is relatively well conserved across related herpesviruses, and all family members share a set of core genes involved in basic metabolic and structural functions (Albà *et al.*, 2001). Genes that are specific of a given virus or group of viruses are typically involved in immune evasion and show often significant similarity to genes in the host genome (Holzerlandt *et al.*, 2002). Herpesviruses share the following properties of biological relevance:

1. They establish latency within their host and are able to reactivate and cause disease after latency.
2. Viral DNA replication and capsid assembly takes place in the host nucleus.
3. They encode for enzymes that are important in the metabolism of nucleic acids.

Herpesviruses infect a wide range of animals and are highly adapted to their host. To date there are 130 herpesviruses identified, of which several are clinically relevant human viruses such as Human cytomegalovirus (HCMV, HHV-5), Herpes simplex

virus-1 and -2 (HSV-1 and -2), Epstein-Barr virus (EBV), Kaposi's sarcoma-associated herpesvirus (KSHV) and Varicella zoster virus (VZV). HCMV, chimpanzee cytomegalovirus (CCMV) and rhesus cytomegalovirus (RhCMV) belong to the *Cytomegalovirus* genus of the *β-herpesviruses*. Cytomegaloviruses are highly species specific adapted pathogens that establish a lifelong relationship with its host. The highly host specific HCMV has the largest genome of any characterized human viruses (Chee *et al.,* 1990). HCMV infection was initially documented as early as 1881 by Ribbert as cells containing abnormal ´protozoan-like´ cells. In 1921, Goodpasture and Talbert suggested that 'cytomegalia' (cell enlargement) may be caused by a virus and not by a *protozoa*. The name 'cytomegalovirus' was proposed by Kluge *et al.* in 1960 after isolating the virus from the urine of congenitally infected infants (Reviewed in Mocarski *et al.,* 2007).

1.1 Transmission and Pathogenesis of HCMV

Seroprevalence of HCMV ranges between 40-100% depending on the socioeconomic status, geographical location and age of the population. HCMV infection is often acquired during childhood by transmission through breast-feeding or contact with secretions. Upon infection, the virus persists in a life-long latent state, which is interrupted by periodical phases of reactivation, resulting in virus shedding and transmission to new hosts. Increased viral DNA and specific T cells have been detected in peripheral blood during these reactivation periods (Sissons *et al.*, 2002). However, large numbers of specific T cells can also be found in peripheral blood during latency periods, reflecting the impact of the virus on the T cell compartment (Reddehase *et al., 2002*).

Transplacental transmission during pregnancy results in congenital HCMV infection that may cause severe neurological disorders (Fowler *et al.,* 2006). Infection in immunocompetent children and adolescents is normally asymptomatic. An effective immune response is responsible of clearing the virus during primary infection and controlling the reactivations of the virus from a latent state. In contrast, immunocompromised individuals such as, transplant patients under immunosuppression and patients undergoing immunodeficiencies (i.e. AIDS) are

unable to control viral replication and the virus spreads to different organs and tissues as a result of reactivation, primary infection or re-infection. The severity of the disease often correlates with the degree of immunosuppression (Griffiths *et al.,* 2006).

The most severe manifestation of HCMV disease in HIV-infected patients is retinitis (Crough and Khanna, 2009). Following solid-organ allograft transplantation, HCMV infection is considered to contribute to graft rejection. Common manifestations of HCMV disease in these patients are pneumonia, hepatitis, enteritis and encephalitis. In allogeneic bone marrow transplantation HCMV may cause interstitial pneumonia with a high mortality rate. Moreover, HCMV infection is believed to be one of the factors contributing to atherosclerosis and restenosis after coronary angioplasty (Zhou *et al.,* 1996). Currently approved antiviral therapies against active HCMV infection have limited efficiency. Thus far, the results of vaccine development using attenuated strains (i.e. Towne) or a subunit vaccine using HCMV glycoprotein gB have been unsatisfactory.

1.2 Molecular Virology of HCMV

1.2.1 Genome Organization and Genetic Content

The linear 240-Kbp double-stranded DNA genome of HCMV has a class E structure, and is organized in long (L) and short (S) unique sequences (U_L and U_S) flanked by inverted repeats, referred to as TR_L/IR_L and the IR_S/TR_S (terminal repeat long and internal repeat long) This results in a overall genome configuration TR_L–U_L–IR_L–IR_S–U_S–TR_S (Figure 1). The L and S genome segments can invert during replication resulting in four different sequence arrangements or isoforms of the viral genome (Chee *et al.*, 1990; reviewed in Mocarski and Courcelle, 2001), which is unique among the so far characterized β-herpesviruses.

HCMV genome (reviewed in Murphy and Shenk, 2008) is predicted to encode between 252 open reading frames (ORFs) (Murphy *et al.,* 2003) and 165 ORFs (Davison *et al.,* 2003b), including at least 14 viral microRNAs and 150 viral proteins. Despite the fact that the first complete genome sequence of an HCMV strain has now

been accessible for 20 years (Chee *et al.,* 1990), the majority (~ two-thirds) of the predicted viral products are of unknown function.

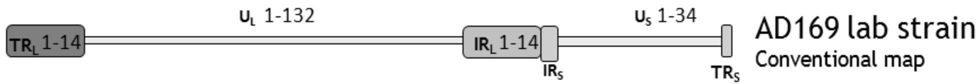


Figure 1. Conventional map of HCMV double stranded DNA genome in AD169 laboratory strain. Two unique regions, unique long (U_L) and unique short (U_S), are divided by inverted repeat regions: internal repeat long and short (IR_L and IR_S) and terminal repeat long and short (TR_L and TR_S). This figure was adapted from Crough and Khanna, 2009.

Most of the microarray and mutagenesis studies in HCMV employ the AD169 and Towne laboratory strains, these strains were originally developed as vaccine candidates by extensive cell culture passage of clinical isolates. HCMV laboratory strains are known to differ both genetically and biologically from clinical isolates of HCMV (Cha *et al.,* 1996; Murphy *et al.,* 2003). Mutagenesis studies of HCMV AD169 laboratory strain have demonstrated that about 40 genes, which are located in the central region of the viral genome within the U_L domain, are essential for viral replication and assembly into virions *in vitro* (e.g structural and regulatory proteins). A large number of genes (~88 genes) encode for dispensable proteins, presumably involved in the pathogenesis of the virus infection (Yu *et al.,* 2003; Shenk *et al.,* 2003). Many of these are involved in immune evasion and are located within the terminal regions, including a large portion of the Us segment. The transcription unit *UL122-UL123* encodes for the key immediate-early transcriptional regulatory proteins IE1 and IE2. Proteins important for entry into epithelial and endothelial cells are encoded by *UL128-UL131* (Hahn *et al.,* 2004) and genes blocking apoptosis are located in *UL36-UL38* region (Dolan *et al.,* 2004). Sequence analysis of the HCMV genome shows that as many as 70 putative glycoprotein-encoding ORFs are found in clinical isolates (Cha *et al*, 1996) and 57 in the AD169 strain (Chee *et al.*, 1990).

1.2.2 HCMV Strains and BAC Mutagenesis

The DNA sequence heterogeneity between HCMV strains is estimated to a sequence identity between 90-95% (Huang *et al.,* 1976; Pritchett, 1980). Some sequenced HCMV

strains are available for *in vitro* experimentation, including laboratory strains AD169 and Towne, which have been extensively passaged in fibroblasts, and some clinical isolates, which have been passaged to a limited extent in the laboratory (e.g. Toledo, FIX, PH, and TR). Many of the sequenced HCMV genomes have been cloned as infectious bacterial artificial chromosomes (BAC) (Brost *et al.,* 1999) such as the complete and genetically intact genome of Merlin strain (Stanton *et al*, 2010), and the endotheliotropic strain TB40/E (Sinzger *et al.*, 1999) reconstituted from a BAC clone TB40/E-BAC4 (Sinzger *et al.*, 2008b). BAC mutagenesis techniques (Figure 2) have provided invaluable resources in the study of CMVs and other infections viruses by allowing mutations and insertions into viral genomes.

Extensive cell culture passage of the strains AD169 and Towne resulted in a large deletion of 15-Kbp in the unique region UL/b´ at the right end of UL region and numerous other subtle mutations. The UL/b´ region encodes for at least 19 ORFs, which include *UL146,* encoding for viral CXCL chemokine, *UL144,* encoding for tumour necrosis factor receptor), several genes with natural killer (NK) cell evasion functions (e.g. *UL141* and *UL142*) and a regulator of latency encoded by *UL138*. These highly-passaged HCMV strains are also mutated in the *UL128-UL131* region, that encodes for proteins forming complexes with glycoproteins gH and gL, mediating endothelial and epithelial cell tropism (Ryckman *et al.,* 2006; Ryckman *et al.,* 2008).

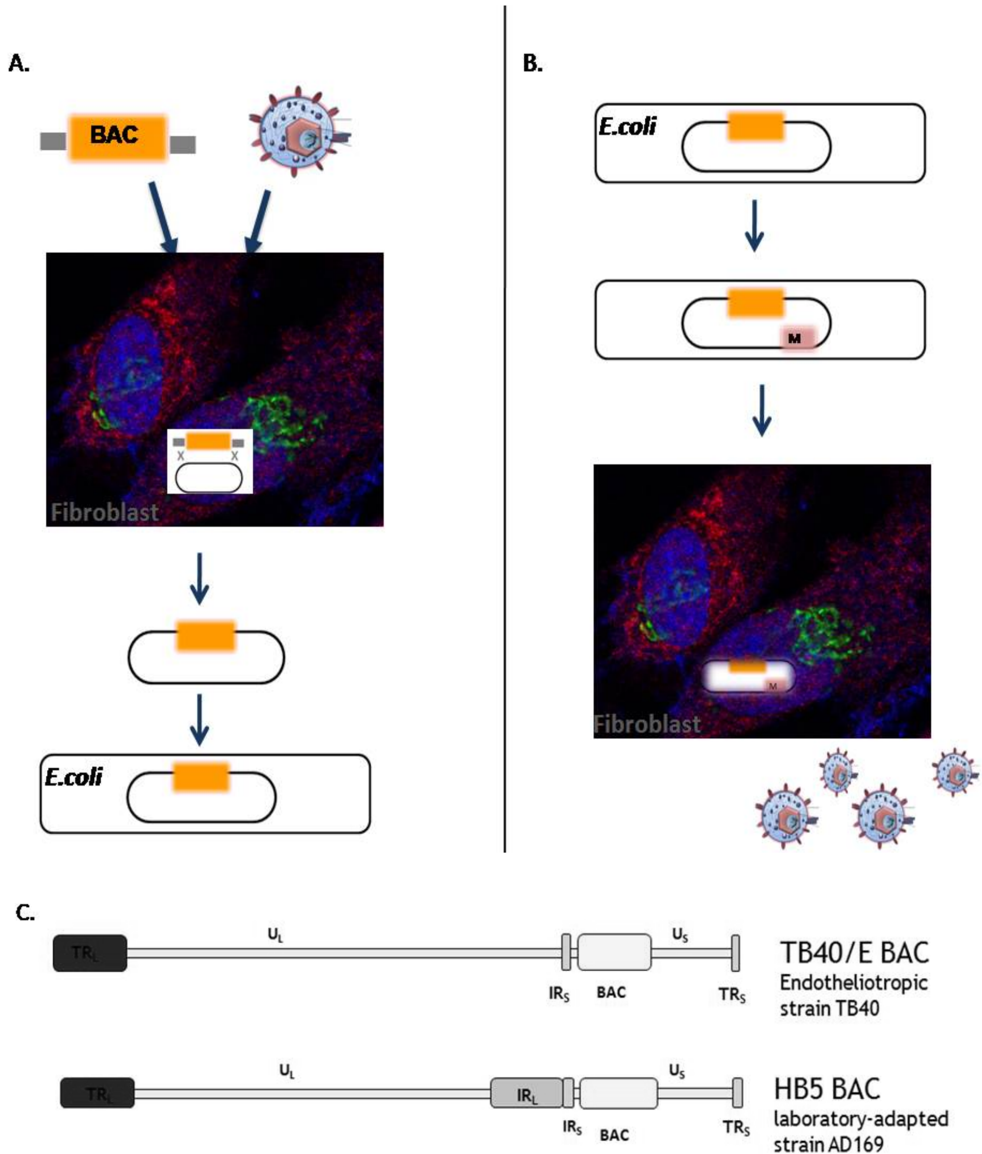


Figure 2. (A-B) Scheme of the strategy for cloning of HCMV genome as a BAC adapted from Brune *et al.*, 2006 (Book: Cytomegaloviruses Molecular Biology and Immunology, Reddehase (ed.)). Fibrobasts are transfected by the BAC vector (yellow) flanked by viral sequences and infected with HCMV. By homologous recombination viral genomes are incorporated into BAC, these BAC viruses are isolated and transferred to *E.coli*. (B)Mutant HCMVs maintained as BAC in *E.coli* can subsequently be used to introduce mutation and insertions into the HCMV genome. The mutant virus BACs are reconstituted by transfection into fibroblasts. (C) Conventional map of endotheliotropic HCMV-TB40/E-derived BAC4 (TB40/E BAC4) and the laboratory adapted AD169 BAC (pHB5) in *E. coli*.

1.3 HCMV Virion

The HCMV virion has been structurally divided into three major regions: the capsid, the tegument, and the envelope (Figure 3). The first internal layer is the nucleocapsid which consists of proteins enclosing the double-stranded viral DNA genome. The capsid is surrounded by a less structured tegument protein layer, which is surrounded by an envelope that contains a cellular lipid bi-layer with viral and host proteins.

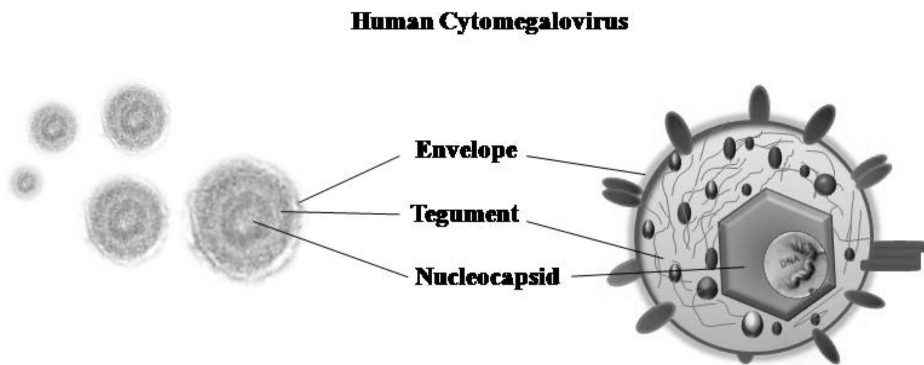


Figure 3: HCMV virion: the capsid containing the double-stranded DNA is surrounded by an asymmetric and amorphous like structure, called the tegument. The tegument is enveloped by a lipid bi-layer with both viral and cell derived proteins. Electron microscopy picture of HCMV infectious virus particle on the left hand side and schematic representation of the virion on the right hand side.

1.3.1 The Capsid

The HCMV virion with ~230 nm in diameter is composed of a 100 nm icosahedral (T=16) capsid contains a linear dsDNA. The capsid appears in three different forms, the A-, B- and C-capsid. These forms represent three different stages of virion maturation. The A capsid structure lacks viral DNA, the B capsid is mainly found in the host nucleus, and also lacks viral DNA but contains viral assembly and scaffolding proteins. The fully mature nucleocapsid is formed into a C type capsid. The capsid is composed of several protein, the major capsid protein pUL86, comprises the main component (Chee *et al.,* 1989). The capsid pentamers and hexamers are built up by the minor capsid proteins encoded by *UL85* (Baldick and Shenk, *1996)* and *UL46 (*Gibson

et al., 1996). Further, capsid assembly and packaging are carried out by proteins encoded by *UL80*, *UL80a* and *UL80.5* (Robson and Gibson, *1989).*

1.3.2 The Tegument

The tegument region, which is unique to herpesviruses, contains ~30 virus-encoded proteins (Baldick and Shenk, 1996; Gibson, 1996). The tegument is described as an amorphous and asymmetric structure, with associated proteins which are phosphorylated and many have unknown functions. Tegument proteins are delivered immediately upon infection with the nucleocapsid onto the infected cell and have the potential to function even before the viral genome is activated. Tegument proteins usually regulate viral gene expression or interfere with host cell responses against viral infection. The most abundant phosphoproteins are pp150 (*UL32*) and pp65 (*UL83*). pp71 (*UL82*) is a membrane-associated myristylated protein, reported to be a critical regulator of the major immediate–early promoter (MIEP), which initiates a productive infection and transactivates IE1 gene (Bresnahan and Shenk, 2000; Reviewed in Kalejta, 2008). Whereas some tegument proteins accumulate in the nuclei of infected cells (e.g. pp71), pp28 (*UL99*) locates in the cytoplasm, partially co-localizing at the site of virion assembly with other tegument proteins (pp150 and pp65) and envelope glycoproteins (glycoprotein gH and gB) (Sanchez *et al.,* 2000; Silva *et al.,* 2003).

1.3.3 The Envelope

The envelope contains a host derived lipid bi-layer with both viral glycoproteins and host-derived proteins including annexin II (Wright *et al.,* 1995) and CD13. HCMV´s glycoprotein composition is the most complex among herpesviruses and is incompletely defined (Baldick and Shenk, 1996; Varnum *et al.,* 2004). A growing number of viral glycoproteins has been reported as component of the HCMV envelope, including molecules conserved between members of the *Herpesviridae* family (gB, gH, gL, gM, gN) and others which are HCMV specific (e.g. gO [Huber and Compton, 1998], gpTRL10 [Spaderna *et al.*, 2002], gpRL13 [Stanton *et al.*, 2010],, UL4/gp48 [Chang *et al.*, 1989], gpUL128 [Wang and Shenk, 2005], gpUL130 [Patrone *et al.*, 1996], gpUL132 [Spaderna *et al.*, 2005], *US27* [Fraile-Ramos *et al.,* 2002], and

UL33 [Margulies *et al.*, 1989], [Varnum *et al.,* 2004]). It is believed that genera specific envelope glycoproteins, whose expression is restricted to β-herpesviruses, contribute to the broad cell/tissue tropism of HCMV and its pathogenicity (Dunn *et al.*, 2003).

gB is the most abundant glycoprotein of the envelope and it is highly conserved among members of the family *Herpesviridae*. Many key functions in the HCMV biology have been ascribed to it, including mediating entry into the host cell, targeting of progeny virus during egress (Compton *et al.,* 1993), and intercellular spread (Bold *et al.,* 1996). Envelope glycoproteins of herpesviruses have a number of different functions during the replication cycle. They mediate attachment and fusion with the target cell, and have roles in cell-to-cell spread, envelopment and virus egress. Moreover, some envelope glycoproteins have a role in modulation of host immune responses.

Many of the HCMV envelope glycoproteins form high molecular weight disulfide-linked oligomers, including gCI complex composed of homodimers of gB (Britt *et al.*, 1984); gCII complex is a heterodimeric complex of gM, gN (Mach *et al.*, 2000); gCIII complex is a heterotrimeric complex composed of gH, gL and gO, a glycoprotein unique for HCMV (Huber and Compton, 1998), which appears to enhance HCMV entry or virus spread between fibroblasts (Huber and Compton, 1998). While the gH/gL complex association with *UL128* locus encoded proteins, forming the pentameric gH/gL/UL128/UL130/UL131 envelope complex, mediate the entry into epithelial and endothelial cells and involves endocytosis and low pH (Ryckman *et al.,* 2007).

1.3.4 Other components related to the virion

Several host cell proteins and viral transcripts have been associated with the HCMV virion, including actin, actin-related protein 2/3 (Arp2/3), heat-shock protein 70, CD13 (aminopeptidase N) CD46, CD59, β2-microglobulin, protein phosphatase I, annexin II (Baldick and Shenk, 1996; Varnum *et al.,* 2004). In addition to infectious virions, HCMV infection of cells in culture generates non-infectious enveloped particles (NIEPs) and dense bodies (DBs). NIEPs are composed of the same viral proteins as infectious virions but lack viral DNA. DBs mainly composed of pp65, are

non-replicating enveloped particles and are uniquely characteristic of HCMV infection (Varnum *et al.,* 2004).

1.4 HCMV replication cycle

In vivo, HCMV replicates within a broad cell range including fibroblasts, endothelial cells, epithelial cells, monocytes, neutrophils, neuronal cells, hematopoietic cells and smooth muscle cells (Reviewed in Sinzger *et al.,* 2008a). The ability to infect such a wide range of cell types is part of the reason why HCMV infection gives rise to many diseases. Infection of epithelial cells represents the route of host-to-host transmission. Infection of endothelium in the blood vessels serves as a route for virus spread to monocytes and lymphocytes, which facilitate the distribution of the virus throughout the body. Infection of endothelium monocytes and lymphocytes, facilitate the distribution of the virus throughout the body. Infection of ubiquitously distributed cell types such as fibroblasts and smooth muscle cells allows the virus to replicate in virtually every organ. The tropism for these cell types differs among different HCMV strains (Reviewed in Sinzger *et al.,* 2008a). *In vitro,* a wide range of primary cell cultures such as skin and lung fibroblasts, retina and vascular epithelial cells, placental trophoblast cells, hepatocytes, neuronal and glial brain cells, monocyte-derived macrophages and dendritic cells support a productive HCMV replication (Reviewed in Sinzger *et al.,* 2008a).

Cytomegaloviruses do not arrest the host cell protein synthesis during their replication cycle as other viruses; instead they promote a favorable environment for its gene expression and replication by modulating metabolic pathways of the host cell and regulating the transcription profile of many cellular genes (Reviewed in Mocarski and Courcelle, 2001). The HCMV replication cycle is initiated by virus binding and entry into the host cell, followed by release of the nucleocapsid into the cytoplasm. The nucleocapsid is uncoated at the nuclear pore and the dsDNA is introduced into the nucleus (Figure 4). HCMV replication cycle takes ~ 72 h and is conventionally categorized into three phases of viral gene expression: immediate-early (IE), early (E), or late (L). The cycle starts with the IE genes transcription within the first few hours

post infection of permissive cells. The IE expression is not dependent of the expression of any other viral genes and IE proteins are critical for the subsequent expression of E and L genes, the activation of the lytic infection cycle and changes in the metabolic pathways of the host cell. IE1 and IE2 proteins return to the nucleus and transactivate the early gene promoters, transcribing viral proteins that are involved in replication process (< 24h). E genes products are insensitive to inhibitors of viral DNA synthesis (e.g. phosphonoacetate and ganciclovir) and are necessary for viral DNA replication, cleavage and packaging of the viral genome, assembly of the virus particles, createing a favorable environment for viral gene expression and replication. Viral DNA synthesis begins 18 to 24 h post infection and is followed by the transcription of L genes that typically encode for structural components of the virion, such as gH and pp28.

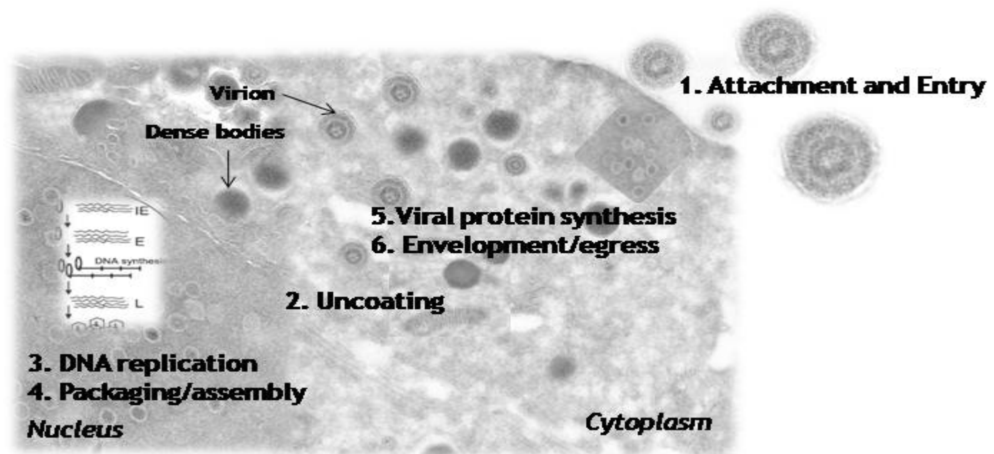


Figure 4. Overview of HCMV life cycle in a human cell. 1) HCMV enters through direct fusion or through the endocytic pathway. Via interactions between cellular receptors and viral envelope glycoproteins, the virus attaches and fuses its envelope to the cell and consequently releases the nucleocapsid and the tegument content into the cytoplasm. 2) The nucleocapsids are translocated into the nucleus, uncoated at the nuclear pore and the dsDNA is released into the nucleus. 3-5) The virus replicates in parallel with the primary envelopment of the viral DNA containing capsids at the nuclear envelope. 6) The final envelopment and tegumentation of the nucelocapsid takes place in a cytoplasmic site, where the nucleocapsid fuses with the membranes containing envelope glycoproteins and cellular proteins. Finally the mature virion (~ 96 h p.i.) egresses via the cellular endocytic pathway into the extracellular medium. This figure is made from images obtained from immunoelectron microscopy of HCMV infected fibroblasts, adapted from a model suggested by Crough and Khanna (2009).

1.4.1 Entry into the host cell

The cell-type dependent HCMV entry is presumably mediated by the interaction of multiple envelope glycoproteins with several cell surface receptors. HCMV enters through direct fusion of the viral envelope and cell plasma membrane at neutral pH in fibroblasts (Revieved in Compton*,* 2004) or through receptor-mediated internalization into endocytic vesicles at low-pH in endothelial and epithelial cells (Ryckman *et al.,* 2006). The entry initiates with the attachment of HCMV to host cells, mediated by an initial interaction between viral gB and/or gM and cell surface heparan sulfate proteoglycans (HSPGs), which are required but not sufficient to mediate HCMV entry. After the gB mediated attachment and tittering, the virus moves to a docking state as gB interacts with different proposed cellular receptors (including HSPG, Annexin II, CD13, 92,5 KDa receptor, EGFR and integrins).

HCMV entry is sensed by Toll-like receptors (TLR-1 and TLR-2) through a physical interaction with gB and gH, which results in the activation of cellular inflammatory cytokine secretion and NF-$_{\kappa}$B (Boehme *et al.,* 2006*).* This fusion event is mediated by gB, and the gH/gL complex that consequently lead to the fusion of the virion envelope with the cell plasma membrane. The entry process ends by the release of tegument proteins and capsid into the cytoplasm and results in the alteration of the transcriptional profile of hundreds of cellular genes (Figure 5).

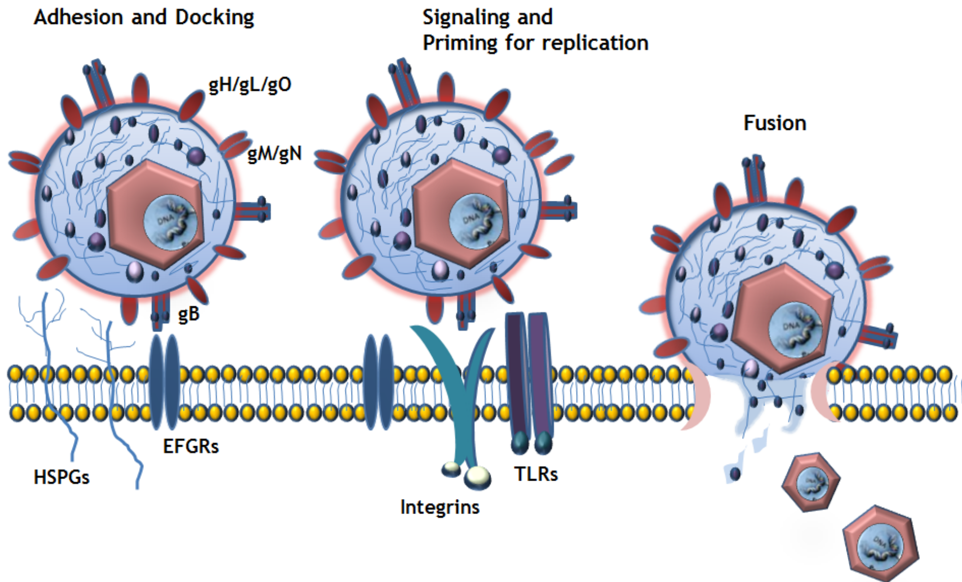


Figure 5. Schematic model of HCMV entry process, adapted from a model suggested by Boehme *et al.,* 2006.

1.4.2 Maturation and Egress

The *in vitro* assembly model of HCMV particles proposes the nucelocapsids formed in the nucleus of the infected cell are released into the cytoplasm after budding across the nuclear membranes (Skepper *et al.,* 2001). These newly formed nucleocapsids mature and accumulate within the nucleus prior to the subsequent process of envelopment and de-envelopment at the inner and outer nuclear membrane, respectively. The nucleocapsids undergo a second stage of envelopment at the cytoplasmic cisternae, which produces particles with a double membrane (Reddehase, 2006). The site of virus factory, where viral particles acquire their final envelope and several tegument and envelope proteins of HCMV accumulate late in infection, is believed to be located within a juxta-nuclear cytoplasmic compartment (Sanchez *et al.*, 2000). After the envelopment, virions are transported in vacuoles to the plasma membrane and the mature progeny is released to the extracellular medium.

1.5 Immune responses to HCMV

1.5.1 Innate and Adaptive Immunity

HCMV infection induces innate and adaptive, humoral and cell mediated immune responses. Innate immune response is observed during the entry process mediated by the interaction of viral glycoproteins gB and gH with TLR-2 on the cell surface (Boehme *et al.,* 2006; Compton *et al.,* 2003). This response is characterized by the upregulation of NF-κB, enhanced inflammatory cytokine secretion, activation of interferon sensitive genes (ISG) through interferon regulatory factor 3 (IRF3) (DeFilippis *et al.,* 2006), changes in Ca^{2+} homeostasis and the activation of mitogen-activated protein kinase (MAPK) signalling pathway (Billstrom *et al.,* 1998; Johnson *et al.,* 2000). Moreover, phosphatidylinositol-3-OH kinase (PI3-K) (Johnson *et al.,* 2001), activation of phospholipases C and A2, as well as an increased release of arachidonic acid and its metabolites are observed. Cytosolic DNA sensors, including DAI and AIM2 have been shown to recognize CMV DNA, inducing a type I interferon response and inflammasome assembly (Pyzik *et al.,* 2010).

NK-cells play an important role in the innate immunity against CMV infection. Studies performed in mice demonstrate their importance in MCMV clearance (Reviwed in Biron *et al.,* 1997; Arase *et al.,* 2002*)*. Furthermore, adoptive transfer of NK cells provides protection *in vivo* (Bukowski *et al.,* 1985) against the virus. Studies performed with different mouse strains have revealed a variation in the susceptibility to MCMV, a virus infection resistance due to the presence of the *cmv-1* genetic locus (contained in the NK gene complex on mouse chromosome 6). This resistance is mediated by the murine NK cell activation receptor Ly-49H and its direct binding to MCMV m157 protein, which activate NK cell effector functions for virus elimination. Plasmacytoid dendritic cells (plasmacytoid DC), which are the main producers of type I IFNs, are another important innate immunity component against CMV infection. The role of DC/NK cross-talk contributing to the *in vivo* clearance of the MCMV has been shown in several studies. Following MCMV infection, plasmacytoid DC produce IFN-α, IL-12, and TNF-α through engagement of TLR-9. These cytokines are capable of activating NK cell cytotoxicity and IFN-γ secretion that directly inhibits MCMV

replication (Krug *et al.,* 2004). NK cells have also been shown to directly kill MCMV-infected cells through perforin secretion (Loh *et al.,* 2005).

In humans, the role of NK cells in the early control of HCMV infection has received attention (Guma *et al.,* 2006). In support of NK cell importance, an individual selectively lacking NK cells was reported to exhibit enhanced susceptibility to primary herpes virus infection, including HCMV disease (Biron *et al.,* 1989). Also, a patient lacking functional T cells was reported to reduce the virus load in the bloodstream during primary HCMV infection through the involvement of an NK cell mediated response (Kuijpers *et al.,* 2008). *In vitro* experiments using HCMV infected fibroblasts suggest that the loss of MHC class I molecules that occurs in infected cells to avoid T cell recognition may impair the engagement of NK cell inhibitory receptors, thus promoting the activation of NK cell effector functions. The importance of NK cells in the innate immune response to HCMV is further strengthened by multiple HCMV encoded proteins that control NK cell activation and cytotoxicity by inhibiting activating immunoreceptors and providing inhibitory signals. The influence of the interaction between DC and NK cells during HCMV infection is poorly defined. However a recent study by Magri *et al.* (2010) show that NK cells efficiently react against HCMV-infected monocyte derived DCs, overcoming viral immune evasion strategies.

Cell-mediated immunity, including CD4+ helper, CD8+ cytotoxic and γδ T cells, play key roles in the control of HCMV infection (Jackson *et al.,* 2010). CD4+ T-cells are able to secrete cytokines that activates CD8+ T-cells, which in turn can activate B-cells. Upon activation the CD8+ T-cells can develop into cytotoxic T-cells (CTL), which have the capacity to release cytolytic proteins and lyse infected cells. CD8+ T cells that recognize HCMV antigens are essential for clearing HCMV (Reusser *et al.,* 1991). During primary infection, specific CD4+ T cells are identified in peripheral blood shortly after detection of CMV DNA, followed by specific antibodies and CD8+ T cells. Moreover, T cell responses show a broad specificity against HCMV antigens, among them the immunodominant antigens pp65, gB and IE1 elicit the dominant of T cell responses (Sylwester *et al.,* 2005).

Numerous essential genes are conserved between the species-specific viruses MCMV and HCMV, and the use of mouse models of infection has provided valuable insight into the conserved functions and pathogenic mechanisms of CMV, however the diversity of gene products and differences in replication make it difficult to extrapolaite direct conclusions from a phenotype observed with the animal model to the behavior of HCMV *in vivo*.

1.5.2 Immunoregulatory mechanisms

Despite a strong primary immune response, HCMV is still able to establish latency and the following long term immune response is unable to clear the virus. Consequently, it persists life-long within its host. HCMV genome's large protein coding capacity has allowed the virus to evolve many mechanisms to interfere with the host immune response, avoid detection and elimination. The immunoregulatory mechanisms include the following strategies (Reviewed in Loenen *et al.,* 2001; Crough and Khanna, 2009):

1. HCMV encodes for several genes that modulate MHC class I antigen presenting pathway affecting the cell-surface expression of MHC class I. These glycoproteins include: US2, US3, US6, US8, US10 and US11. US2 and US11 translocate MHC class I molecules from the ER to the cytoplasm promoting proteasomal degradation and consequently reduce their cell surface expression and antigen presentation (Ahn *et al.*, 1996; Machold *et al.,* 1997). HCMV US3, an IE protein, binds to tapasin in the peptide loading complex and thereby interferes with the formation of stable MHC complexes (Park *et al.,* 2004). US6 impairs peptide-loading of MHC class I molecules through binding to the transporter associated with antigen presentation (TAP) in the endoplasmatic reticulum (ER) (Hengel *et al.,* 1997) and down-regulates the expression of the non classical MHC class I molecule HLA-E (Llano *et al.,* 2003). The US10 gene product, E protein gp21, acts selectively impairing the expression of of the non-classical MHC-I molecule HLA-G (Park *et al.,* 2010).

2. Other HCMV encoded proteins that interfere with MHC class I antigen presenting pathway are the structural phosphoproteins pp65 and pp71, released into the

cytoplasm during the final virus entry phase. pp71 interferes with the expression of MHC class I complexes on the cell surface (Trgovcich *et al.,* 2006), and the tegument protein pp65 has been reported to prevent processing of IE1 for MHC presentation (Besold *et al.,* 2007). By reducing the MHC I expression on the surface of the infected cell, all these proteins interfere with the recognition by HCMV-specific T cells. Moreover, HCMV avoids $CD4^{+}$ T-cells recognition by down-regulating MHC II expression, associated with the presentation of antigen by DCs and macrophages. US2 and US3 have both been described to interfere with MHC class II presentation (Reviewed in Johnson and Hegde, 2002).

3. Since a low level of MHC I expression is a signal for NK cell activation, HCMV has developed several mechanisms controlling NK cell mediated lysis and cytokine productin by expressing molecules that prevent triggering by activating receptors or engage inhibitory receptors (Reviewed in Wilkinson *et al.,* 2008). A UL40–derived leader sequence peptide bound to HLA-E provides a ligand for the inhibitory NK-cell receptor CD94/NKG2A (Tomasec *et al.,* 2000). HCMV *UL16*-, *UL142*-encoded proteins and HCMV´s miRNA miR-*UL112* respectively impair the expression of ligands for the activating NK-cell receptors NKG2D. HCMV glycoprotein UL18 shows homology to class I MHC heavy chain, associates with β2-microglobulin and interacts with the ILT2 (LILRB1) inhibitory receptor (Cosman *et al.,* 1997; Beck and Barrell, 1988). Interestingly, LILRB1 has been shown to bind UL18 with 1000-fold higher affinity than its physiological ligands (Willcox *et al.,* 2003) but the consequence of this interaction during infection and the uncertain cell surface expression of UL18 remains to be verified (Guma *et al.,* 2006).

4. HCMV also encodes for genes that interfere with the humoral immune responses, e.g. by incorporating host complement inhibitory proteins (that normally protects cells from complement lysis) CD46 and CD55, into the virion envelope (Spiller *et al.*, 1996). HCMV encodes for at least two viral Fc-receptors, *UL119-118* and *TRL11/IRL11*, which are believed to protect the HCMV-infected cell from antibody-dependent cell mediated cytotoxicity and complement binding to the

IgG-Fc receptors (Lilley *et al.,* 2001; Atalay *et al.,* 2001).

5. Another mechanism that HCMV uses to interfere with the immune defense is modulation of extracellular signals, including cytokines and chemokines that are involved in inflammatory reactions and in activation and recruitment to the sites of infection of T cells, NK cells, neutrophils, and monocytes. HCMV interfere with these host factors by encoding for cytokine-binding proteins, cytokine-receptor antagonists and G-protein coupled receptor (GCRs) homologs (*UL33, UL78, US27,* and *US28*). A functional homologue of IL-10, cmvIL-10 (*UL111A*), has been proposed to downregulate macrophage and T cell responses and thereby has an anti-inflammatory role (Kotenko *et al.*, 2000). Another mechanism is the alpha chemokine, vCXCL-1 (*UL146*) that attracts neutrophils via CXCR2 (Penfold *et al.*, 1999).

6. Several HCMV genes are involved in modulation of IFN signaling pathway (Reviewed in McCormick and Mocarski, 2007), a major line of immune defense. pp65 regulates the interferon response factor 3 (IRF3) (Abate *et al.,* 2004) and the NF-κB subunit p65 (Browne and Shenk, 2003), IE1 protein (IE72), prior to IFNα/β treatment, reduces the expression of the interferon stimulated gene ISG54 and the dsRNA binding proteins, TRS1/IRS1, have been shown to prevent virus-induced transcription factor activation (Hakki *et al.,* 2005).

HCMV gene	Potential role
UL18	MHC I homolog, Binds LIR-1 and thus interfers with NK cell recognition inhibitory effect on cells carrying the ligand CD94/NKG2.
UL33	CCR1 Chemokine receptor homolog, present in viral envelop, role in inflammatory response.
UL78	CCR1 Chemokine receptor with similarity to the fMLP receptor, role in inflammatory response, leukocyte migration
US28	Chemokine receptor; sequesters CC chemokines and may assist in viral dissemination
UL111A	IL-10 homolog, Inhibits MHC class I/II expression and lymphocyte proliferation
UL144	TNF receptor homolog, Most members bind TRAFs, upregulate NF-kB and kinases and stimulate differentiation of many subsets of lymphocytes.
UL146 & UL147	IL-8 Chemokine homolog, Role in inflammatory response, leukocyte migration
US27 & US28	CCR1 Chemokine receptor-like protein Role in inflammatory response, leukocyte migration
UL83	Targets MHC I, pp65 blocks presentation of IE1 peptide
US3	IE protein, Intracellular retention of MHC I in ER
US2	E protein, Dislocation of MHC I complex to cytoplasm and degradation Degradation HLA-DR-alpha and DM-alpha
US11	E protein, Dislocation of MHC I to cytoplasm and degradation
US6	E/L protein, targets MHC I, Inhibition TAP-dependent peptide translocation
UL40	E protein, targets MHC Ib, TAP-independent gpUL40 peptide presented by HLA-E allows potential interaction with CD94/NKG2 on APC, T and NK cells
TRL11 & UL118	Serves as viral Fc receptor to block ADCC
UL16	Bindes to nonclassical MHC proteins (MICB, ULBP1 and ULBP2) and thus blocks NKG2D-mediated NK-cell activation

Table1. Summary of HCMV immunnemodulation mechanisms (Reviewed in Loenen *et al.,* 2001; Crough and Khanna, 2009).

1.6 Primate CMV RL11 gene family

Most of HCMV genes lack introns and are believed to have arisen from insertion of a cDNA copy of a cellular mRNA or pre-mRNA into the viral genome subsequently modified during co-evolution. This process called gene capture has occurred throughout herpesviruses co-evolution with their hosts. Many HCMV genes have counterparts in their host genes, such as homologs of Uracil-DNA glycosidase, major histocompatibility complex (MHC) class-I molecules, cytokines and chemokines (Raftery *et al.,* 2000; Holzerlandt *et al.,* 2002). These captured genes may mimic the original host protein or, alternatively, they may perform novel functions while preserving some structural features.

Gene duplication has been the main mechanism behind the genetic diversity observed in large eukaryotic DNA viruses and their hosts. HCMV genome comprises 12 multigene families (Chee *et al.*, 1990; Davison *et al.*, 2003b) that probably arose by gene capture and subsequent gene duplication. One of them, the RL11 multigene family, a family of putative glycoproteins with characteristic signal peptide, transmembrane domain and N-glycosylation sites, is encoded by primate CMV genomes, including HCMV, chimpanze CMV (CCMV) and rhesus CMV (RhCMV). The RL11 gene family is a rapidly evolving and highly variable virus gene family, both in terms of gene members as well as in terms of sequence divergence across related viruses or strains (Dolan *et al.,* 2004). This family has presumably arisen due to gene duplication of an ancestral gene captured from the host genome (Davison *et al.*, 2003a).

In RhCMV, RL11 family is represented by the following putative glycoproteins, *Rh19 (UL7*-homolog), *Rh20 (UL6*-homolog*), Rh17, Rh21-23 (UL11*-homologs*)* and *Rh25 (UL9)* (Hansen *et al.,* 2003). In the HCMV genome, RL11 genes are located in the TR_L-U_L junction region and include 11 conserved genes, namely *TRL/IRL11, TRL/IRL12, TRL/IRL13, UL4, UL5, UL6, UL7, UL8, UL9, UL10* and *UL11* (Chee *et al.,* 1990, Figure 6). Comparison of HCMV and CCMV genomes identified several HCMV genes that were absent from CCMV (Davison *et al.,* 2003b). These genes cluster in the genomic region containing the RL11 gene family and are *UL1, RL6, RL5A* and *UL3*. An additional gene not found in CCMV, *UL111A,* encodes a protein with homology to human interleukin 10 (IL-10), and seems to be secondarily lost in CCMV as it is present in rhesus cytomegalovirus (Lockbridge *et al.,* 2000). *UL3* encodes a 105 amino acid-long putative protein (close to the 100 amino acid length threshold for ORF annotation) and shows no homology to other virus or host genes. *UL1, RL5A and RL6* (224, 104 and 111 amino acids long respectively) are homologues to other members of the RL11 HCMV gene family**.**

The members of the RL11 gene family were initially assembled due to the presence of a defined motif in their sequence resembling the cellular Thy-1 whithin a region conserved between other Ig superfamily members (Chee *et al.*, 1990). Subsequently, in a more detailed analysis, this shared core motif was further examined in the context of the neighboring sequences allowing the characterization of an RL11

domain (RL11D) (Davison *et al.*, 2003a). RL11D consists of a region of variable length (65-82 amino acids) containing three characteristic conserved residues (a tryptophan and two cysteines) and several potential N-linked glycosylation sites.

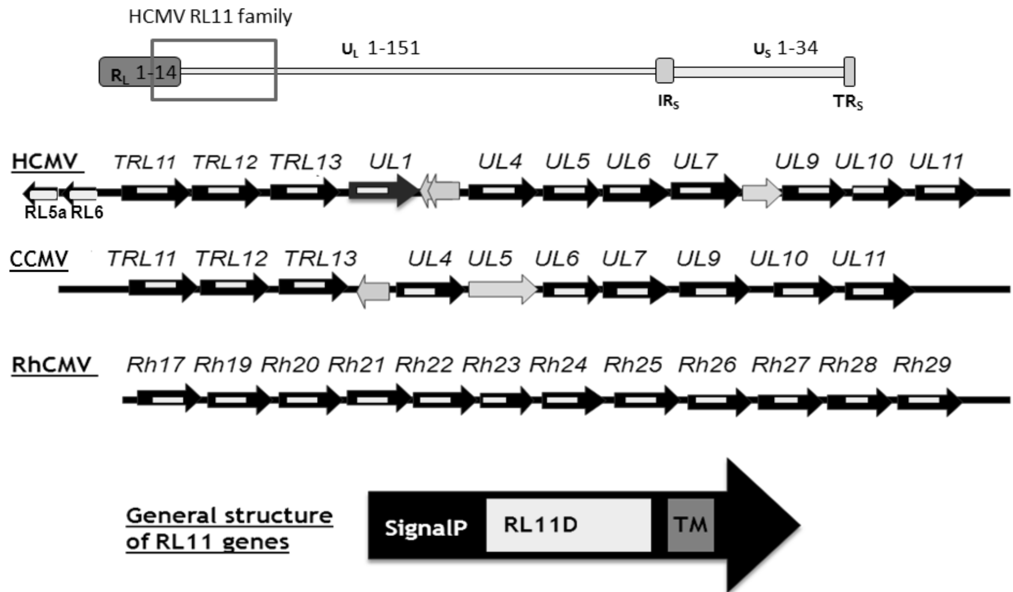


Figure 6. Genomic arrangement of RL11 genes in HCMV, CCMV and RhCMV. Diagram showing the locations of RL11D (yellow box) in HCMV (AD169), CCMV and RhCMV RL11 genes. The HCMV specific gene UL1 is shown in red. This figure is adapted from Davison *et al.*, 2003.

Regarding the distribution and functions of the *RL11* family members, gp48 (*UL4*) and gpRL13 have been described as virion envelope glycoproteins. The interest in gp48, a nonessential component of the viral envelope (Hobom *et al.,* 2000), has been limited due to the complex transcriptional regulation of its expression. *UL10* and *RL13* are so-called temperance genes, with cell type-specific virus-growth inhibition functions (Dunn *et al.,* 2003; Stanton *et al.,* 2010). *TRL/RL11* (Lilley *et al.,* 2001; Atalay *et al.,* 2001) encodes for a IgG-Fc binding glycoprotein, that is expressed on the surface of HCMV-infected cell, and which has been shown to bind all four human IgG subclasses. *UL11* encodes for a hypervariable protein expressed on the cell surface of infected cells (Hitomi *et al.*, 1997). Of note, RL11 genes, dispensable for virus growth in cell culture, are among the most variable HCMV genes (Murphy *et al.*, 2003, Dolan *et*

al., 2004) and *RL5A, RL13* and *UL9*, are members of a restricted set of HCMV hypervariable genes (Dolan *et al*., 2004). *RL6* and *RL5A* are transcribed but direct evidence for protein expression is not available (Davison *et al*., 2003a). Due to sequence similarities of *RL6* and *RL5A* to RL11 genes and high variability of *RL6* is presumed that these two genes are translated into small RL11D proteins. No function has been demonstrated yet for UL6, but structural predictions have indicated that, in addition to being a transmembrane glycoprotein UL6 contains a subtilisin fold and may therefore act as a protease (Novotny *et al.,* 2001). *UL7,* is a glycoprotein with a structural homology to the signaling lymphocyte-activation molecule (SLAM)-family receptor CD229, with the capacity to mediate adhesion to leukocytes and interfere with cytokine production (Engel *et al.,* 2011).

Though there is no information on the expression and function of UL1, a significant sequence similarity was found with members of the human carcinoembryionic antigen (CEA) family (Holzerlandt *et al.,* 2002). The CEA a structural sub-group of the immunoglobulin superfamily (IgSF) is composed of 29 members located on chromosome 19q13.2. Members of the family are broadly divided into CEA-related cell adhesion molecule (CEACAM) and the pregnancy specific glycoprotein (PSG) subgroups (Reviewed in Kuespert *et al.,* 2006), involved in immune responses and reproduction. Interestingly, CEACAMs have been reported as receptors for bacterial or viral pathogens, such as CEACAM-1 shown to target Opa proteins of *Neisseria* species (Boulton and Gray-Owen, 2002). In the present study we aimed to better understand the origin and evolution of HCMV-specific *UL1* and explore in more detail the putative implications of its sequence similarity with the CEA family molecules that are involved in homotypic and heterotypic adhesion mechanisms.

RL11 gene	Protein	Potential role
TRL/RL11	Glycoprotein, expressed on the cell surface of HCMV-infected cells	viral Fc receptor, bas been shown to bind all four human IgG subclasses
TRL/RL13	Envelope glycoprotein	So-called temperance genes, with cell type-specific virus-growth inhibition functions, members of a restricted set of HCMV hypervariable genes
UL1	Putative transmembrane glycoprotein	sequence similarity with CEA family
UL4	Envelope glycoprotein	
UL6	Putative transmembrane glycoprotein	Contains a subtilisin fold and may therefore act as a protease
UL7	Glycoprotein, expressed on the cell surface of HCMV-infected cells	Capacity to mediate adhesion to leukocytes and interfere with cytokine production , SLAM-family receptor CD229 homolog
UL9	Putative transmembrane glycoprotein,	Member of a restricted set of HCMV hypervariable genes
UL10	Putative transmembrane glycoprotein,	So-called temperance genes, with cell type-specific virus-growth inhibition functions
UL11	Glycoprotein, expressed on the cell surface of HCMV-infected cells	
RL5A & RL6	n/a	Members of a restricted set of HCMV hypervariable genes

Table 2. HCMV RL11 family, structural and functional roles.

2. AIMS

The main objectives of this research project have been the following:

1) To study the evolution and origin of *UL1*, a HCMV-specific member of the RL11 family, using a bioinformatics approach.

2) To examine the expression and biochemical properties of HCMV UL1 using eukaryotic recombinant expression system.

3) To explore the expression and functional role of UL1 protein in the context of HCMV infection.

PART II MATERIAL AND METHODS

3. MATERIAL AND METHODS

3.1 Sequence analysis and evolutionary studies

Primate CMV genes and proteins were retrieved from NCBI (Pruitt *et al.,* 2009) and human proteins were retrieved from Ensembl Human database (Hubbard *et al*, 2007). The viruses and their NCBI accession number are as follows: HCMV AD169 strain, NC001347 (Chee *et al.*, 1990; Davison *et al.*, 2003b); HCMV Merlin strain, NC006273.2; Chimpanzee cytomegalovirus (CCMV), NC003521.1 and Rhesus cytomegalovirus (strain 68-1, RhCMV)**,** NC006150.1. HMMs corresponding to member of immunoglobulin domain containing family, including V-set (variable), C1-set (constant-1), C2-set (constant-2) and I-set (intermediate) subsets were downloaded from PFAM (Finn *et al.,* 2008). Ig-domain containing human and HCMV proteins were obtained using the program HMMsearch (Eddy, 1998) with default parameters (E-value cut-off < 10^{-3}). The BLASTP program was used to search for homologous human and HCMV proteins (Altschul *et al.*, 1990). Sequence alignments of homologous Ig-domain containing HCMV proteins and cellular proteins were generated with the help of the CLUSTALW program (Thompson *et al.*, 1994).

Evolutionary studies of RL11 family in primate CMV: The number of amino acid substitutions per site (K) was calculated for HCMV and CCMV orthologous pairs of proteins using the PHYLIP programs Protdist with default parameters (Felsenstein, 1989). Protein predictions: N-glycosylation sites were predicted using NetNGlyc 1.0 (Gupta and Brunak, 2002). Signal sequences were predicted using SignalP 3.0 (Nielsen *et al.*, 1997) and transmembrane domains predicted with TMHMM 2.0 (http://www.cbs.dtu.dk/services/). The prediction of potential globular regions was performed with GlobPlot (Linding *et al.,* 2003). The Genbank accession numbers for UL1 protein sequences in different HCMV strains are as follows: Merlin: Q6SWC8, HAN13: C8BKG1, 6397: Q6SX88, 3301: Q6SX71, AD169: B8YE39 and Towne: B9VXH4.

3.2 Cell lines

The adherent human embryonic kidney cell line (HEK293), African green monkey kidney cell line (COS-7), the MRC-5 human fibroblast cell line (ATCC CCL-171, passages 5–15), the HEL299 human lung fibroblast cell line (ATCC CCL-137), and the U373-MG cell line derived from glioblastoma (ECACC 89081403) were grown in DMEM (Gibco) supplemented with 10% (v/v) heat-inactivated fetal bovine serum (FBS), penicillin (100 U/ml) and streptomycin (10 μg/ml). The retinal pigment epithelium (RPE) cells were grown in DMEM-Ham's F-12 (1:1) medium (Gibco) supplemented with 5% (v/v) FBS, penicillin (50 U/ml), streptomycin (50 μg/ml), 0.25% (v/v) sodium bicarbonate, and 2 mM glutamine. Cell lines were maintained at 37 ºC in a humidified atmosphere containing 5% CO_2.

3.3 Antibodies

Commercial mAbs specific for IE1 (B810R, Chemicon) and pp28 (ab6502, Abcam) were employed. gB-specific mAb 27-78 and mAb 58-15, (generously provided by William J. Britt, University of Alabama-Birmingham) were respectively used for immunofluoresence microscopy and immunoblotting. Reagents specific for Calnexin (610524, BD Bioscience), GM-130 (558712, BD Bioscience), TGN-46 (AHP500G, AbDseroTec) and ERGIC-53 (E6782, Sigma-Aldrich) were generously provided by Julia V. Blume (Center for genomic regulation, Barcelona). Rabbit anti-HA (H6908, Sigma-Aldrich) was used to detect UL1-HA. Secondary antibodies used in immunofluorescence microscopy were Alexa Fluor® 555 goat anti-mouse IgG (H+L) (A21422, Invitrogen) and Alexa Fluor® 488 goat anti-rabbit IgG (H+L) (A11008, Invitrogen). Anti-mouse or anti-rabbit IgG horseradish peroxidase-labelled antibodies (Amersham Bioscience, Buckinghamshire, UK) were used for immunoblotting.

The following antibodies were used for flow cytometry analysis: anti-Myc (9E10), anti-HA (12CA5); anti-IREM1 (UP-D2); isotypic controls (IgG and IgG2a) from Becton Dickinson; secondary FITC- or PE-conjugated rabbit anti-mouse $(Fab')_2$ DAKO (Glostrup, Denmark) and FITC-conjugated goat anti-rat $IgG_{1/2}$ (BD

Pharmigen™). Single colour immunofluoresence analysis were performed as previously described and samples were analysed in FACSscan using CellQuest software (Becton Dickinson Bioscience-BDB, San Jose, USA). Secondary antibodies anti-human (Fc-specific) horseradish peroxidise antibody (Sigma, A0170) and anti-mouse IgG horseradish peroxidise linked whole antibody (Amersham Bioscience, Buckinghamshire, UK) were used for immunoblotting.

3.4 Preparation of eukaryotic expression constructs and DNA transfection

pCI-neoUL1

Full-length *UL1*, a 682 bp long nucleotide sequence, was obtained by PCR using DNA extracted from HCMV AD169 virions as template and the primers 5'-AAC ATG GGC ATG CAA TGC-3´ and 5'-GTA ATC ACG CTA TGC GAT T-3'. The PCR product was inserted into an intermediate vector pGEMT (Promega, catalog no. A1360) and thereafter the complete *UL1* was then cloned into the *SmaI*-digested pCI-neo (Promega, catalog no. E1841) dephosphorylated expression vector to generate pCI-neoUL1.

pDisplay-$UL1_{Myc-HA}$

Complete *UL1* sequence was cloned into pDispaly expression vector (Invitrogen, catalog no. V660-20) fused to HA- and Myc-epitopes (pDisplay-$UL1_{Myc-HA}$). The construction of *UL1* fused to Myc-tag and HA-tag was preformed as following: *UL1* was used amplified from pCI-neoUL1 by PCR and *SmaI* restriction site were introduced, using the primer pairs 5'-CCC GGG ATG GGC ATG CAA TGC AAC A-3´ and 5'-CCC GGG CGC TAT GCG ATT TAC GTT G-3´. *UL1* fragment was further cloned into pGEM-T vector. Site-directed mutagenesis was performed introducing a stop codon between Myc-tag and PDGFR TM domain in pDisplay vector using the primer pair 5´-CAT CTC AGA AGA GGA TCT GTA AGC TGT GGG CCA GGA CAC-3´ and 5´-GTG TCC TGG CCC ACA GCT TAC AGA TCC TCT TCT GAG ATG-3´. After the PCR reaction, the mixture was treated with *DpnI*

and transformed into DH5α *E.coli* bacterial strain. An additional Myc epitope was introduced into pDisplay vector between *SacII* and *SalI* restriction site, using sense and anti-sense pimers 5'-TCC CCG CGG GAT GGA GCA AAA GCT CAT TTC TGA AGA GGA CTT GAA TGA AGT CGA CGT CGG CCA TAG CGG CCG CGG AA-3' and 5'- TTC CGC GGC CGC TAT GGC CGA CGT CGA CTT CAT TCA AGT CCT CTT CAG AAA TGA GCT TTT GCT CCA TCC CGC GGG GA-3'. Finally, the *UL1* fragment was cloned into the mutated pDisplay vector at *SmaI* restriction site and used for transformation of DH5α bacteria. Positive clones were identified and verified by sequencing.

pDisplay-$UL1_{HA}$

UL1 without signal peptide sequence was cloned into pDisplay vector (pDisplay-$UL1_{HA}$) using following primer pairs 5´-CCC GGG CAT GAA CAA AAA AAG GCG TTT TAC-3´ and 5´-CCC GGG TCA CGC TAT GCG ATT TAC GTT GTT-3´, amplifying the sequence of *UL1* without signal peptide and introducing *SamI* restriction site. First, *UL1* sequence was inserted into an intermediate vector pGEM-T vector (Promega, catalog no. A1360). Subsequently, the *UL1* fragment was cloned into pDisplay vector at *SmaI* restriction site. After transformation of DH5α bacteria, pDisplay-$UL1_{HA}$ was digested with *kpnI* restriction enzyme, confirming correct orientation of *UL1* insert followed by sequencing.

To avoid re-ligation of the vector used for cloning, they were dephosphorylated at 5'-end using calf intestinal alkaline phosphatase, as suggested by the manufacturer. The dephosphorylated DNA vectors were purified using a Qiagen silica-gel–membrane (QIAquick PCR purification). DNA fragments were produced by digestion of DNA with each restriction endonuclease. The fragments and the target vector were then ligated in a reaction volume of 10 µl, the fragment to be inserted was given in a 5-fold excess to the vector DNA (100 ng). The ligation reaction using T4 DNA ligase was performed following manufacturer's protocol at 16° C overnight.

Transient transfection

HEK293 and COS-7 cells were transiently transfected with UL1 cDNA or control plasmids at 50-70% confluency using the calcium phosphate method (15-24 ug of UL1 cDNA) or Superfect transfection method (Qiagen). The cells were harvested between 8 to 72 h after transfection and analysed by immunofluorescence (cells surface or intracellular staining) and fluorescence microscopy. Transfected cells were also lysed in 100 mM Tris-HCL at pH 7.4, 150 mM NaCl, 2 mM sodium orthovanadate, 5 mM EDTA and 1% Triton X-100, followed by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) and western blot analysis using anti-Myc or anti-HA antibodies.

3.5 Expression and production of UL1-Fc fusion proteins

pSeqTag-$UL1_{Fc}$

The UL1 IgV-like fragment (HCMV AD169) was amplified by PCR and cloned into the pSecTag ABC/mouse IgG2a vector (Invitrogen) between *BamHI* and *HindIII* restriction sites. The obtained clones were digested with *kpnI* restriction enzyme, confirming the correct orientation of *UL1* insert.

pCR3-$UL1_{Fc}$

IgV-like fragment of UL1 (TB40 strain) was cloned into *BamHI* and *SalI* sites and fused to human IgG1-Fc in pCR3 expression vector. pCR3 contains the Fc part is from human IgG1 including the hinge, CH_2 and CH_3 domains. This expression vector was a kind gift from M. Messerle (Hannover Medical School, Germany).

Recombinant proteins were expressed by transient transfection of HEK293 cells using the calcium phosphate method; 18 h later, cells were washed twice and cultured on serum free medium (EX-cell ACF CHO Medium, Sigma, St. Louis, MO). At day 3 and 6 the supernatants were recovered and the soluble molecules purified by affinity chromatography with protein A Sepharose CL-4B (Amersham Pharmacia; Piscataway, NJ). UL1-Fc protein was eluted with 0.1 M sodium citrate, pH 4.0. Column fractions were neutralized with 1 M Tris-HCl, pH 9.0, and analyzed by SDS-

PAGE and immunoblotting for positive fractions. Supernatants collected were tested for the presence of fusion protein by immunopreciptation (IP) using protein A Sepharose beads. Briefly, the supernatant (concentrated or not-concentrated) of transfected cells were pre-cleared for 2 h at 4 ºC on protein A Sepharose beads (GE Healthcare Bio-Sciences AB, Uppsala, Sweden) and then incubated for 3 h at 4ºC with primary antibody (fusion protein-containing supernatant). Proteins were solubilized in loading buffer and boiled for 5 minutes before resolution by SDS-PAGE (12% acrylamide gels) under non-reducing and reducing (β-mercaptoethanol) conditions. The samples were transferred to polyvinylidene difluoride (PVDF) membrane (Millipore Corporation, Bedford, MA, USA) using semidry blotting apparatus. Membranes were blocked with PBS containing 0.05% Tween-20 and 5% milk powder. Antibodies were diluted in PBS containing 0.05% Tween-20. For detection of primary antibody binding, horseradish peroxidase-conjugated goat anti-human or anti-mouse IgG followed by enhanced chemiluminescence detection system using West Pico Supersignal kit (Pirce, Rockford, IL, USA) according to the manufacturer's instructions.

3.6 Viruses

The following HCMV strains were used: Towne (ATCC VR-977), Toledo (Quinnan *et al.,* 1984), the endotheliotropic strain TB40/E (Sinzger *et al.*, 1999), TB40/E reconstituted from respective BAC clone TB40/E-BAC4 (Sinzger *et al.*, 2008b) and AD169 reconstituted from respective BAC clone pHB5 (Borst *et al.,*1999). Infectious viral stocks were propagated as follows, MRC-5 cells were infected at low multiplicity of infection, and the supernatant of infected cells which exhibited 100% cytopathic effect were cleared of cellular debris by centrifugation at 6000 x g for 10 min. The cell free supernatant was centrifuged at 20,000 x g for 3 h to pellet the virus particles. Pelleted virus was resuspended in serum free DMEM, stored at -80ºC. The amount of infectious virus was determined by counting viral IE1 (MAb B810R, Chemicon) positive MRC-5 cells. Briefly, MRC-5 cells were incubated for 2 h with virus-

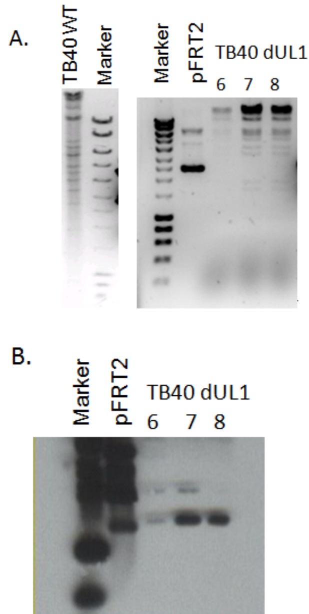
containing supernatant, after the adsorption period the inoculums were washed away with phosphate-buffered saline rinse, and the cells were incubated with DMEM supplemented with 3% FBS. After 20 h cells were fixed for 20 min at room temperature with 4% paraformaldehyde (PFA). Samples were blocked with PBS containing 4% bovine serum albumin, permeabilized in 0.5% Triton X-100, and incubated with the IE1 MAb and Alexa 488–conjugated rabbit anti-mouse $(Fab')_2$ fragments. Nuclei were counterstained with DAPI (4'-6'-diamidine-2-phenylindole; Sigma-Aldrich) for 5 min at room temperature. Slides were examined with a Leica DM6000B fluorescence microscope. Images were analyzed with the Leica fluorescence workstation software (Leica FW4000).

3.7 Construction and generation of recombinant viruses

HCMV bacterial chromosome (BAC) mutagenesis was carried out by homologous recombination between a linear DNA PCR fragment and the endotheliotropic HCMV-TB40/E-derived BAC4 (TB40/E BAC4) or the laboratory adapted AD169 BAC (pHB5) in *Escherichia coli*. The mutagenesis was performed as previously described by Wagner *et al.* (2002). TB40/E BAC4 (Sinzger *et al.*, 2008b) was used for the generation of *UL1* deletion viruses, removing the complete *UL1* and pHB5 was used to generate the recombinant AD169 UL1-HA virus. To delete the *UL1* sequence in the TB40/E BAC4 and generate TB40/E BAC4 ΔUL1 the following oligonucleotides were used: (5´-GTG GTT TTT ATA GGT TAA CCA CTT ATG GTG TAA AGT AGG ATA TTC ATA GTT ATT GAA AAC CGG CCA GTG AAT TCG AGC TCG GTA CCC-3´) and (5´-TGA GAG CTA CAC AAA GTA GCA CAA ATT TAT CCT GAA TAA ACT AGC GTC GAT ACT TTG TAA GAC CAT GAT TAC GCC AAG CTC CC-3´). For construction of *UL1* HA-epitope tagged in pHB5: forward primer containing the HA-tag epitope sequence followed by a stop codon (5´-CGT TCC GGC AAC TCT GAG ACA CAA ACT ACG AAC TAG AAA CAA CGT AAA TCG CAT AGC GGC CTA CCC ATA CGA TGT TCC AGA TTA CGC TTA ACC AGT GAA TTC GAG CTC GGT AC-3´) and reverse primer containing homology

region downstream of *UL1* (without stop codon) 5´-GAG AAC CGC ACG GAG TAA TCA AAT TTT ATC TTG GAT AAA TTA GTG TCG ATA CTT TGT AAG ACC ATG ATT ACG CCA AGC TCC-3´ were used. PCR fragments using the plasmid pSLFRTKn as template containing a kanamycin resistance (Kan^R) gene (flanked by minimal FRT sites) were inserted into TB40/E BAC4 or pHB5, respectively by homologous recombination in *E. coli* (Wagner *et al.,* 2002). TB40/E BAC4 ΔUL1 and pHB5 UL1-HA mutants were digested with *XbaI* and the pattern of the restriction enzyme was compared to the wild-type TB40/E BAC4 or wild-type pHB5. The presence of Kan^R was further confirmed by probing the Southern transfer with a specific kanamycin cassette probe. Kan^R was excised from both viruses by FLP-mediated recombination (Wagner *et al.,* 2002), and the correct mutagenesis was confirmed by restriction pattern followed by Southern blotting and PCR analysis.

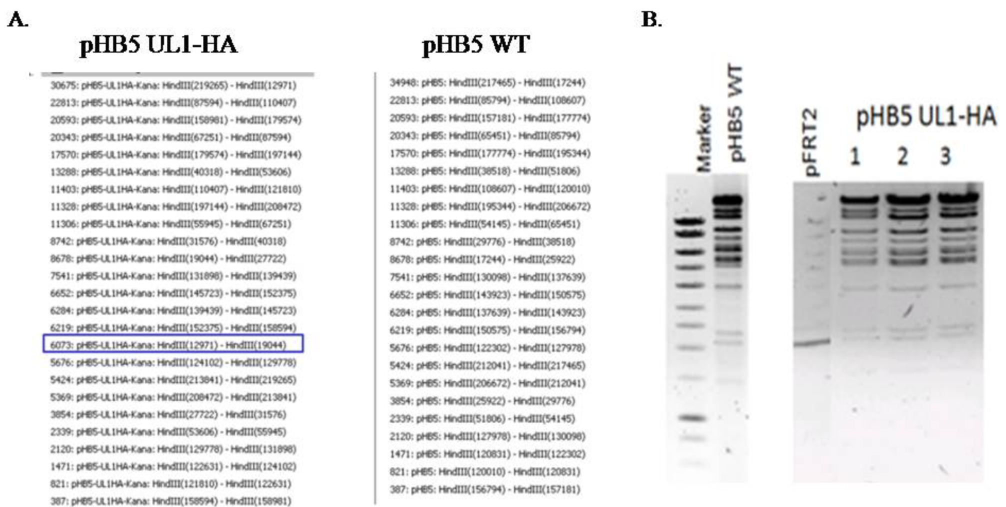
TB40/E BAC ΔUL1 and pHB5 UL1-HA mutants were digested with *XbaI* and the pattern of the restriction enzyme was compared to respective parental TB40/E BAC4 or pHB5 viruses. The difference between TB40/E BAC4 wild-type and TB40/E BAC4 ΔUL1 Kan^R digested with *XbaI* is a band at 1629bp. The presence of Kan^R was further confirmed by probing after Southern blotting with a kanamycin cassette specific probe. A band at the expected size of kanamycin gene was detected in the analyzed mutant clones (Figure 7). pHB5 UL1-HA recombinant viruses were digested with *HindIII* and the pattern of the restriction enzyme was compared to the wild-type pHB5. *HindIII* digestion, compared between pHB5 wild-type and pHB5 UL1-HA kanamycin the difference is a band seen at 6073bp (Figure 8).



Difference between TB40/E BAC4 wild-type (kanamycin) and TB40/E BAC4 dUL1 is in band at 1629 bp

3363: MS-TB40/E-BAC4-dUL1Kana: XbaI(8097) - XbaI(11460)
1875: MS-TB40/E-BAC4-dUL1Kana: XbaI(180482) - XbaI(182357)
1679: MS-TB40/E-BAC4-dUL1Kana: XbaI(13647) - XbaI(15326)
1629: MS-TB40/E-BAC4-dUL1Kana: XbaI(12018) - XbaI(13647)
795: MS-TB40/E-BAC4-dUL1Kana: XbaI(226400) - XbaI(227195)
558: MS-TB40/E-BAC4-dUL1Kana: XbaI(11460) - XbaI(12018)
279: MS-TB40/E-BAC4-dUL1Kana: XbaI(226121) - XbaI(226400)
18: MS-TB40/E-BAC4-dUL1Kana: XbaI(108453) - XbaI(108471)

Figure 7. (A) *In silico XbaI* digestion, compared between TB40/E BAC4 wild-type and TB40/E BAC4 ΔUL1 (TB40 ΔUL1) the difference is a band seen at 1629bp. The BAC-derived viruses digested with *XbaI* and separated using 0.8% agarose gel electrophoresis and visualized using ethidium bromide staining (B) Southern blot detecting Kanamycin cassette in TB40/E BAC ΔUL1 (clone 6, 7, 8).


A.

pHB5 UL1-HA

30675: pHB5-UL1HA-Kana: HindIII(219265) - HindIII(12971)
22813: pHB5-UL1HA-Kana: HindIII(87594) - HindIII(110407)
20593: pHB5-UL1HA-Kana: HindIII(158981) - HindIII(179574)
20343: pHB5-UL1HA-Kana: HindIII(67251) - HindIII(87594)
17570: pHB5-UL1HA-Kana: HindIII(179574) - HindIII(197144)
13288: pHB5-UL1HA-Kana: HindIII(40318) - HindIII(53606)
11403: pHB5-UL1HA-Kana: HindIII(110407) - HindIII(121810)
11328: pHB5-UL1HA-Kana: HindIII(197144) - HindIII(208472)
11306: pHB5-UL1HA-Kana: HindIII(55945) - HindIII(67251)
8742: pHB5-UL1HA-Kana: HindIII(31576) - HindIII(40318)
8678: pHB5-UL1HA-Kana: HindIII(19044) - HindIII(27722)
7541: pHB5-UL1HA-Kana: HindIII(131898) - HindIII(139439)
6652: pHB5-UL1HA-Kana: HindIII(145723) - HindIII(152375)
6284: pHB5-UL1HA-Kana: HindIII(139439) - HindIII(145723)
6219: pHB5-UL1HA-Kana: HindIII(152375) - HindIII(158594)
6073: pHB5-UL1HA-Kana: HindIII(12971) - HindIII(19044)
5676: pHB5-UL1HA-Kana: HindIII(124102) - HindIII(129778)
5424: pHB5-UL1HA-Kana: HindIII(213841) - HindIII(219265)
5369: pHB5-UL1HA-Kana: HindIII(208472) - HindIII(213841)
3854: pHB5-UL1HA-Kana: HindIII(27722) - HindIII(31576)
2339: pHB5-UL1HA-Kana: HindIII(53606) - HindIII(55945)
2120: pHB5-UL1HA-Kana: HindIII(129778) - HindIII(131898)
1471: pHB5-UL1HA-Kana: HindIII(122631) - HindIII(124102)
821: pHB5-UL1HA-Kana: HindIII(121810) - HindIII(122631)
387: pHB5-UL1HA-Kana: HindIII(158594) - HindIII(158981)

pHB5 WT

34948: pHB5: HindIII(217465) - HindIII(17244)
22813: pHB5: HindIII(85794) - HindIII(108607)
20593: pHB5: HindIII(157181) - HindIII(177774)
20343: pHB5: HindIII(65451) - HindIII(85794)
17570: pHB5: HindIII(177774) - HindIII(195344)
13288: pHB5: HindIII(38518) - HindIII(51806)
11403: pHB5: HindIII(108607) - HindIII(120010)
11328: pHB5: HindIII(195344) - HindIII(206672)
11306: pHB5: HindIII(54145) - HindIII(65451)
8742: pHB5: HindIII(29776) - HindIII(38518)
8678: pHB5: HindIII(17244) - HindIII(25922)
7541: pHB5: HindIII(130098) - HindIII(137639)
6652: pHB5: HindIII(143923) - HindIII(150575)
6284: pHB5: HindIII(137639) - HindIII(143923)
6219: pHB5: HindIII(150575) - HindIII(156794)
5676: pHB5: HindIII(122302) - HindIII(127978)
5424: pHB5: HindIII(212041) - HindIII(217465)
5369: pHB5: HindIII(206672) - HindIII(212041)
3854: pHB5: HindIII(25922) - HindIII(29776)
2339: pHB5: HindIII(51806) - HindIII(54145)
2120: pHB5: HindIII(127978) - HindIII(130098)
1471: pHB5: HindIII(120831) - HindIII(122302)
821: pHB5: HindIII(120010) - HindIII(120831)
387: pHB5: HindIII(156794) - HindIII(157181)

Figure 8. (A) *Insilico HindIII* digestion, compared between pHB5 wild-type and pHB5 UL1-HA the difference is a band seen at 6073bp. (B) The BAC-derived viruses digested with *HindIII* and separated using 0.8% agarose gel electrophoresis and visualized using ethidium bromide staining. The pattern of the restriction enzyme, *HindIII*, compared between pHB5 wild-type (kanamycin) and pHB5 UL1-HA.

Further proof of the correct replacement of *UL1* by the Kanamycin gene was obtained by PCR analysis. Amplification of DNA in TB40/E BAC parental or *UL1* deletion mutants using UL1-specific primers resulted in the expected fragment in TB40/E BAC whereas no amplified fragment was observed using TB40/E BAC ΔUL1 as template (Figure 9A). Control PCR reactions using primers specific for IE1 (*UL132*) were positive with both TB40/E BAC parental and mutant viruses (Figure 9B).

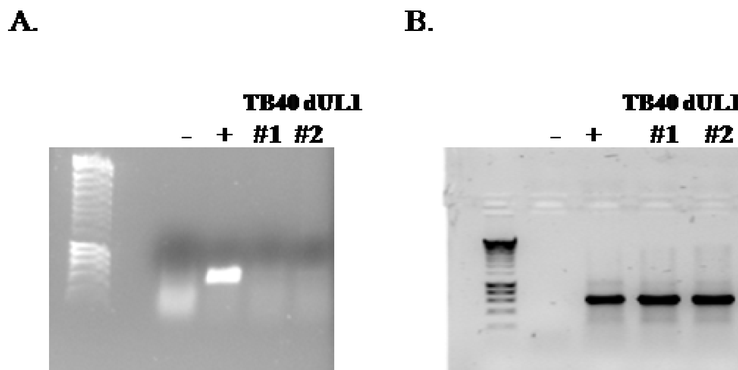


Figure 9. (A) PCR, checking if *UL1* is deleted in TB40/E BAC. (B) PCR, checking for viral DNA using IE1 primers. In (A-B) (-): a control reaction contains all primers and no template; (+): wild-type TB40/E BAC used as template.

Kan^R was excised from both viruses by FLP-mediated recombination, and the correct mutagenesis was confirmed by restriction pattern followed by southern blotting and PCR analysis (Figure 10).

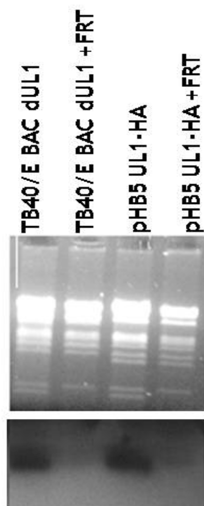


Figure 10. Upper panel shows *XbaI* digestion and FRT recombination and lower panes shows southern blot detecting Kanamycin cassette.

Recombinant HCMV was reconstituted from HCMV BAC mutants by the Superfect transfection method as described by Borst *et al,* 1999. Briefly infectious viruses were recovered from transfection of HCMV BAC DNA or BAC mutant into MRC-5 cells. MRC-5 cells were plated into 6 well tissue culture plates at 70-80% confluency a day before transfection. 1 µg of the BAC DNA, purified using Nucleobond PC 100 columns was added to 10 µl Superfect transfection reagents (QIAGEN, Hilden, Germany). Transfection mixture was removed 8 h post transfection and replaced with complete DMEM. Cells were spit and cultured until appearance of virus plaques.

Sequencing of the reconstructed viruses, confirming correct mutagenesis and that TB40/E ΔUL1 and parental virus have not acquired additional detrimental mutations during propagation in *UL128* locus was performed with DNA extracted from infected cells after reconstitution of the viruses. Extraction of viral DNA was carried out using the QIAamp DNA Mini Kit (Qiagen). The extracted viral genomic DNA was used as the template for amplification of *UL128* locus using primers previously described by Sun *et al.,* 2010. The purified PCR products were sequenced using the ABI Prism BigDye Terminator Cycle Sequencing kit (PE Applied Biosystems) and analyzed on an ABI 377 automated DNA sequencer.

3.8 UL1 RNA analysis

For Northern blot analysis, whole-cell RNA was isolated from mock-infected or infected MRC-5 cells at 8, 24, 48, and 72 h p.i. using QIAshredder columns and RNeasy Mini kit (Qiagen, Hilden, Germany). For gel electrophoresis, 10 μg of each sample was used and the preparations were separated in 1% agarose gels containing 2% formaldehyde. Northern blot analysis was carried out according to standard procedures. Hybridization and detection were performed using the DIG System (Roche) according to the manufacturer's instructions. Briefly, *UL1*-specific DNA hybridization probes were labelled with digoxigenin-dUTPs using DIG DNA labelling kit (Roche). Probes were prepared by PCR with *UL1* specific primers (5´-GCG AAC TTT GTG GAT GGA ACG G-3´ and 5´-ACG TTA TTT CTA GTT CGT AGT-3´), using pHB5 as a template. Late phase gene expression was inhibited by the use of phosphonoacetic acid (PAA; Sigma-Aldrich). 250 μg/m1 of PAA was added to the medium at the time of infection and maintained throughout infection.

For northern blot analysis of *UL1* transcripts in transiently transfected cells, whole-cell RNA was isolated from HEK293 cells transfected with pDispaly-$UL1_{Myc\text{-}HA}$ or pDispaly-$UL1_{HA}$. Non transfected HEK293 cells were used as negative control and RNA isolated form AD169 infected MRC-5 cells was used as positive control. RNA extraction and northern blot analysis was carried out as described above.

For the semi-quantitative analysis, checking putative splice sites (SS) in *UL1*, from mRNA transcripts from total RNA was extracted using the OneStep RT-PCR Kit (Qiagen). 5 μl from these samples were further analyzed in agarose gel electrophoresis. Semi-quantitative RT-PCR was performed using following AD169 UL1-specific primers: upstreamORF1 (forward) 5´-GAC TGC AAG GAC ACC TCC C-3´; upstreamORF2 (forward): 5´-GCT GTA TGG GCA GGC GTA GTG G-3´; ORFstart (forward): 5´-ATG GGC ATG CAA TGC AAC AC-3´; SS1 (reverse): 5´-GCT GGA ATA ATT ACA GGG ATT ATT GCT AGC-3´; downstreamSS3 (reverse): 5´-CGC AAT TAC CGT GGT ACC-3´; ORFend (reverse): 5´-TCA CGC TAT GCG ATT TAC G-3´. To rule out possible DNA contamination, a control

reaction without reverse transcriptase and containing RNA from all the preparations was carried out.

3.9 Western blot analyses and N-glycosidase F treatment of HCMV infected cells

Virus-infected cell proteins were obtained from AD169 UL1-HA infected MRC-5 cells grown in 6 well tissue culture dishes. Cell lysates of mock-infected MRC-5 cells grown in parallel with infected cells were used as controls. Cells were lysed under reducing conditions and heated to 100°C. Proteins were separated by SDS-PAGE (10-15%) and transferred to nitrocellulose membranes that were blocked with PBS containing 0.05% Tween-20 and 3% BSA (or 5% milk powder). Antibodies were diluted in PBS containing 0.05% Tween-20. For detection of primary antibody binding, horseradish peroxidase-conjugated goat anti-rabbit or anti-mouse IgG followed by enhanced chemiluminescence detection system (West Pico Supersignal kit, Pierce, Rockford, USA) was used according to the manufacturer's instructions. Removal of N-linked oligosaccharides was carried out using *N*-glycosidase F, recombinant (PNGase F, Roche) according to the manufacturer's specification. Briefly, the whole cell lyates were denatured in glycoprotein-denaturing buffer at 100°C for 10 min and cooled to 4°C for 5 min. The samples were then digested for 4 h at 37°C with PNGase F before being analyzed by immunoblotting.

3.10 Purification of HCMV particles and analysis of the virion protease-sensitivity

HCMV particles were further purified by negative-positive glycerol-tartrate gradient centrifugation as previously described (Talbot and Almeida, 1977). The isolated gradient-tartrate purified AD169 virions were treated with Proteinase K. Protease digestion was performed with 10 µg of Proteinase K (Roche) per ml for 20 min at 37°C. Some virion samples were solubilised with sodium dodecyl sulfate (SDS) added

to a final concentration of 0.5% (w/v). Virion preparations were subjected to SDS-PAGE and immunoblot analysis.

3.11 Immunoprecipitation of UL1 protein

MRC-5 were grown on p100, the monolayer was infected MOI 5 with AD169 UL1-HA or AD169 wild-type. Cells were lysed in lysis buffer (10 mM Tris, 150 mM NaCl pH 7.5, 1 mg/ml BSA, 1% Triton X-100 and protease inhibitor cocktail) for 30 min at 4°C and the cellular debris was removed from the lysates by high-speed centrifugation. The lysates were further processed for the IP as follows: the lysates were pre-cleared for 1 h at 4°C on a rotator with protein A Sepharose beads. Separately 50 µl of Protein A Sepharose beads were incubated with the rabbit anti-HA specific antibody or with isotype control antibody (rabbit IgG anti-NFAT5) for 2 h at 4°C. Protein A Sepharose beads and antibody were added to the pre-cleared lysate and incubated for 2 h at 4°C. The Sepharose pellet was washed and resuspended in 40 µl of 4x SB and incubated for 5 min at 95° C. The samples were cooled on ice and then centrifuged for 1 min at 13.000 rpm. The samples were loaded to SDS-PAGE for further analysis. After the confirmation of the IP of UL1-HA with rabbit anti-HA, the specific IP was immunoblotted with immune serum from individuals with HCMV infection, searching for UL1-specific antibodies.

3.12 Immunofluorescence microscopy

MRC-5 cells were grown in 24-well tissue culture plates containing a 13-mm-diameter coverslip. When cells had reached at least 90% confluence, they were infected with AD169 UL1-HA for 72 h. The coverslips were harvested by first washing the cells with PBS and then fixing them for 20 min at room temperature (RT) in paraformaldehyde (4%) freshly prepared in PBS. The coverslips were blocked with PBS containing 2% BSA for 20 min at RT and permeabilized with 0.5% Triton X-100 in PBS for 5 min. After washing with PBS, the coverslips were incubated for 1 h at RT with primary

antibody diluted in PBS containing 2% BSA, followed by 45 min incubation with fluorochrome-conjugated secondary antibody in PBS containing 2% BSA. Coverslips were mounted in Mowiol solution and analyzed under confocal laser scanning microscope (Confocal SP2, Leica), at a magnification of x40.

3.13 Viral growth kinetics

MRC-5, HEL299, RPE, and U373-MG cells seeded in 24-well plates were infected with TB40/E or TB40/E ΔUL1 at low moi (0.025 for MRC-5 cells, 0.1 for RPE cells, and 0.3 for U373-MG cells). After 2 h adsorption period, the inoculums were removed, infected monolayers were rinsed with PBS, and incubated with DMEM supplemented with 3% FBS. At various times post infection, the supernatants of the infected cells were harvested, cleared of cellular debris, and frozen at -70ºC. In the case of U373-MG cells, intracellular viral titers were measured. Quantification of infectious virus was determined by a fluorescence focus unit (FFU) assay counting HCMV IE1 positive cells as indicated before. The number of infected cells (fluorescence focus) was determined by microscopy.

3.14 Measurement of IE1 mRNA levels

RPE cells plated on 24-well plates were mock-infected or infected with TB40/E or TB40/E ΔUL1. Quantification of infectious virus was determined by a FFU assay counting HCMV IE1 positive cells as indicated before. During the adsorption phase TB40/E or TB40/E ΔUL1 were added at MOI 1 for 90 min at 4ºC in DMEM without carbonate, supplemented with 25mM Hepes. After adsorption period the inoculums were washed away with PBS and cells were incubated at 37ºC, 5% CO_2 with DMEM supplemented with 3% FBS. Whole cell RNA isolation was performed at 4 h p.i., cells using High pure RNA isolation kit (Roche). 200 ng of RNA quantified in a NanoDrop (ND-1000 Spectrophotometer) was retrotranscribed to cDNA using SuperScript III

reverse transcriptase and random primers (Invitrogen). For real-time quantitative polymerase chain reaction (RT-qPCR), Light Cycler 480 SYBR Green I Master (Roche) and a Roche LightCycler 480 detection system (Roche Molecular Biochemical) were used, according to the manufacturer´s instructions. All the samples were normalized to glyceraldehyde-3-phosphate dehydrogenase (GAPDH) mRNA levels and standard curves obtained with a serial 10-fold dilution of cDNA were generated to determine the level of each amplified transcript. Samples were analyzed in triplicate and the results were averaged. Primers used for quantifying the expression of GAPDH mRNAs were as follows: 5'-GCC ATC AAT GAC CCC TTC ATT-3' and 5'-TTG ACG GTG CCA TGG AAT TT-3'. To amplify a 217-bp product within the HCMV *IE1* (Kondo *et al.,* 1994) 5'-CAA TAC ACT TCA TCT CCT CGA AAG G-3' and 5'-CAA CGA GAA CCC CGA GAA AGA TGT C-3' were used. Amplification of an IE1 product of correct size was confirmed by agarose gele electrophoresis.

PART III RESULTS

4. THE UL1 PROTEIN IS A HCMV-SPECIFIC MEMBER OF THE RL11 FAMILY

4.1 Rapidly evolving and highly variable RL11 family

The primate CMV RL11 gene family, located on the extremity of the CMV genomes, that comprises 11 genes conserved in HCMV and chimpanzee CMV (CCMV) (i.e. *TRL/RL11-13, UL4-11*) and HCMV-specific genes (i.e. *UL1, RL5A* and *RL6*). In rhesus CMV (RhCMV) RL11 family is represented by putative glycoproteins *Rh19 (UL7*-homolog), *Rh20 (UL6*-homolog*), Rh17, Rh21-23 (UL11*-homologs*),* and *Rh25 (UL9)* (Hansen *et al.,* 2003). Genes in RL11 family encode a characteristic Ig-like domain type called RL11D, and most of them also contain a predicted transmembrane region, a signal peptide and glycosylation signals (Davison *et al.,* 2003a). RL11D has also been found in proteins encoded by adenovirus genomes. Many of these glycoproteins are also predicted to contain transmembrane region and a signal peptide (Figure 6). The conserved RL11D domain consists of a region of variable length (65–82 residues). There are several conserved residues including a tryptophan (W), two cysteines (C) and several potential N-glycosylation sites (Figure 11). The alignment of RL11 proteins also contains several conserved S and T residues, indicative of O-glycosylation (Chee *et al.*, 1990). The two disulphide linked cysteine residues characteristic of certain Ig containing proteins are not present in RL11D, but the two cysteine residues conserved in the central region of RL11D may form an alternative bridge.

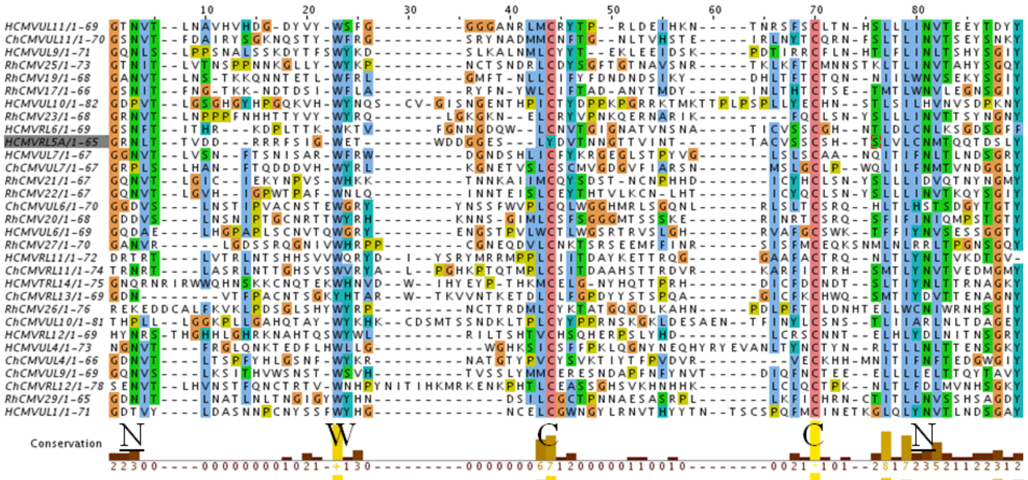


Figure 11. RL11D multiple sequence alignment. The conserve d RL11D domain consists of several conserved residues (W, two C) and several potential N-glycosylation sites (N). Sequences obtained from Genbank were aligned using ClustalW.

The RL11 gene family is a rapidly evolving virus gene family both in terms of gene members as well as in terms of sequence divergence across related viruses or strands (Dolan *et al.,* 2004). The estimated number of amino acid substitutions per site (K) between RL11 HCMV-CCMV orthologous is, on average 1.5, about 15 times higher than the average amino acid substitution rate of a prototypically well-conserved herpesvirus gene, UL54, encoding DNA polymerase (Table 3). The UL18 glycoprotein, which is an MHC class I homolog (Beck and Barrell, 1988), also shows elevated amino acid substitution rates.

Orthologues HCMV-CCMV RL11	K (aa subs/site)
TRL11	0.7
TRL12	1.6
TRL13	1.2
UL4	2.5
UL6	1.2
UL7	1.3
UL9	2
UL10	2.3
UL11	1
UL54	0.1
UL18	0.7

Table 3. Amino acid substitution per site (K) in orthologues HCMV-CCMV RL11 protein.

Primate CMV RL11 proteins can be divided into three clusters: RL5a-6 cluster consisting of *RL5A, RL6*; TRL cluster: *TRL11-14* and *UL1* and UL-cluster: *UL4-11* (Figure 12). *RL5A* and *RL6* are HCMV specific, encode for hort peptide sequences, mainly the RL11D domain with conserved potential N-linked glycosylation sites. These genes do not encode for signal peptide and transmembrane domain. Given the rapid evolution of the clusters, no cluster-specific motifs can be identified. Hypervariability has already been documented for several RL11 genes, including *RL6, TRL13* and *UL10* (Davison *et al.,* 2003a; Cha *et al.,* 1996; Hitomi *et al.,* 1997). The HCMV-specific genes, *UL1, RL5A* and *RL6,* have probably originated by gene duplication of other gene family members in the last 6-5 Mya, which are separating humans and chimps. These rapidly evolving RL11 genes, with high degree of variation putatively encode for glycoproteins that are secreted or membrane-associated, and therefore exposed to selection by the immune system. Proteins with such characteristics may have an important role in entry and egression of HCMV (Dolan *et al.,* 2004).

Figure 12. Multiple sequence alignments of the three gene clusters (RL5a-6; TRL cluster and UL-cluster) within RL11 family in HCMV and CCMV.

4.2 Members of RL11 family show sequence similarity to several cellular proteins

We investigated the similarity of the members of HCMV RL11 family Ig-like domain to the whole set of human Ig domains using BLAST searches against a complete collection of human Ig protein sequences (Figure 13). Hidden Markov Models (HMMs) corresponding to members of the Ig domain containing family were downloaded from PFAM (Finn *et al.,* 2008). They were used to search for Ig-domain containing human and HCMV proteins using the program HMMsearch (Eddy, 1998). Once the set of Ig-containing human and virus proteins had been obtained, the BLASTP program was used to extract highly similar human and HCMV proteins (Altschul *et al.,* 1990, E-value cutoff: 0.01). Pairwise sequence alignments of Ig-domain containing HCMV proteins and cellular proteins were generated with the help of the CLUSTALW program (Thompson et al., 1994).

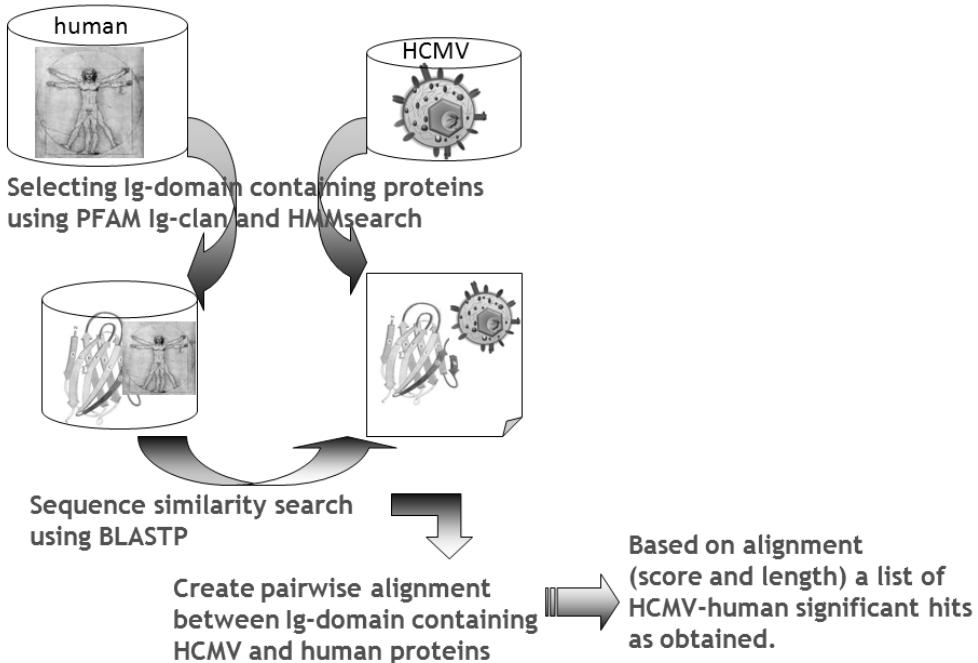


Figure 13. Searching for human and HCMV homologs: whole set of human and HCMV Ig domains obtained with HMMsearch program and BLAST searches.

As depicted in Figure 14 several RL11 proteins show high sequence similarity to cellular Ig-domain containing proteins (redundant hits to members of the same gene family are not shown in this Figure). High similarity may indicate that the viral protein is mimicking the host protein, as a way to interfere with its function (Elde and Malik, 2009). Scores above 22 are considered highly significant (E-value < 10^{-3}). The level of sequence similarity between members of RL11 family and cellular Ig-containing proteins is compared to the previously reported homology between MHC class I antigen and HCMV UL18 Ig-like domain. HCMV UL7 shows similarity with several cellular Ig containing proteins, such as the signaling lymphocyte-activation moleculer (SLAM) family. In contrast, TRL11, a known viral Fc gamma receptor, only shows weak sequence similarity with human Fc gamma receptors. HCMV UL1 had significant sequence similarity (score > 22) with several members of the CEA family. The genes that encode for RL11 proteins have probably derived from a single ancestral Ig-like gene, capture from the host genome, but the nature of this gene has been obscured by the long co-evolution with the host and the rapid evolution of RL11 genes.

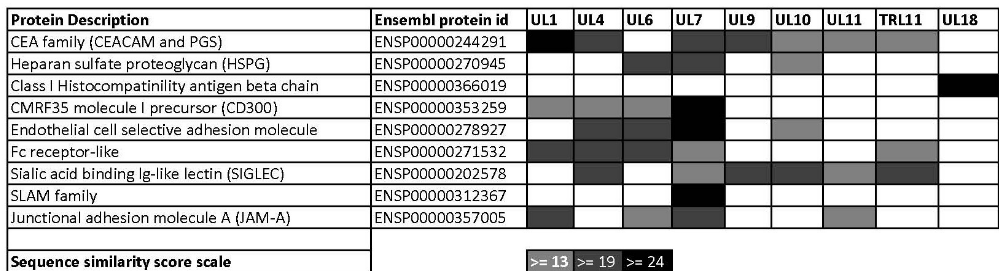


Figure 14. Human proteins showing high sequence similarity to HCMV RL11 proteins. Level of sequence similarity (score) is represented with different colours scales.

4.3 UL1 shows significant similarity to CEA proteins

The *UL1* gene is predicted to encode a 224-amino acid type I transmembrane glycoprotein, comprising a long signal peptide (amino acids 1 to 46), an Ig-like domain (amino acids 47 to 151) with 9 potential N-linked glycosylation sites, a transmembrane region (amino acids 184 to 206), and a short cytoplasmic tail (amino acids 207 to 224)

without apparent signaling motifs (Figure 15A). Since a sequence similarity of pUL1 to the pregnancy-specific glycoprotein 5 and other members of the human CEA family (Holzerlandt *et al.,* 2002) was previously noted, we sought to explore this finding in more detail. In fact, the region of significant sequence similarity was found to cover the Ig-like domain in pUL1 and the N-terminal IgV-like domain in the CEA family members (shown for CEACAM-1, -3, -5 and -6 in Figure 15B). As an example, the pUL1 Ig-like domain exhibits 61% amino acid similarity and 28% amino acid identity with the human N-terminal CEACAM-1 IgV domain (E-value = 0.0004) (Figure 15B). The alignment of the pUL1 Ig-like domain and the three-dimensional crystal structure of the soluble ecto-domain of the murine CEACAM-1a isoform (Tan *et al.*, 2004) indicated that the formation of Ig-like β-sheets is conserved and the conserved residues are mostly buried in the hydrophobic core, maintaining the Ig fold. CEACAM positions that appear to be critical for the formation of the typical β-sheets in IgV-like domains are present in pUL1 (Figure 15C). Furthermore, multiple alignments of the Ig-like domain of UL1 and the N-terminal IgV domain of several CEA proteins showed that a number of well-conserved residues in members of the CEA family are also present in pUL1, whereas they are absent from other RL11 proteins (Figure 15B and data not shown).

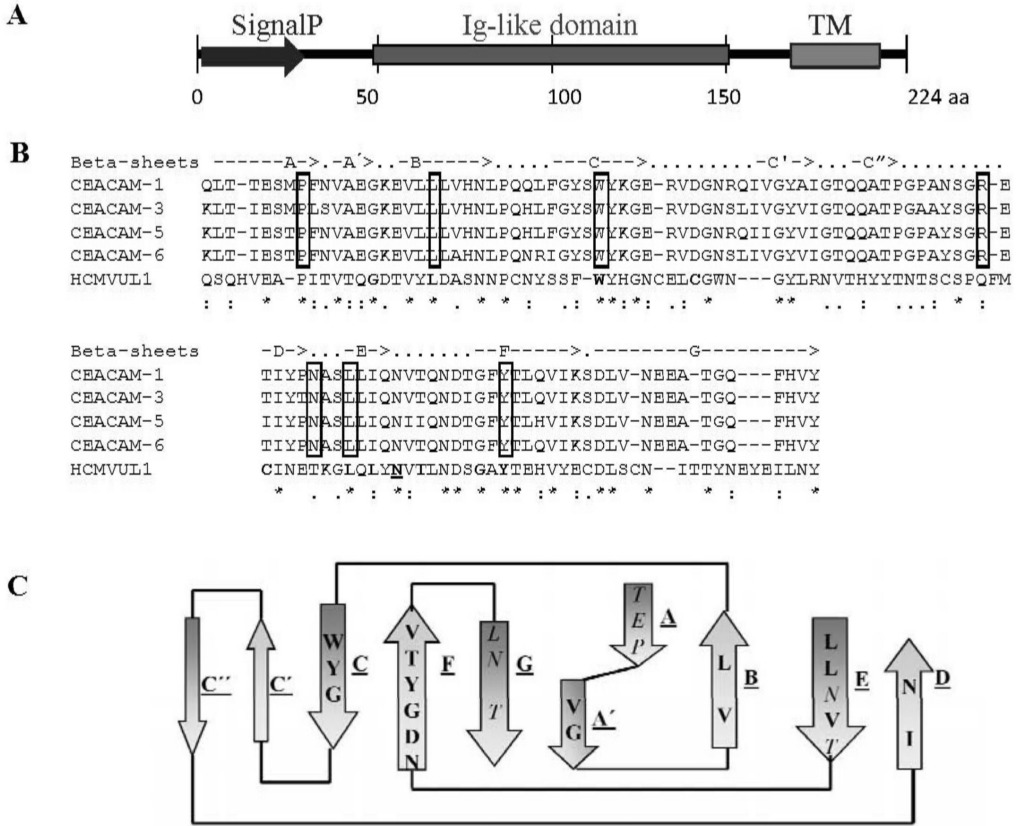


Figure 15. UL1 shares key conserved residues with CEA proteins. (A) Structural predictions of HCMV UL1. A schematic drawing of the 224-amino-acid UL1 protein is shown, with the signal peptide (signalP: amino acids 1 to 26), the Ig-like domain (Ig-like domain: amino acids 45 to 152), and the transmembrane region (TM: amino acids 184 to 206) indicated. (B) Multiple sequence alignment showing the sequence similarity between HCMV UL1 protein Ig-like domain (showing amino acid 45 to 152) and the N-terminal IgV domain of four representative members of the CEA family, human CEACAM-1, -3, -5, -6. Bold indicates highly conserved residues in RL11-family, and the potential *N*-glycosylated residue in the RL11-family is underlined (Davison *et al.,* 2003a). Highly conserved residues in CEACAM-1 and other IgSF members that appear to be critical for the formation of the β-sheets in IgV-like domains are boxed (Kammerer and Zimmermann, 2010). The arrows above the sequence alignment indicated the β-sheet structure of CEACAM-1a as determined by crystal structure analysis (Tan *et al.,* 2004). An asterisk (*) indicates identical amino acids, a colon (:) conserved amino acids, and a period (.) less conserved amino acids. (C) Alignment of the pUL1 RL11D domain and the three-dimensional crystal structure of murine CEACAM1a. Conserved residues are indicated on the β-sheets, bold indicates conserved buried residues and italic indicates conserved surface residues.

CEA family consists of the CEA-related cell adhesion molecules (CEACAMs) and the pregnancy specific glycoproteins (PSGs) subgroups. The level of sequence similarity between UL1 and CEACAMs is comparable to previously reported HCMV cellular homologs, such as the homology between MHC class I antigen and HCMV UL18 Ig-like domain (Figure 14). Glycoprotein UL18 shows homology to class I MHC heavy chain and interacts with the ILT2 (LIR-1) inhibitory receptor (Beck and Barrell, 1988; Cosman *et al.,* 1997). This is one of many mechanisms evolved by HCMV to interfere with the host immune response expressing molecules that interact with inhibitory immunoreceptors.

CEACAMs are broadly expressed immune receptors, including inhibitory and activating receptors, members of this family have been reported as critical modulators of several physiological processes (Kammerer and Zimmermann, 2010). CEACAMs are generally composed of one N-terminal Ig variable (IgV)-like domain (N domain), followed by a variable number of two different types of Ig constant domains. The extracellular domain of CEACAMs is involved in homotypic and heterotypic interactions (Hammarstrom, 1999). Interestingly, N-domains of CEACAMs have been found to be exploited by bacterial or viral pathogens in humans and mice, to which they adhere to and infect their target cells (Kuespert *et al.,* 2006) and most positions that are important for interaction between CEACAMs and their pathogen ligands are located in the highly conserved IgV domain (CFG face). PSGs, are mainly expressed in the placenta and are secreted into the maternal circulation. Trophoblast-specific PSGs are essential for a successful pregnancy and are thought to regulate maternal immune system responses (Hau *et al.,* 1985; Waterhouse *et al.*, 2002). Therefore, similarity of UL1 to PSGs could be related to HCMV pathogenicity during pregnancy.

4.4 The UL1 protein is a HCMV-specific member of the RL11 family

Comparison of the available amino acid sequences in a number of HCMV strains deposited in GenBank, including some unpassaged clinical isolates (3301, W, AF1, JP, U8, and U11), low-passage clinical isolates (Toledo, Merlin, 3157, 6397, HAN13, HAN20 and HAN38) and two BAC-cloned isolates (TR and PH), endotheliotropic strain TB40/E, and laboratory-adapted strains (AD169, Towne) (Figure 16). Pairwise alignments show that the UL1 sequences among different strains share an identity of 52–99,5%. Moreover, intra-strain studies show that UL1 has the highest degree of variation in the Ig-like domain compared with the other RL11 family Ig-like domains (Sekulin *et al.,* 2007) and comparation of UL1 coding sequences from different HCMV strains show putative proteins between 215-224 amino-acids long.
Moreover, comparing UL1 sequences from different strains, we observed that due to an insertion the prediction of a UL1 signal peptide (predicted for several strains such as Merlin and HAN13) was remarkably reduced in AD169, 6397 and 3301.

UL1 is highly divergent between different viral strains (showing high K-value, a measure of nonsynonymous nucleotide substitution, Dolan *et al.*, 2004). It has been shown that several clinical isolates (4 of 31) contain in-frame mutations in *UL1* (a STOP codon at amino acid position 76, corresponding to an N) and in the fibroblast-adapted Towne strain UL1 is truncated due to a deletion (Sekulin *et al.,* 2007). Towne strain is a high-passaged strain and is known to have both lost virulence and experienced substantial genome alternation during cell culture passage, and comparing the sequence of UL1 Towne with the low-passaged 6397 clinical isolate we observed that they were identical except for the transmembrane domain missing in UL1 Towne due to a four nucleotide in-frame deletion, both Towne and 6397 were originally isolated from the urine of a congenitally infected infant. The wild-type form of Towne UL1 was probably lost during cell culture passage. Another observation made was that

due to an insertion the prediction of a signal peptide in UL1, compared to wild type Merlin, was greatly reduced in several strains (e.g. AD169, 6397 and 3301). The fact the UL1 is highly variable between different viral strains, may imply that HCMV can tolerate significant amino acid variation of this protein.

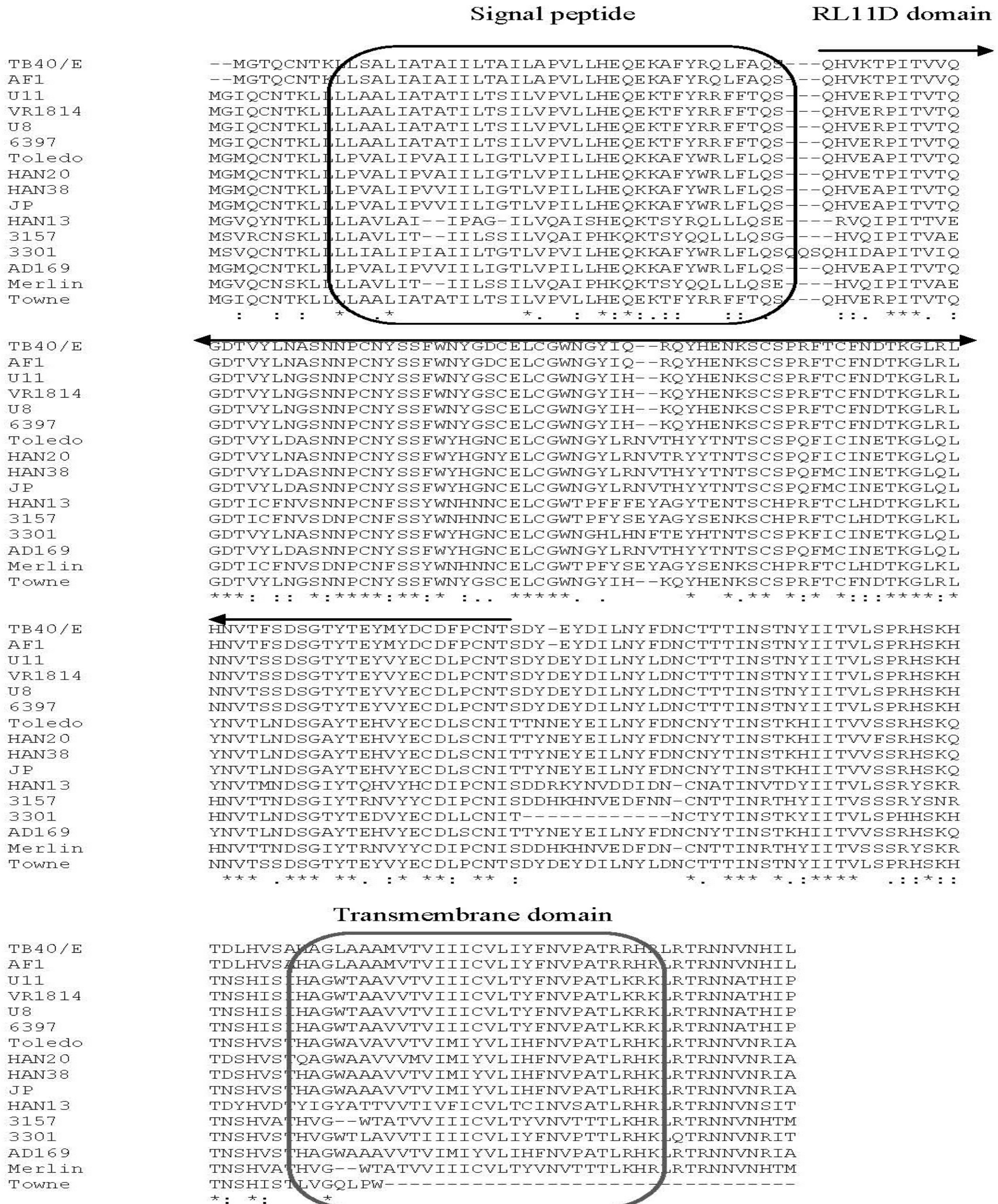


Figure 16. Alignment of UL1 sequence variation of the available amino acid sequences in a number of HCMV strains deposited in GenBank, including some unpassaged clinical isolates,

low-passage clinical isolates, laboratory-adapted strains and endotheliotropic strain TB40/E show an overall low sequence similarity. The putative signal peptide and membrane anchor sequence of UL1 polypeptides are indicated by boxes and the Ig-like domain by arrows.

4.5 Sequence similarity between UL1 and other HCMV RL11 family members

To better understand the process of formation and divergence of UL1 we examined the amino acid substitutions per site (K) between UL1 and the rest of RL11 gene family members (Table 4). K values were calculated using PHYLIP program Protdist with default parameters (Felsenstein, 1989). The lowest K values are for the alignments of UL1 with TRL11, TRL12 and TRL13 (1.9 in the three cases), indicating that UL1 is evolutionary closer to the TRL-cluster (*TRL11, TRL12* and *TRL13*) than to other RL11 gene family members. This, together with the fact that is located just next to *TRL13*, suggests that *UL1* may have arisen via gene duplication event from a gene from the TRL-cluster. The lack of a counterpart of *UL1* in CCMV indicates that this probably took place after the speciation of human and chimp. *UL1* appears to have diverged very quickly, although *UL1* is a young RL11 copy, the K values with its closest homologues (other genes in the TRL cluster) are higher than those typically observed between orthologous HCMV and CCMV genes.

Next, using the PAML program *codeml* (Yang, 1997), we estimated the non-synonymous (Ka, amino acid altering) and synonymous (Ks, silent) substitution rates between UL1 and the rest of RL11 genes. However, the overall low degree of sequence conservation among the RL11 family makes evolutionary analysis of this family difficult and in general the Ka and Ks values were too saturated (Ka > 2 and/or Ks > 2) to be reliable. Exceptions were UL1 and TRL11, which showed Ka=1,51 and Ks=1,21. Interestingly, Ka/Ks > 1, is an indication of positive selection, that amino acid changing mutations have accumulated more rapidly than expected by genetic drift alone, indicating that some amino acid replacements have been fixed in the virus population due to their beneficial effects. For example, such replacement may have helped the virus escape the host immune system.

Paralogues UL1-HCMV RL11	K (aa subs/site)
RL5a	3.5
RL6	3.3
TRL11	1.9
TRL12	1.9
TRL13	1.9
UL4	2.8
UL6	3.1
UL7	2.3
UL9	2.6
UL10	2
UL11	2.7

Table 4. Amino acid substitution per site (K) in paralogues UL1-HCMV RL11 protein.

5. RECOMBINANT EXPRESSION OF HCMV UL1

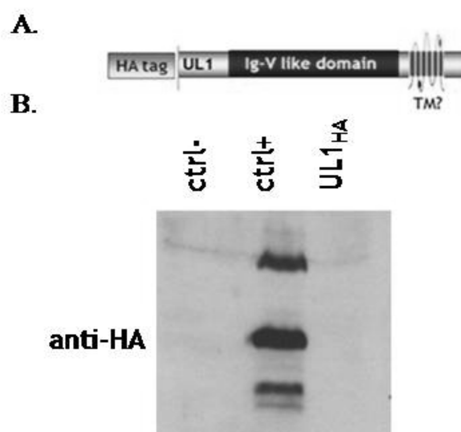
UL1 is a member of HCMV RL11 family of glycoproteins and many of these proteins are also predicted to encode for a C-terminal transmembrane region (TM) and a signal peptide (Davison *et al.,* 2003a). In order to obtain information of the structure, intracellular distribution and trafficking of UL1 protein in absence of HCMV infection, the coding sequence from HCMV AD169 UL1 was inserted into mammalian expression vector pDisplay. Following transient transfection of UL1 plasmids into HEK293 cells or COS-7 cells, UL1 protein expression was analyzed by immunoblotting and flow cytometry analysis.

UL1 shares a striking similarity to the N-terminal IgV domain cellular CEA family molecules that are involved in homotypic and heterotypic adhesion with each other. CEACAM subfamily includes inhibitory and activating receptors widely expressed in different cell types, and members of this family have been reported as critical modulators of several physiological processes. Multiple alignments showed that a number of well-conserved residues in several members of the CEA family are also present in UL1, CEACAM positions that appear to be critical for the formation of the typical β-sheets in Ig-like domains are also conserved. Based on the sequence similarity of UL1 Ig-like domain to the N-terminal domain of CEAs, we hypothesized that UL1 could potentially interact with members of CEA family. For this purpose, the expression and production of UL1 as a secretary fusion protein was undertaken by inserting the sequence corresponding to UL1 Ig-like domain (from AD169 or TB40) into the mammalian expression vectors pSecTag (UL1 fused to mouse IgG_{2a} Fc) and pCR3 (UL1 fused to human IgG_1 Fc).

5.1 UL1 sequence fused to HA and Myc tags

The *UL1* gene is predicted to encode a 224-amino acid type I transmembrane glycoprotein, comprising a long signal peptide (amino acids 1 to 46), an Ig-like domain (amino acids 47 to 151) with 9 potential N-linked glycosylation sites, a transmembrane region (amino acids 184 to 206), and a short cytoplasmic tail (amino acids 207 to 224) without apparent signaling motifs. The complete sequence from HCMV AD169 *UL1* was inserted into mammalian expression vector pDisplay. Prior to the cloning of *UL1* into pDisplay, an additional Myc- epitope and a stop codon (between Myc-tag and PDGFR TM domain) were introduced into the vector, thereafter *UL1* was fused to C-terminal Myc-tags and N-terminal HA-tag epitopes. Following transient transfection with $UL1_{Myc\text{-}HA}$ cDNA in HEK293 cells or COS-7 cells, the tagged UL1 protein was analyzed by immunoblots and flow cytometry analysis. Expression of the $UL1_{Myc\text{-}HA}$ protein could not be detected by immunoblotting using Myc-specific antibodies neither with HA-specific antibodies (Data not shown). The protein was also undetectable by intracellular and cell surface staining with anti-HA and anti-Myc antibodies.

Moreover, UL1 without the signal peptide was then excised and inserted in-frame with the HA at the N-terminal end of the UL1 protein into the mammalian expression vector pDisplay ($UL1_{HA}$ cDNA, Figure 14A). Expression of the $UL1_{HA}$ was undetectable by immunoblotting (Figure 17B), and by intracellular and cell surface staining (Figure 17C) using anti-HA.



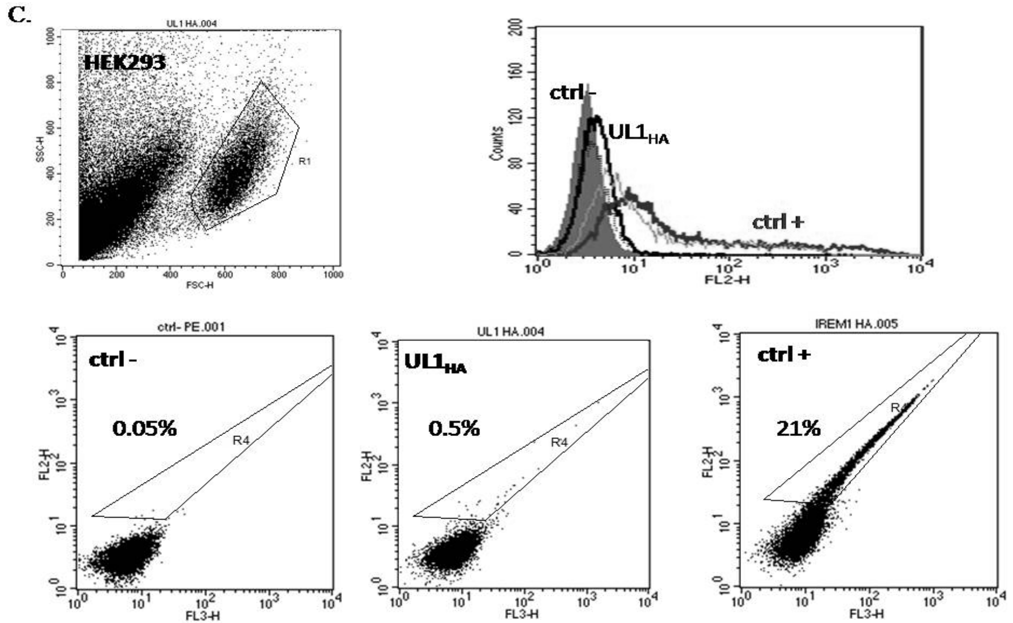


Figure 17. (A) Model of UL1 protein fused to N-terminal HA-tag and two C-terminal Myc-tags. (B) Immunoblotting of cell lysate of HEK293 cells transfected with pDisplay-UL1$_{HA}$ and control plasmids, using anti-HA. (C) Cell surface staining of HEK293 cells with anti-HA.

Searching for an explanation for the lack of UL1 expression using mammalian expression vectors, the presence of *UL1* transcripts in transiently transfected cells was examined. Whole cell RNA was extracted from HEK293 cells transfected with UL1$_{Myc\text{-}HA}$ or UL1$_{HA}$ cDNA. MRC-5 cells infected with HCMV AD169 strain were used as a control of *UL1* expression and non-transfected HEK293 were used as negative control. The RNA was subjected to northern blot analysis using *UL1* (AD169 strain) specific sequence as a probe. As depicted in figure 18, *UL1* transcript from transiently transfected cells could not be detected using northern blotting analysis.

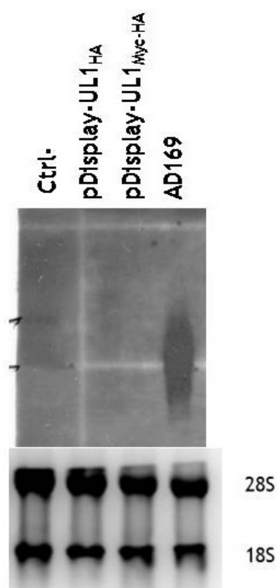


Figure 18. Northern blot with UL1-specific sequence probe. RNA was extracted from HEK293 cells transfected with pDispaly-UL1$_{myc\text{-}HA}$, pDispaly-UL1$_{HA}$; non transfected HEK293 cells (ctrl-) and AD169 infected MRC-5 cells (ctrl+). The lower panel shows ethidium bromide staining of 28S (4.8-Kbp) and 18S (1.8-Kbp) ribosomal RNA species used as loading control.

5.2 Expression and production of UL1-Fc fusion proteins

In general, CEA family receptores possess a leader sequence, one N-terminal IgV-like domain, and a variable number of two different types of Ig constant (IgC)-like domains. Members of the family are broadly divided into CEACAM and PSG subgroups, involved in immune responses and reproduction. Primarily, CEACAMs modulate immune responses by establishing both homophilic (for example, CEACAM1 and CEACAM3) and heterophilic (for example, CEACAM1- CEACAM5 and CEACAM6-CEACAM8) interactions with other members of the same family through their highly conserved N-terminal IgV domains. To explore the possibility of interactions between UL1 and members of CEA family, the expression and production of UL1 as a secretory fusion protein was undertaken by inserting the sequence corresponding to UL1 Ig-like domain (from AD169 or TB40) into the mammalian expression vectors pSecTag (UL1 fused to Fc region of mouse IgG2a) and pCR3 (UL1

fused to the Fc region of human IgG1). UL1-Fc fusion protein was also constructed for the production of the anti-UL1 monoclonal antibodies.

Prior to testing the binding ability of UL1-Fc the presence of the soulble protein (pSecTag-UL1$_{Fc}$ and pCR3-UL1$_{Fc}$) in the supernatant (purified or non-purified) of transfected cells was tested by enzyme-linked immunosorbent assay (ELISA), immunoprecipitation (IP) and immunoblot analysis (Figure 19). The soluble molecules were purified by affinity chromatography with protein A sepharose beads. The supernatant from pSecTag-UL1$_{Fc}$ and control plasmids (ILT4 or ILT5 cloned into pSecTag expression vector) transiently transfected cells, were tested by ELISA detecting mouse IgG (Figure 19A). Mouse IgG fusion proteins were detected in the supernatant of cells transfected with control plasmids (ILT4-Fc and ILT5-Fc cDNA), but mouse IgG was undetectable in the supernatant from UL1-Fc cDNA transiently transfected cells.

IP analysis with purified supernatant from UL1-Fc cDNA transfected HEK293 cell was performed (Figure 16B). The control fusion protein ILT4-Fc produced in a dimeric form was detected as expected under both non-reducing (NR) and reducing (R) conditions. UL1 protein has eight putative N-glycosylation sites and the excepted MW for UL1-Fc is ~ 120-kDa under non-reducing conditions and 60-kDa under reducing conditions (Fc: 30-kDa and UL1 IgV like domain ~30-kDa). UL1-Fc protein was undectable by IP. Further, IP analysis with purified supernatant from pCR3-UL1-Fc cDNA (UL1 fused to the Fc region of human IgG1) transfected cells was performed. pCR3-Fc (empty vector) was used as transfection and expression control. In an IP with supernatant from pCR3-Fc transfected cells, the Fc protein was detected as expected at 30-kDa, but UL1-Fc fusion protein was undetectable (Figure 19C).

The secretory UL1-Fc fusion protein was primarily aimed to be tested for its binding capability to cells transiently transfected with constructs expressing different members of the CEACAM family (e.g. COS-7 transiently transfected with CEACAM-3/CEACAM-6; Hela transiently transfected with CEACAM1-4L; AsPC-1 cells constitutively expressing CEACAM1-4L). UL1-Fc proteins were also constructed for

the generation of UL1-specific monoclonal antibodies. Briefly, we were planning to generate the specific antibodies by fusing the murine myeloma cell line NS-1 with spleen cells of Balb/c mouse immunized with UL1-Fc recombinant proteins. Taken together, UL1-Fc proteins could not be generated using different mammalian expression vectors, insert origin (UL1 sequence from either AD169 or TB40), transfection and detection methods.

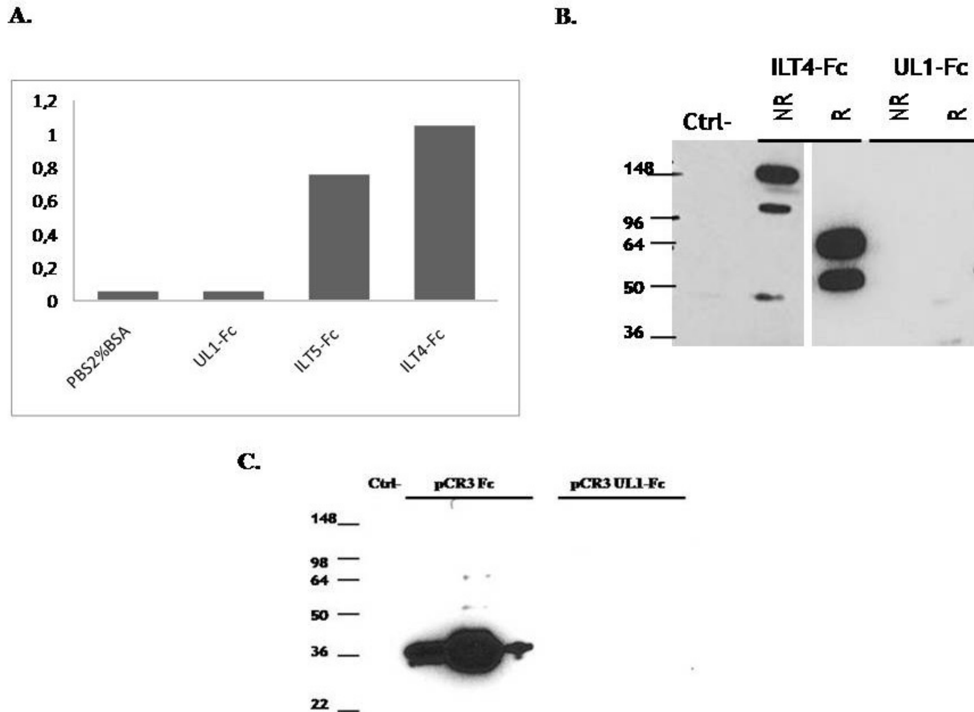


Figure 19. (A) ELISA detecting mouse IgG comparing the optical density values of the supernatant from cells transfected with UL1-Fc cDNA or ILT4-Fc/ILT5-Fc cDNA (control plasmids). PBS 2%BSA was used as negative control. (B) IP of the supernatant from UL1-Fc cDNA or ILT4-Fc cDNA (control plasmid) transiently transfected HEK293 cell. Proteins were solubilized under reducing (R) and non-reducing conditions (NR). The negative control (ctrl-) corresponds to the supernatant of non transfected cells. (C) IP of supernatant from UL1-Fc cDNA (pCR3-UL1$_{Fc}$) or empty vector (pCR3$_{Fc}$) transfected cells. The negative control (ctrl-) corresponds to the supernatant of non transfected cells.

6. UL1 PROTEIN HCMV INFECTED CELLS AND VIRIONS

In the present study, we have approached the characterization of the HCMV specific *UL1*. To this end a HCMV (AD169-derived HB5 background) recombinant with an HA-epitope tagged UL1 and a mutant with a full *UL1* deletion in the endotheliotropic HCMV TB40/E strain were generated.

6.1 UL1 is transcribed with a late gene kinetics

In the HCMV AD169 genome, *UL1* expands from nucleotide 11835 to 12509 (675 base pairs). To analyze transcripts and the kinetics of the RNA originating from *UL1*, total RNA was isolated from MRC-5 cells at different times after infection with HCMV AD169 and subjected to Northern blot analysis using a *UL1*-specific probe. Large *UL1* transcripts in AD169-infected cells could be detected at 48-72 hours post infection (h p.i., Figure 20A).

HCMV lytic gene expression is conventionally divided into three major kinetic classes of viral genes, immediate-early (IE), early (E), and late (L). The replication cycle starts with the IE gene transcription within the first hours p.i. of permissive cells. The IE proteins are critical for the subsequent expression of E and L genes, and together with E proteins they provide changes of host cell environment optimal for viral gene expression and DNA replication. E gene products are insensitive to inhibitors of viral DNA synthesis (e.g. phosphonoacetic acid and ganciclovir) and are necessary for several processes such as viral DNA replication, packaging of the viral genome, and assembly of the viral particles. Viral DNA synthesis begins 18-24 h p.i. and is followed by the transcription of L genes which typically encode the structural components of the HCMV virion, such as *UL75* and *UL99*.

Transcription of *UL1* RNA in AD169-infected cells was completely blocked by the viral DNA polymerase inhibitor phosphonoacetic acid (PAA), consistent with the classification of *UL1* as a true late phase gene, as previously described and

reproduced here for *UL99* (Figure 20B, Kohler *et al.,* 1994). The presence of *UL1* transcripts was studied in MRC-5 cells infected with a panel of HCMV strains, including laboratory strains AD169, Toledo and the endotheliotropic strain TB40/E. Total RNA from cells infected with HCMV TB40/E, AD169, Toledo and Towne were isolated 72 h p.i. and analyzed by Northern blot using a *UL1* (AD169)-specific probe. Towne strain, previously reported to be *UL1* defective due to a deletion in this ORF (Sekulin *et al.*, 2007) was used as negative control. *UL1* transcripts were detected in cells infected by HCMV TB40/E, AD169 and Toledo, the main transcript sizes detected at 72 h p.i. and compared with 28S and 18S ribosomal RNA species are approximately 5.8-Kbp and 3.4-Kbp (Figure 20C). The signal intensity was relatively lower in Toledo-infected cells and *UL1* transcripts were, as expected, undetectable upon infection with Towne.

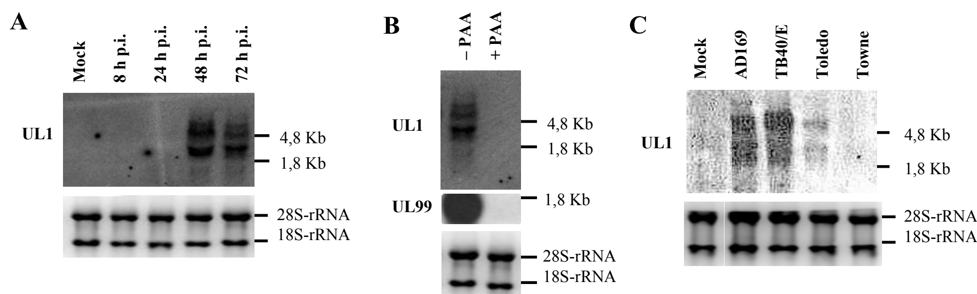
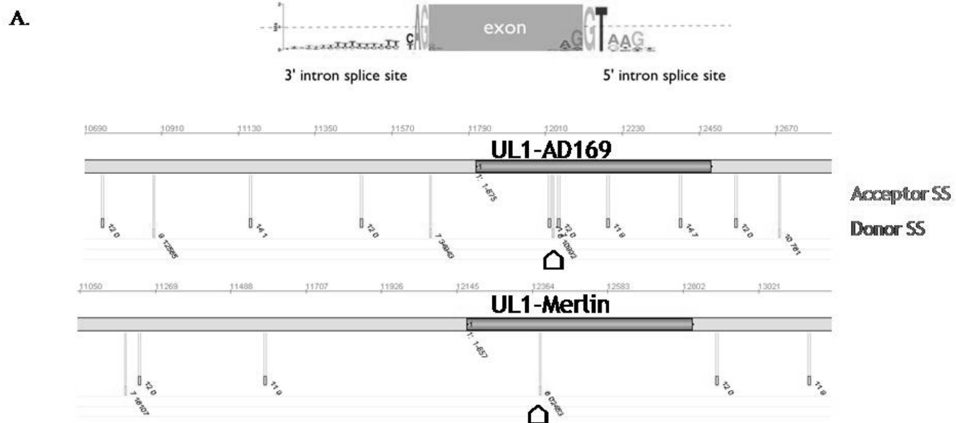


Figure 20. Kinetics of UL1 gene expression. (A) Whole-cell RNA was isolated, from uninfected (mock) MRC-5 cells or cells infected at an MOI of 2 with HCMV AD169, at the indicated hours post infection (h p.i.). Equivalent amounts of RNA were subjected to agarose gel electrophoresis and subsequently analyzed by northern blot using a HCMV-UL1-specific probe. (B) UL1 transcription was inhibited by the use of phosphonoacetic acid (+PAA). Whole-cell RNA was isolated from HCMV AD169 infected cells at 72 h p.i. incubated with or without PAA. In addition to the UL1-specific probe, RNA was hybridized with a probe specific for the late HCMV gene *UL99*. (C) UL1 transcripts were analyzed in MRC-5 cells infected with different HCMV strains (AD169, Toledo, TB40/E and Towne) for 72 h, as described earlier. In (A-C) the lower panel shows ethidium bromide staining of 28S (4.8-Kbp) and 18S (1.8-Kbp) ribosomal RNA species.

UL1 is annotated at 675 bp, but larger transcripts, of approximately 5.8-Kbp and 3.8-Kbp were detected, different RNA bands that are observed may be due to

splicing events in *UL1*. Putative 5'-donor and 3'-acceptor splice sites in *UL1* (AD169 and Merlin) were predicted using *in-silico* splice site prediction tool MEM (based on maximum entropy modeling, developed by C.B. Burge, 2004). A conserved putative acceptor splice site was found in *UL1* (Figure 21). RT-PCR using a combination of UL1-specific primer upstream and downstream of the putative acceptor splice site was performed. To rule out possible DNA contamination, a control reaction without reverse transcriptase and containing RNA was carried out. The presence of conserved putative splice sites in *UL1* could not be confirmed and the complete *UL1* could be detected by RT-PCR (Figure 21) and its nucleotide sequence as annotated was confirmed by sequencing.

A.



B.

Conserved acceptor splice site in UL1 ORF:

```
TB40/E        GTTTACCTTAACGCTAGTAATAACCCCTGCAACTATTCCAGCTTCTGGAACTATGGCGAT
Towne         GTTTACCTGAACGGTAGTAATAATCCCTGCAACTATTCCAGTTTCTGGAACTACGGCAGT
Toledo        GTCTACCTAGACGCTAGCAATAATCCCTGTAATTATTCCAGCTTTTGGTACCACGGTAAT
AD169         GTCTACCTAGATGCTAGCAATAATCCCTGTAATTATTCCAGCTTTTGGTACCACGGTAAT
Strain-3157   ATTTGCTTTAACGTTAGTGATAACCCCTGCAACTTTTCTAGTTACTGGAATCACAATAAC
Merlin        ATTTGCTTTAACGTTAGTGATAACCCCTGCAACTTTTCTAGTTACTGGAATCACAATAAC
                 *  *  *  *   *  *  ***   ****  *****  **  *  ***  **  *   ***  *   *
```

C.

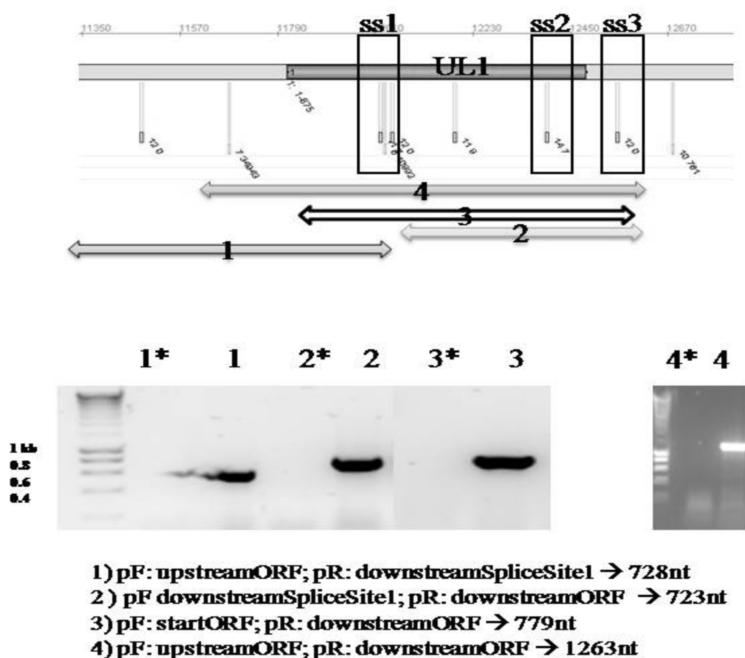


Figure 21. (A) Putative donor and acceptor splice sites in *UL1* (AD169 and Merlin strain), conserved acceptor splice site is indicated by ⌂. (B) Multiple sequence alignment of *UL1* from different HCMV strains, showing the sequence conservation in the putative acceptor splice site. (C) RT-PCR, Checking for N- and C-terminal putative splice sites in *UL1* using different combinations of UL1-specific primer pairs (1, 2, 3 and 4). 72h p.i. AD169 RNA (DNAase treated), * indicates control RT-PCR performed without cDNA synthesis.

6.2 Expression of the UL1 protein in HCMV infected cells

The 224 amino acid long UL1 (Mr 25.5-kDa) polypeptide is predicted to be expressed as a type I glycoprotein with a signal peptide and a membrane anchor. We investigated the expression kinetics of the protein product of *UL1* during productive HCMV infection. For this purpose, we generated a HCMV (AD169-derived HB5 background) recombinant with an HA tag fused to the *UL1* C-terminal end. The reconstituted AD169 UL1-HA was used to infect MRC-5 cells and the kinetics of pUL1 expression was followed by immunoblot analysis using an anti-HA antibody in cell extracts harvested at different time points ranging from 4 to 72 h p.i. HCMV-specific mAbs reactive with the major immediate protein IE1 pp72 (*UL123*) and the pp28 (*UL99*)

late phosphoprotein were used as controls. As shown in Figure 22A, the anti-HA antibody visualized two protein species, a faster migrating non-specific band which was present in both HCMV-infected and mock-infected MRC-5 cells, and a PAA-sensitive band of ~ 55-kDa protein which was weakly detectable at 48 h p.i. but was abundantly found at 72 h p.i. This 55-kDa HA-tagged UL1 band was absent in immunoblots of AD169 wild-type infected cells (data not shown) further confirming its specific detection. The 55-kDa band was first detected at 48 h p.i. and showed maximal levels at 72 h p.i., coincident with the expression kinetics of pp28 (Kohler *et al.*, 1994, Figure 22A). At time points after infection when signals for pUL1 were not observed, IE1 pp72 was present as expected, being already expressed after 4 h p.i and remaining throughout the duration of the HCMV replication cycle. Consistent with the observed kinetic pattern and the results obtained at the transcriptional level, metabolic blockade with PAA resulted in the disappearance of the 55-kDa band (Figure 22B) indicating that the UL1 protein is expressed late during HCMV replication.

Many of the HCMV structural glycoproteins form high molecular weight disulfide-linked oligomers, such as the gCI complex composed of homodimers of gB (Britt *et al.*, 1984), therefore we analyzed UL1 protein migration pattern under reducing (with β-mercaptoethanol) and non-reducing conditions. pUL1 protein was β-mercaptoethanol resistant implying that it is not part of a disulfide-linked oligomer in infected MRC-5 cells (data not shown). The high number of putative N-linked glycan sites (n=9) present in pUL1, the apparent molecular weight (55-kDa) observed for this protein by immunoblotting and the calculated molecular mass (25.5-kDa) of the 224 aa polypeptide backbone suggested that pUL1 undergoes an extensive posttranslational glycosylation process. Indeed, digestion with PNGase F, that removes complex and high-mannose N-linked sugars, enhanced the migration of the specific band from 55-kDa band to 26-kDa (Figure 22C). These results confirmed that pUL1 is expressed as a glycoprotein.

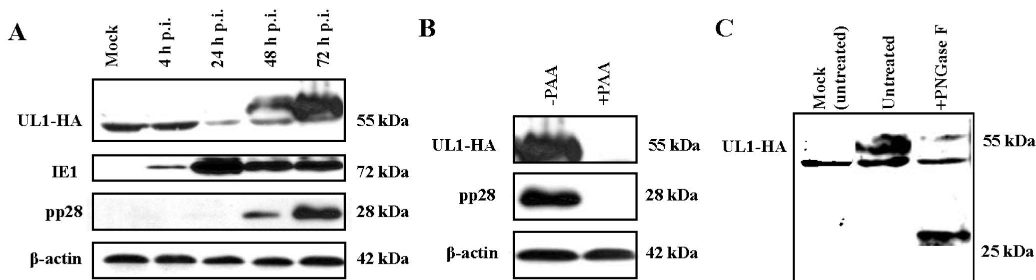


Figure 22. Expression of the UL1 glycoprotein in infected cells. (A) Uninfected (mock) and AD169 infected cells at an MOI of 5 were harvested at the indicated times (h p.i.). Equivalent amounts of cell lysates were subjected to SDS-PAGE and analyzed by immunoblotting with antibodies reactive with HA or HCMV IE1 and pp28 proteins. β-actin detection was used as protein loading control. (B) Experiments were performed as in (A) assessing the effect of PAA treatment (+PAA) on late phase protein expression. (C) Whole cell lysates of AD169 UL1-HA infected MRC-5 cells (72 h p.i.) were left untreated or digested with PNGase F. The preparations were subjected to immunoblot analysis using an anti-HA antibody.

6.3 Subcellular localization of pUL1 in HCMV infected cells

To analyze the localization of pUL1 in HCMV infected cells, the HA epitope of the AD169 UL1-HA virus was employed to trace its subcellular distribution by confocal microscopy when comparing it with a panel of molecular markers of the exocytic and endocytic pathways. In MRC-5 cells following fixation at 72 h p.i. UL1-HA accumulates in a large cytoplasmic compartment (Figure 24). This compartment was not detectable in uninfected and AD169 wild-type infected MRC-5 cells, thus confirming the specificity of the observated staining pattern (Figure 23). UL1-containing structures failed to co-localize with several intracellular markers, including the endoplasmatic reticulum (ER) marker Calnexin (Figure 24A) and ERGIC-53 (Figure 24B), an integral protein localized in the ER-Golgi intermediate compartment (ERGIC). GM-130 staining, a *cis*-Golgi marker appeared in a ring structure in the proximity of UL1 but did not either overlap (Figure 24C). As observed by other investigators, alterations in the morphology of Golgi and ERGIC could be seen during HCMV infection (Sanchez *et al.*, 2000).

Furthermore, UL1 co-localization with the lysosomal protein LAMP-1 was also ruled out (data not shown). It has been shown that HCMV during its final

envelopment acquires membranes containing endosomal and *trans*-Golgi network markers, while lysosomes, *cis*- and *medial*-Golgi are not present and surround the virus envelopment and assembly site (Cepeda *et al.,* 2009). Markers for the endocytic pathway such as the early endosomal marker, EEA-1, and recycling endosomes (Transferrin receptor) did not co-localize with UL1. Yet, a partial co-localization of UL1-containing structures and the *trans*-Golgi marker TGN-46 (Figure 24D) was observed indicating that the protein undergoes post translational modifications.

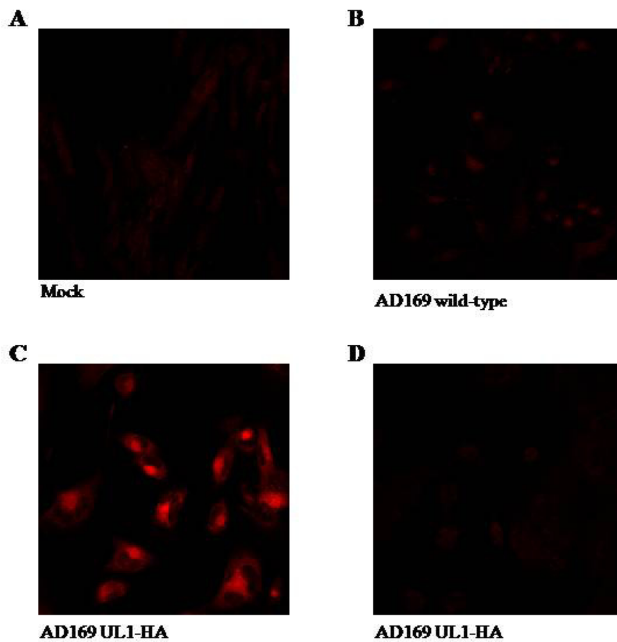


Figure 23. MRC-5 cells grown on glass coverslips were infected with AD169 wild-type (B) or AD169 UL1-HA (C and D) at MOI 5 and fixed at 72 h p.i. Uninfected (mock) MRC-5 cells (A) were processed similarly. Cells were stained with anti-HA antibody. (D) AD169 UL1-HA cells were stained with isotype control. In (A-D) primary anti-body was detected with goat anti-rabbit IgG Alexa 555.

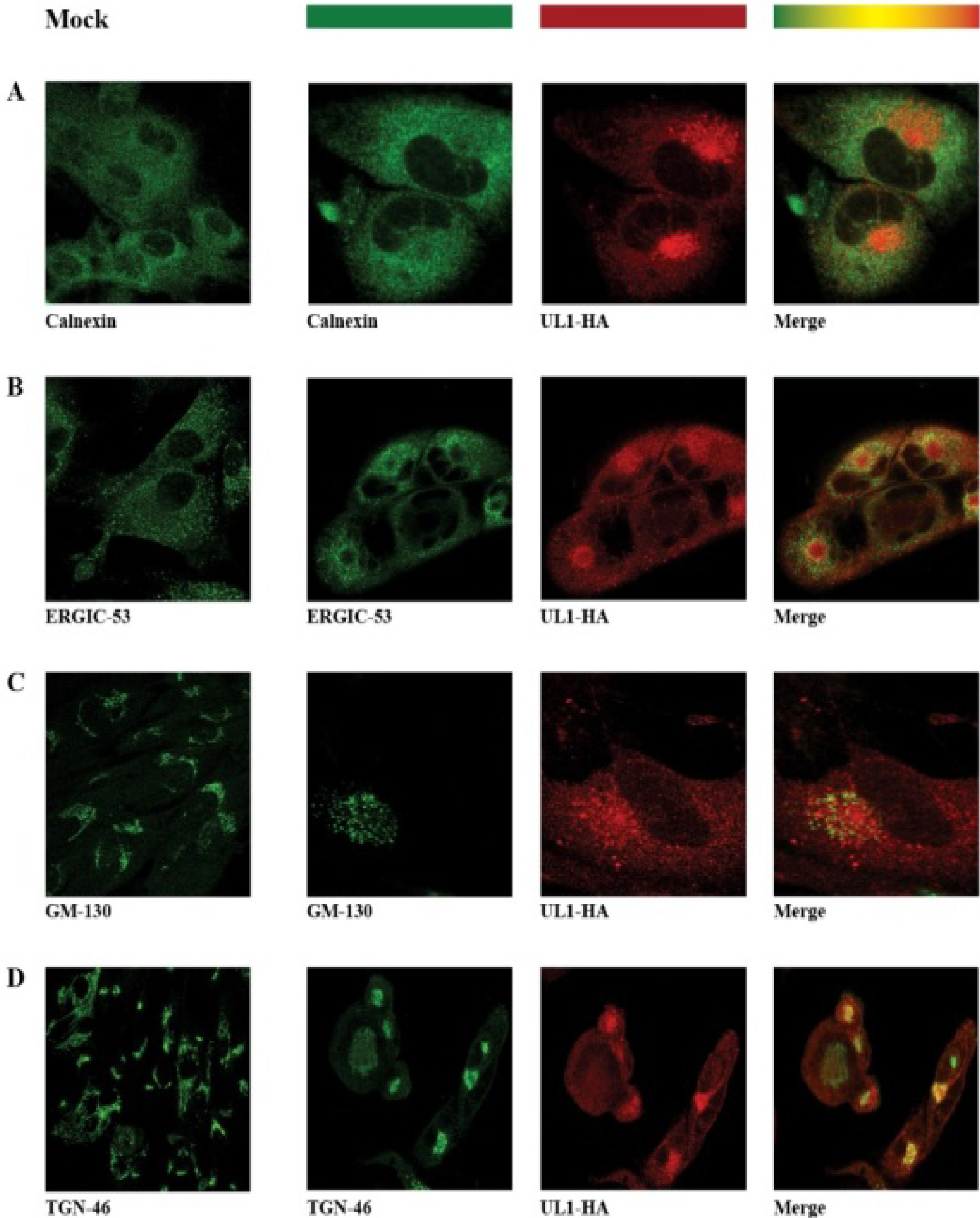


Figure 24. Accumulation of UL1-HA in a cytoplasmic compartment, resistant to treatment with brefeldin A. MRC5 cells grown on glass coverslips were infected with HB5 UL1-HA at MOI 5 and fixed on day 3 post infection. Uninfected MRC-5 cells were processed similarly (far left panel). The various cellular compartments in the secretory system were stained with the

following antibodies: (A) Calnexin for the ER, (B) ERGIC-53 for the ERGIC, (C) GM130 for the Golgi apparatus, and (D)TNG-46 for the *trans*-Golgi Network. (F) Infected MRC-5 cells were incubated in the presence of brefeldin A (2mg/ml) for 30min or left untreated and then processed as described above. In (A-F) the cellular markers were detected with goat anti-mouse IgG alexa 488, and UL1-HA was detected goat anti-rabbit IgG alexa 555. Co-localization is indicated by a yellow signal in the merge channel.

During late stages of the infectious cycle, pUL1 did not colocalize with markers of the *cis*- or *medial*-Golgi compartment (ERGIC-53 and GM-130). To further confirm that the UL1-containing compartment did not co-localize with Golgi nor ERGIC the infected cells were incubated with brefeldin A (BFA) and distribution of UL1-HA, GM-130 and ERGIC-53 positive membranes was compared (Figure 25). BFA interferes with the exocytic cellular transport and the treatment resulted in the vesiculation of the ring-structured GM-130 and ERGIC-53 positive membranes, whereas such changes were not observed in the appearance or redistribution of the UL1-HA containing structure. These results suggest that intracellular UL1 is not associated with the *cis*-Golgi apparatus.

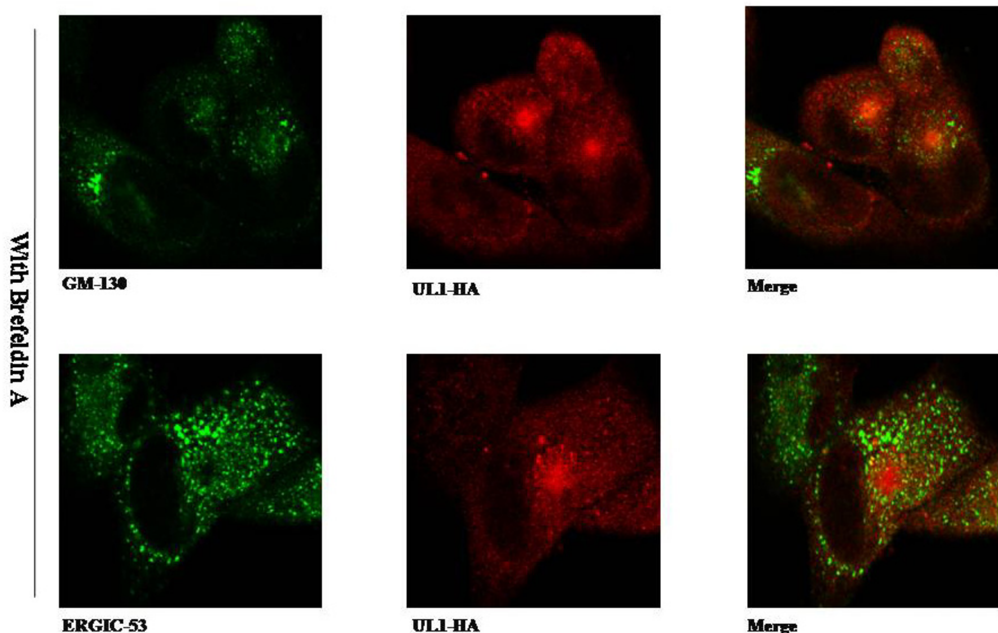


Figure 25. Infected MRC-5 cells were incubated in the presence of brefeldin A (2mg/ml) for 30min or left untreated (see Figure 3B, 3C) and then processed as described in figure 3. The cellular markers were detected with goat anti-mouse IgG Alexa 488, and UL1-HA was detected with goat anti-rabbit IgG Alexa 555.

6.4 Co-localization of pUL1 with tegument and envelope HCMV proteins

Based on the cytoplasmic compartmentalization of UL1, we explored whether other structural HCMV proteins gathered at the UL1-containing site. To this end antibodies reactive with the tegument phosphoprotein pp28 and a major virion envelope glycoprotein gB (*UL55*-encoded) were employed. HCMV pp28, has been reported to co-localize with other tegument and viral envelope proteins at the virus assembly site during the late phase of the infectious cycle (Sanchez *et al.*, 2000) and it is also acquired by the virion in the cytoplasm. Interestingly, UL1 co-localized with pp28 (Figure 26A) and gB (Figure 26B), consistent with its presence at the site of viral assembly and final envelopment, thus suggesting that it might be integrated in the HCMV virion.

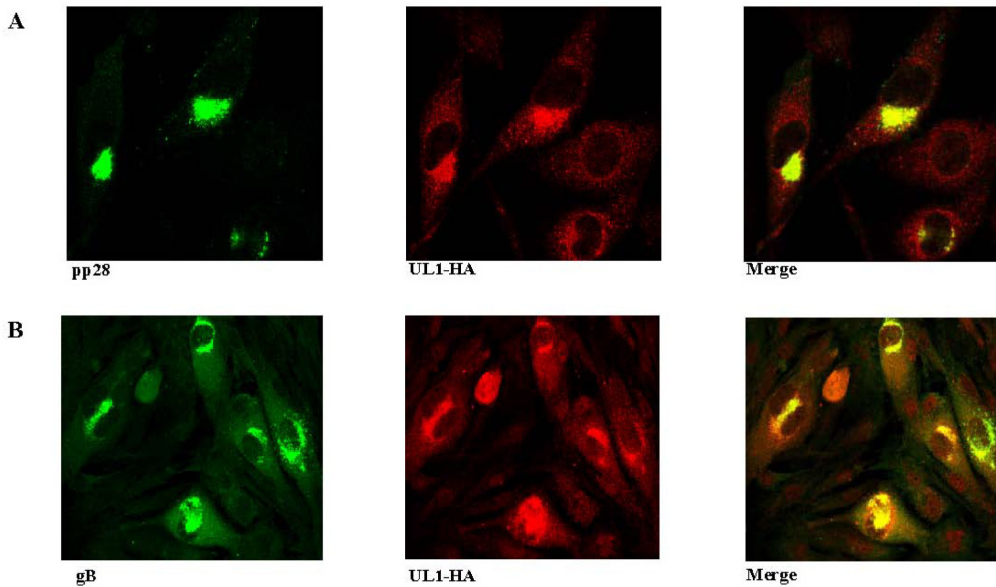


Figure 26. Co-localization with tegument and envelope HCMV proteins. MRC-5 cells grown on glass coverslips were infected with HB5 UL1-HA at MOI 5 and fixed at day 3 p.i. Following fixation, samples were incubated with MAbs specific for: (A) tegument protein pp28 and HA or (B) envelope glycoprotein gB and HA. pp28 and gB were detected with goat anti-mouse IgG Alexa 488 and UL1-HA was detected goat anti-rabbit IgG Alexa 555. Co-localization is indicated by a yellow signal in the merge channel.

6.5 pUL1 is a component of the HCMV virion envelope

The HCMV virion is divided into three major structural layers. The first layer is the nucleocapsid containing the double-stranded viral DNA genome, which is surrounded by a less structured tegument protein layer. The tegument is surrounded by an envelope that consists of a cellular lipid bilayer with both viral and cell derived proteins (Baldick and Shenk, 1996; Mocarski and Courcelle, 2001).

To explore whether pUL1 is a virion structural glycoprotein, we purified AD169 UL1-HA particles by negative viscosity-positive glycerol-tartrate gradient centrifugation and analyzed them by immunoblotting with the anti-HA antibody. This purification process facilitates the isolation of intact virions, separated from non-infectious enveloped particles (NIEPs) and dense bodies (DBs) and removal of contaminating host cell debris. Purified wild-type AD169 virions were used as a negative control due to the absence of the HA epitope. In these experiments, the 55-kDa HA-tagged UL1 specific band observed in lysates of AD169 UL1-HA infected cells was used as positive control the detection of which was reproduced both in lysates (Figure 27A, lane 1) and in AD169 UL1-HA virion preparations (Figure 27A, lane 3), in parallel with the pp28 structural protein but not IE1 pp72, a non-structural protein present only in infected cells. In order to define the virion incorporation of of pUL1, additional experiments were performed. Previous studies have shown that extracellular virions envelope glycoproteins are sensitive to protease digestion, while capsid and tegument proteins are protected (Baldick and Shenk, 1996; Yao and Courtney, 1992). Thus, we analyzed the sensitivity of virion pUL1 to protease digestion. AD169 UL1-HA virions were treated with detergent (SDS) to disrupt the envelope and then digested with proteinase K (Figure 27B). When samples were untreated, the tegument protein pp28 and envelope glycoprotein gB were detectable (Figure 27B, lane 1). Upon proteinase K digestion and in the absence of detergent, pp28 remained intact, whereas pUL1 was no longer detected and the highly abundant gB was largely digested (Figure 27B, lane 2). In samples only treated with SDS, both tegument and envelope proteins remained intact (Figure 27B, lane 3). By contrast,

when initially solubilized and then digested with proteinase K, pUL1, gB and pp28 were no longer detected (Figure 27B, lane 4). Altogether, these findings provided evidence that pUL1, in a similar manner as gB, is sensitive to protease treatment and likely exposed on the HCMV envelope.

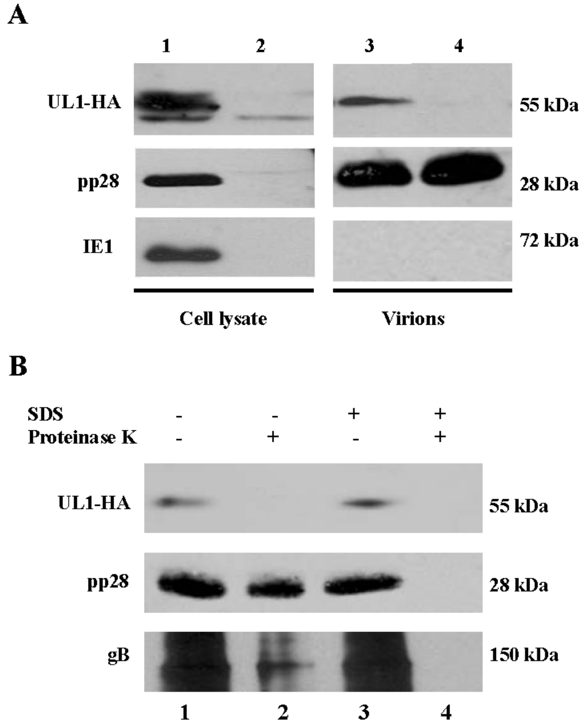


Figure 27. UL1 is a structural glycoprotein that forms part of the HCMV envelope. (A) Analysis of UL1 in cell lysates of AD169 UL1-HA infected MRC-5 cells (lane 1) and AD169 UL1-HA virions (lane 3). Mock cells (lane 2) and AD169 WT virions (lane 4) were used as negative controls. Western blots were developed using anti-HA, anti-pp28 or anti-IE1 antibodies. (B) AD169 UL1-HA virions were treated with SDS (lanes 3 and 4) or left untreated (lanes 1 and 2). One half of each virus preparation was subsequently digested with proteinase K (lanes 2 and 4) or left undigested (lanes 1 and 3). The preparations were subjected to immunoblot analysis, using anti-HA, anti-pp28 or anti-gB antibodies.

6.7 Analysis of UL1 mutant viruses

The aim of these experiments is to characterize the importance of UL1 glycoprotein in the viral replication cycle. The envelope glycoproteins of herpesviruses have a number of different functions during the replication cycle. During the initial phase of infection they mediate attachment and fusion with the target cell. During later stages of infection, the glycoproteins have roles in cell-to-cell spread, envelopment and egress. Moreover, they also have a role in the modulation of host immune responses.

6.7.1 Deletion of *UL1* results in HCMV growth defects in epithelial cells

A number of the HCMV envelope glycoproteins have been shown to play a role in critical steps of the the viral life cycle, i.e. processes such as cell entry, assembly and envelopment of virions or cell-to-cell spread. To assess the impact that the pUL1 envelope glycoprotein may have on the HCMV replication cycle we constructed a HCMV mutant lacking the *UL1* ORF using the TB40/E derived BACmid (Sinzger et al. 2008b). Unlike the fibroblast adapted AD169 strain, TB40 virions exhibit an intact cell tropism for endothelial as well as epithelial cells the entry of which is mediated by the pentameric gH/gL/UL128/UL130/UL131 envelope complex which is deficient in AD169 (Ryckman *et al.,* 2008). The complete *UL1* sequence was removed from the genome of the endotheliotropic strain TB40/E, generating the recombinant virus TB40/E ΔUL1. We then infected human MRC-5 cells, retinal pigment epithelial (RPE) cells and U373-MG cells derived from glioblastoma at a low MOI with TB40/E ΔUL1 or the parental TB40/E. A moi of 0.025 was used to infect the fully permissive fibroblasts, while RPE and U373-MG (which support HCMV replication to a lower extent) were infected at a moi of 0.1 and 0.3, respectively.

The production of extracellular (MRC-5 and RPE cells) or cell associated (U373-MG) virus was analyzed over the course of the replicative cycle. As shown in Figure 28, in RPE cells the growth levels of the UL1-deficient virus were significantly reduced (up to 100-fold) in comparison to those obtained with the parental virus

(p<0.05 at the different time points). Consistent with this finding, TB40/E ΔUL1 produced smaller plaques relative to those obtained after TB40/E infection of this cell type (data not shown). In contrast, under these conditions of multistep growth, infections of MRC-5 and U373-MG demonstrated that TB40/E ΔUL1 and parental TB40/E grew with comparable kinetics. These results indicate that pUL1 is required for efficient HCMV growth in a cell type-specific manner.

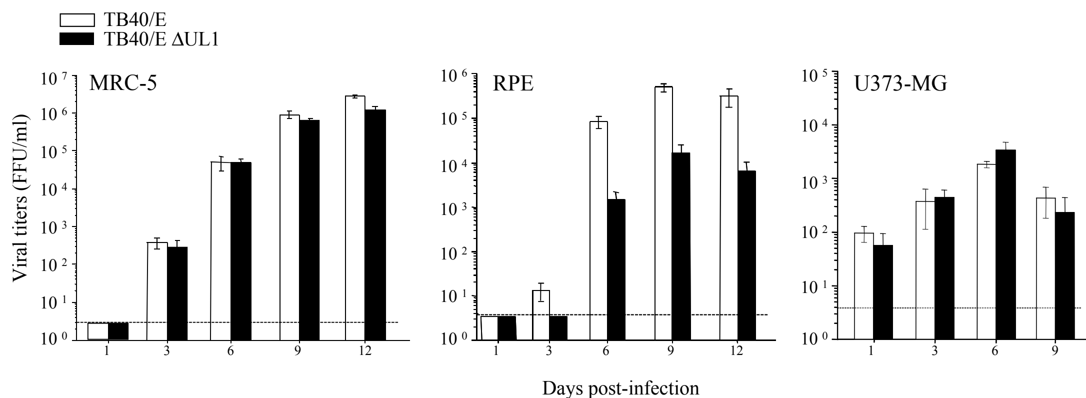


Figure 28. Growth kinetics of TB40/E ΔUL1 in HEL299, RPE and U373 cells. HEL299, RPE and U373 cells were infected with TB40/E or TB40/E ΔUL1 at a low moi. At the designated days pi, cell supernatants were collected and yields of infectious virus determined by FFU assays. Each data point represents the average and standard deviation of three separate cultures. The dashed line indicates the limit of detection.

6.7.2 Entry of TB40/E *UL1* deletion viruses

HCMV entry in different cell-types may require different envelope glycoproteins composition and different entry pathways. To date, two different entry processes have been described for HCMV. In fibroblasts, HCMV enters through direct fusion of the viral envelope and cell plasma membrane at neutral pH (Reviewed in Compton, 2004). However in epithelial and endothelial cells, HCMV entry mediated by receptor endocytosis (Ryckman *et al.,* 2006). We then sought to determine the step of the viral replication cycle that is compromised in the absence of UL1. The potential role of UL1 during an early step of viral replication cycle was tested by checking the entry of TB40/E ΔUL1 compared with parental TB40/E onto RPE cells. Measuring IE1 mRNA by quantitative RT-PCR shortly after infection, we observed significant

differences (P value < 0,05) between the entry of TB40/E ΔUL1 and parental TB40/E onto RPE cells (Figure 29). These results indicate that UL1 glycoprotein may have a role in the entry of HCMV onto RPE cells.

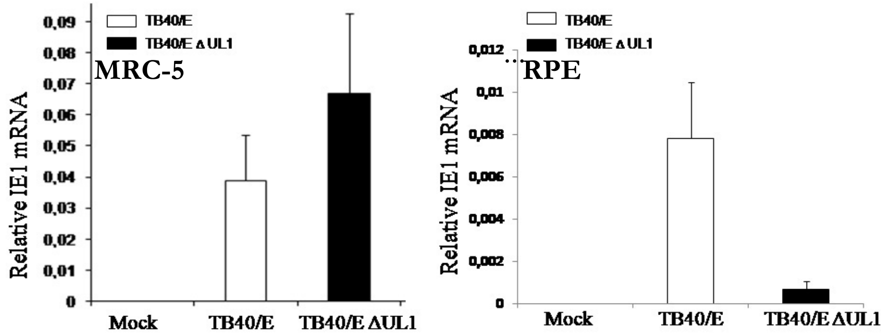


Figure 29. Entry of TB40/E ΔUL1 in MRC-5 and RPE cells. (A) MRC-5 and (B) RPE cells were unifected (mock) or infected with TB40/E or TB40/E **Δ**UL1 infected at moi 1. IE1 mRNA was quantified by RT-qPCR 4 h p.i.. Error bars show the mean ± SD; results are representative of 4 experiments. MRC5 p value=0,14 and RPE p value=0,018 (*).

PART IV DISCUSSION

7. DISCUSSION

HCMV has the largest genome of any characterized human viruses and with this large protein coding capacity the virus has evolved a variety of mechanisms to interact with host cells and interfere with the host immune response; some of these may have evolved from captured host cellular genes. The identification and characterization of these genes and their cellular homologs should improve our knowledge of the virus-host interaction. Here we investigated the similarity of the HCMV-specific glycoprotein UL1 to the N-terminal IgV domain of members of the CEA family that are involved in homotypic and heterotypic adhesion mechanisms. Moreover, we characterized the UL1 glycoprotein, transcribed in the AD169 strain and in passaged clinical isolates such as Toledo and TB40/E. UL1 is a PNGase F sensitive glycoprotein, with a putative C-terminal transmembrane anchor domain, detected late during HCMV infection. In productively infected human fibroblasts, UL1 accumulates in a cytoplasmic site together with several HCMV tegument and envelope proteins, consistent with its location within the site of viral assembly and envelopment. UL1 is detected in purified AD169 virions and, according to its pattern of sensitivity to protease treatment, appears exposed in the HCMV envelope. Based on these results, we propose that UL1 constitutes a novel envelope protein associated with the HCMV virion that may play a functional role in modulating cellular tropism.

The *UL1* gene is predicted to encode a 224-amino acid type I transmembrane glycoprotein, comprising a long signal peptide (amino acids 1 to 46), an Ig-like domain (amino acids 47 to 151) with 9 potential N-linked glycosylation sites, a transmembrane region (amino acids 184 to 206), and a short cytoplasmic tail (amino acids 207 to 224) without apparent signaling motifs. Since a sequence similarity of pUL1 to the pregnancy-specific glycoprotein 5 and other members of the human CEA family (Holzerlandt *et al.,* 2002) was previously noted, we sought to explore this finding in more detail. In fact, the region of sequence similarity was found to cover the Ig-like domain in pUL1 and the N-terminal IgV-like domain in the CEA family members. The alignment of the pUL1 Ig-like domain and the three-dimensional crystal structure of

the soluble ecto-domain of the murine CEACAM-1a isoform indicated that the formation of Ig-like β-sheets is conserved and the conserved residues are mostly buried in the hydrophobic core, maintaining the Ig fold. CEACAM positions that appear to be critical for the formation of the typical β-sheets in IgV-like domains are present in pUL1. Furthermore, multiple alignments of the Ig-like domain of UL1 and the N-terminal IgV domain of several CEA proteins showed that a number of well-conserved residues in members of the CEA family are also present in pUL1, whereas they are absent from other RL11 proteins.

To date, several HCMV genes homologues to cellular genes that regulate immune responses have been identified, among them are gpUL18: MHC class I; *UL111a*: IL-10; *UL144*: TNFR; *UL14 and UL147*: supergene family IL-8; *UL33, UL78, US27 and US28:* G protein coupled receptor; *UL119-118, TRL11 and TRL12*: Fc receptor homologs. Here we describe the striking similarity between UL1 and the N-terminal IgV domain cellular CEA family molecules that are involved in homotypic and heterotypic adhesion with each other. CEACAM subfamily includes inhibitory and activating receptors widely expressed in different cell types, and members of this family have been reported as critical modulators of several physiological processes. CEA family consists of the CEA-related cell adhesion molecule (CEACAM) and the pregnancy specific glycoprotein (PSG) subgroups. CEA family is a rapidly evolving gene family involved in immune responses and reproduction (Kammerer and Zimmermann, 2010). Of note, only a few CEACAM homologues are present in rodents, whereas the family has expanded considerably during primate evolution and is still diversifying (Hammarstrom and Baranov, 2001; Zhou *et al.,* 2001). .

The sequence similarity between UL1 and host CEA proteins raises the possibility that it may constitute a molecular mimicry strategy providing some biological advantage to HCMV (Elde and Malik, 2009). As other members of the *RL11* family do not resemble the CEA proteins beyond the RL11D domain, we hypothesize that pUL1 may have evolved to mimic members of CEA family by sequence convergence, gaining CEA-like features along the million years of virus-host co-

evolution. The molecular targets and the biological significance of the sequence similarity between pUL1 and members of CEA family remain to be explored. Notably, a number of pathogens take advantage of the CEA extracellular domains to bind and infect their target cells, e.g. mouse CEACAM1a is targeted by the mouse hepatitis virus (Dveksler *et al.*, 1991). Moreover, *Neisserial* pathogens have been shown to use CEACAM-1 in bronchial epithelial cells to suppress TLR-2 signaling (Slevogt *et al.,* 2008). TLR-2 has been described to sense HCMV entry through a physical interaction with gB and gH, which results in the activation of NF-kB and cellular inflammatory cytokine secretion (Boehme *et al.,* 2006). Thus, pUL1 might contribute to HCMV access to host cells by interacting with members of the CEA family.

To explore the possibility of interactions between UL1 and members of CEA family, the expression and production of UL1 as a secretory fusion protein was undertaken by fusing the sequence corresponding to UL1 Ig-like domain to the Fc region of human IgG. UL1-Fc fusion protein was also constructed for the production of the anti-UL1 monoclonal antibodies. Unfortunatelly, using UL1 sequence of different origin (TB40 or AD169) and cloning into different constructs, the steady state expression in mammalian cells was unsuccessful. This may be due to a problem at transcription level (e.g. complex transcription), as no UL1 RNA was detected in cells transiently transfected with UL1 containing expression vectors. It is also possible that UL1 protein gets miss-folded and degraded and/or it may require other viral proteins for its expression. The unsuccessful production of UL1 fusion proteins have been a major drawback in setting up the appropriate functional studies and generating the tools that could allow us to examine the expression and biochemical properties of UL1 in a steady state or in the context of HCMV infection. Future studies will be aimed to the expression of UL1 recombinant proteins by baculovirus-mediated infection (BEVS) of insect cells and cloning of UL1 from HCMV cDNA will be considered.

A growing number of viral glycoproteins has been reported as component of the HCMV envelope, including molecules conserved between members of the *Herpesviridae* family (gB, gH, gL, gM, gN) and others which are HCMV specific e.g.

gpTRL10 (Spaderna *et al*., 2002), gpRL13 (Stanton *et al*., 2010), gpUL132 (Spaderna *et al*., 2005), UL74-encoded gO (Huber and Compton, 1998), UL4/gp48 (Chang *et al*., 1989), *US27* (Fraile-Ramos *et al.,* 2002), *UL33* (Margulies *et al*., 1996), gpUL128 and gpUL130 (Wang and Shenk, 2005; Patrone et al., 2005). It is believed that genera specific envelope glycoproteins contribute to the broad cell/tissue tropism of HCMV and its pathogenicity. In this study we have characterized the HCMV *UL1* encoded glycoprotein, transcribed in the fibroblast-adapted strains AD169 and Toledo as well as the endotheliotropic strain TB40/E. pUL1 is a PNGase F sensitive glycoprotein, with a putative C-terminal transmembrane anchor domain, which is synthesized late during productive HCMV infection. In infected human fibroblasts, pUL1 is targeted to a cytoplasmic site together with several HCMV tegument and envelope proteins, consistent with its final destination at the previously described distinct compartment of virion assembly and final envelopment (reviewed in Britt *et al.,* 2007). Consequently pUL1 can be incorporated into virions and becomes exposed in the HCMV envelope. Based on these results, we propose that pUL1 constitutes a novel HCMV structural envelope glycoprotein.

UL1 belongs to the highly variable and rapidly evolving RL11 gene family. The pUL1 sequence is exclusively found in all HCMV genomes but lacking a counterpart in other CMV lineages and β-herpesviruses. Therefore, we hypothesize that *UL1* is originating from a gene duplication event that probably took place after the speciation of human and chimpanzee. Of note, two other HCMV *RL11* family members, *RL6* and *RL5A*, are not found in the chimpanzee CMV neither in the more distantly related rhesus macaque CMV, and are thus likely to have been formed within the *RL11* cluster as well in the last 5-6 Mya. Adaptable RL11 genes, specific for primate CMV, show a high degree of sequence variation and may therefore have an important role in HCMV cell tropism or pathogenicity (Wang *et al.,* 2008). UL1 may have originated by gene duplication of a gene from RL11 TRL-cluster (*TRL11-13*) and subsequently diverged at a high rate, as sequence analysis studies indicate that UL1 is evolutionary closer to these genes than to the reset of RL11 genes. Indeed, a substantial heterogeneity of pUL1 sequences has been appreciated among different HCMV strains (Dolan *et al*.,

2004; Sekulin *et al.,* 2007). Phylogenetic analysis of *UL1* in a variety of clinical HCMV isolates allowed the differentiation of three distinct genotypes, with the existence of a considerable number (12.9%, 4/31) of isolates containing impaired versions of the *UL1* sequence due to internal stop codon mutations (Sekulin *et al.,* 2007). In addition, the fibroblast-adapted Towne strain, that has lost virulence and experienced substantial genome alteration during extensive *in vitro* passaging (Quinnan *et al.,* 1984), bears a deletion in *UL1* (Sekulin *et al.,* 2007). The hypothesis that the *UL1* polymorphisms existing in clinical isolates may have an effect on HCMV tissue tropism and pathogenesis remains to be substantiated in the future.

Both the transcriptional analysis of *UL1* and the time course of pUL1 synthesis in HCMV infected MRC-5 cells proved exclusive expression in the late phase of the replication cycle. The absence of *UL1* transcription and translation after treatment with the potent viral DNA polymerase inhibitor PAA corroborated that *UL1* belongs to the late class of CMV genes. *UL1* is 675 bp, but during northern blot analysis larger transcripts, of approximately 5.8-Kbp and 3.8-Kbp were detected, similar results have been observed for other genes in the proximity of *UL1*, such as *TRL/IRL10* (Spaderna *et al.*, 2002), and *TRL/IRL13* (Yu *et al.,* 2002). Of note, a canonical polyadenylation (polyA) site is predicted between *TRL13* and *UL1,* an AATAAA signal, nt 11714-11719 upstream of *UL1* with a cleavage site for poly(A) addition (Rawlinson and Barrell, 1993). Although, these genes are probably driven by a different promoter as *TRL13* is characterized as an early/late gene, insensitive to PAA (Yu *et al.,* 2002) and *UL1* is a late gene. Moreover, different RNA bands that are observed may be due to splicing events in *UL1*. A conserved putative acceptor splice site was predicted in *UL1*, but its presence could not be confirmed and the complete *UL1* could be detected by RT-PCR. However, other splicing events in *UL1* cannot be ruled out. To more precisely map the *UL1*-specific transcripts, SMART RACE on RNA from infected cells by using two overlapping primers located inside the *UL1* needs to be performed.

In infected MRC-5 cells, pUL1 does not seem to form higher order disulphide-linked complexes. We show that gpUL1 acquires a molecular weight of 55-kDa. The glycoprotein is sensitive to PNGase F producing a 26-kDa polypeptide. Based on these results, one could conclude that most, if not all of the nine potential N-glycosylation sites that UL1 bears are modified by N-linked carbohydrates, eight of them residing in the Ig-like domain and the remaining one in the 32 amino acid stretch that connects the Ig-like domain to the transmembrane sequence,. It must also be noted that in this pUL1 region expanding from amino acid 152 to 183 several potential O-glycosylation sites are present. After entering the secretory pathway the UL1 glycoprotein exits from the ER to accumulate in close vicinity with other components of the HCMV particle, such as the tegument protein pp28 and the envelope glycoprotein gB, at the juxtanuclear cytoplasmic site of virion assembly. The HCMV envelope is still incompletely defined (Britt, 2007). Based on previous mass spectrometric studies (Varnum *et al.,* 2004) at least 19 structural glycoproteins of AD169 virions could be identified. It is interesting to note that pUL1 among several other bona fide virion glycoproteins (Spaderna *et al.,* 2002; Stanton *et al.,* 2010) was not resolved by this approach. We find that after gradient-tartrate viral particle purification pUL1 can be detected as a virion constituent. Moreover, biochemical analysis of purified virions revealed the same features of pUL1 as demonstrated for gB, allowing to assign pUL1 to the virion envelope.

UL1 is dispensable for HCMV growth in cell culture (Ripalti and Mocarski, 1991). Our results indicate however that pUL1 might be of significant relevance for HCMV to establish efficient infections in certain cell-types. As a virion envelope glycoprotein pUL1 has the potential to modulate HCMV host cell tropism. We demonstrate that a HCMV TB40/E mutant lacking *UL1* displayed an attenuated growth in RPE cells, not observed in MRC-5 or U373-MG cells. Notably, an involvement in cell tropism has been reported for other members of the RL11 glycoprotein family; a *UL10* deletion mutant grew better than the parental virus in an epithelial cell line (Dunn *et al.,* 2003) and the RL13 envelope glycoprotein has been shown to repress HCMV replication in certain cell types (Stanton *et al.,* 2010). The

nature of the inefficient growth associated with *UL1* deficiency is at present unknown. A possibility is that the UL1 glycoprotein may have a role in the entry of HCMV when replicating in epithelial cells as represented by RPE. As indicated above, and in line with the homology of the UL1 glycoprotein with members of the CEACAM family, an attractive hypothesis to test is whether this viral protein could be interacting with a member of this family at the host cell surface. In addition, the possibility that pUL1 may represent a relevant target of humoral immune responses during HCMV infection should be considered. Future studies will be aimed at clarifying the function(s) of pUL1 during HCMV replication.

PART V CONCLUSIONS

8. CONCLUSIONS

1. Sequence similarity searches against human Ig-containing proteins confirmed that HCMV UL1 shows significant similarity to the cellular CEA protein family. CEACAM positions that appear to be critical for the formation of the typical β-sheets in Ig-like domains are conserved.

2. Sequence analysis studies indicate that *UL1* may have originated by gene duplication of a gene from RL11 TRL-cluster (*TRL11-13*) and subsequently diverged at a high rate.

3. *UL1* is transcribed during the late phase of the HCMV replication cycle, both in fibroblast-adapted and endotheliotropic viral strains.

4. UL1 is a expressed as a glycoprotein and detected late during HCMV infection.

5. In infected human fibroblasts, UL1 accumulates in a cytoplasmic site together with several HCMV tegument and envelope proteins, consistent with its location within the site of viral assembly and envelopment.

6. UL1 is detected in purified AD169 virions and, according to its pattern of sensitivity to protease treatment, appears exposed on the HCMV envelope.

7. Deletion of *UL1* in HCMV TB40/E resulted in a reduced growth in a cell type-specific manner, suggesting that pUL1 may be implicated in regulating HCMV cell tropism.

ANNEX I

Parts of this thesis have been submitted for publication

The HCMV-specific *UL1* gene encodes a late phase glycoprotein incorporated in the virion envelope

Medya Shikhagaie[1], Eva Mercé-Maldonado[2], Elena Isern[3], Aura Muntasell[4], M. Mar Albà[4,5,6], Miguel López-Botet[1,4], Hartmut Hengel[2], and Ana Angulo[3*]

[1]*Immunology Unit, Pompeu Fabra University, Barcelona, Spain;*

[2]*Institut für Virologie, Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany;*

[3]*Institut d'Investigacions Biomèdiques August Pi i Sunyer, Barcelona, Spain;*

[4]*IMIM-Hospital del Mar Research Institute, Barcelona, Spain*

[5]*Research Programme in Biomedical Informatics, Pompeu Fabra University, Barcelona, Spain*

[4]*Catalan Institution for Research and Advanced Studies (ICREA), Barcelona, Spain*

*Corresponding author:

Mailing address: Dr. Ana Angulo
Institut d'Investigacions Biomèdiques August Pi i Sunyer
C/ Villarroel 170. Barcelona 08036, Spain
Phone: 34 647 450269
Fax: 34 93 4021907
E-mail: aangulo@ub.edu

Running title: HCMV UL1 envelope glycoprotein

REFERENCES

REFERENCES

Abate D. A., Watanabe S., Mocarski E. S. 2004. Major human cytomegalovirus structural protein pp65 (ppUL83) prevents interferon response factor 3 activation in the interferon response. J. Virol.; 78:10995–11006.

Ahn K., Angulo A., Ghazal P., Peterson P.A., Yang Y., Früh K. 1996. Human cytomegalovirus inhibits antigen presentation by a sequential multistep process. Proc Natl Acad Sci U S A. Oct 1;93(20):10990-5.

Alba, M. M., R. Das, C. A. Orengo, and P. Kellam. 2001. Genomewide function conservation and phylogeny in the Herpesviridae. Genome Res. **11**:43-54.

Altschul, S. F., T. L. Madden, A. A. Schaffer, J. Zhang, Z. Zhang, W. Miller, and D. J. Lipman. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 25:3389-3402.

Arase, H., Mocarski,E.S., Campbell,A.E., Hill,A.B. & Lanier,L.L. 2002. Direct recognition of cytomegalovirus by activating and inhibitory NK cell receptors. Science 296, 1323-1326.

Atalay, R., A. Zimmermann, M. Wagner, E. Borst, C. Benz, M. Messerle, and H. Hengel. 2002. Identification and expression of human cytomegalovirus transcription units coding for two distinct Fcgamma receptor homologs. J.Virol. 76:8596-8608.

Baldick, C. J. and T. Shenk. 1996. Proteins associated with purified human cytomegalovirus particles. J.Virol. 70:6097-6105.

Billstrom, M. A., G. L. Johnson, N. J. Avdi, and G. S. Worthen. 1998. Intracellular signaling by the chemokine receptor US28 during human cytomegalovirus infection. J.Virol. 72:5535-5544.

Biron CA. 1997. Activation and function of natural killer cell responses during viral infections. Curr Opin Immunol. (1):24-34. Review.

Boehme, K. W., M. Guerrero, and T. Compton. 2006. Human cytomegalovirus envelope glycoproteins B and H are necessary for TLR2 activation in permissive cells. J.Immunol. 177:7094-7102.

Bold, S., M. Ohlin, W. Garten, and K. Radsak. 1996. Structural domains involved in human cytomegalovirus glycoprotein B-mediated cell-cell fusion. J.Gen.Virol. 77 (Pt 9):2297-2302.

Boulton IC, Gray-Owen SD. 2002. Neisserial binding to CEACAM1 arrests the activation and proliferation of CD4+ T lymphocytes. Nat Immunol. Mar;3(3):229-36.

Borst, E. M., G. Hahn, U. H. Koszinowski, and M. Messerle. 1999. Cloning of the human cytomegalovirus (HCMV) genome as an infectious bacterial artificial chromosome in Escherichia coli: a new approach for construction of HCMV mutants. J.Virol. **73**:8320-8329.

Bukowski JF, Warner JF, Dennert G, Welsh RM. 1985. Adoptive transfer studies demonstrating the antiviral effect of natural killer cells in vivo. J Exp Med. 161:40-52.

Bresnahan, W. A. and T. Shenk. 2000. A subset of viral transcripts packaged within human cytomegalovirus particles. Science 288:2373-2376.

Britt, W. J. and M. Mach. 1996. Human cytomegalovirus glycoproteins. Intervirology 39:401-412.

Britt, W. J. and L. G. Vugler. 1992. Oligomerization of the human cytomegalovirus major envelope glycoprotein complex gB (gp55-116). J.Virol. 66:6747-6754.

Britt, B. 2007. CMV maturation and egress, p. 311-323. *In* A. Arvin, G. Campadelli-Fiume, E. Mocarski, P. S. Moore, B. Roizman, R. Whitley, and K. Yamanishi (ed.), Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press. Cambridge, United Kingdom.

Cha, T. A., E. Tom, G. W. Kemble, G. M. Duke, E. S. Mocarski, and R. R. Spaete. 1996. Human cytomegalovirus clinical isolates carry at least 19 genes not found in laboratory strains. J.Virol. 70:78-83.

Chang, C. P., D. H. Vesole, J. Nelson, M. B. Oldstone, and M. F. Stinski. 1989. Identification and expression of a human cytomegalovirus early glycoprotein. J.Virol. 63:3330-3337.

Chee, M. S., A. T. Bankier, S. Beck, R. Bohni, C. M. Brown, R. Cerny, T. Horsnell, C. A. Hutchison, III, T. Kouzarides and J. A. Martignetti. 1990. Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. Curr.Top.Microbiol.Immunol. 154:125-169.

Chi, J. T., H. Y. Chang, G. Haraldsen, F. L. Jahnsen, O. G. Troyanskaya, D. S. Chang, Z. Wang, S. G. Rockson, M. van de Rijn, D. Botstein, and P. O. Brown. 2003. Endothelial cell diversity revealed by global expression profiling. Proc.Natl.Acad.Sci.U.S.A 100:10623-10628.

Compton, T., D. M. Nowlin, and N. R. Cooper. 1993. Initiation of human cytomegalovirus infection requires initial interaction with cell surface heparan sulfate. Virology 193:834-841.

Compton, T., E. A. Kurt-Jones, K. W. Boehme, J. Belko, E. Latz, D. T. Golenbock, and R. W. Finberg. 2003. Human cytomegalovirus activates inflammatory cytokine responses via CD14 and Toll-like receptor 2. J.Virol. 77:4588-4596.

Compton, T. 2004. Receptors and immune sensors: the complex entry path of human cytomegalovirus. Trends Cell Biol. 14:5-8.

Cosman D., Fanger N., Borges L.1997. A novel immunoglobulin superfamily receptor for cellular and viral MHC class I molecules. *Immunity*. 7(2):273-282.

Crough, T. and R. Khanna. 2009. Immunobiology of human cytomegalovirus: from bench to bedside. Clin.Microbiol.Rev. 22:76-98.

Cunningham, C., D. Gatherer, B. Hilfrich, K. Baluchova, D. J. Dargan, M. Thomson, P. D. Griffiths, G. W. Wilkinson, T. F. Schulz, and A. J. Davison. 2010. Sequences of complete human cytomegalovirus genomes from infected cell cultures and clinical specimens. J.Gen.Virol. 91:605-615.

Davison, A. J., P. Akter, C. Cunningham, A. Dolan, C. Addison, D. J. Dargan, A. F. Hassan-Walker, V. C. Emery, P. D. Griffiths, and G. W. Wilkinson. 2003a. Homology between the human cytomegalovirus RL11 gene family and human adenovirus E3 genes. J. Gen.Virol. 84:657-663.

Davison, A. J., A. Dolan, P. Akter, C. Addison, D. J. Dargan, D. J. Alcendor, D. J. McGeoch, and G. S. Hayward. 2003b. The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. J. Gen.Virol. 84:17-28.

DeFilippis, V. R., B. Robinson, T. M. Keck, S. G. Hansen, J. A. Nelson, and K. J. Fruh. 2006. Interferon regulatory factor 3 is necessary for induction of antiviral genes during human cytomegalovirus infection. J.Virol. 80:1032-1037.

Dolan, A., C. Cunningham, R. D. Hector, A. F. Hassan-Walker, L. Lee, C. Addison, D. J. Dargan, D. J. McGeoch, D. Gatherer, V. C. Emery, P. D. Griffiths, C. Sinzger, B. P. McSharry, G. W. Wilkinson, and A. J. Davison. 2004. Genetic content of wild-type human cytomegalovirus. J.Gen.Virol. 85:1301-1312.

Dunn, W., C. Chou, H. Li, R. Hai, D. Patterson, V. Stolc, H. Zhu, and F. Liu. 2003. Functional profiling of a human cytomegalovirus genome. Proc.Natl.Acad.Sci.U.S.A 100:14223-14228.

Dveksler, G. S., M. N. Pensiero, C. B. Cardellichio, R. K. Williams, G. S. Jiang, K. V. Holmes, and C. W. Dieffenbach. 1991. Cloning of the mouse hepatitis virus (MHV) receptor: expression in human and hamster cell lines confers susceptibility to MHV. J.Virol. 65:6881-6891.

Eddy, S. R. 1998. Profile hidden Markov models. Bioinformatics. 14:755-763.

Engel, P., N. Pérez-Carmona, M. M. Albà, K. Robertson, P. Ghazal, and A. Angulo. 2011. Human cytomegalovirus UL7, a homologue of the SLAM-family receptor CD229, impairs cytokine production. Immunol. Cell Biol. 1-14.

Felsenstein, J. 1989. Mathematics vs. Evolution: Mathematical Evolutionary Theory. Science 246:941-942.

Finn, R. D., J. Mistry, J. Tate, P. Coggill, A. Heger, J. E. Pollington, O. L. Gavin, P. Gunasekaran, G. Ceric, K. Forslund, L. Holm, E. L. Sonnhammer, S. R. Eddy, and A. Bateman. 2010. The Pfam protein families database. Nucleic Acids Res. 38:D211-D222.

Fowler, K. B. and R. F. Pass. 2006. Risk factors for congenital cytomegalovirus infection in the offspring of young women: exposure to young children and recent onset of sexual activity. Pediatrics 118:e286-e292.

Fraile-Ramos, A., A. Pelchen-Matthews, T. N. Kledal, H. Browne, T. W. Schwartz, and M. Marsh. 2002. Localization of HCMV UL33 and US27 in endocytic compartments and viral membranes. Traffic. 3:218-232.

Guma, M., A. Angulo, C. Vilches, N. Gomez-Lozano, N. Malats, and M. Lopez-Botet. 2004. Imprint of human cytomegalovirus infection on the NK cell receptor repertoire. Blood. 104:3664-3671.

Guma, M., A. Angulo, and M. Lopez-Botet. 2006. NK cell receptors involved in the response to human cytomegalovirus infection. Curr.Top.Microbiol.Immunol. 298:207-223.

Gupta, R. and S. Brunak. 2002. Prediction of glycosylation across the human proteome and the correlation to protein function. Pac.Symp.Biocomput.310-322.

Hakki M, Geballe AP. 2005. Double-stranded RNA binding by human cytomegalovirus pTRS1. J Virol. Jun;79(12):7311-8.

Hahn, G., M. G. Revello, M. Patrone, E. Percivalle, G. Campanini, A. Sarasini, M. Wagner, A. Gallina, G. Milanesi, U. Koszinowski, F. Baldanti, and G. Gerna. 2004. Human cytomegalovirus UL131-128 genes are indispensable for virus growth in endothelial cells and virus transfer to leukocytes. J.Virol. 78:10023-10033.

Hammarström S, 1999. The carcinoembryonic antigen (CEA) family: structures, suggested functions and expression in normal and malignant tissues. Semin Cancer Biol. 9(2):67-81.

Hammarström S., Baranov V. 2001. Is there a role for CEA in innate immunity in the colon? Trends Microbiol. 9(3):119-25. Review.

Hansen, S. G., L. I. Strelow, D. C. Franchi, D. G. Anders, and S. W. Wong. 2003. Complete sequence and genomic analysis of rhesus cytomegalovirus. J.Virol. 77:6620-6636.

Hengel H., Koopmann J.O., Flohr T., Muranyi W., Goulmy E., Hämmerling G.J., Koszinowski U.H., Momburg F. 1997. A viral ER-resident glycoprotein inactivates the MHC-encoded peptide transporter. Immunity. 6(5):623-32.

Hitomi, S., H. Kozuka-Hata, Z. Chen, S. Sugano, N. Yamaguchi, and S. Watanabe. 1997. Human cytomegalovirus open reading frame UL11 encodes a highly polymorphic protein expressed on the infected cell surface. Arch.Virol. 142:1407-1427.

Holzerlandt, R., C. Orengo, P. Kellam, and M. M. Alba. 2002. Identification of new herpesvirus gene homologs in the human genome. Genome Res. 12:1739-1748.

Huang, E. S., B. A. Kilpatrick, Y. T. Huang, and J. S. Pagano. 1976. Detection of human cytomegalovirus and analysis of strain variation. Yale J.Biol.Med**.** 49:29-43.

Hubbard, T. J., B. L. Aken, K. Beal, B. Ballester, M. Caccamo, Y. Chen, L. Clarke, G. Coates, F. Cunningham, T. Cutts, T. Down, S. C. Dyer, S. Fitzgerald, J. Fernandez-Banet, S. Graf, S. Haider, M. Hammond, J. Herrero, R. Holland, K. Howe, K. Howe, N. Johnson, A. Kahari, D. Keefe, F. Kokocinski, E. Kulesha, D. Lawson, I. Longden, C. Melsopp, K. Megy, P. Meidl, B. Ouverdin, A. Parker, A. Prlic, S. Rice, D. Rios, M. Schuster, I. Sealy, J. Severin, G. Slater, D. Smedley, G. Spudich, S. Trevanion, A. Vilella, J. Vogel, S. White, M. Wood, T. Cox, V. Curwen, R. Durbin, X. M. Fernandez-Suarez, P. Flicek, A. Kasprzyk, G. Proctor, S. Searle, J. Smith, A. Ureta-Vidal, and E. Birney. 2007. Ensembl 2007. Nucleic Acids Res. 35:D610

Huber, M. T. and T. Compton. 1998. The human cytomegalovirus UL74 gene encodes the third component of the glycoprotein H-glycoprotein L-containing envelope complex. J.Virol. 72:8191-8197.

Johnson, R. A., S. M. Huong, and E. S. Huang. 2000. Activation of the mitogen-activated protein kinase p38 by human cytomegalovirus infection through two distinct pathways: a novel mechanism for activation of p38. J.Virol. 74:1158-1167.

Johnson D.C., Hegde N.R. 2002. Inhibition of the MHC class II antigen presentation pathway by human cytomegalovirus. Curr Top Microbiol Immunol. 269:101-15. Review.

Kluge, R. C., R. S. Wicksman, and T. H. Weller. 1960. Cytomegalic inclusion disease of the newborn: report of case with persistent viruria. Pediatrics. 25:35-39.

Kohler, C. P., J. A. Kerry, M. Carter, V. P. Muzithras, T. R. Jones, and R. M. Stenberg. 1994. Use of recombinant virus to assess human cytomegalovirus early and late promoters in the context of the viral genome. J.Virol. 68:6589-6597.

Kotenko, S. V., S. Saccani, L. S. Izotova, O. V. Mirochnitchenko, and S. Pestka. 2000. Human cytomegalovirus harbors its own unique IL-10 homolog (cmvIL-10). Proc.Natl.Acad.Sci.U.S.A 97:1695-1700.

Kuespert, K., S. Pils, and C. R. Hauck. 2006. CEACAMs: their role in physiology and pathophysiology. Curr.Opin.Cell Biol. 18:565-571.

Krug A, French AR, Barchet W et al. 2004. TLR9-dependent recognition of MCMV by IPC and DC generates coordinated cytokine responses that activate antiviral NK cell function. Immunity. 2004;21:107-119.

Landini, M. P. and P. M. La. 1991. Humoral immune response to human cytomegalovirus proteins: a brief review. Comp Immunol.Microbiol.Infect.Dis. 14:97-105.

Loenen W.A., Bruggeman C.A., Wiertz E.J. 2001. Immune evasion by human cytomegalovirus: lessons in immunology and cell biology. Semin Immunol. 13(1):41-9. Review

Lilley, B. N., H. L. Ploegh, and R. S. Tirabassi. 2001. Human cytomegalovirus open reading frame TRL11/IRL11 encodes an immunoglobulin G Fc-binding protein. J.Virol. 75:11218-11221.

Linding, R., R. B. Russell, V. Neduva, and T. J. Gibson. 2003. GlobPlot: Exploring protein sequences for globularity and disorder. Nucleic Acids Res. 31:3701-3708.

Llano, M., M. Guma, M. Ortega, A. Angulo, and M. Lopez-Botet. 2003. Differential effects of US2, US6 and US11 human cytomegalovirus proteins on HLA class Ia and HLA-E expression: impact on target susceptibility to NK cell subsets. Eur.J.Immunol. 33:2744-2754.

Lockridge, K. M., S. S. Zhou, R. H. Kravitz, J. L. Johnson, E. T. Sawai, E. L. Blewett, and P. A. Barry. 2000. Primate cytomegaloviruses encode and express an IL-10-like protein. Virology. 268:272-280.

Lopez-Botet, M., A. Angulo, and M. Guma. 2004. Natural killer cell receptors for major histocompatibility complex class I and related molecules in cytomegalovirus infection. Tissue Antigens 63:195-203.

Mach, M., B. Kropff, M. P. Dal, and W. Britt. 2000. Complex formation by human cytomegalovirus glycoproteins M (gpUL100) and N (gpUL73). J.Virol. 74:11881-11892.

Machold RP, Wiertz EJ, Jones TR, Ploegh HL. 1997. The HCMV gene products US11 and US2 differ in their ability to attack allelic forms of murine major histocompatibility complex (MHC) class I heavy chains. J Exp Med. 185:363-366.

Magri, G., A. Muntasell, N. Romo, A. Saez-Borderias, D. Pende, D. E. Geraghty, H. Hengel, A. Angulo, A. Moretta, and M. Lopez-Botet. 2011. NKp46 and DNAM-1 NK-cell receptors drive the response to human cytomegalovirus-infected myeloid dendritic cells overcoming viral immune evasion strategies. Blood 117:848-856.

Margulies, B. J., H. Browne, and W. Gibson. 1996. Identification of the human cytomegalovirus G protein-coupled receptor homologue encoded by UL33 in infected cells and enveloped virus particles. Virology. 225:111-125.

McCormick A.L., Mocarski Jr. E.S. 2007. Viral modulation of the host response to infection.In: Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K, editors. Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge: Cambridge University Press; 2007. Chapter 21.

Mocarski E.S., Courcelle C.T. 2001. Cytomegaloviruses and their replication. In: Knipe DM HPGDLR, ed. fields virology. Philadelphia: Lippincott Williams & Wilkins. 2629-2673.

Mocarski E.S, Shenk T, Pass RF. 2007. Cytomegaloviruses. In: Knipe DM, Howley PM (eds). Fields Virology, vol. 2. Lippincott Williams & Wilkins: Philadelphia, pp. 2701–2772.

Murphy, E., D. Yu, J. Grimwood, J. Schmutz, M. Dickson, M. A. Jarvis, G. Hahn, J. A. Nelson, R. M. Myers, and T. E. Shenk. 2003. Coding potential of laboratory and clinical strains of human cytomegalovirus. Proc.Natl.Acad.Sci.U.S.A 100:14976-14981.

Murphy E, Shenk T. 2008. Human cytomegalovirus genome. Curr Top Microbiol Immunol. 325:1-19.

Nielsen, H., J. Engelbrecht, S. Brunak, and H. G. von. 1997. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. Protein Eng 10:1-6.

Park B., Kim Y. and Shin J. 2004. Human cytomegalovirus inhibits tapasin-dependent peptide loading and optimization of the MHC class I peptide cargo for immune evasion. Immunity. 20:71-85.

Park B., Spooner E., Houser B.L., Strominger J.L., Ploegh H.L. 2010. The HCMV membrane glycoprotein US10 selectively targets HLA-G for degradation. J Exp Med. 207(9):2033-41.

Patrone, M., M. Secchi, L. Fiorina, M. Ierardi, G. Milanesi, and A. Gallina. 2005. Human cytomegalovirus UL130 protein promotes endothelial cell infection through a producer cell modification of the virion. J. Virol. 79:8361-8373.

Penfold, M. E., D. J. Dairaghi, G. M. Duke, N. Saederup, E. S. Mocarski, G. W. Kemble, and T. J. Schall. 1999. Cytomegalovirus encodes a potent alpha chemokine. Proc.Natl.Acad.Sci.U.S.A 96:9839-9844.

Pruitt, K. D., T. Tatusova, W. Klimke, and D. R. Maglott. 2009. NCBI Reference Sequences: current status, policy and new initiatives. Nucleic Acids Res. 37: D32-36.

Quinnan, G. V., Jr., M. Delery, A. H. Rook, W. R. Frederick, J. S. Epstein, J. F. Manischewitz, L. Jackson, K. M. Ramsey, K. Mittal, S. A. Plotkin, and . 1984. Comparative virulence and immunogenicity of the Towne strain and a nonattenuated strain of cytomegalovirus. Ann.Intern.Med. 101:478-483.

Reusser, P., S. R. Riddell, J. D. Meyers, and P. D. Greenberg. 1991. Cytotoxic T-lymphocyte response to cytomegalovirus after human allogeneic bone marrow transplantation: pattern of recovery and correlation with cytomegalovirus infection and disease. Blood 78:1373-1380.

Ripalti, A. and Mocarski, E. S. 1991. The products of human cytomegalovirus genes UL1–UL7, including gp48, are dispensable for growth in cell culture. In Prog Cytomegalovirus Research: Proc Third International Cytomegalovirus Workshop, pp. 57–62. Edited by M. P. Landini. Amsterdam: Elsevier.

Robson L., Gibson W. 1989. Primate cytomegalovirus assembly protein: genome location and nucleotide sequence. J Virol. 63(2):669-7

Ryckman B.J., Jarvis M.A., Drummond D.D., Nelson J.A., Johnson D.C. 2006. Human cytomegalovirus entry into epithelial and endothelial cells depends on genes UL128 to UL150 and occurs by endocytosis and low-pH fusion. J Virol. 80(2):710-22.

Ryckman, B. J., B. L. Rainish, M. C. Chase, J. A. Borton, J. A. Nelson, M. A. Jarvis, and D. C. Johnson. 2008. Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells. J.Virol. 82:60-70.

Sanchez, V., K. D. Greis, E. Sztul, and W. J. Britt. 2000a. Accumulation of virion tegument and envelope proteins in a stable cytoplasmic compartment during human cytomegalovirus replication: characterization of a potential site of virus assembly. J.Virol. 74:975-986.

Sanchez, V., E. Sztul, and W. J. Britt. 2000b. Human cytomegalovirus pp28 (UL99) localizes to a cytoplasmic compartment which overlaps the endoplasmic reticulum-golgi-intermediate compartment. J.Virol. 74:3842-3851.

Sekulin, K., I. Gorzer, D. Heiss-Czedik, and E. Puchhammer-Stockl. 2007. Analysis of the variability of CMV strains in the RL11D domain of the RL11 multigene family. Virus Genes 35:577-583.

Silva, M. C., Q. C. Yu, L. Enquist, and T. Shenk. 2003. Human cytomegalovirus UL99-encoded pp28 is required for the cytoplasmic envelopment of tegument-associated capsids. J.Virol. 77:10594-10605.

Sinzger, C., M. Digel, and G. Jahn. 2008a. Cytomegalovirus cell tropism. Curr.Top.Microbiol.Immunol. 325:63-83.

Sinzger, C., G. Hahn, M. Digel, R. Katona, K. L. Sampaio, M. Messerle, H. Hengel, U. Koszinowski, W. Brune, and B. Adler. 2008b. Cloning and sequencing of a highly productive, endotheliotropic virus strain derived from human cytomegalovirus TB40/E. J.Gen.Virol. 89:359-368.

Sinzger, C., K. Schmidt, J. Knapp, M. Kahl, R. Beck, J. Waldman, H. Hebart, H. Einsele, and G. Jahn. 1999. Modification of human cytomegalovirus tropism through propagation in vitro is associated with changes in the viral genome. J.Gen.Virol. 80 (Pt 11):2867-2877.

Slevogt, H., S. Zabel, B. Opitz, A. Hocke, J. Eitel, P. D. N'guessan, L. Lucka, K. Riesbeck, W. Zimmermann, J. Zweigner, B. Temmesfeld-Wollbrueck, N. Suttorp, and B. B. Singer. 2008. CEACAM1 inhibits Toll-like receptor 2-triggered antibacterial responses of human pulmonary epithelial cells. Nat. Immunol. 9:1270-1278.

Spaderna, S., H. Blessing, E. Bogner, W. Britt, and M. Mach. 2002. Identification of glycoprotein gpTRL10 as a structural component of human cytomegalovirus. J.Virol. 76:1450-1460.

Spaderna, S., G. Hahn, and M. Mach. 2004. Glycoprotein gpTRL10 of human cytomegalovirus is dispensable for virus replication in human fibroblasts. Arch.Virol. 149:495-506.

Spaderna, S., B. Kropff, Y. Kodel, S. Shen, S. Coley, S. Lu, W. Britt, and M. Mach. 2005. Deletion of gpUL132, a structural component of human

cytomegalovirus, results in impaired virus replication in fibroblasts. J.Virol. 79:11837-11847.

Spiller OB, Morgan BP, Tufaro F, Devine DV. 1996. Altered expression of host-encoded complement regulators on human cytomegalovirus-infected cells. Eur J Immunol. 26(7):1532-8.

Stanton, R. J., K. Baluchova, D. J. Dargan, C. Cunningham, O. Sheehy, S. Seirafian, B. P. McSharry, M. L. Neale, J. A. Davies, P. Tomasec, A. J. Davison, and G. W. Wilkinson. 2010. Reconstruction of the complete human cytomegalovirus genome in a BAC reveals RL13 to be a potent inhibitor of replication. J.Clin.Invest 120:3191-3208.

Sun Z, Ren G, Ma Y, Wang N, Ji Y, Qi Y, Li M, He R, Ruan Q. 2010. Transcription pattern of UL131A-128 mRNA in clinical strains of human cytomegalovirus. J Biosci. Sep;35(3):365-70.

Talbot, P. and J. D. Almeida. 1977. Human cytomegalovirus: purification of enveloped virions and dense bodies. J.Gen.Virol. 36:345-349.

Tan, K., B. D. Zelus, R. Meijers, J. H. Liu, J. M. Bergelson, N. Duke, R. Zhang, A. Joachimiak, K. V. Holmes, and J. H. Wang. 2002. Crystal structure of murine sCEACAM1a[1,4]: a coronavirus receptor in the CEA family. EMBO J. 21:2076-2086.

Thompson, J. D., T. J. Gibson, and D. G. Higgins. 2002. Multiple sequence alignment using ClustalW and ClustalX. Curr.Protoc.Bioinformatics.

Tomasec, P., V. M. Braud, C. Rickards, M. B. Powell, B. P. McSharry, S. Gadola, V. Cerundolo, L. K. Borysiewicz, A. J. McMichael, and G. W. Wilkinson. 2000. Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. Science 287:1031.

Trgovcich, J., C. Cebulla, P. Zimmerman, and D. D. Sedmak. 2006. Human cytomegalovirus protein pp71 disrupts major histocompatibility complex class I cell surface expression. J.Virol. 80:951-963.

Varnum, S. M., D. N. Streblow, M. E. Monroe, P. Smith, K. J. Auberry, L. Pasa-Tolic, D. Wang, D. G. Camp, K. Rodland, S. Wiley, W. Britt, T. Shenk, R. D. Smith, and J. A. Nelson. 2004. Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. J.Virol. 78:10960-10966.

Wagner, M., A. Gutermann, J. Podlech, M. J. Reddehase, and U. H. Koszinowski. 2002. Major histocompatibility complex class I allele-specific cooperative and competitive interactions between immune evasion proteins of cytomegalovirus. J. Exp. Med. 196:805-816.

Wang, D., and T. Shenk. 2005. Human cytomegalovirus virion protein complex required for epithelial and endothelial cell tropism. Proc. Natl. Acad. Sci. U.S.A. 102: 18153-18158.

Waterhouse R., Ha C., Dveksler G.S. 2002. Murine CD9 is the receptor for pregnancy-specific glycoprotein 17. J Exp Med. 195(2):277-82.

Willcox B.E., Thomas L.M, Bjorkman P.J. 2003. Crystal structure of HLA-A2 bound to LIR-1, a host and viral major histocompatibility complex receptor. Nat Immunol. 4:913-919.

Wilkinson G.W., Tomasec P., Stanton R.J., Armstrong M., Prod'homme V., Aicheler R., McSharry B.P., Rickards C.R., Cochrane D., Llewellyn-Lacey S., Wang E.C., Griffin C.A., Davison A.J. 2008. Modulation of natural killer cells by human cytomegalovirus. J Clin Virol. 41(3):206-12. Review.

Wright, J. F., A. Kurosky, E. L. Pryzdial, and S. Wasi. 1995. Host cellular annexin II is associated with cytomegalovirus particles isolated from cultured human fibroblasts. J.Virol. 69:4784-4791.

Yang Z. 1997. PAML: a program package for phylogenetic analysis by maximum likelihood. Comput Appl Biosci. 13(5):555-6.

Yao, F. and R. J. Courtney. 1992. Association of ICP0 but not ICP27 with purified virions of herpes simplex virus type 1. J.Virol. 66:2709-2716.

Yu, D., M. C. Silva, and T. Shenk. 2003. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. Proc.Natl.Acad.Sci.U.S.A 100:12396-12401.

Yu, D., G. A. Smith, L. W. Enquist, and T. Shenk. 2002. Construction of a self-excisable bacterial artificial chromosome containing the human cytomegalovirus genome and mutagenesis of the diploid TRL/IRL13 gene. J.Virol. 76:2316-2328.