Title
Proteomic Study of Human Cytomegalovirus Using an Epitope Tag System

Author
Yang, Edward

Publication Date
2011

Peer reviewed|Thesis/dissertation

Proteomic Study of Human Cytomegalovirus Using an Epitope Tag System

By

Edward Yang

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Comparative Biochemistry

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Fenyong Liu, Chair
Professor George Sensabaugh
Professor Laurent Coscoy

Fall 2011

Abstract

Proteomic Study of Human Cytomegalovirus Using an Epitope Tag System

by

Edward Yang

Doctor of Philosophy in Comparative Biochemistry

University of California, Berkeley

Professor Fenyong Liu, Chair

Human Cytomegalovirus (HCMV) is a ubiquitous pathogen that can cause significant morbidity in neonates and immunodeficient individuals such as, AIDS patience and organ transplant recipients. With a genome size of 190-230 kilobases in size and an estimated 160 or more protein-encoding genes, HCMV is currently the largest known virus to infect humans. The functions of many of HCMV's genes remain unknown and only by understanding their role in viral infection and replication will we be able to successfully develop therapies to combat the disease.

Using a Bacterial Artificial Chromosome (BAC) containing the genome of HCMV, we constructed over 100 recombinant viruses, each with a single open reading frame (ORF) expressing its protein with an epitope tag at the C terminus end. Using these viruses, we were able to identify subcellular localization for 82 ORFs at 72 hours post infection. Our study found 17 proteins to localize to the nucleus, 51 to a juxtanuclear structure in the cytoplasm, 11 to the cytoplasm, and three to an uncharacterized structure in the cytoplasm.

We then conducted an extensive study of HCMV ORF US20, which exhibited a unique cytoplasmic localization in the previous global localization study. The US20 ORF is a member of the US12 gene family and is present only in cytomegaloviruses strains that infect rhesus monkeys, chimpanzees, and humans, suggesting that the ORF is an important factor for the infection of upper primates. We found the US20 ORF to express a 26-28kDa- seven transmembrane (TM) domain protein during the early/delayed early phase of viral replication, which localized to a cytoplasmic structure during the late phase of replication. This localization was found to be independent of late gene expression. Through immunofluorescence (IFA) studies, we found the US20 protein to be present in early endosomes but not in the ER or TGN, suggesting that the protein is present in membranes of endosomes. However, we do not believe that the protein traffics to the surface of the cell, which we showed by selective membrane permeabilization. The US20 protein was also determined to form homo-dimers, however the function of the dimerization is still unknown. Using viruses expressing truncated forms of US20, which replicated at levels similar to the wild-type, we found that deleting TM5-TM7 significantly reduced US20 protein levels, deleting TM4-TM7 abolished dimerization, and deleting TM2-TM7 destabilized the localization of US20 in the cytoplasm. Through immunoprecipitation and mass spectrometry, we found US20 to associate with valosin containing protein, sodium potassium ATPase, and succinate dehydrogenase. Currently, under

the context of infection we have confirmed the interaction of US20 with valosin containing protein (VCP), which has been found to be associated with some neurodegenerative diseases. This interaction presents the hypothesis that US20 may play a role in HCMV congenital infection of the brain.

Lastly, we examined the role of a predicted SUMOylation site located near the C terminus tail of the HCMV processivity factor UL44, which forms a C clamp around the viral DNA. SUMOylation is a post-translation modification that has been found to be involved in various cellular systems, including cell cycle regulation, transcription, cellular localization, degradation, and chromatin organization. In our study, we found that mutating the lysine residue of the conserved SUMOylation motif, ψKxE, at position 410 of UL44 to an alanine did not impair viral DNA synthesis or viral replication. This result indicates that the lysine residue of UL44 is not important for replication of HCMV in fibroblasts.

The results of this dissertation bring new insight into many of HCMV's genes, specifically US20, which prior to this study had not been extensively characterized. The data I present here provides new information about US20 and generates new hypotheses relating to its function. Furthermore, my work here can serve as a foundation for the study of other genes in the US12 gene family.

This work is dedicated to my parents, Ted and Angela Yang, for their love and support and for demonstrating through their own lives that anything can be achieved with hard work and perseverance. You guys are truly an inspiration for me.

Table of Contents

List of Figures and Tables

Chapter 1

Chapter 2

Chapter 3

Figures (Continued)

Appendix Tables

Chapter 4

Figures

Acknowledgments

First and foremost I would like to thank my advisor, Professor Fenyong Liu, for providing me with a great environment to pursue my research interests. Words cannot express how much I appreciate this experience that started with me entering the lab as an undergraduate researcher and leaving with a Ph.D. education.

Special thanks to my committee members, Professor George Sensabaugh, Bing Jap, Barry Shane, Caroline Kane, and Laurent Coscoy for their time and feedback.

Thanks must be given to Paul Rider, who has been a "brother in arms" that kept me mentally sane when experiments were not working and made sure my immaturity during my early graduate years stayed in check.

I'm grateful to Sean Umamoto for being a great friend, who was always willing to lend a hand.

I would also like to thank my first graduate mentor, Rong Hai, for being a great patient teacher at the bench and in the hood.

Special thanks to all of the undergraduate researchers I've had the opportunity to work with, Heather Pines, Angela Zhou, Tess Wiskel, Karen Chang, Ronald Lee, Tiffany Wu, Gary Chan, Yejin Choi, and John Kim. Much of this work would not have been possible without their hard work and dedication.

I am extremely grateful to Mike Meighan for always having a teaching position open for me and the staff of Bio1AL, Erol Kepkep, Jim Sharkey, and Amy DeHart, for creating a fun teaching environment that helped me become a better instructor.

Many thanks to my parents Ted and Angela Yang for their continuous support and unwavering belief in my ability to finish this work, even when I had my own doubts.

Thanks to my sister, Jennifer Yang, for not only laboring beside me for one summer but for being an awe-inspiring person in general.

Last, but definitely not least, I would like to thank my fiancé, Zhimin Lin, for her love, support and patience. Her belief in me kept me strong and for that this dissertation is as much mine as it is hers.

Chapter 1

Introduction

ABSTRACT

Human Cytomegalovirus (HCMV) is a ubiquitous human pathogen that causes significant morbidity and mortality in immunodeficient individuals, including AIDS patients, transplant recipients, and neonates. Despite the significant amount of work accomplished to understand the virus' genes and the role they play in infection, the function of many of the 160 or more open reading frames (ORFs) remains unknown. This review will present an overview of the *herpesviridae* family followed by a brief review of HCMV, covering the virus' associated diseases and ORFs in relation to its impact on viral replication in a tissue culture model. Lastly, this review will also examine the approaches taken by researchers in the field to construct mutants in an effort to dissect the function of ORFs, their transcripts, and protein products.

INTRODUCTION

Human Cytomegalovirus (HCMV) is a ubiquitous pathogen that can cause serious morbidity and mortality in neonates and individuals with weakened immune systems. Given the severity of the diseases associated with the virus, it is critical for researchers to understand the function of viral genes, their transcripts, and protein products. Only by understanding the mechanisms governing infection and replication of the virus at the molecular level will researchers be able to successfully develop therapies and vaccines to combat HCMV.

HUMAN HERPESVIRUSES

The first human herpesvirus to be isolated was herpes simplex virus (HSV) by Wilhelm Grüter, who demonstrated the infectious nature of HSV in a series of animal studies in the late 19th Century [1]. This initial discovery, coupled with advancements in deoxyribonucleic acid (DNA) sequencing, lead to an explosion in discovery of other human and non-human infecting herpesviruses during the 20th Century [2]. Currently, more than 200 species of herpesvirus have been identified and although they may differ in sequence, primary host, and clinical manifestation of diseases, there are biological properties that they all share [2]. These properties include: the expression of virally-encoded proteins involved in nucleic acid metabolism, DNA synthesis, and protein processing; the synthesis of viral DNA and capsid assembly in the nucleus with final processing in the cytoplasm; and the destruction of the infected cell following production of virion [2]. The most important characteristic shared by all members of the *herpesviridae* family is the ability of the virus to enter a state of infection called latency, which occurs when the genome of the virus is maintained in the infected cell with minimal viral gene expression and no virion production [2]. This has made treatment of herpesvirus infection very difficult. Latency is linked with another phase of infection called reactivation where, the latently infected cells begin to shed virus again [2]. The properties of latency and reactivation are of extreme interest to researchers, especially to those studying the clinical manifestations associated with herpesvirus capable of infecting humans. As of yet, eight herpesviruses have been identified to infect humans, these include herpes simplex 1 (HSV-1), herpes simplex virus 2 (HSV-2), varicella-zoster virus (VZV), Epstein-Barr virus (EBV), human cytomegalovirus (HCMV), herpesvirus 6 and 7, and Kaposi's sarcoma-associated herpesvirus [2]. Individuals who become infected with any of the eight herpesviruses will become carriers for life.

The human herpesviruses have been further classified into subfamilies: *alphaherpesvirinae* (α), *betaherpesvirinae* (β) or *gammaherpesvirinae* (γ) (Table 1.1). *Alphaherpesvirinae* herpesviruses are characterized by their variable host range, relatively short reproductive cycle, rapid spread in culture, efficient destruction of infected cells, and capacity to establish latent infections primarily but not exclusively in sensory ganglia [2]. Members of the *betaherpesvirinae* have a restricted host range, a long reproductive cycle, and a slow infection progress in culture. Furthermore, members of this subfamily can establish a latent infection in secretory glands, lymphoreticular cells, kidneys, and other tissues [2]. The last subfamily is the *gamma herpesviruses* which are characterized by their ability to replicate and establish latency in lymphoblastoid cells and cause malignancies in persistently-infected individuals [2, 3].

Despite varying in size from 120 to 260 nanometers in diameter, all herpesvirus virion share a similar composition [2]. Each virion contains an electro-opaque core containing the double-stranded-linear viral DNA, an icosahedral capsid surrounding the core, a proteinaceous

Designation	Vernacular name	Subfamily
Human herpesvirus-1	Herpes Simplex virus 1	α
Human herpesvirus-2	Herpes Simplex virus 2	
Human herpesvirus-3	Varicella-zoster virus	
Human herpesvirus-5	Cytomegalovirus	β
Human herpesvirus-6	HHV-6 variant A	
Human herpesvirus-7	HHV-6 variant B	
Human herpesvirus-4	Epstein-Barr virus	γ
Human herpesvirus-8	Kaposi's sarcoma associated herpesvirus	

Table 1.1 **Table of human herpesvirus subfamily and members**

layer surrounding the capsid called the tegument, and an outer lipid bilayer called the envelope [2]. Interestingly, in addition to viral proteins, host-derived proteins and RNAs have also been identified in the virion although their role in infection is not well understood [4, 5].

HUMAN CYTOMEGALOVIRUS

Human Cytomegalovirus (HCMV), also known as Human Herpesvirus 5, is one of the eight known herpesviruses capable of causing diseases in humans. While the virion was first isolated independently by Margaret Smith, Thomas Weller, and Wallace Rowe in the 1950s, the typical intranuclear inclusions associated with the virus were observed as early as 1881 by German scientists [6]. Following isolation, the name "cytomegalovirus" was then coined by Weller and his coworkers in 1960 [7, 8].

HCMV associated diseases

HCMV infection is considered to be ubiquitous with developed countries such as the United States and Europe having a 40-60% seropositive rate and less developed countries, including Southeastern Asia, having a seropositive rate closer to 100% [6]. The route of infection of children commonly occurs from either virus shedding in the cervix of the mother during birth or exposure to the virus in breast milk, saliva, and urine during childhood [6]. In adults, transmission is more common through exchange of sexual fluids or blood [9]. As previously mentioned, individuals infected with herpesviruses will become carriers for life and HCMV is no exception. However, despite its prevalence, individuals will remain asymptomatic as long as they remain immunocompetent [9].

Cytomegalovirus infection can become a significant concern for those individuals who are immunocompromised, such as patients with acquired immune deficiency syndrome (AIDS) or organ transplant recipients [6]. For individuals infected with AIDS, the weakened immune system is believed to cause reactivation of the virus leading to clinical manifestations such as HCMV retinitis, which can cause vision loss and blindness [10]. In organ transplant recipients, HCMV-associated disease can be caused by the combination of receiving an organ from a seropositive donor and using immunosuppressive treatments, which may lead to reactivation of the latent virus in the transplanted organ or the recipient [9]. The reactivation and subsequent lytic infection can cause cytomegalovirus (CMV) pneumonia in patients [11] and has been linked to dysfunction or rejection of the transplanted organ [9].

Congenital cytomegalovirus infection is also a major concern and a recognized public health problem given that CMV is "*estimated to be the leading cause both of sensorineural deafness and of infectious brain damage in children in the United States*" [9]. Observed clinical manifestations in these infected neonates include mental retardation, impaired vision, cerebral palsy, and hearing loss [9]. Given the severity of the diseases associated with the HCMV, research to understand the molecular mechanisms governing infection and replication of the virus is critical to the development of treatments.

HCMV Genome and Genes

With a genome size ranging from 196 to 241 kilobase pairs, HCMV has the largest genome of any virus known to infect humans and is estimated to have more than 160 protein-

coding genes [9]. HCMV gene expression, like all herpesviruses, progresses in a cascade fashion, with each class dependent on the previous class. The immediate early (IE) genes are expressed immediately after viral entry and then followed by the early (E) / delayed early(DE) genes which are typically expressed 4-6 hours post-infection (hpi) and continue through to 24 hpi [9]. The Late (L) genes are then expressed from 24 hpi and beyond. From viral entry to release of progeny at 48 to 72 hours, HCMV's replication cycle is consider relatively slow when compared to HSV-1's 20 hour replication cycle [12].

Recombinant HCMV with gene knockouts have been studied in an *in vitro* cell culture model to determine their importance in HCMV's replication cycle. This review presents a brief overview of the gene knockout study involving the HCMV Towne strain [13], which was initially isolated from the urine of a 2-month-old infant [14]. If an essential open reading frame (ORF) was deleted, no plaque formation or viral progeny production would occur. Depending on the growth phenotypes in cultured fibroblasts, the deletion mutants that were nonessential for viral replication *in vitro* were classified into four subclasses. Those mutants that showed defective growth relative to the parental strain were divided into two groups: "severe defective growth" and "moderate defective growth". Mutants with deletions that had no consequence on viral growth were classified as "dispensable". Finally, mutants that showed an increase in viral growth relative to the parental strain were labeled "enhanced growth" (Table 1.2) [13]. Classifying the genes in this manner can help prioritize therapeutic targets for researchers focused on drug development.

Essential genes

Currently, 45 ORFs of the HCMV Towne strain have been identified to be essential for replication in cultured fibroblasts (Table 1.1). Similar ORFs were also found to be essential for growth for the HCMV AD169 strain by Yu *et al*., who used a transposon mutagenesis approach [15]. Minor differences between the two studies may be a result of either the genomic differences between the AD169 and Towne strains, the impact of deleting portions of the target ORF's neighboring genes, or the difference in mutagenesis strategies employed to generate the recombinant virus. Of the identified essential ORFs, 35 are core genes that are conserved in all herpesviruses, seven are conserved within the beta-herpesvirus family, and three are unique to CMV [9]. In addition, the function of many of these genes have been identified by either individual studies of the gene or extrapolated from studies of their HSV-1 homologous counterparts [16]. A vast majority of the proteins expressed from these genes have also been found to be present in the virion itself [4]. Here I present a brief overview of the identified essential ORFs and their roles in viral replication and infection of human foreskin fibroblasts (Figure 1.1).

<u>Capsid and Capsid Assembly</u>

As expected, the core capsid proteins are essential for replication since the lack of any of these proteins would result in an incomplete formation of the nucleocapsid. The roughly 1300 Å diameter icosahedral nucleocapsid [17] is made up of five unique proteins: UL46 (TRI1), UL85 (TRI2), UL86 (MCP), UL48A (SCP), and UL104 (PORT) [9]. In addition, the UL80 ORF that encodes the protease (PR)-assembly protein (AP) precursor (pPR), which facilitates capsid assembly, is also necessary for replication. This "no growth" phenotype was also confirmed by

ORF	Function/Comments	Function Family	Homology	Virion Location
No growth (Essential)				
UL 32	Encodes pp150, virion maturation in Assembly Compartment	Cytoplasmic Maturation	β-herpes	Tegument
UL 34	Represses US3 expression		CMV	
UL 37.1	Antiapoptosis	Antiapoptosis	β-herpes/CMV	
UL 44	Processivity factor	DNA replication	Core	Tegument
UL 46	Capsid Protein (TRI1)	Capsid/ Capsid Assembly	Core	Nucleocapsid
UL 48	Deubiquitinating protease	(Unknown)	Core	Tegument
UL 48A	Small Capsid Protein(SCP)	Capsid/ Capsid Assembly	Core	Nucleocapsid
UL 49	Unknown Function	(Unknown)	Core	In Virion
UL 50	Inner Nuclear Membrane Protein	Nuclear Egress	Core	Tegument
UL 51	Terminase Component	DNA Encapsidation/Cleavage	Core	In Virion
UL 52	Cleaveage/Packaging of viral genome	DNA Encapsidation/Cleavage	Core	
UL 53	Nuclar Matrix Protein	Nuclear Egress	Core	Tegument
UL 54	DNA Polymerase	DNA replication	Core	Tegument
UL 55	Glycoprotein B, viral entry, cell-to-cell spread	Glycoprotein / Viral entry	Core	Envelope
UL 56	Large Subunit of HCMV Terminase / DNA packaging	DNA Encapsidation/Cleavage	Core	
UL 57	ssDNA binding protein	DNA Replication	Core	Tegument
UL 60	Unknown Function	(Unknown)	CMV	
UL 70	DNA helicse primase subunit	DNA replication	Core	
UL 71	Unknown Function	(Unknown)	Core	Tegument
UL 73	Virion envelope glycoprotein (gN)	Glycoprotein / viral entry	Core	Envelope
UL 75	Glycoprotein gH/ complexes with gO	Glycoprotein / viral entry	Core	Envelope
UL 76	Virion Associated regulatory protein / inhibits viral replication	Regulatory Protein	Core	Tegument
UL 77	Portal Capping Protein	DNA Encapsidation/cleavage	Core	Tegument
UL 79	Unknown function	(Unknown)	Core	Tegument
UL 80	Precursor of maturational protease	Capsid Assembly	Core	In Virion
UL 84	UTPase,RNA-binding within oriLyt, regulatory, shuttling protein	DNA replication	β-herpes	Tegument
UL 85	Capsid Protein (TRI2)	Capsid Protein	Core	Nucleocapsid
UL 86	Major Capsid Protein (MCP)	Capsid Protein	Core	Nucleocapsid
UL 87	TATT- binding protein / function unknown	(Unknown)	Core	
UL 89.1	Small subunit of HCMV terminase	DNA Encapsidation/cleavage	Core	In Virion
UL 90	Unknown Function	(Unknown)	CMV	
UL 91	Unknown Function	(Unknown)	β-herpes	
UL 92	Unknown Function	(Unknown)	β-herpes	
UL 93	Unknown Function	(Unknown)	Core	Tegument
UL 94	unknown function/possible nucleic acid binding	(Unknown)	Core	Tegument
UL 95	Unknown Function	(Unknown)	Core	
UL 96	Unknown Function	(Unknown)	β-herpes	Tegument
UL 98	Unknown Function	DNA Encapsidation/cleavage?	Core	
UL 99	Cytoplasmic egress tegument protein	Cytoplasmic Egress	Core	Tegument
UL 100	GlycoproteinM (gM) / gM:gN	Glycoprotein/ Entry	Core	Envelope
UL 102	Primase-associated factor	DNA replication	Core	
UL 104	Portal protein (PORT)	DNA Encapsidation/cleavage	Core	Nucleocapsid
UL 105	DNA Helicase	DNA replication	Core	
UL 115	Glycoprotein L (gL) / gO:gL	Glycoprotein/ Entry	Core	Envelope
UL 122	IE2, interacts transcriptional machinery	Gene Regulation	β-herpes	In VIrion

ORF	Function/Comments	Function Family	Homology	Virion Location
Severe Growth Defect (reduction of titers by 2 x 10^4)				
UL 21	Unknown function	(Unknown)	CMV	
UL 26	Stimulate major IE enhancer-promoter	Gene Regulation	CMV	Tegument
UL 28	Unknown function	(Unknown)	β-herpes	
UL 30	Unknown function	(Unknown)	CMV	
UL 69	nuclear export of RNA, induce cell cycle arrest in G1 phase	Gene Regulation	Core	Tegument
UL 82	virion transactivator/pp71	Gene Regulation	β-herpes	Tegument
UL 112	Major early protein	DNA Replication	β-herpes	Tegument
UL 113	Major early protein	DNA Replication	β-herpes	
UL 117	promote development of nuclear replication compartments	(Unknown)	β-herpes	
UL 123	Mjaor immediate early 1 co-transactivator	Gene Regulation	CMV	
UL 124	Latent transcript	(Unknown)	CMV	
US 26	Unknown function	(Unknown)	β-herpes	

Table 1.2. **HCMV Towne strain ORFs organized by their respective deletion virus phenotypes in human foreskin fibroblasts.**

ORF	Function/Comments	Function Family	Homology	Virion Location
Moderate growth defect (reduction of titers between 10^1 - 10^4)				
UL 2	Unknown Function	(Unknown)	CMV	
UL 11	Unknown Function/ transmembrane protein	(Unknown)	CMV	
UL 12	Unknown Function	(Unknown)	CMV	
UL 14	Unknown Function/ putative membrane protein	(Unknown)	CMV	
UL 20	Unknown Function / T-cell receptor homolog	(Unknown)	CMV	
UL 29	Unknown Function	(Unknown)	β-herpes	
UL 31	dUTPase Family	(Unknown)	β-herpes	
UL 35	Modulate IE expression/particle formation	Gene Regulation	β-herpes	Tegument
UL 38	suppress apoptosis	Antiapoptosis	β-herpes	Tegument
UL 47	Release of DNA from Capsid/Intracellular transport	Capsid Transport?	Core	Tegument
UL 65	Unknown Function	(Unknown)	CMV	
UL 72	Unknown Function/ inactive UTPase	(Unknown)	Core	Tegument
UL 74	virion-envelope glycoprotein O/role in virus release	Secondary Envelopment	β-herpes	Envelope
UL 88	Unknown Function	(Unknown)	β-herpes	Tegument
UL 97	Kinase/role in DNA synthesis/enhancement of gene expression	DNA packaing/ DNA replication / nuclear egress	Core	Tegument
UL 103	Unknown Function	(Unknown)	Core	Tegument
UL 108	Unknown Function	(Unknown)	CMV	
UL 114	Uracil-DNA glycosylase/DNA repair	DNA Replication?	Core	
UL 129	Unknown Function	(Unknown)	CMV	
UL 132	Unknown Function/ transmembrane protein	(Unknown)	CMV	In Virion
US 13	Unknown Function/ transmembrane protein	(Unknown)	CMV	
US 23	Unknown Function	(Unknown)	β-herpes	In Virion
TRS 1	dsRNA binding protein/transcription activator	Gene Regulation	CMV	Tegument

ORF	Function/Comments	Function Family	Homology	Virion Location
Growth like Wildtype				
UL 3	Unknown Function	(Unknown)	CMV	
UL 4	Unknown Fucntion	(Unknown)	CMV	
UL 5	Unknown Function	(Unknown)	CMV	In Virion
UL 6	Unknown Fucntion	(Unknown)	CMV	
UL 7	Unknown Function	(Unknown)	CMV	
UL 8	Unknown Fucntion	(Unknown)	CMV	
UL 10	Unknown Function	(Unknown)	CMV	
UL 13	Unknown Fucntion	(Unknown)	CMV	
UL 15	Unknown Function	(Unknown)	CMV	
UL 16	suppress NK cell recognition	Immunomodulation	CMV	
UL 17	Unknown Function	(Unknown)	CMV	
UL 18	suppress NK cell recognition/mimic MHC I	Immunomodulation	CMV	
UL 19	Unknown Function	(Unknown)	CMV	
UL 24	Unknown Function	(Unknown)	β-herpes	Tegument
UL 25	Unknown Function	(Unknown)	β-herpes	Tegument
UL 27	Unknown Function/Maribavir resistance	(Unknown)	β-herpes	
UL 33	Unknown Function/vGPCR	(Unknown)	β-herpes	In VIrion
UL 36	viral inhibitor of caspase-8 activation [vICA]	Antiapoptosis	β-herpes	Tegument
UL 37.3	novel MHC-like protein	Immunomodulation	β-herpes	
UL 39	Unknown Function	(Unknown)	CMV	
UL 42	Unknown Function	(Unknown)	CMV	
UL 43	Unknown Function	(Unknown)	β-herpes	Tegument
UL 45	R1 ribonucleotide reductase(RR) homolog	(Unknown)	Core	Tegument
UL 59	Unknown Function	(Unknown)	CMV	
UL 62	Unknown Function	(Unknown)	CMV	
UL 64	Unknown Function	(Unknown)	CMV	
UL 67	Unknown Function	(Unknown)	CMV	
UL 78	vGPCR	(Unknown)	CMV	
UL 83	suppress NK cell recognition/pp65	Immunomodulation	β-herpes	Tegument
UL 89.2	Terminase component	DNA encapsidation	Core	In Virion
UL 109	Unknown Function	(Unknown)	CMV	
UL 110	Unknown Function	(Unknown)	CMV	

Table 1.2 (*Continued*)

ORF	Function/Comments	Function Family	Homology	Virion Location
Growth like Wildtype (Continued)				
UL 110	Unknown Function	(Unknown)	CMV	
UL 111a	viral IL-10 homolog	Immunomodulation	CMV	
UL 116	Unknown Function	(Unknown)	CMV	
UL 119	binds Fc Region of IgG	Immunomodulation	CMV	In VIrion
UL 121	Unknown Function	(Unknown)	CMV	
UL 127	Unknown Function	(Unknown)	CMV	
UL 130	binds gH and gL/epithelial and endothelial tropism	Entry	CMV	
UL 146	CXC chemokine/ attract neutrophils	(Unknown)	CMV	
UL 147	hCXCR2-specific CXC chemokine	(Unknown)	CMV	
IRS	dsRNA binding protein/transcription activator	Gene Regulation	CMV	Tegument
US 14	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 15	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 16	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 17	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 18	oral mucosal tropism/ putative transmembrane protein	(Unknown)	CMV	
US 19	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 20	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 21	Unknown function/ putative transmembrane protein	(Unknown)	CMV	
US 22	Unknown Function	(Unknown)	β-herpes	Tegument
US 24	Unknown Function	(Unknown)	CMV	Tegument
US 25	Unknown Function	(Unknown)	CMV	
US 27	Chemokine receptor/ vGPCR	(Unknown)	CMV	In Virion
US 28	Chemokine receptor/ vGPCR	(Unknown)	CMV	
US 29	Unknown Function	(Unknown)	CMV	
US 31	Unknown Function	(Unknown)	CMV	
US 32	Unknown Function	(Unknown)	CMV	
US 33	Unknown Function	(Unknown)	CMV	
US 34	Unknown Function	(Unknown)	CMV	
RL 1	Unknown Function	(Unknown)	CMV	
RL 2	Unknown Function	(Unknown)	CMV	
RL 4	Unknown Function	(Unknown)	CMV	
RL 10	Transmembrane protein	(Unknown)	CMV	Envelope
RL 11	Transmembrane protein/ binds IgG Fc;	Immunomodulation	CMV	
RL 13	Transmembrane protein	(Unknown)	CMV	

ORF	Function/Comments	Function Family	Homology	Virion Location
Enhanced Growth (Increase of titers by 10 fold)				
UL 9	Unknown Function	(Unknown)	CMV	
UL 20a	CC chemokine binding protein	(Unknown)	CMV	
UL 23	Unknown Function	(Unknown)	β-herpes	Tegument
US 30	Transmembrane Protein	(Unknown)	CMV	

Table 1.2 (*Continued*)

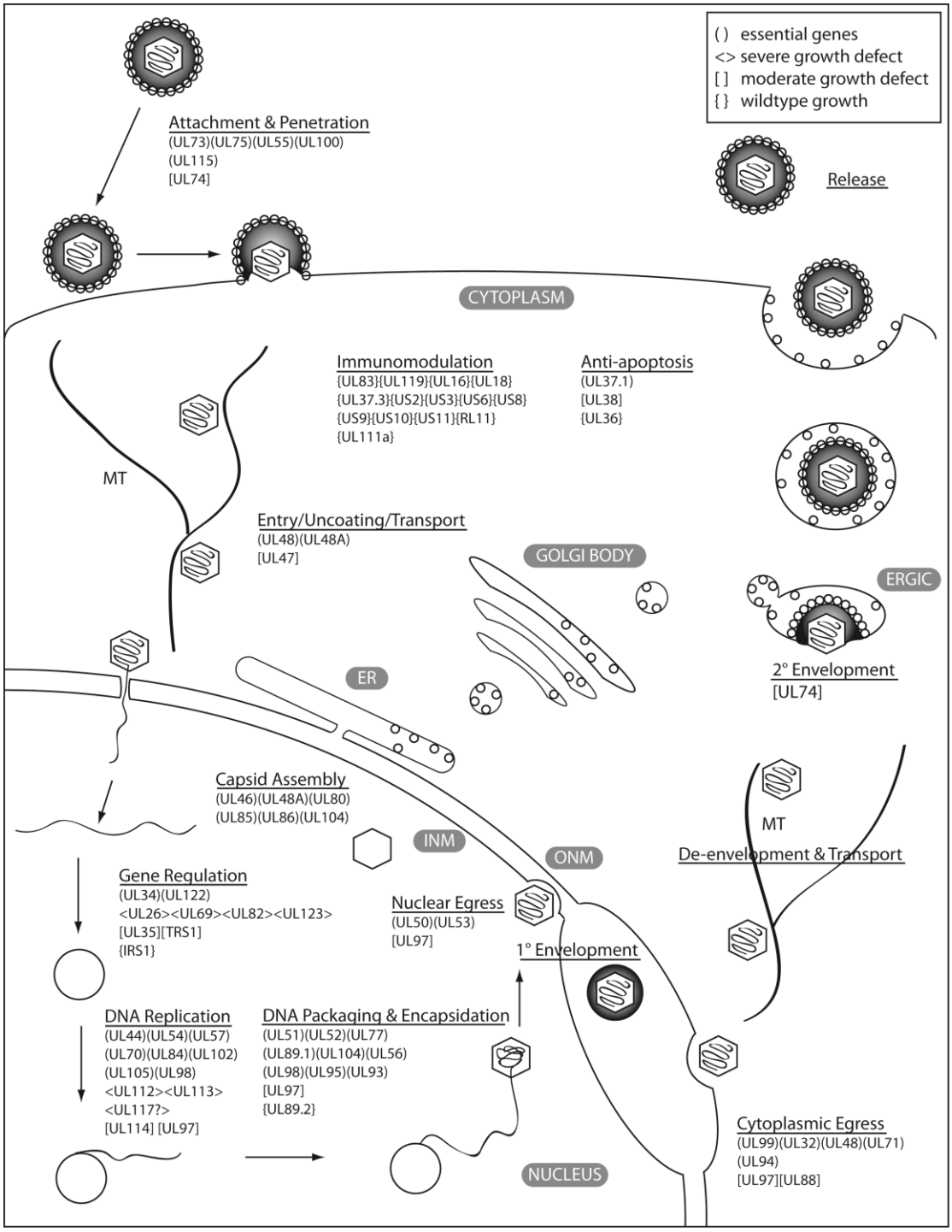


Figure 1.1. **HCMV replication cycle in fibroblasts**
The virion attaches to the cellular surface via glycoproteins in the envelope and then penetrates the cell. It is hypothesized that the capsid travels to the nucleus via the microtubule (MT) network, where it docks and releases the viral DNA at a nuclear pore complex. Upon release of DNA into the nucleus, products responsible for various functions including but not limited to, viral replication, immunomodulation, and inhibition of apoptosis, are then expressed.

Figure 1.1 (*Continued*)
The linear viral DNA circularizes and replicates via a rolling circle mechanism. The new DNA strands are then packaged into capsids formed within the nucleus. Nuclear egress of the nucleocapsid occurs by a primary envelopment and de-envelopment, which releases the nucleocapsid into the cytoplasm. In the cytoplasm, the nucleocapsid obtains its tegument proteins and buds into the Endoplasmic Reticulum-Golgi Intermediate Compartment (ERGIC) to obtain its envelope. After secondary envelopment, the mature virion-containing vesicles then fuse with the plasma membrane, resulting in the release of the progeny into the periphery. The roles of the ORFs in the replication cycle have either been shown or implicated in the literature. The growth phenotypes presented are based upon the infection of human foreskin fibroblasts by the ORF's respective deletion virus.

another study which used RNase P ribozyme-based inhibition to knock down pPR mRNA [18]. Despite the high homology that the core proteins share with their counterparts, it cannot always be assumed that if a core gene is identified as essential for one herpesvirus, its homologues are then essential for growth in other herpesviruses. This observation is exemplified by the herpesvirus small capsid protein (SCP). While UL48A, the HCMV SCP, was identified as essential for growth *in vitro* [13, 19], its HSV-1 homologue was found to be dispensable for virion maturation in tissue culture [20], thus indicating that while core proteins are conserved throughout all herpesviruses, their importance in viral infection, structure, and replication may not be the same.

Viral DNA Replication

The ORFs responsible for viral DNA replication are critical for the production of viral progeny. These ORFs include, UL84 and UL122 (IE2), which express transcriptional transactivators that form protein complexes across the *ori*Lyt to start viral replication [9]. Following the complex formed by UL84 and IE2, six core replication proteins are then recruited to the site of DNA synthesis. These proteins include: DNA polymerase (UL54); C-clamp DNA processivity factor (UL44); helicase (UL105); single-stranded DNA-binding protein (UL57), which is believed to aid in strand separation during DNA replication; primase (UL70); and primase-associated factor (UL102) [21]. The deletion of any of these genes results in a lack of viral progeny generation [13] .

DNA cleavage, Encapsidation and Egress

It is generally believed that HCMV DNA replication occurs by a rolling circle mechanism, which produces a concatemeric DNA strand that is then cleaved and threaded into the preformed capsid [22]. As the DNA is being replicated, both UL93 and UL52 are believed to facilitate the transport of the preformed capsid to the DNA replication compartment [9]. Following transport, the protein products of UL89 (TER1) and UL56 (TER2) interact with UL104 (PORT) to thread the viral DNA concatemer into the preformed capsid [9]. Once a single genome length has been inserted, the terminase complex formed by TER1 and TER2 will then cleave the DNA strand [9]. The portal is then covered by the portal capping protein homologue, UL77, marking the end of encapsidation [9]. Phosphorylation by UL97, the viral serine-threonine protein kinase, is believed to regulate the process of cleavage and encapsidation [9]. Although their function is not well characterized, UL51, UL95, and UL103 are also believed to participate in the encapsidation process. Of the 10 ORFs that participate in this process, all have been found to be essential for replication and viral growth *in vitro*, except UL97, whose respective deletion virus was able to grow in culture, but at significantly reduced levels [13].

Following encapsidation, egress from the nucleus is facilitated by two essential genes: UL50, a type 2 membrane protein, and UL53, a tegument protein, although the mechanism is still not well understood [9]. Once inside the cytoplasm, the nucleocapsid continues the maturation process by obtaining the tegument followed by the envelope. UL32 (pp150) and UL99, both of which play a role in virion maturation in the cytoplasm and cytoplasmic egress from the cell [9], have also been found to be essential for viral replication [13].

Envelope Glycoproteins

The HCMV virion has several glycoproteins that are embedded in the envelope, which is obtained during the secondary envelopment when the capsid buds into cytoplasmic vesicles [23]. The deletion of the ORFs encoding these glycoproteins that play significant roles in viral entry has been shown to result in no production of virion [13]. These ORFs included: UL115 (gL), which forms a complex with UL75 (gH); UL100 (gM), which interacts with UL73 (gN); and UL55 (gB), which is the major heparan sulfate binding protein, a cell surface protein that is conserved in herpesvirus entry pathways [9]. Furthermore, recent studies of gB show that while the protein is needed for viral entry and cell-to-cell spread, it is not needed for virion attachment, assembly, or egress [24].

Regulatory Proteins

Deletions in some ORFs that encode transcription factors have also been shown to inhibit production of virions [13]. These ORFs include: UL76, which binds to the major immediate-early promoter/enhancer (MIEP) [25] and has been shown to be a negative gene regulator [26]; UL34, which was shown to repress expression of US3, an immune evasion gene [27]; and the previously mentioned UL122, which encodes IE2. Besides UL122, neither the mechanisms that utilize these ORFs nor their role in viral replication are well understood. In addition to these three regulatory proteins, anti-apoptotic regulatory gene UL 37.1 has also been shown to be essential for viral replication and growth [13, 28].

Severe Defective Growth-related ORFs

12 ORFs have been identified to cause a reduction in plaque titers of their respective deletion viruses by at least 10,000-fold relative to the wild-type strain (Table 1.2) [13]. Of these 12, only seven genes have been characterized (Figure 1.1). These include the three transcription-factor-expressing ORFs: UL82 (pp71), a well-studied tegument protein that regulates the activity of MIEP [9]; UL26, which influences phosphorylation of tegument proteins [29] and transactivation of HCMV MIEP [30]; and UL123 (IE1), which is responsible for a range of activity, including delayed early and late protein expression and disruption of ND-10, a matrix-assisted nuclear structure [9].

Reductions in viral growth among viruses with deletions of either UL69, which has a role in viral RNA nuclear export and virally induced cell cycle arrest [31], or UL117, which has been found to promote the development of nuclear replication compartments [32], have also been observed [13]. Lastly, the severe defective growth phenotype was observed for viruses with deletions in UL112 and UL113 loci, which express multiple peptides that recruit UL44 for viral replication [33].

Moderate Defective Growth Mutants-related ORFs

A phenotype labeled “moderate defective growth” has been observed for 23 ORFs that when deleted caused peak titers of 10-10,000 times less than then the wild-type in fibroblasts (Table 1.2) [13]. Similar to the “severe defective growth” ORFs, these genes may play a role in mechanisms that enhance expression of the essential genes or provide additional stability to the

virion. Of the 23 ORFs, seven have been determined or predicted to be glycoproteins (UL2, UL11, UL14, UL38, UL74, UL132 and US13), although many have not been well characterized. Recent studies indicate that UL38 functions as an inhibitor of apoptosis [34], while UL74 (gO) encodes a virion envelope protein that contributes to secondary envelopment [35]. A decrease in viral replication was also observed in viruses with deletions of UL31, which encodes a protein with dUTPase motifs [36] or UL114, a DNA repair enzyme uracil-DNA glycosylase (UNG) that was recently suggested to play a role in viral DNA replication [9, 13, 37].

Deletions of genes with roles in transcriptional activation have also been observed to reduce viral growth, these include: TRS1, an immediate early protein [9], and UL29 and UL35, both of which can transactivate the MIEP [13, 38, 39]. In addition to transcription activation, TRS1 is able to block the shutoff of translation by the antiviral dsRNA-dependent protein kinase R response pathway [40], which combined with gene knock out studies [13], suggests that this pathway may not be sufficient to shut down viral replication. UL97, the serine-theronine protein kinase shown to play a multitude of roles in CMV replication, including DNA synthesis, DNA packaging, and nuclear egress [9], was also found to play a significant role in viral replication. While UL12, UL20, UL47, UL65, UL88, UL103, UL108, UL129, and US23 have not been well characterized, the defective phenotype observed by their respective deletion viruses suggests that they may influence stabilization or enhancement of the essential mechanisms required by HCMV for a successful and efficient infection.

ORFs dispensable for growth in cultured fibroblasts

65 ORFs have been found to be dispensable for growth in fibroblasts [13] (Table 1.2) (Figure 1.1). Many of these ORFs have previously been found to play a role in suppressing either the innate or adaptive immune response of the host. Thus, it is expected that these ORFs would be nonessential for viral replication in cultured human fibroblasts. These immune evasion ORFs include a set of viral genes found within the region encompassing US1 through US12, which have been deleted in many viruses produced from infectious bacterial artificial chromosomes (BAC) in order to offset the packaging constraints imposed by the insertion of the BAC cassette in the viral genome [41]. This region of the genome has been shown to be responsible for the downmodulation of the major histocompatibility complex class I (MHC class I) [42] and potentially the MHC class II molecules [43, 44], although the latter requires further study. In addition to this region, the protein product of UL37x3 (UL37.3) was also recently predicted to have a MHC-like domain, suggesting that there may be other genes outside of the US1 through US12 region regulating expression of MHC molecules [45]. Besides MHC regulation, HCMV is also able to modulate other parts of the immune system through ORFs that were dispensable for viral growth *in vitro*. These ORFs include: UL16, UL18, and UL83, which have been shown to suppress natural killer (NK) cell recognition [46], and UL119 and RL11, which can bind the Fc domain of immunoglobulin G and potentially block the antiviral activities mediated by the domain [47].

Besides ORFs with roles in immunomodulation, UL146 and UL147, which are predicted to be chemokines-secreted proteins that control leukocyte migration and trafficking [48], have also been found to be nonessential in fibroblasts [13]. Furthermore, ORFs US27 and US28, which have been shown to be chemokine receptors [9], are also dispensable. While the function of a vast majority of these ORFs is still unknown, preliminary studies of the deletion viruses in

different cell lines suggest that a subset of these genes may have important roles in cell-specific infection.

Enhanced Growth- related ORFs

One of the most interesting findings in HCMV research has been the identification ORFs that encode temperance factors for fibroblast infection and replication. When each of these four ORFs (UL9, UL20a, UL23, and US30) were deleted, plaque assays showed an increase in viral growth by at least ten-fold relative to the wild-type [13] . While the functions of these four ORFs have not been well studied, these genes may be part of an unknown mechanism used by the virus to self-regulate its replication. It is speculated that this "self-imposed control" could be a method employed by the virus to protect the infected cell from the host's immune system. Furthermore, these genes may prove to be important in the establishment of latency in the host.

Tropism Factors

One of the hallmarks of HCMV pathogenesis is the virus' ability to infect a wide range of cell types within its human host, suggesting that the virus may encode tropism factors that aid in the specific infection of different cells types. Supporting this observation is the previously demonstrated decrease in endothelial, epithelial, and dendritic cell tropism when the ORFs within the UL128-131 region of endothelial propagated HCMV strains were deleted [49-51]. Additional defective growth phenotypes have also been observed with other ORFs in specific cell types. Both UL64 and US29 deletion virus have been found to grow 100-fold less than the wild-type strain in retinal pigment epithelial (RPE) cells, suggesting that both ORFs may encode epithelial cell specific tropism factors [13]. Support for tropism factors specific to epithelial cells has also been found in rhesus cytomegalovirus (RhCMV), where mutations in viral genes orthologous to HCMV TRL1, UL148, UL132, and US22 caused replication defects in rhesus RPEs [52]. An additional tropism factor has also been suggested for endothelial cells, with the observation of the deletion of UL24, a US22 gene family membrane, causing a reduction of viral growth in human microvascular endothelial cells (HMVEC) [13]. While UL24 has not been well characterized, except that it is a tegument protein [53], murine cytomegalovirus (MCMV) homologues of US23 (M140) and US24 (M141), which are also members of the US22 gene family, have been shown to encode tropism factors for macrophage infection [54]. This suggests that the US22 gene family may encode various tropism factors that facilitate efficient CMV infection of a wide range of tissue types.

Tissue specific Temperance Factors

In addition to the temperance factors, UL9, UL20a, UL23, and US30, which were identified in cultured fibroblasts, other temperance factors unique to particular cell types have also been found [13]. Both UL10 and UL16's respective deletion viruses had enhanced growth in RPEs, while US16 and US19 viruses showed an increase in titers in HMVECs relative to the wild-type [13]. The mechanism governing this phenomenon is still not well understood. While the concept of the virus expressing proteins to suppress its own replication appears counterintuitive from the perspective of the virus, this view is consistent with the observation that HCMV exhibits various growth rates in different cell types [9]. During *in vivo* infection, the

virus may express these genes in an effort to reduce viral levels to prevent or slow down cellular death and thus allow for better dissemination and enhance viral survival. These factors may also play a role in facilitating persistent and latent infections, a mechanism that has been difficult to investigate due to the lack of suitable cellular models. Precedent for this phenomenon has been observed in other viruses, such as spumavirus, a retrovirus which encodes a factor that can reactivate the virus from its latent state [55].

HCMV MUTAGENESIS STRATEGIES AND APPROACHES

The ability to create genomic mutations is critical to dissect and define the functions of genes, their transcripts and expressed protein product. Over the last 30 years, various methods have been developed and employed to create mutant forms of cytomegalovirus and other herpesviruses in an effort to better understand the mechanisms that govern the pathology of the virus at the molecular level.

Genome manipulation in eukaryotic cells

Forward genetic techniques have been used to generate mutations in HCMV via a method called *in vivo* mutagenesis, which involves the identification of temperature sensitive (Ts) mutants. These mutants are a result of a missense point mutation in the protein that causes a loss of function when the mutant virus is grown at a higher, restrictive temperature [56]. While growth phenotypes can be observed using this method, the genetic mappings of the mutants have proven to be difficult [57]. Following the generation of Ts mutants, the next phase of mutant generation came in the form of homologous recombination in eukaryotic cells. Briefly, a marker gene containing a positive selection marker, either β-galactosidase [58], neomycin resistance gene [59], or the xanthine guanine phosphoribosyltransferase (gpt) gene [60], was introduced to the viral genome in infected cells resulting in the mutation or deletion of the target gene [57]. While it became easier to map the mutation in the genome, the low recombination efficiency and the presence of the wild-type viral genome make this approach difficult in obtaining clonal, un-contaminated mutants.

To overcome this challenge, an alternative technique was developed using cosmid vectors containing overlapping portions of the viral genome. These cosmids would then be co-transfected into mammalian cells resulting in the reconstitution of the virus. To generate genomic mutations in the virus, nucleotide changes in the cosmids could be made using *in vitro* techniques established in bacteria [57]. While the cosmid strategy eliminated the issue of having the wild-type genome contaminating the mutant clones, the genomic instability of the system became another problem for researchers [61].

Genome manipulation in bacteria

Bacterial Artificial Chromosomes (BAC)

Current HCMV and herpesvirus mutagenesis approaches are now based upon the use of the bacterial artificial chromosomes (BAC) system, which utilizes *Escherichia coli* and its single-copy plasmid F factor to perform genetic manipulation on the genome [62, 63]. The technology was first pioneered in 1997 when Messerle et al. successfully cloned the murine

cytomegalovirus (MCMV) genome into a BAC [63]. Because of the BAC's ability to maintain low copy numbers, show sequence stability in recombination deficient *E.Coli* strains and maintain DNA fragments greater than 300 kilobases [57], the BAC system has been successfully used to clone various genomes of human herpesviruses [16, 64, 65], including various strains of HCMV (e.g. Towne, AD169) [15, 41, 66]. When compared to the yeast artificial chromosome (YAC) system, which has a larger coding capacity of 2 Mbps, the BAC system is superior since it is less prone to undesired rearrangements and deletions [57]. To overcome packaging constraints caused by the insertion of the BAC vector sequence into the CMV genome, nonessential regions in the genome were deleted in order to accommodate the additional genomic material [41]. In addition to the origin of replication and the genes that regulate its copy number, the BAC vector also contains an antibiotic resistance gene to ensure that the large plasmid is maintained in the host cell [41]. Furthermore, the use of *E. Coli* as a host for mutagenesis of the BAC-containing HCMV genome is greatly beneficial since it allows for genetic manipulation utilizing tools that have been well established in bacteria, including homologous recombination and transposon-based methods.

Allelic Exchange Mutagenesis

Allelic exchange mutagenesis, which is based upon ET recombination, has been successfully applied to the construction of HCMV mutants [13, 19]. This technique, adapted from the methodology developed by Zhang *et al.* and Muyrers *et al.*, exploits the RecET proteins of *E. Coli* to perform homologous recombination on the BAC [67, 68]. Briefly, the approach involves the construction of a linear DNA fragment containing an antibiotic resistance gene flanked by arms of homology that target the insertion site (Figure 1.2). To facilitate recombination, the deletion cassette is then delivered to transformation-competent *E. Coli* harboring the HCMV BAC and then placed under antibiotic selection pressure [13].

While the original *E. Coli*-based recombination systems used bacteria strains that harbored plasmids to express the necessary recombination proteins, newer systems used bacterial strains containing a defective prophage system, which was initially developed by Yu *et al.* [69] and later modified by Lee *et al*. [70]. The prophage system addresses the potential "leakiness" problem of *gam* and *red* proteins in the plasmid system, which can lead to BAC instability and possible loss of cell viability. Rather than having the recombination proteins expressed from a plasmid, this system has a defective prophage carrying the *gam* and the *red* recombination genes, *exo* and *beta*, which are under the control of a temperature sensitive promoter, thus allowing for transient expression of the recombination proteins. During the recombination process, *gam* inhibits the *E.Coli* RecBCD nuclease from degrading the deletion cassette, while *exo* and *beta* facilitate recombination between the HCMV BAC and the deletion cassette [70]. The newly generated mutant BAC is then propagated in *E. Coli* and then transfected into mammalian cells to obtain viral progeny (Figure 1.2) [13].

Transposon-based mutagenesis

Transposons (Tn) are short mobile DNA sequences that are able to relocate themselves from one DNA molecule to another [71]. This naturally occurring unit has been successfully exploited for genomic manipulation in prokaryotes and higher order eukaryotes, including *Drosophila, Caenorhabditis elegans,* and plants [71]. Given the success of transposon-based

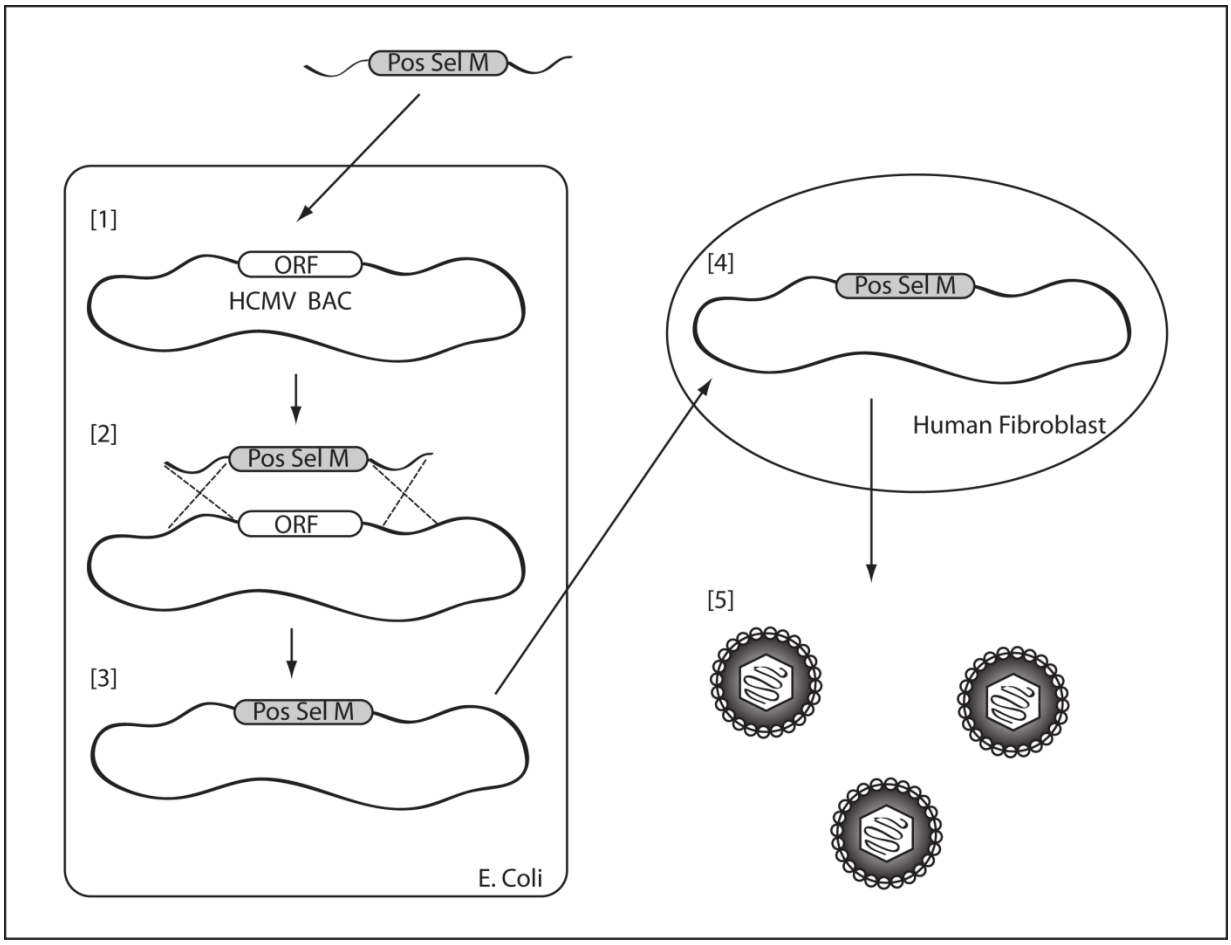


Figure 1.2. **Allelic Exchange Mutagenesis of HCMV BAC**
Mutagenesis begins by the construction of the deletion cassette, which contains a positive selection marker (Pos Sel M) flanked by sequences homologous to the regions flanking the ORF of interest. The deletion cassette is delivered to the competent *E. Coli* strain by transformation (1) and homologous recombination occurs between the deletion cassette and the BAC (2). A successful deletion recombination event occurs by the replacement of the viral ORF with the selection marker (3). The mutant BAC is then harvested and transfected into human fibroblast to generate mutant virus (4). Depending on the ORF deleted, the deletion of a nonessential gene, under the context of *in vitro* infection, will result in viral progeny formation (5), while an essential gene will not.

mutagenesis on studies involving bacteria, the system was then applied to generate HCMV mutant libraries [15]. Briefly, the Tn-donor plasmid is transfected into *E. Coli* harboring the HCMV BAC, whereby the Tn randomly inserts itself into the BAC to generate the mutation. Selection for positive insertions then occurs by incubating the *E. Coli* at the higher restrictive temperature and antibiotic pressure. To determine the insertion site, the mutant BAC is sequenced using primers that sequence outward from the inserted Tn [15]. (Figure 1.3)

CONCLUSION

Given the severity of diseases associated with human cytomegalovirus in neonates and immunocompromised individuals, it is critical to understand the function of its genes and their relation to the mechanisms employed by the virus during infection. The powerful tools that have been developed during the last 30 years have allowed researchers to manipulate the genome of the virus and answer some critical questions in relation to the pathology of the virus. Furthermore, the evolution of these tools has greatly increased the pace of basic HCMV research which may hopefully increase the rate of development of treatments.

Using these tools, I have been able to generate gene deletion virus and recombinant viruses expressing proteins with affinity tags at the C terminus. These constructs have allowed me to perform various studies, including a global subcellular localization study of HCMV proteins, an extensive characterization study of ORF US20, and a brief study characterizing a SUMOylation site in UL44, the HCMV processivity factor. It is my hope that the knowledge obtained from these studies spark new ideas relating to gene families, gene function and mechanisms that one day can be used as a basis for the development of anti-viral therapies against HCMV.

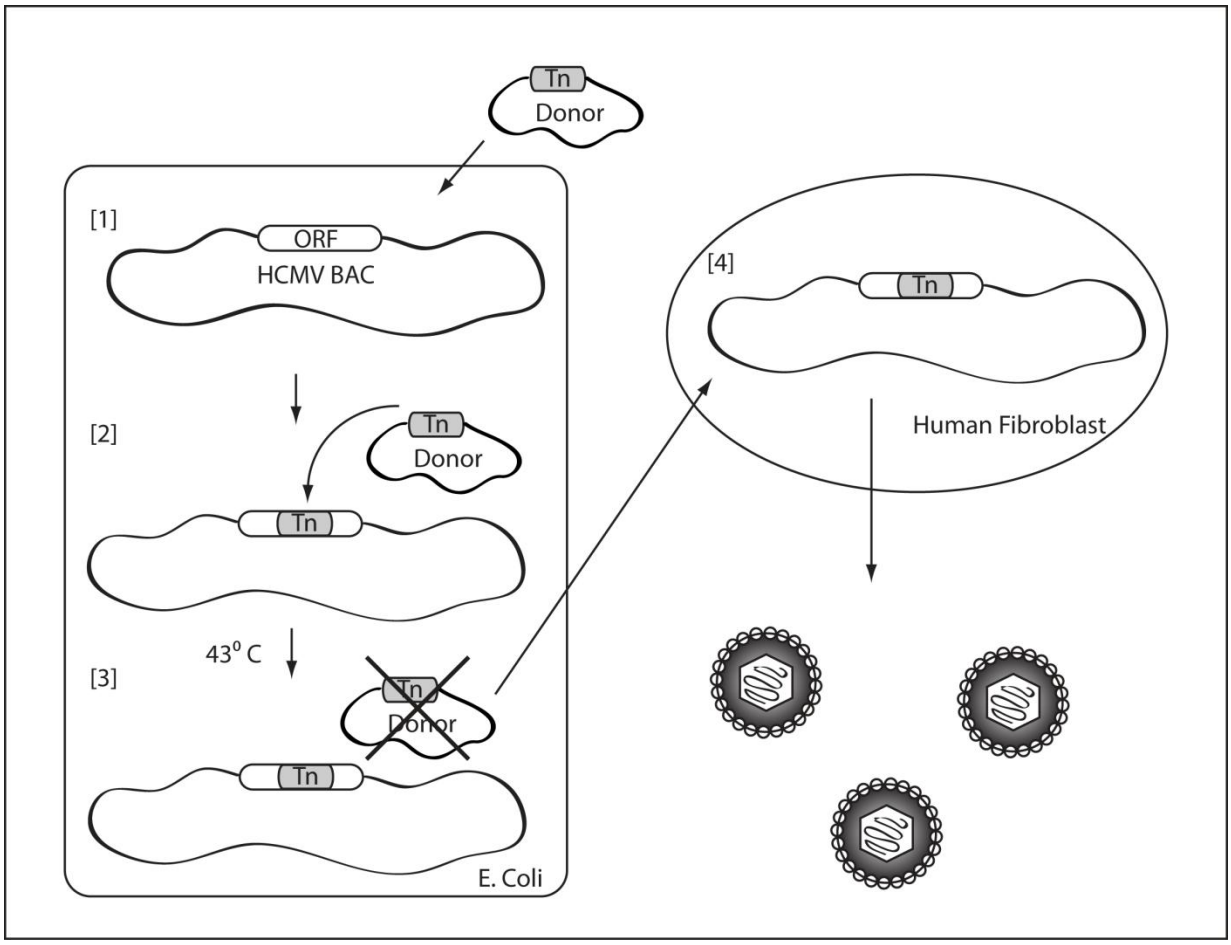


Figure 1.3. **Random Transposon Mutagenesis of HCMV BAC**
A donor plasmid containing the transposon (Tn) is transformed into *E. Coli* harboring the HCMV BAC (1). Transposition then occurs and the Tn is inserted into the BAC at a random site, potentially an ORF of interest (2). The bacteria is then incubated at the higher restrictive temperature (43°C), causing the loss of the donor plasmid (3). Insertions are then selected by positive selection using antibiotics. The insertion site is determined by DNA sequencing and desired clone is then transfected into human fibroblasts to obtain viral progeny (4).

REFERENCES

1. Grüter W. (1924) Das herpesvirus, seine atiologische und klinische bedeutung. *Munch med Wschr* (71): 1058-1060.
2. Pellett PE and Roizman B. (2007) The Family *Herpesviridae:* A Brief Introduction. In: Knipe DM and Howley PM (eds), *Fields' Virology*, pp. 2479-2499. Lippincott Williams & Wilkins, Philadelphia, PA.
3. Hardie DR. (2010) Human gamma-herpesviruses: a review of 2 divergent paths to oncogenesis. *Transfus Apher Sci*. **42**(2): 177-83.
4. Varnum SM, et al. (2004) Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *J Virol*. **78**(20): 10960-6.
5. Terhune SS, Schroer J, and Shenk T. (2004) RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration. *J Virol*. **78**(19): 10390-8.
6. Ho M. (2008) The history of cytomegalovirus and its diseases. *Med Microbiol Immunol*. **197**(2): 65-73.
7. Weller TH. (1970) Review. Cytomegaloviruses: the difficult years. *J Infect Dis*. **122**(6): 532-9.
8. Riley HD, Jr. (1997) History of the cytomegalovirus. *South Med J*. **90**(2): 184-90.
9. Mocarski ES, Shenk T, and Pass R. (2007) Cytomegalovirus. In: Knipe DM, Howley PM, and Griffin DE (eds), *Fields Virology*, pp. 2701-2772. Lippincott Williams & Wilkins, Philadelphia.
10. Bloom JN and Palestine AG. (1988) The diagnosis of cytomegalovirus retinitis. *Ann Intern Med*. **109**(12): 963-9.
11. Peterson PK, et al. (1980) Cytomegalovirus disease in renal allograft recipients: a prospective study of the clinical features, risk factors and impact on renal transplantation. *Medicine (Baltimore)*. **59**(4): 283-300.
12. Roizman B, Knipe DM, and Whitley RJ. (2007) Herpes SImplex Viruses. In: Knipe DM and Howley PM (eds), *Fields Virology*, pp. 2501-2601. Lippincott Willaims & Wilkins, Philadelphia, PA.
13. Dunn W, et al. (2003) Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A*. **100**(24): 14223-8.
14. Plotkin SA, Farquhar J, and Horberger E. (1976) Clinical trials of immunization with the Towne 125 strain of human cytomegalovirus. *J Infect Dis*. **134**(5): 470-5.
15. Yu D, Silva MC, and Shenk T. (2003) Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc Natl Acad Sci U S A*. **100**(21): 12396-401.
16. Roizman B and Knipe DM. (2001) Herpes Simplex Viruses and Their Replication. In: Knipe DM and Howley PM (eds), *Fields Virology*, pp. 2399–2460. Lippincott William & Wilkins, Philadelphia.
17. Chen DH, et al. (1999) Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus. *Virology*. **260**(1): 10-6.
18. Yu X, et al. (2005) Dissecting human cytomegalovirus gene function and capsid maturation by ribozyme targeting and electron cryomicroscopy. *Proc Natl Acad Sci U S A*. **102**(20): 7103-8.
19. Borst EM, et al. (2001) Genetic evidence of an essential role for cytomegalovirus small capsid protein in viral growth. *J Virol*. **75**(3): 1450-8.

20. Desai P, DeLuca NA, and Person S. (1998) Herpes simplex virus type 1 VP26 is not essential for replication in cell culture but influences production of infectious virus in the nervous system of infected mice. *Virology*. **247**(1): 115-24.
21. Pari GS. (2008) Nuts and Bolts of Human Cytomegalovirus Lytic DNA Replication. In: Shenk T and Stinski MF (eds), *Current Topics in Microbiology and Immunology: Human Cytomegalovirus*, pp. 153-166. Springer-Verlag, Berlin-Heidelberg.
22. Anders DG, Kerry JA, and Pari GS. (2006) DNA Synthesis and late viral gene expression. In: Arvin AM, Mocarski ES, and Moore P (eds), *Human Herpesvirus: Biology, Therapy and Immunoprophylaxis*, pp. 292-307. Cambridge Press, Cambridge.
23. Gibson W. (2008) Structure and Function of the Cytomegalovirus Virion. In: Shenk T and Stinski MF (eds), *Human Cytomegalovirus. Current Topics in Microbiology and Immunology*, pp. 187-204. Springer-Verlag, Berlin, Heidelberg.
24. Isaacson MK and Compton T. (2009) Human cytomegalovirus glycoprotein B is required for virus entry and cell-to-cell spread but not for virion attachment, assembly, or egress. *J Virol*. **83**(8): 3891-903.
25. Wang SK, Duh CY, and Chang TT. (2000) Cloning and identification of regulatory gene UL76 of human cytomegalovirus. *J Gen Virol*. **81**(Pt 10): 2407-16.
26. Wang SK, Duh CY, and Wu CW. (2004) Human cytomegalovirus UL76 encodes a novel virion-associated protein that is able to inhibit viral replication. *J Virol*. **78**(18): 9750-62.
27. Biegalke BJ, et al. (2004) Characterization of the human cytomegalovirus UL34 gene. *J Virol*. **78**(17): 9579-83.
28. Goldmacher VS, et al. (1999) A cytomegalovirus-encoded mitochondria-localized inhibitor of apoptosis structurally unrelated to Bcl-2. *Proc Natl Acad Sci U S A*. **96**(22): 12536-41.
29. Munger J, Yu D, and Shenk T. (2006) UL26-deficient human cytomegalovirus produces virions with hypophosphorylated pp28 tegument protein that is unstable within newly infected cells. *J Virol*. **80**(7): 3541-8.
30. Stamminger T, et al. (2002) Open reading frame UL26 of human cytomegalovirus encodes a novel tegument protein that contains a strong transcriptional activation domain. *J Virol*. **76**(10): 4836-47.
31. Toth Z and Stamminger T. (2008) The human cytomegalovirus regulatory protein UL69 and its effect on mRNA export. *Front Biosci*. **13**: 2939-49.
32. Qian Z, et al. (2008) The full-length protein encoded by human cytomegalovirus gene UL117 is required for the proper maturation of viral replication compartments. *J Virol*. **82**(7): 3452-65.
33. Park MY, et al. (2006) Interactions among four proteins encoded by the human cytomegalovirus UL112-113 region regulate their intranuclear targeting and the recruitment of UL44 to prereplication foci. *J Virol*. **80**(6): 2718-27.
34. Terhune S, et al. (2007) Human cytomegalovirus UL38 protein blocks apoptosis. *J Virol*. **81**(7): 3109-23.
35. Jiang XJ, et al. (2008) UL74 of human cytomegalovirus contributes to virus release by promoting secondary envelopment of virions. *J Virol*. **82**(6): 2802-12.
36. Davison AJ and Stow ND. (2005) New genes from old: redeployment of dUTPase by herpesviruses. *J Virol*. **79**(20): 12880-92.

37. Ranneberg-Nilsen T, et al. (2008) Characterization of human cytomegalovirus uracil DNA glycosylase (UL114) and its interaction with polymerase processivity factor (UL44). *J Mol Biol*. **381**(2): 276-88.
38. Schierling K, et al. (2004) Human cytomegalovirus tegument proteins ppUL82 (pp71) and ppUL35 interact and cooperatively activate the major immediate-early enhancer. *J Virol*. **78**(17): 9512-23.
39. Mitchell DP, et al. (2009) Human cytomegalovirus UL28 and UL29 open reading frames encode a spliced mRNA and stimulate accumulation of immediate-early RNAs. *J Virol*.
40. Child SJ, et al. (2004) Evasion of cellular antiviral responses by human cytomegalovirus TRS1 and IRS1. *J Virol*. **78**(1): 197-205.
41. Marchini A, Liu H, and Zhu H. (2001) Human cytomegalovirus with IE-2 (UL122) deleted fails to express early lytic genes. *J Virol*. **75**(4): 1870-8.
42. Jones TR, et al. (1995) Multiple independent loci within the human cytomegalovirus unique short region down-regulate expression of major histocompatibility complex class I heavy chains. *J Virol*. **69**(8): 4830-41.
43. Hegde NR, et al. (2002) Inhibition of HLA-DR assembly, transport, and loading by human cytomegalovirus glycoprotein US3: a novel mechanism for evading major histocompatibility complex class II antigen presentation. *J Virol*. **76**(21): 10929-41.
44. Tomazin R, et al. (1999) Cytomegalovirus US2 destroys two components of the MHC class II pathway, preventing recognition by CD4+ T cells. *Nat Med*. **5**(9): 1039-43.
45. Wyrwicz LS and Rychlewski L. (2008) Cytomegalovirus immediate early gene UL37 encodes a novel MHC-like protein. *Acta Biochim Pol*. **55**(1): 67-73.
46. Wilkinson GW, et al. (2008) Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol*. **41**(3): 206-12.
47. Atalay R, et al. (2002) Identification and expression of human cytomegalovirus transcription units coding for two distinct Fcgamma receptor homologs. *J Virol*. **76**(17): 8596-608.
48. Saederup N and Mocarski ES, Jr. (2002) Fatal attraction: cytomegalovirus-encoded chemokine homologs. *Curr Top Microbiol Immunol*. **269**: 235-56.
49. Hahn G, et al. (2004) Human cytomegalovirus UL131-128 genes are indispensable for virus growth in endothelial cells and virus transfer to leukocytes. *J Virol*. **78**(18): 10023-33.
50. Gerna G, et al. (2005) Dendritic-cell infection by human cytomegalovirus is restricted to strains carrying functional UL131-128 genes and mediates efficient viral antigen presentation to CD8+ T cells. *J Gen Virol*. **86**(Pt 2): 275-84.
51. Wang D and Shenk T. (2005) Human cytomegalovirus UL131 open reading frame is required for epithelial cell tropism. *J Virol*. **79**(16): 10330-8.
52. Lilja AE, et al. (2008) Functional genetic analysis of rhesus cytomegalovirus: Rh01 is an epithelial cell tropism factor. *J Virol*. **82**(5): 2170-81.
53. Adair R, et al. (2002) The products of human cytomegalovirus genes UL23, UL24, UL43 and US22 are tegument components. *J Gen Virol*. **83**(Pt 6): 1315-24.
54. Hanson LK, et al. (2001) Products of US22 genes M140 and M141 confer efficient replication of murine cytomegalovirus in macrophages and spleen. *J Virol*. **75**(14): 6292-302.

55. Meiering CD and Linial ML. (2002) Reactivation of a complex retrovirus is controlled by a molecular switch and is inhibited by a viral protein. *Proc Natl Acad Sci U S A*. **99**(23): 15130-5.
56. Schaffer PA. (1975) Temperature-sensitive mutants of herpesviruses. *Curr Top Microbiol Immunol*. **70**: 51-100.
57. Ruzsics Z and Koszinowski UH. (2008) Mutagenesis of Cytomegalovirus Genome. In: Shenk T and Stinski M (eds), *Human Cytomegalovirus*, pp. 41-62. Springer-Verlag, Berlin Heidelberg.
58. Spaete RR and Mocarski ES. (1987) Insertion and deletion mutagenesis of the human cytomegalovirus genome. *Proc Natl Acad Sci U S A*. **84**(20): 7213-7.
59. Wolff D, Jahn G, and Plachter B. (1993) Generation and effective enrichment of selectable human cytomegalovirus mutants using site-directed insertion of the neo gene. *Gene*. **130**(2): 167-73.
60. Mulligan RC and Berg P. (1981) Selection for animal cells that express the Escherichia coli gene coding for xanthine-guanine phosphoribosyltransferase. *Proc Natl Acad Sci U S A*. **78**(4): 2072-6.
61. Horsburgh BC, et al. (1999) Allele replacement: an application that permits rapid manipulation of herpes simplex virus type 1 genomes. *Gene Ther*. **6**(5): 922-30.
62. Shizuya H, et al. (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A*. **89**(18): 8794-7.
63. Messerle M, et al. (1997) Cloning and mutagenesis of a herpesvirus genome as an infectious bacterial artificial chromosome. *Proc Natl Acad Sci U S A*. **94**(26): 14759-63.
64. Kieff E and Rickinson AB. (2001) Epstein–Barr virus and its replication. In: Knipe DM and Howley PM (eds), *Fields Virology*, pp. 2511–2574. Lippincott William & Wilkins, Philadelphia.
65. Moore PS and Chang Y. (2001) Kaposi's sarcoma-associated herpesvirus. In: Knipe DM and Howley PM (eds), *Fields Virology*, pp. 2803–2834. Lippincott William & Wilkins, Philadelphia.
66. Yu D, et al. (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**(5): 2316-28.
67. Zhang Y, et al. (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet*. **20**(2): 123-8.
68. Muyrers JP, et al. (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res*. **27**(6): 1555-7.
69. Yu D, et al. (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A*. **97**(11): 5978-83.
70. Lee EC, et al. (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics*. **73**(1): 56-65.
71. Hayes F. (2003) Transposon-based strategies for microbial functional genomics and proteomics. *Annu Rev Genet*. **37**: 3-29.

Chapter 2

Proteomic study of Human Cytomegalovirus intracellular protein localization at 72 hours post infection.

Abstract

Human cytomegalovirus (HCMV), a member of the beta subgroup of herpesviruses, is currently the largest known viruses to infect humans. Currently, HCMV is believed to have at least 160 open reading frames (ORFs) capable of encoding proteins with many of these genes not fully characterized. This study presents an intracellular localization study of 82 ORFs using an infectious bacterial artificial chromosome expressing viral proteins tagged at the C terminus with an epitope tag. Our study found 17 proteins to localize to the nucleus, 51 to a juxtanuclear structure in the cytoplasm, 11 to the cytoplasm, and three to uncharacterized cytoplasmic structures at 72 hours post infection in cultured human fibroblasts. Furthermore, we were also able to determine the molecular weight of many of these proteins in the context of infection and show, as a proof of concept for a subset of genes, that our recombinant viruses can be used to determine the protein's expression kinetics during viral replication.

Introduction

Human cytomegalovirus (HCMV), with a genome size of ~240,000 base pairs, is currently the largest known virus to infect humans. Bioinformatic predictions have identified at least 160 open reading frames (ORFs) likely to encode proteins, of which only a small portion have been well studied [1]. While large scale genomic projects have been successfully carried out to study the HCMV transcriptome during infection [2], the proteome of the HCMV virion [3], and the impact of sequentially deleting ORFs on viral replication in fibroblasts [4, 5], there has not been a global study of the intracellular localization of the proteins expressed from these ORFs under the context of infection.

Protein characterization studies have relied on two approaches to identify and probe for the protein of interest. Either an antibody specific to an epitope on the peptide is made or a commercially available antibody is used to probe for an artificial epitope tag that is inserted at either ends of the peptide [6]. Because of the extensive time and costs associated with generating specific antibodies to the protein product of each ORF, we decided to use the latter method for our study since it would only require the purchase of a single antibody to probe for the viral proteins.

This study reports the use of homologous recombination on a bacterial artificial chromosome (BAC) containing the genome of the lab-adapted HCMV Towne strain, which has been shown by our lab to replicate and infect at levels similar to the native Towne strain [4], to construct our recombinant viruses. Of the 132 recombinant BACs we constructed, we were able to obtain 103 recombinant viruses, each with an epitope tag inserted immediately before the predicted stop codon at the 3' end of the ORF of interest. From the 103 recombinant viruses, we were able to identify the subcellular localization of 82 ORFs at 72 hours post infection.

Furthermore, we were also able to use our recombinant viruses to determine the molecular weight of the tagged proteins and, as a proof of concept, we also showed that they could be used in protein kinetic studies to determine if the protein was expressed during either the immediate early, early, or late phase of expression.

Results

Construction of HCMV Recombinant Tag Viruses

Building upon the HCMV BAC mutagenesis technique used previously by our lab [4], we were able to construct 132 recombinant HCMV BACs, each with a different gene tagged at the C terminus with our tandem affinity tag. Briefly, we amplified an insertion cassette containing a tandem affinity tag consisting of the FLAG epitope and the *Staphylococcus aureus*-derived protein A epitope, which have been successfully used in the past for tandem affinity purification applications [7]. Using ET-cloning, which relies on *rec*E and *rec*T proteins for homologous recombination [8], we inserted our cassette immediately before the predicted stop codon of the gene of interest (Figure 2.1). To determine if we successfully constructed the recombinant virus, we first confirmed the insertion of the cassette into the BAC by polymerase chain reaction (PCR) using primers flanking the open reading frame. A successful insertion was indicated by a band shift of 1.1 kilobases relative to the wild-type (Figure 2.2, Figure 2.3). Restriction digests with Hind III enzyme and southern blots using a probe against the Zeocin resistance gene, which was also inserted to be used as a positive selection marker, were used to

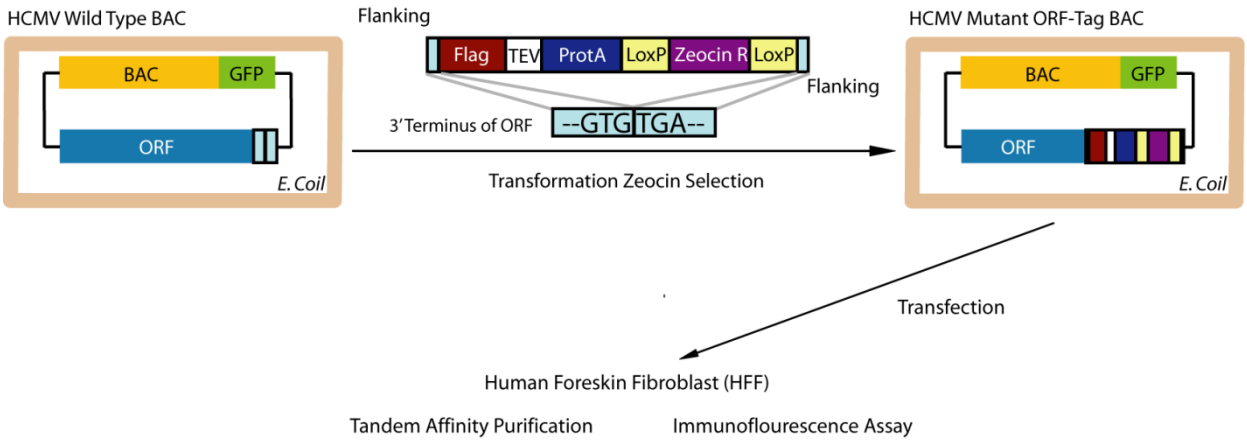


Figure 2.1. **Epitope tag mutagenesis strategy**.
Recombinant human cytomegalovirus was constructed by Red/ET recombination between an infectious BAC clone containing the genome of the HCMV Towne strain and a targeting cassette generated by PCR. Recombination resulted in the insertion of a tandem affinity epitope tag (FLAG and Protein A) at the 3' end of the viral ORF. The recombinant BAC was then transfected into human foreskin fibroblasts to generate virion.

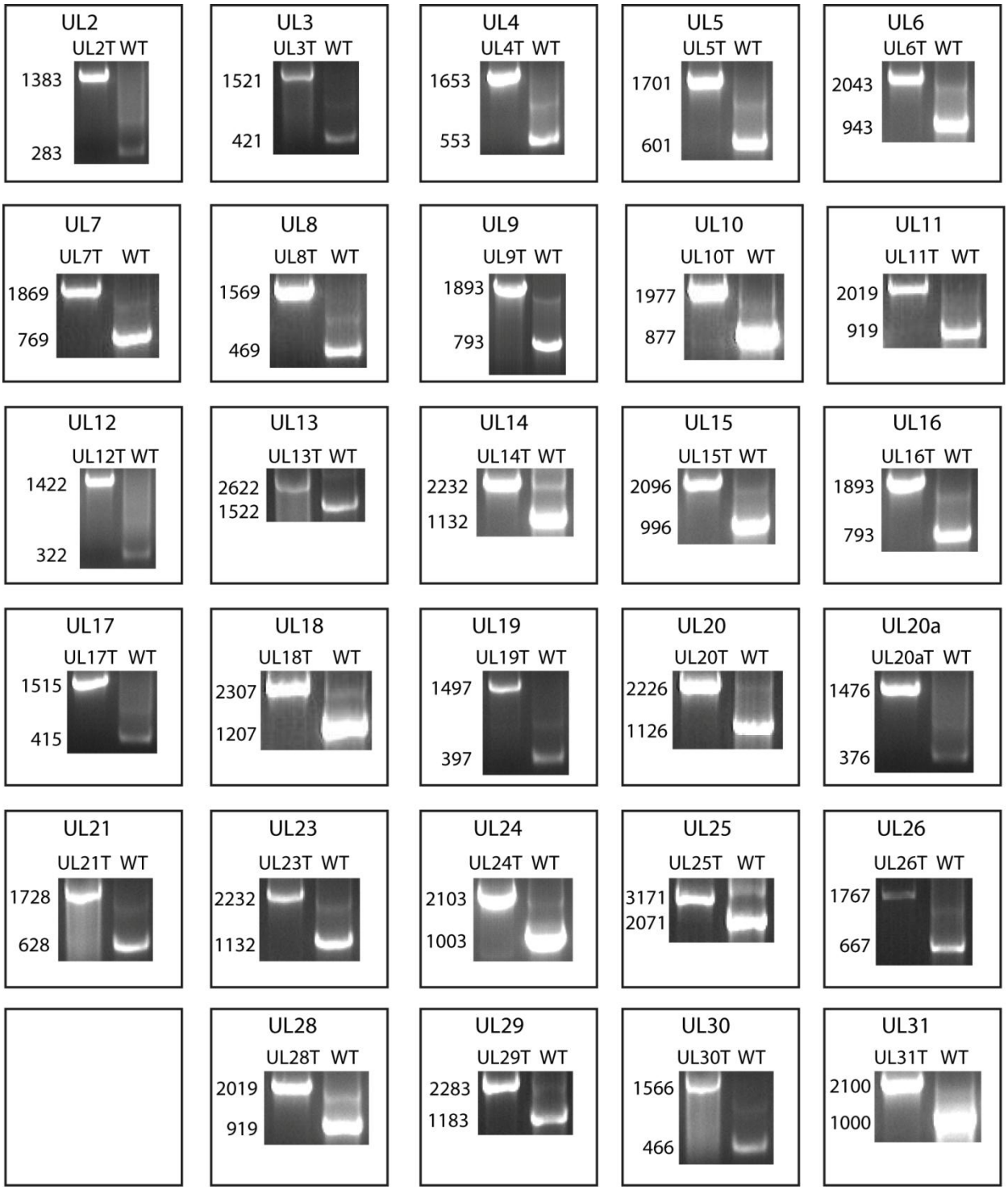


Figure 2.2. **PCR Confirmation of Recombinant HCMV BAC**.
The recombinant BACs were confirmed by PCR using primers that flanked the open reading frame. The numbers indicated are in basepairs.

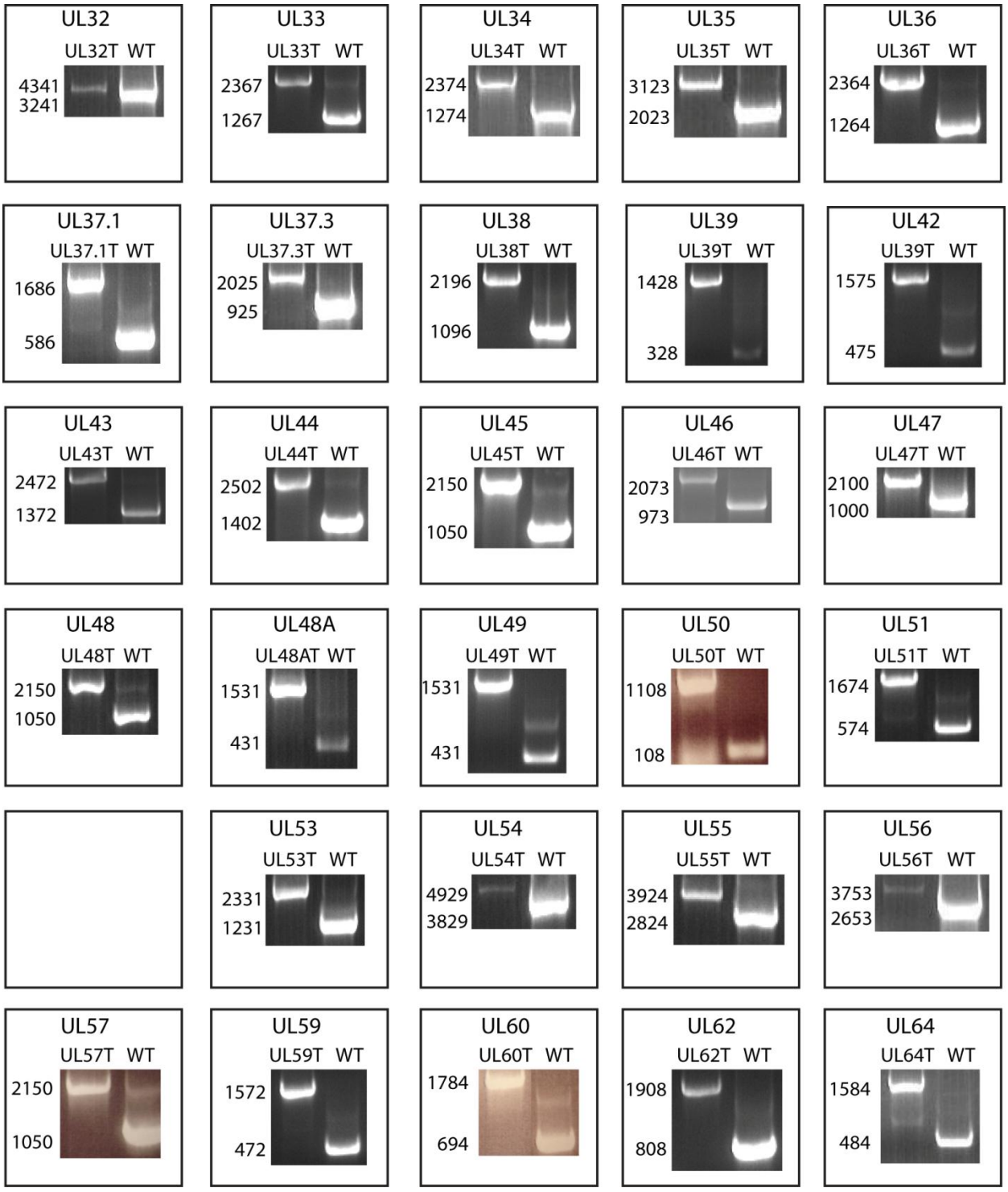


Figure 2.2 (Continued)

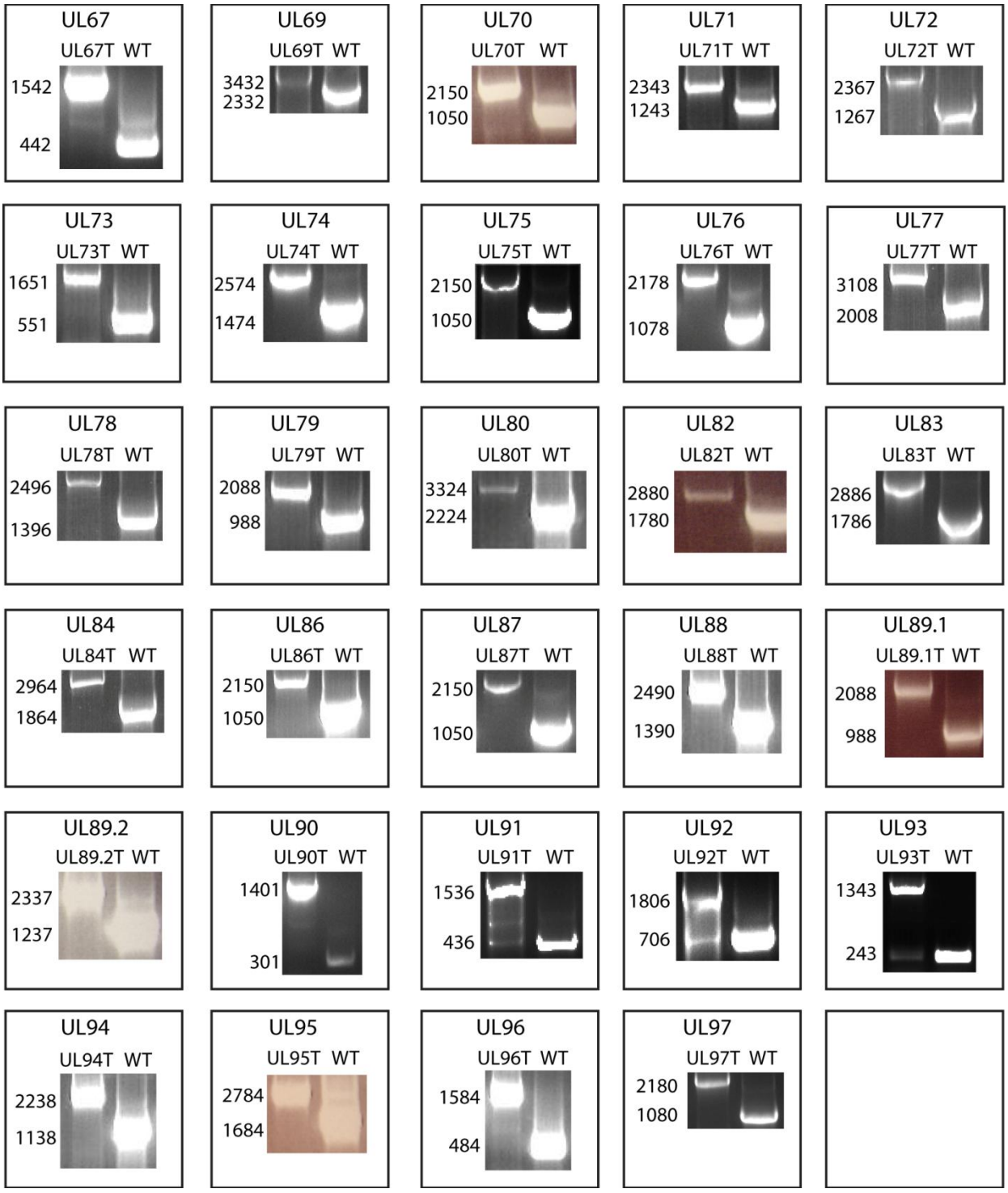


Figure 2.2 (Continued)

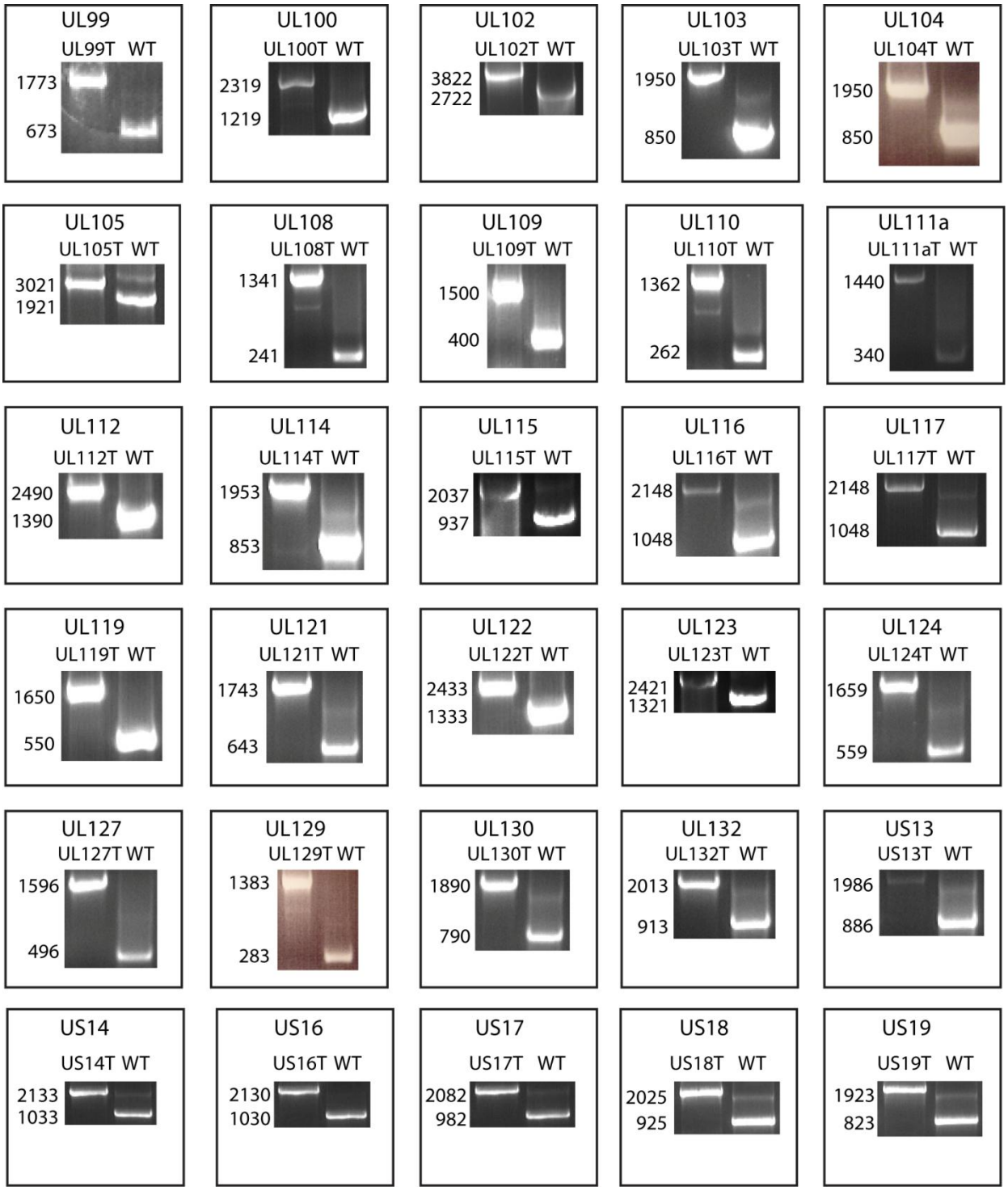


Figure 2.2 (Continued)

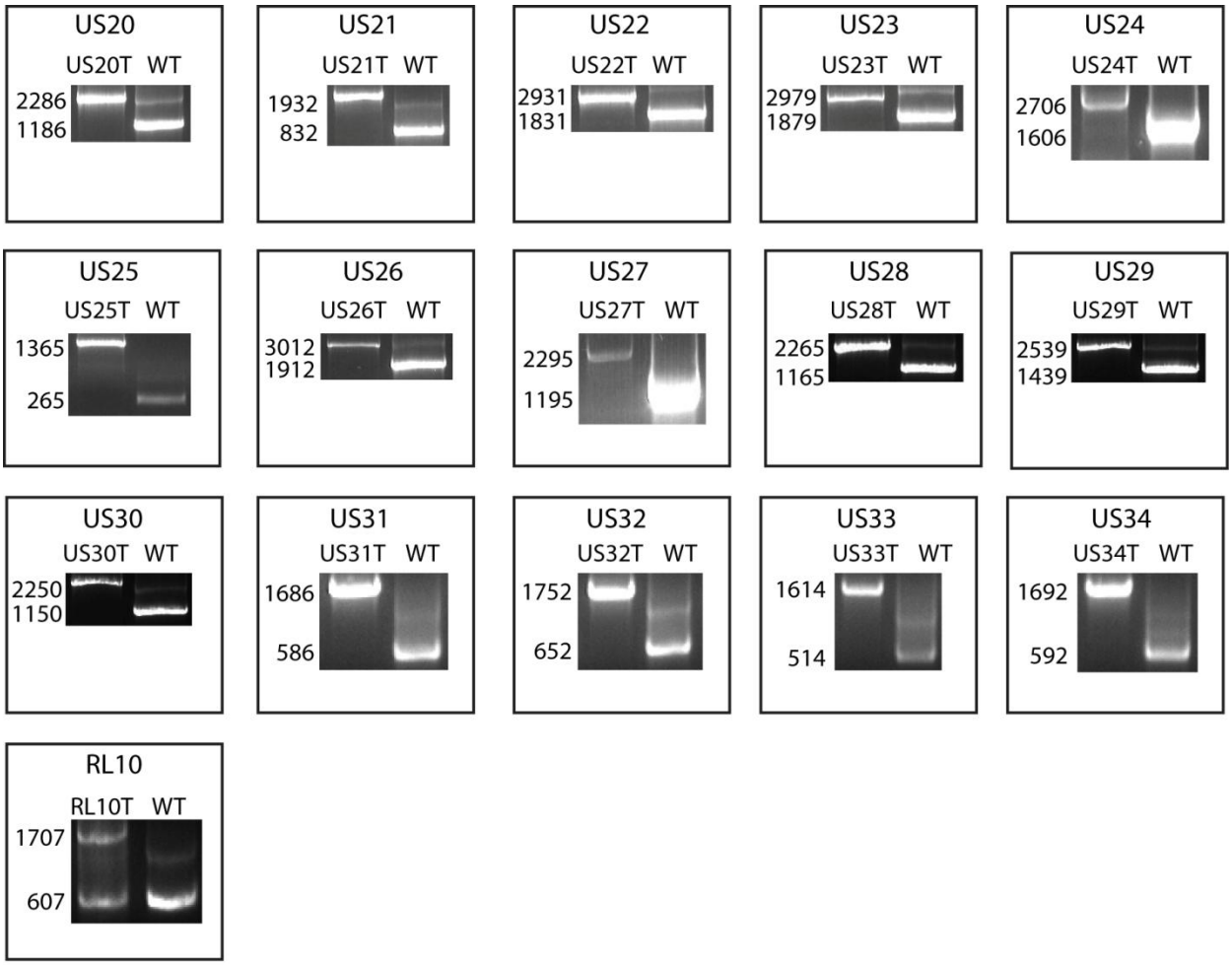


Figure 2.2 (Continued)

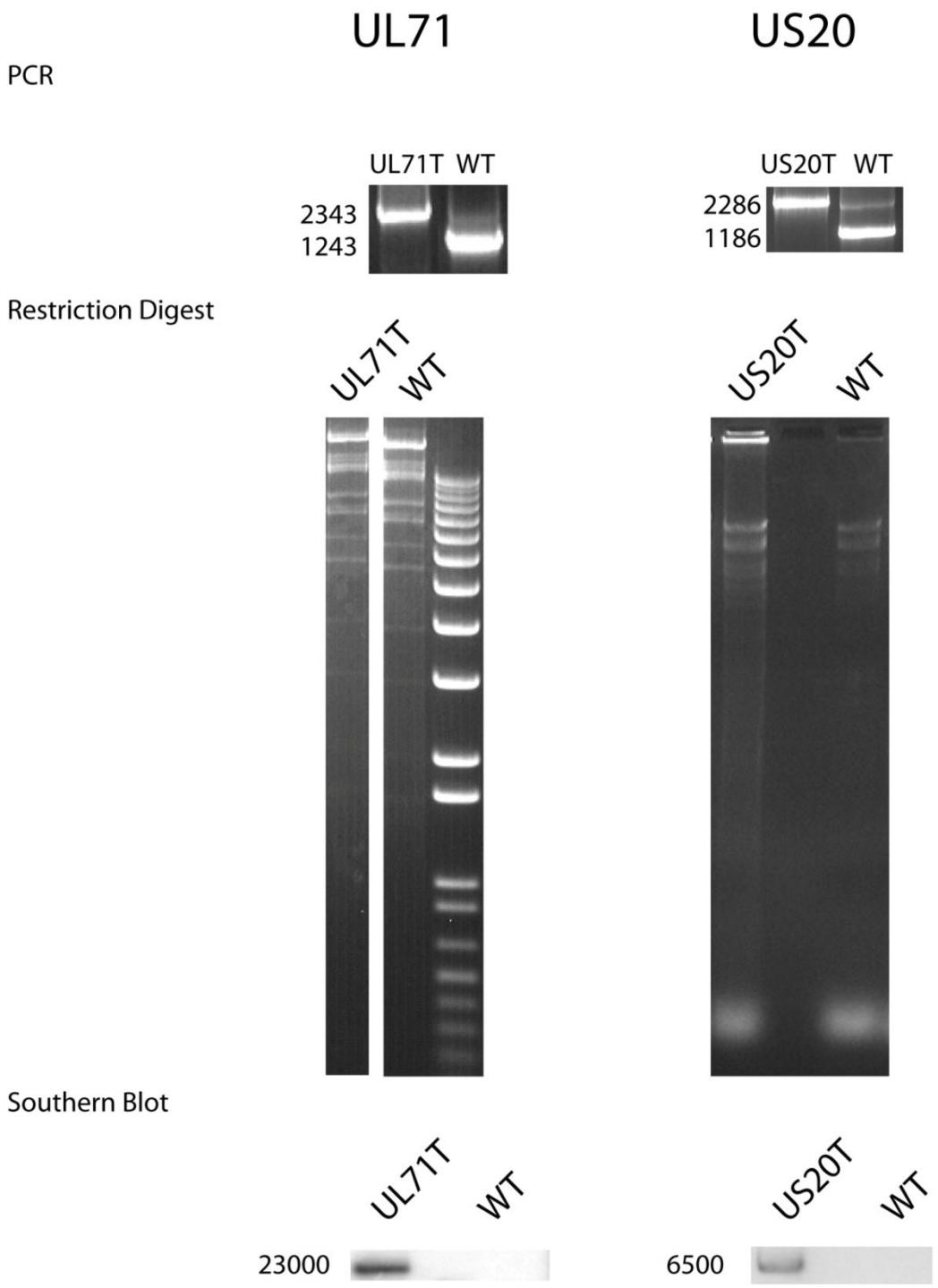


Figure 2.3. **Construction of Recombinant Virus**.
The constructed recombinant BACs of ORF UL71 and US20 were confirmed by PCR using primers flanking the ORF, restriction digestion using Hind III enzyme, and southern blot using a probe specific to the inserted Zeocin resistance gene. The bands of the southern blot represent the restriction enzyme digested fragment of the HCMV genome containing the ORF of interest.

confirm that nonspecific recombination did not occur and that our tag inserted a single time into the genome of the virus (Figure 2.3).

Of the 132 recombinant BACs electroporated into human foreskin fibroblasts (HFFs), we were able to successfully obtain recombinant HCMV virus from 109 of the BACs. The presence of virion was indicated by green florescence from infected cells and the "enlarged" cell morphology associated with the HCMV cytopathic effect. Table 2.1 contains a list of the open reading frames (ORFs), in which the insertion of our epitope tag at the 3' end of the ORF consequently resulted in no virion production in HFFs after electroporation (Table 2.1). Over 90% of these ORFs were found to be essential for viral replication in HFFs previously by our lab and others [4, 5], suggesting that the functional domain of the proteins encoded by these ORFs may be at the C terminus end of the peptide.

Localization study

Of the remaining 109 unique recombinant HCMV epitope tagged viruses, we used immunofluorescence to determine the subcellular localization of the viral proteins expressed by 82 ORFs at 72 hours post infection. Briefly, HFFs were infected with our epitope-tagged virus at a high multiplicity of infection (MOI) and then fixed with methanol after 72 hours. The fixed cells were then stained with anti-FLAG primary antibody. The methanol was able to quench the green fluorescence due to denaturation of GFP expressed by the BAC, thus allowing us to use an anti-mouse fluorescein isothiocyanate (FITC) conjugate secondary antibody for visualization of the tagged viral protein.

We identified 17 ORFs that encoded proteins with fluorescence signaling that overlapped with our 4', 6-diamidino-2-phenylindole (DAPI) staining of the nucleus. Viruses with either, UL3, UL34, UL44, UL69, UL77, UL79, UL83, UL84, UL87, UL97, UL112, UL122, or UL123 tagged all had staining of a sub-nuclear structure in the nucleus, while UL24, UL28, UL31, and UL53 exhibited more uniform nuclear localization (Figure 2.4). We also found a large subset of viral proteins expressed from ORFs: UL4, UL5, UL7, UL8, UL11, UL12, UL13, UL17, UL42, UL43, UL45, UL48, UL67, UL71, UL78, UL82, UL88, UL90, UL94, UL95, UL96, UL99, UL100, UL103, UL109, UL111a, UL115, UL116, UL117, UL119, UL121, UL124, UL127, UL129, UL130, UL132, US14, US17, US18, US19, US22, US23, US25, US26, US27, US28, US29, US31, US32, US33, and US34 to localize to subcellular cytoplasmic structure(s) adjacent to the nucleus (Figure 2.5). In addition, we identified another group of viral that encoded proteins that had a more dispersive localization in the cytoplasm at 72 hpi. This group included UL16, UL18, UL19, UL25, UL29, UL35, UL36, UL37.1, UL72, US30, and RL10 (Figure 2.6). Lastly, we found a small subset of ORFs (US13, US20, and US21) to have a very unique subcytoplasmic localization that differed from the other genes that we profiled (Figure 2.7). Furthermore, members of this specific group were also found to be part of the US12 gene family [9].

Profiling of protein molecular weights

In addition to immunofluorescence studies, we also profiled the molecular mass of the tagged viral proteins. Briefly, we infected HFFs with the recombinant viruses at a high MOI and harvesting the protein lysate after 72hpi. The lysates were then separated by SDS-PAGE and immunoblotted using anti-FLAG antibodies. We then identified the mass of the immunoreactive

Table 1
UL 046
UL 049
UL 050
UL 052
UL 054
UL 056
UL 057
UL 060
UL 070
UL 073
UL 074
UL 075
UL 076
UL 080
UL 086
UL 089.1
UL 089.2
UL 091
UL 092
UL 093
UL 102
UL 104
UL 105

Table 2.1. **List of null ORFs**.
This table lists the ORFs in which we were unable to obtain virus after three attempts to electroporate the respective recombinant BAC into HFFs.

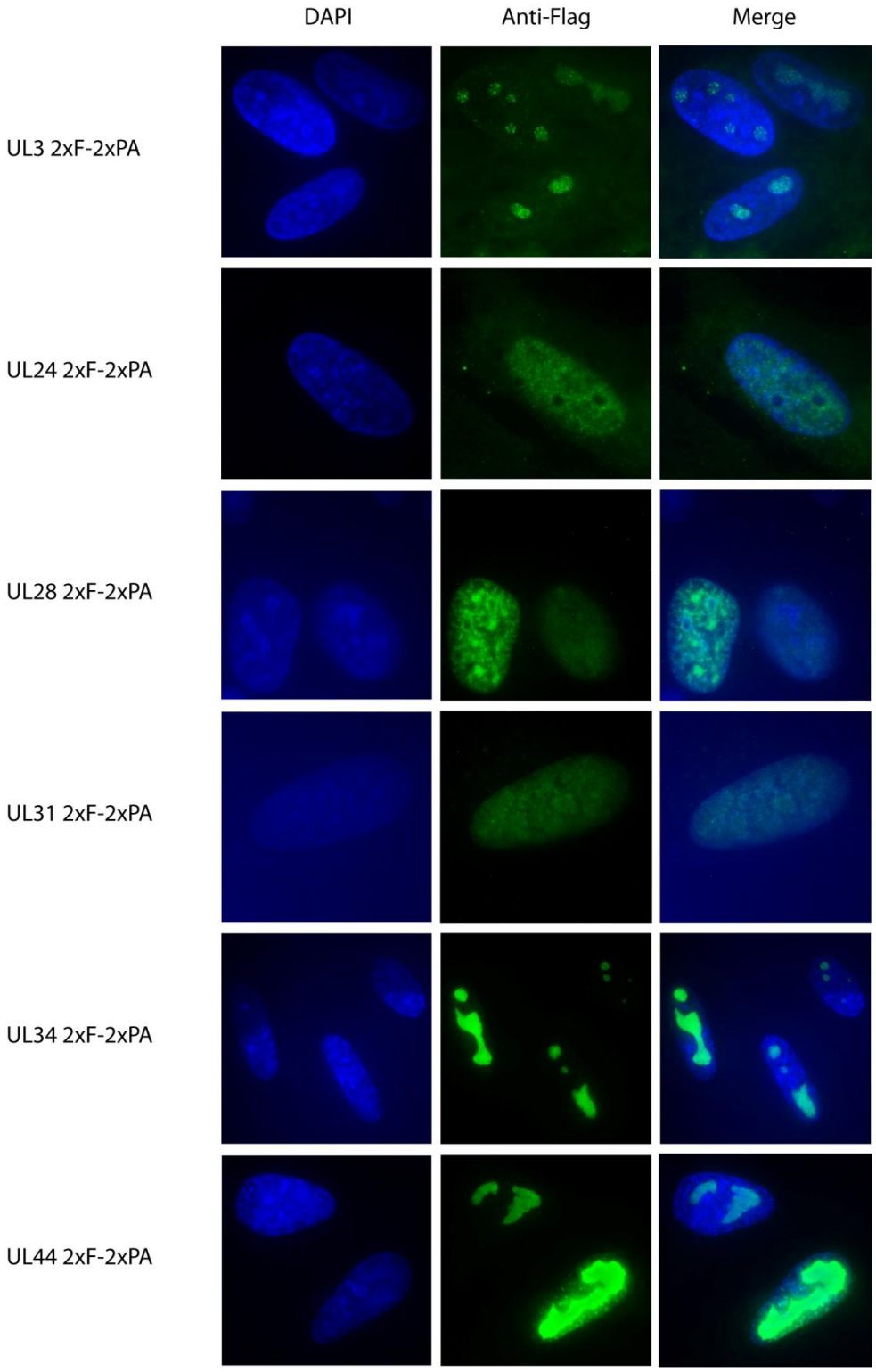


Figure 2.4. **HCMV ORFs with nuclear immunofluorescence localization.**
HFF cells were infected with recombinant virus for 72 hpi. Infected cells were fixed with methanol and stained with primary anti-FLAG antibody and secondary anti-mouse conjugated to FITC. Nucleus was stained with DAPI. Images were taken at 1000x magnification.

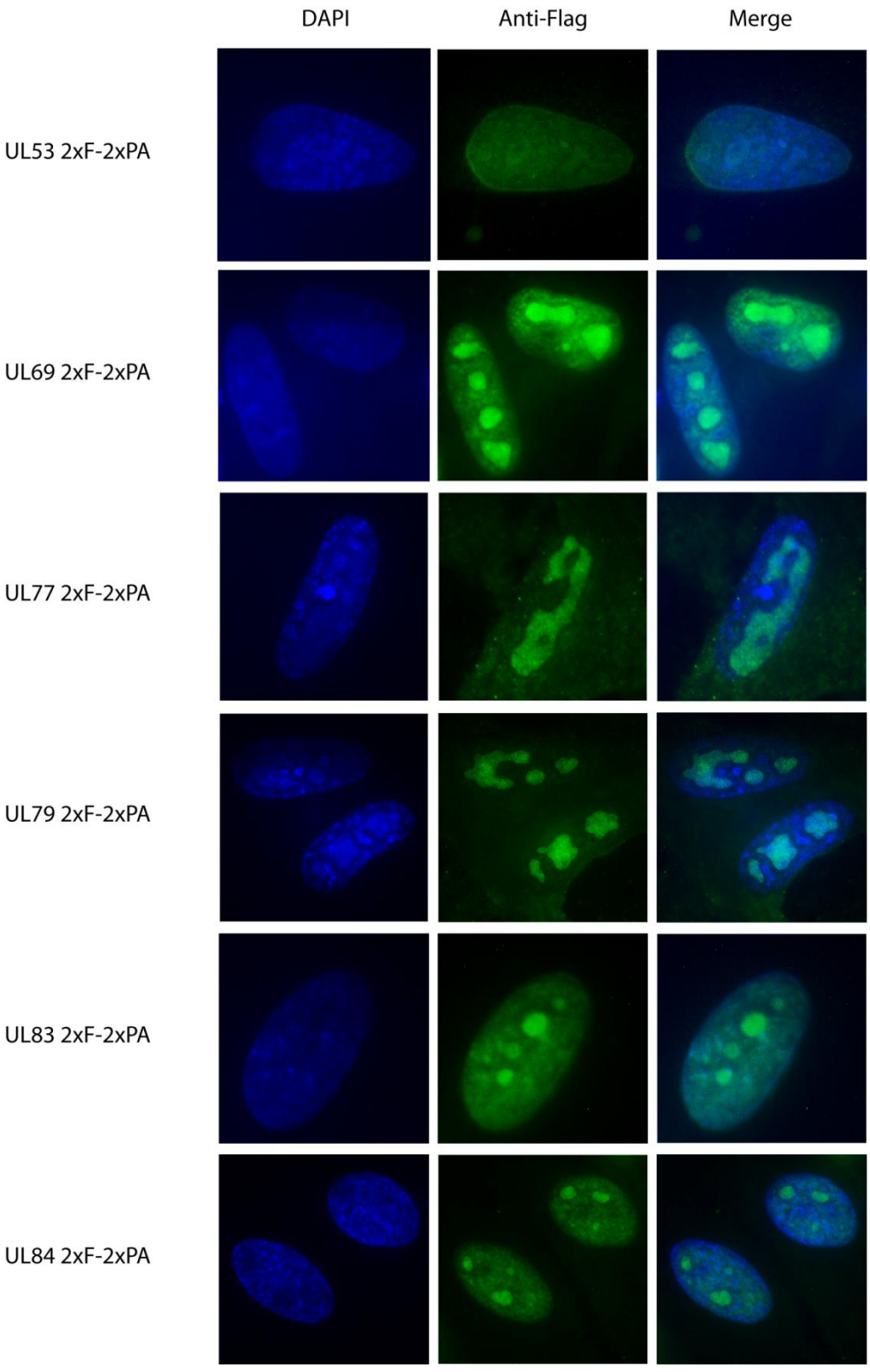


Figure 2.4. (*Continued*)

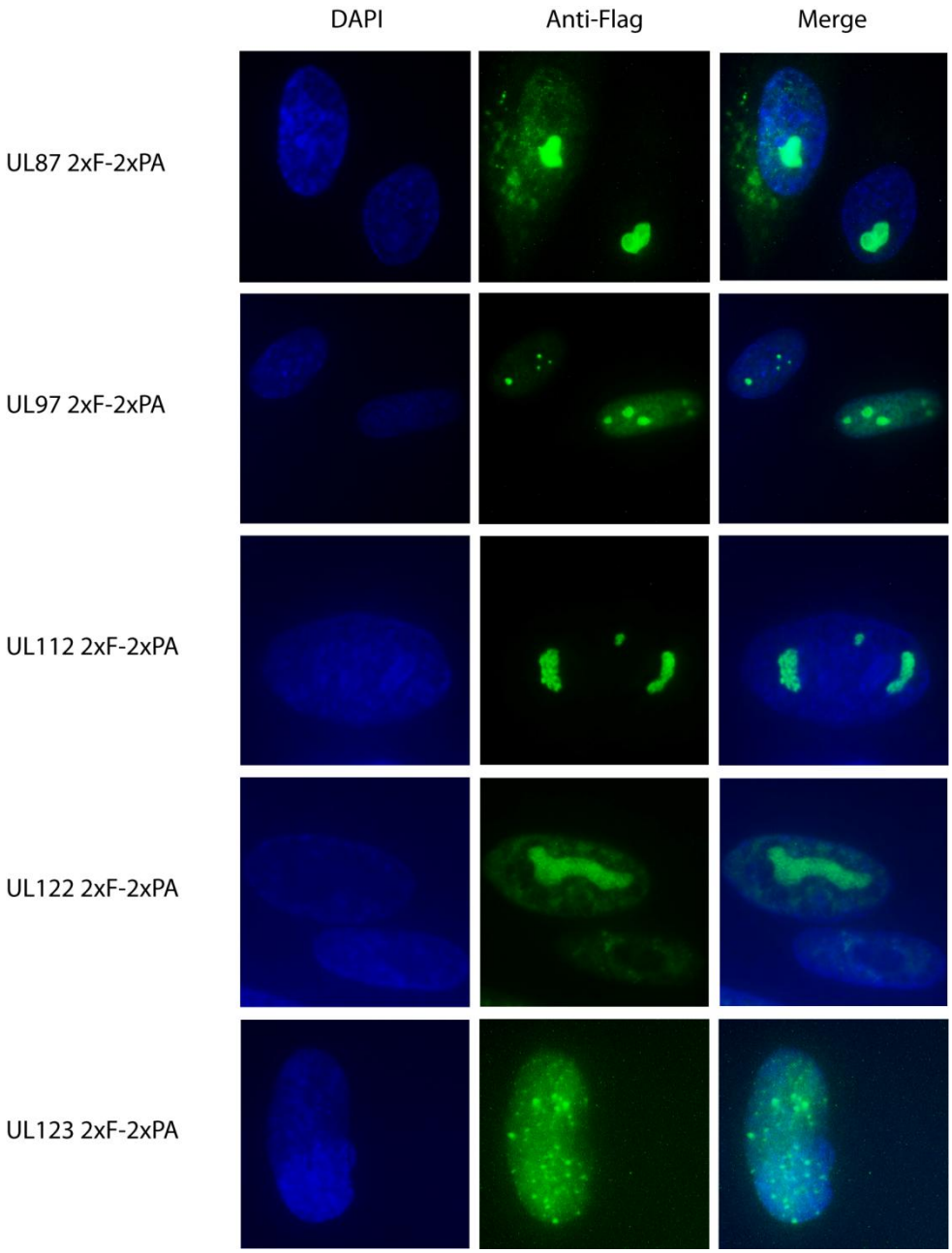


Figure 2.4. (*Continued*)

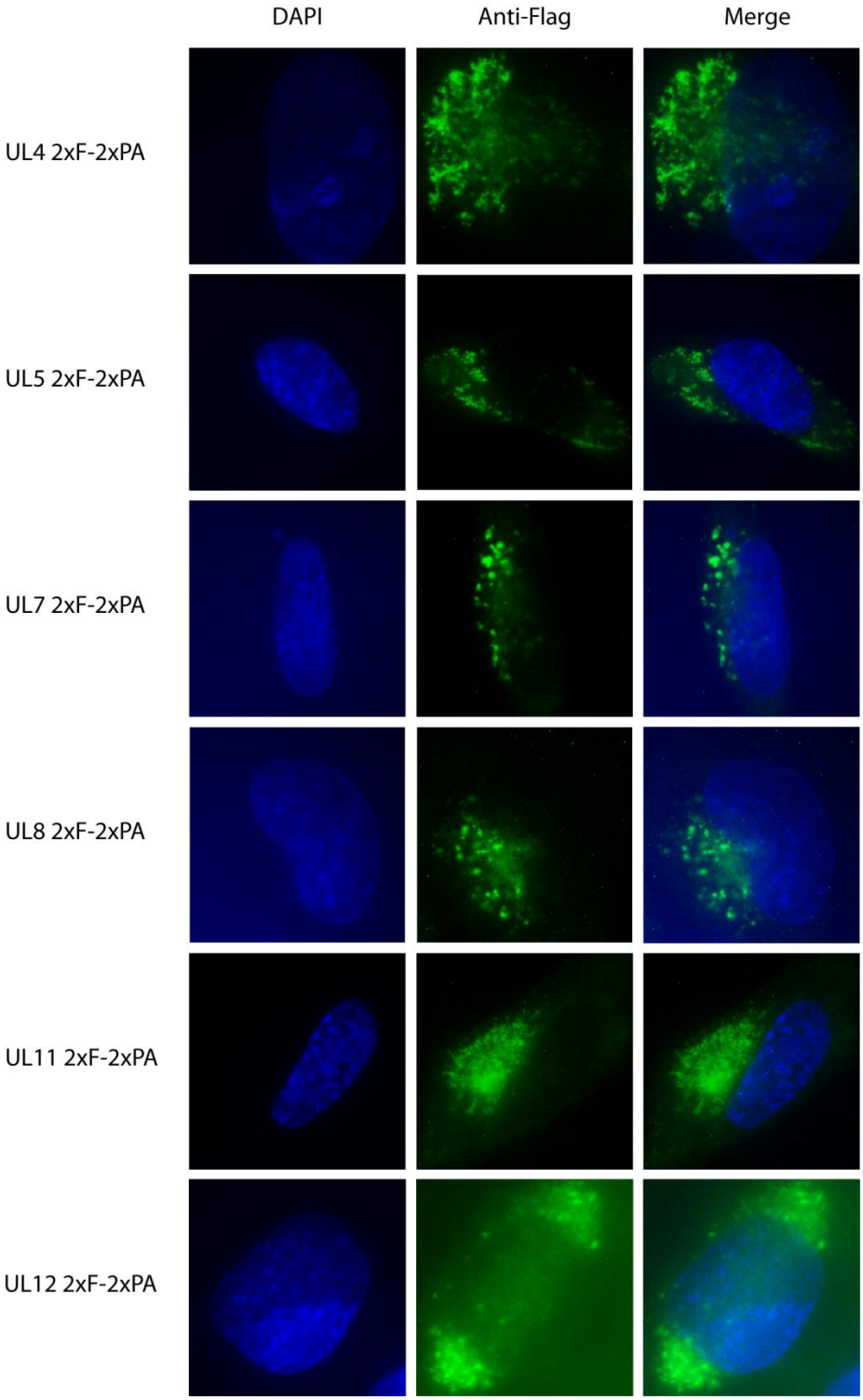


Figure 2.5. **HCMV ORFs with juxtanuclear immunofluorescence localization.**
HFF cells were infected with recombinant virus for 72 hpi. Infected cells were fixed with methanol and stained with primary anti-FLAG antibody and secondary anti-mouse conjugated to FITC. Nucleus was stained with DAPI. Images were taken at 1000x magnification.

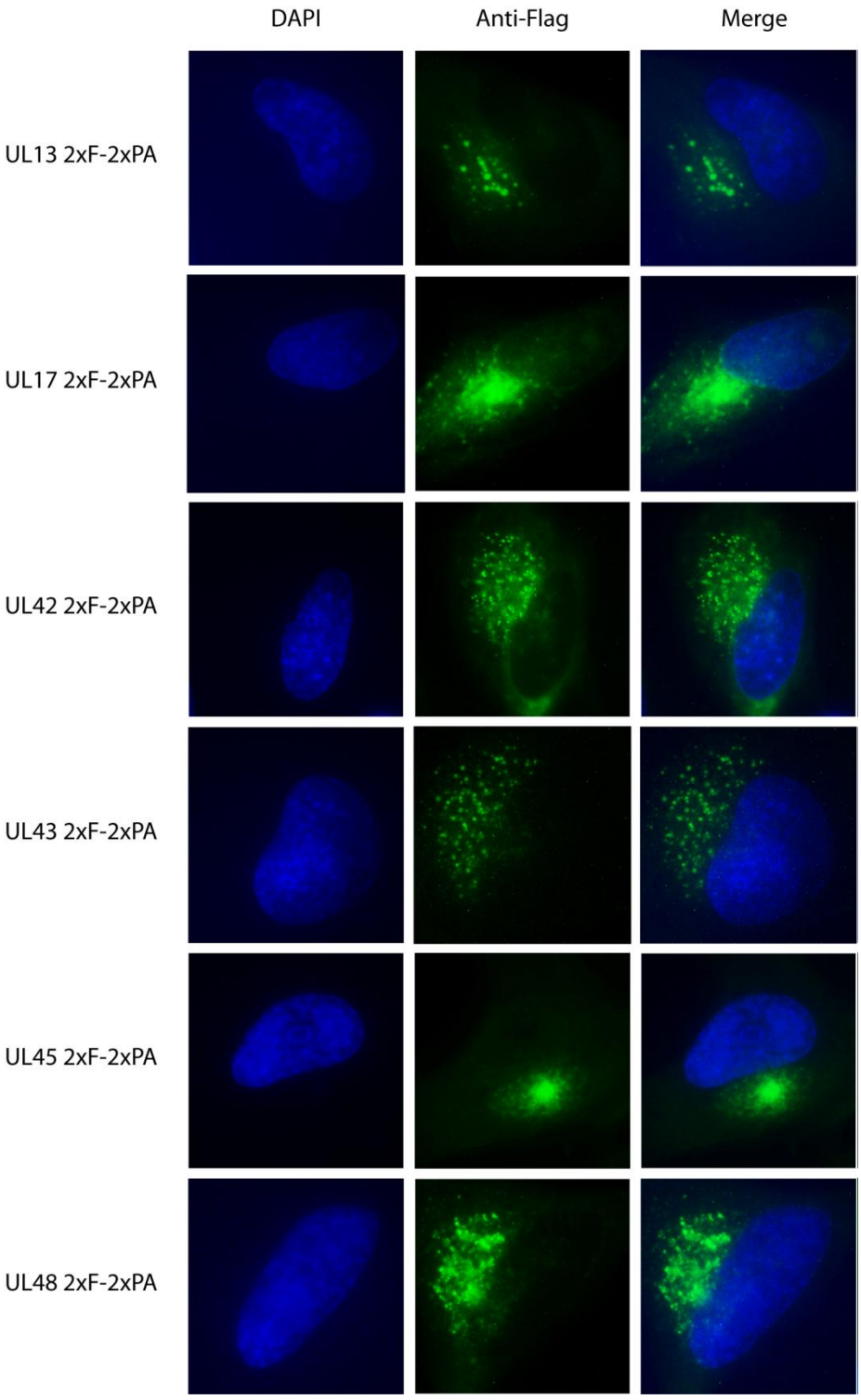


Figure 2.5. (*Continued*)

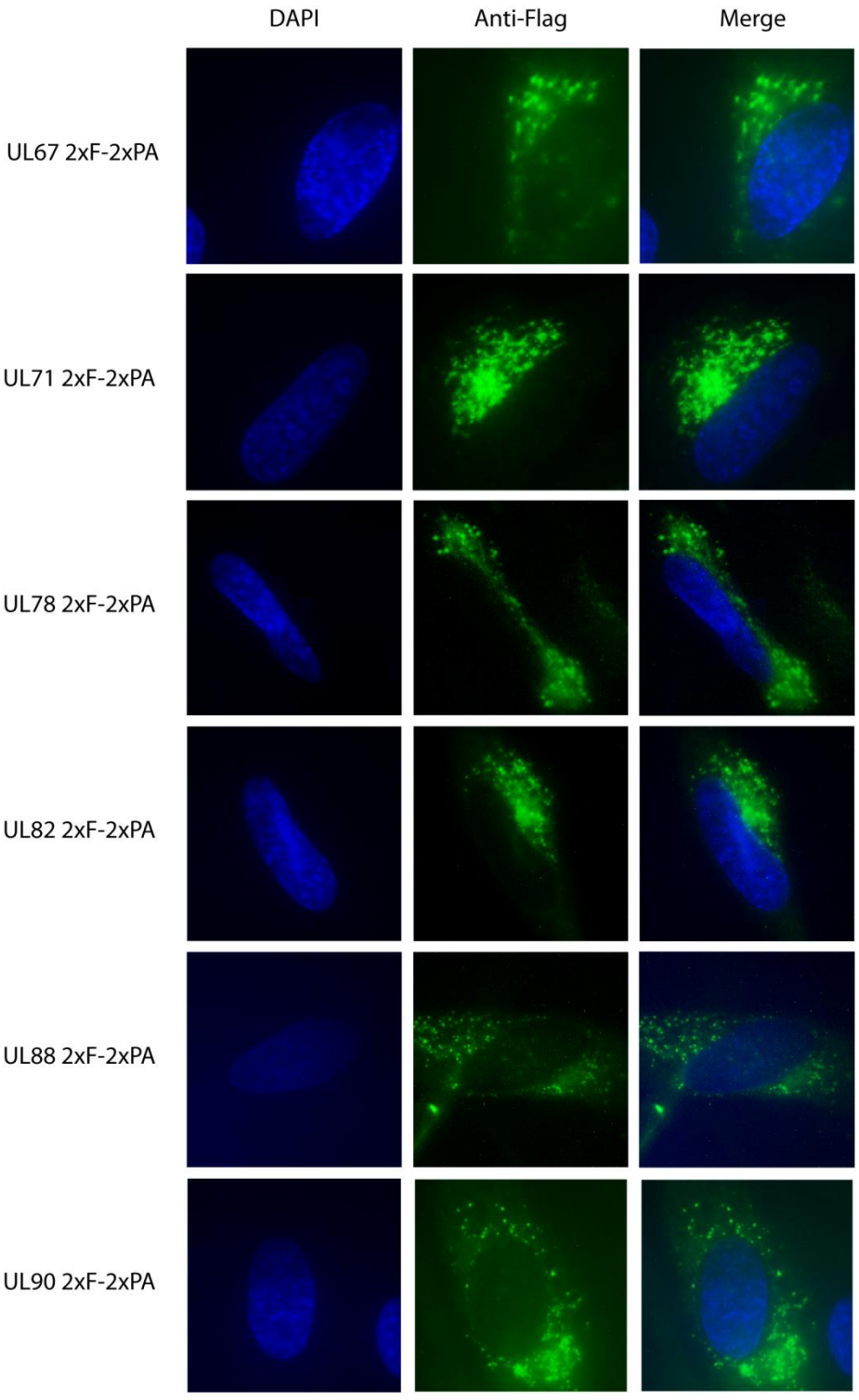


Figure 2.5. (*Continued*)

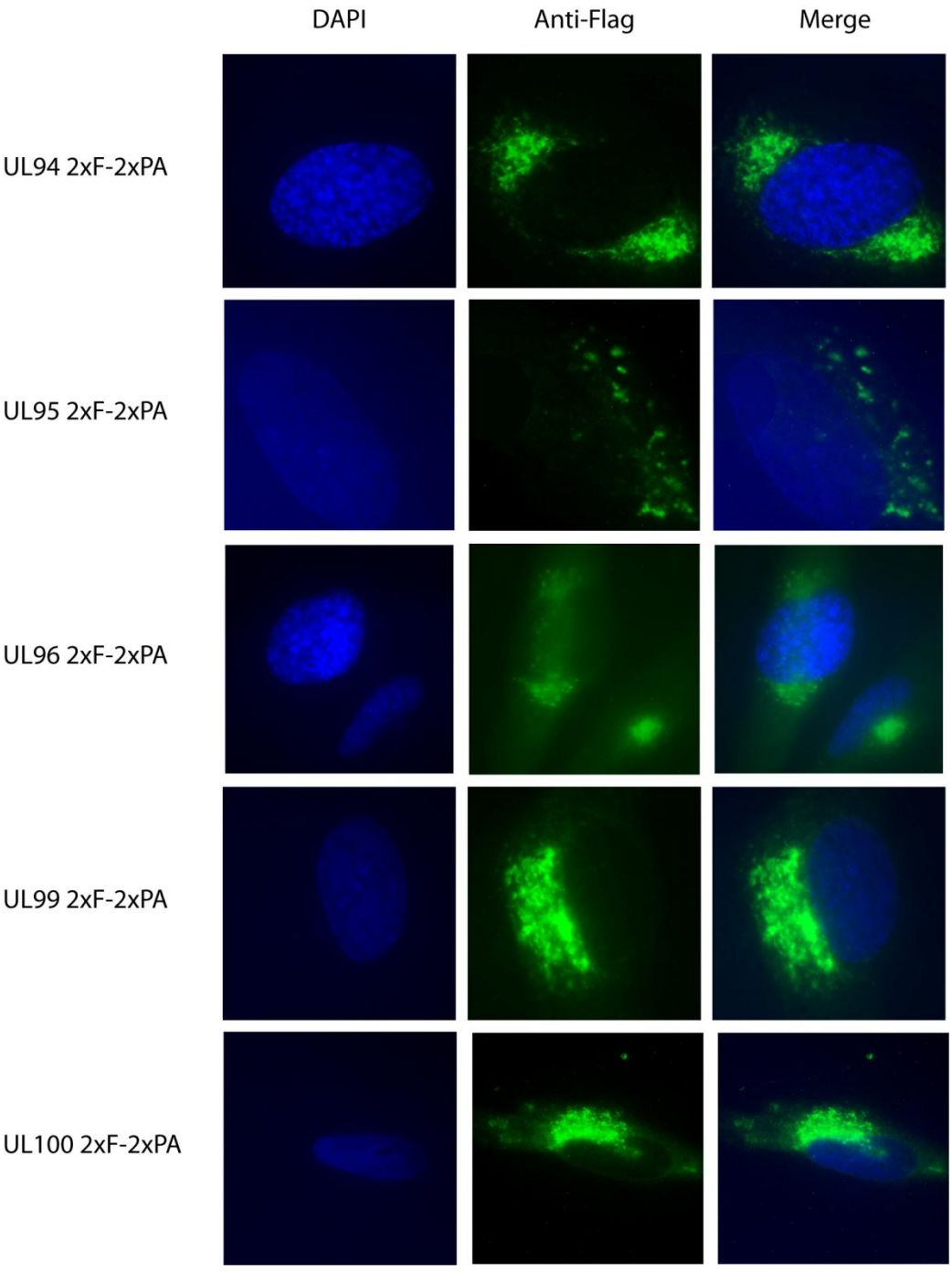


Figure 2.5. (*Continued*)

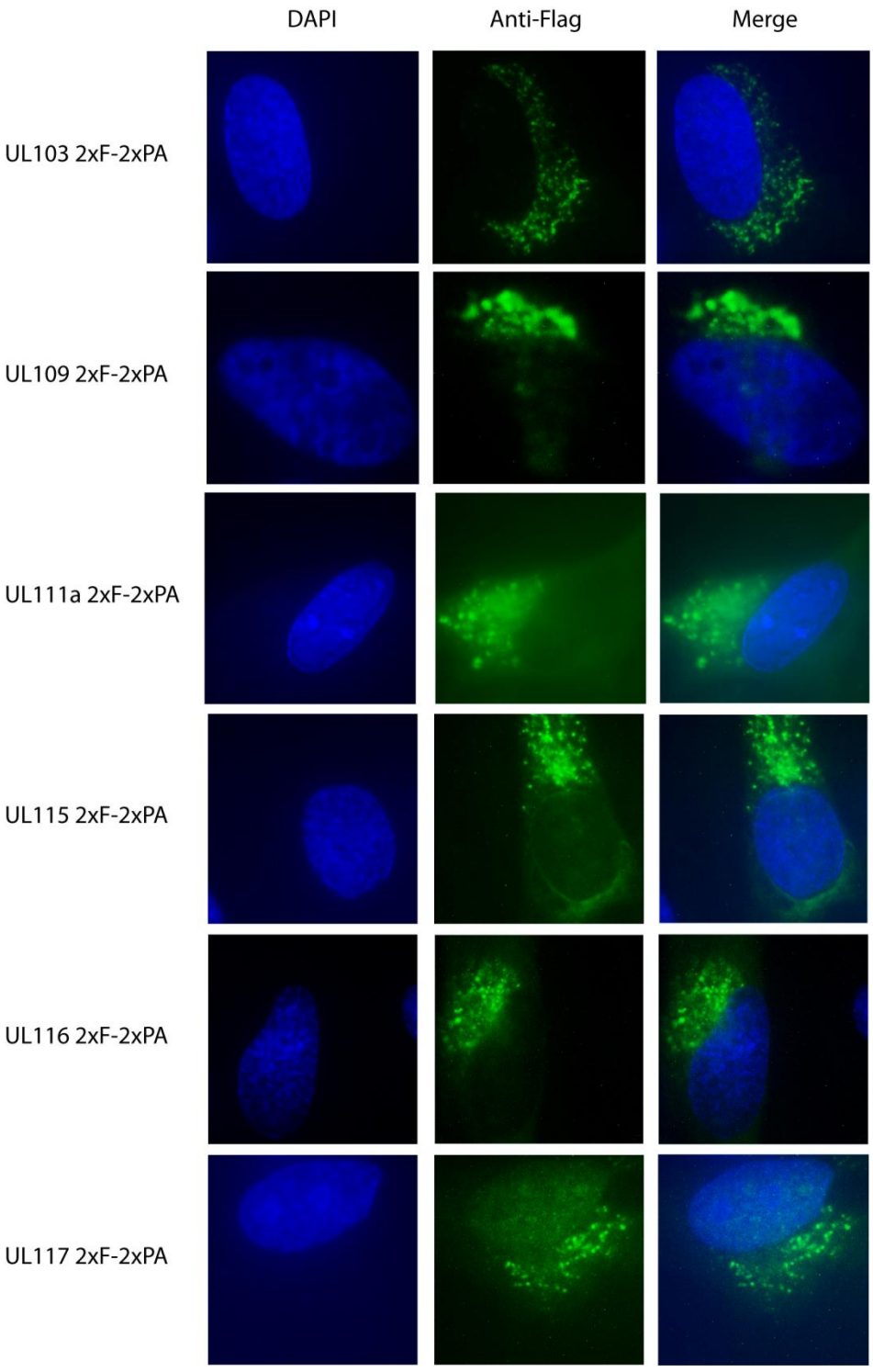


Figure 2.5. (*Continued*)

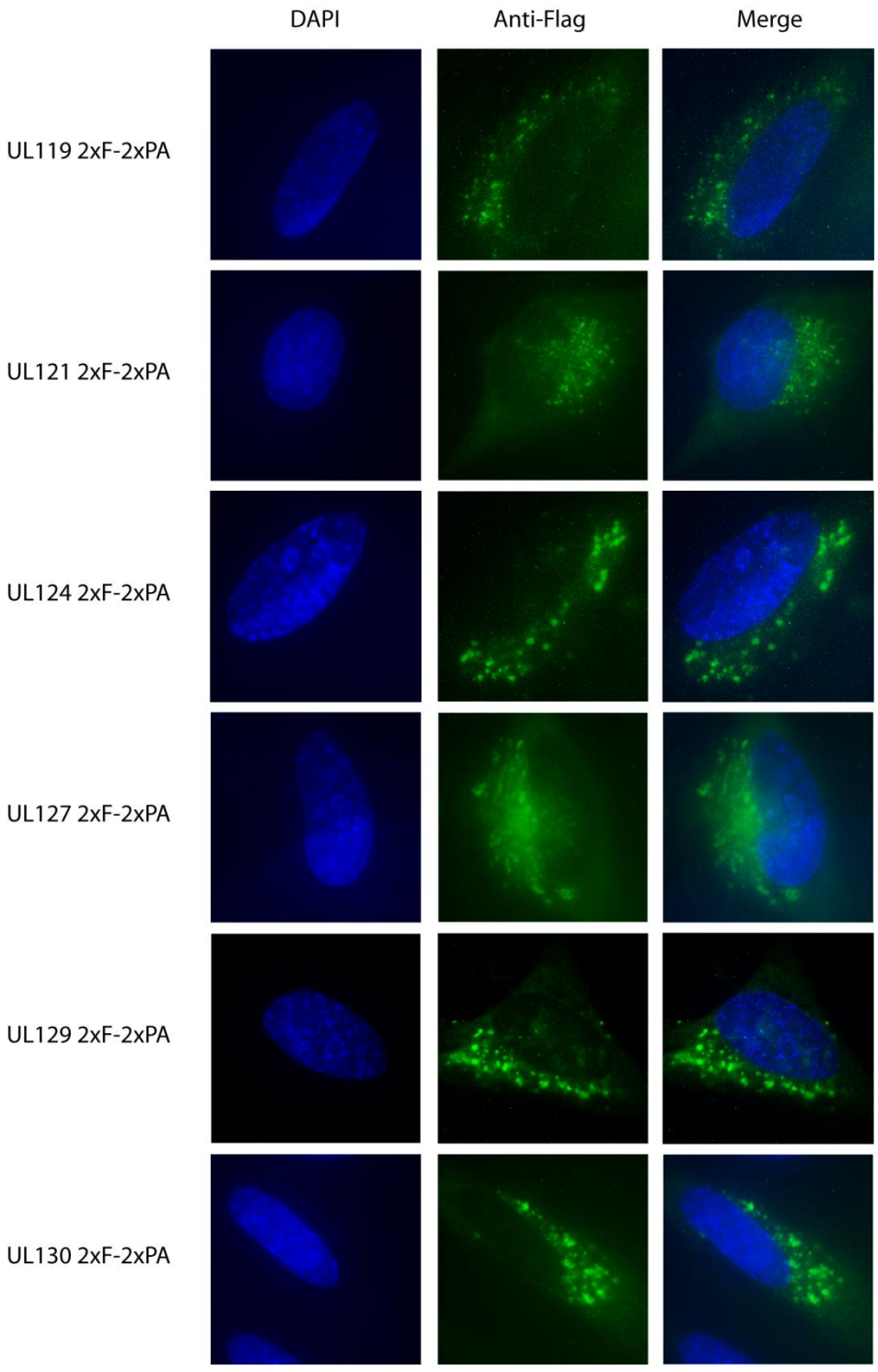


Figure 2.5. (*Continued*)

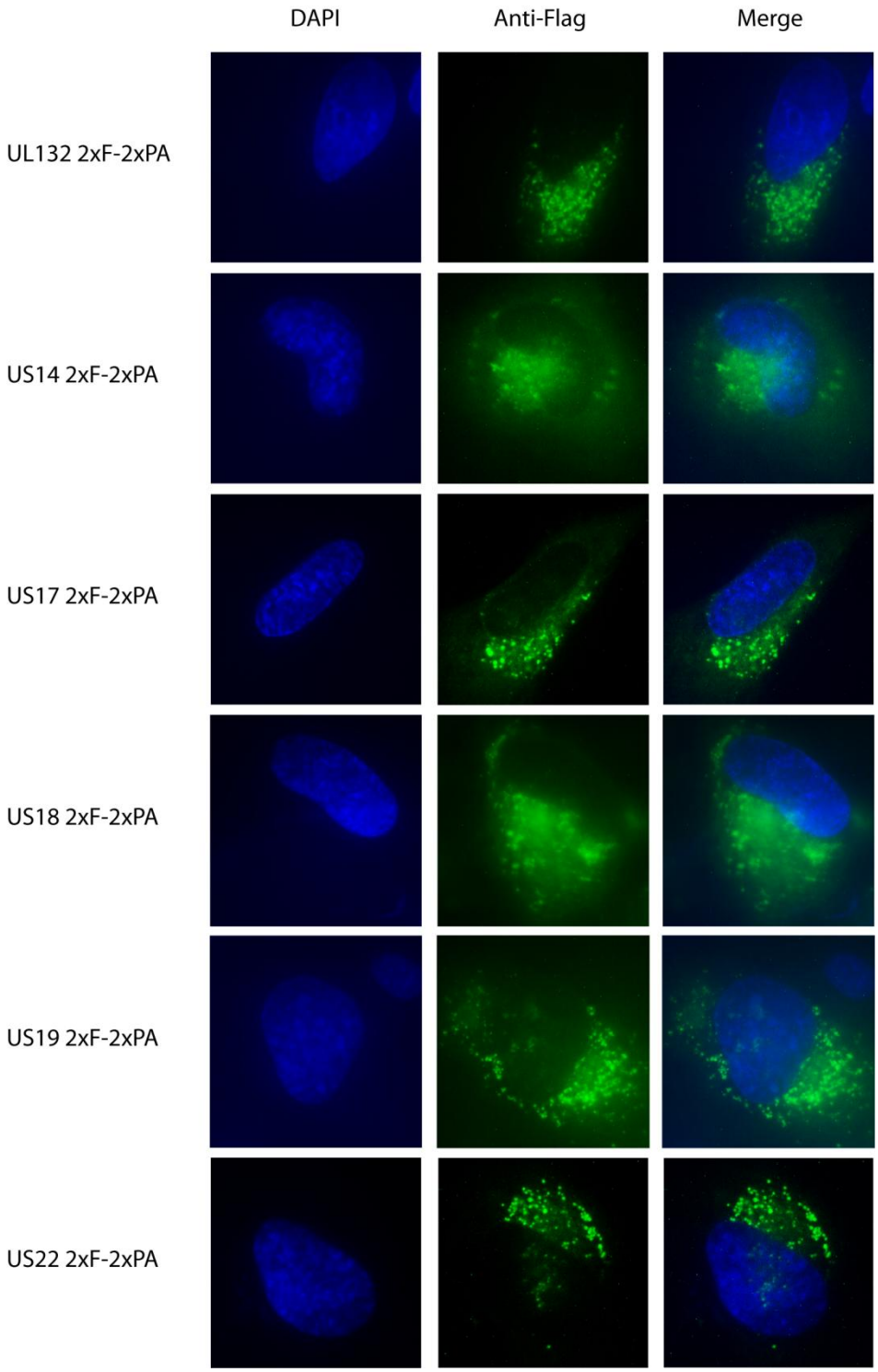


Figure 2.5. (*Continued*)

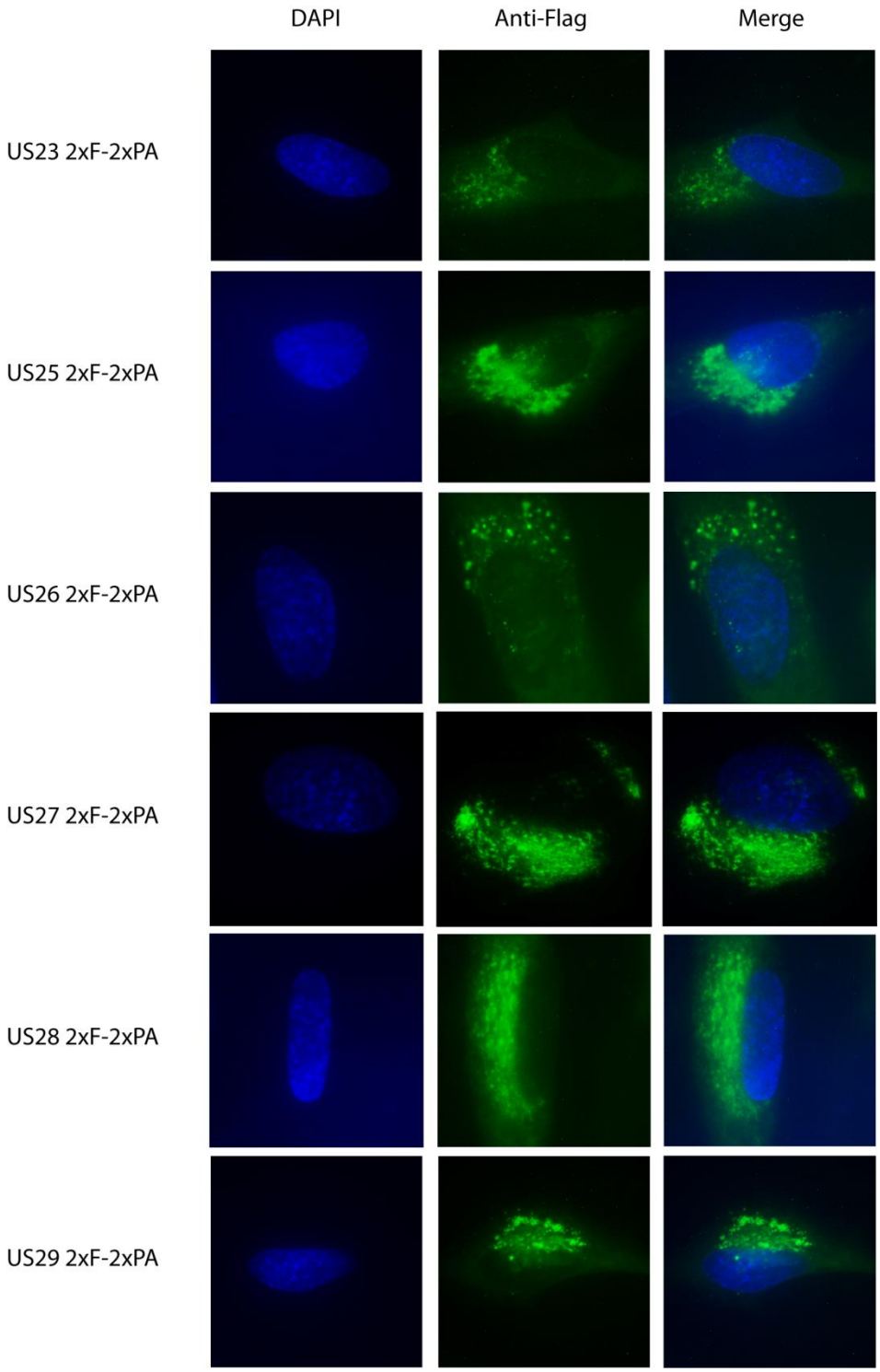


Figure 2.5. (*Continued*)

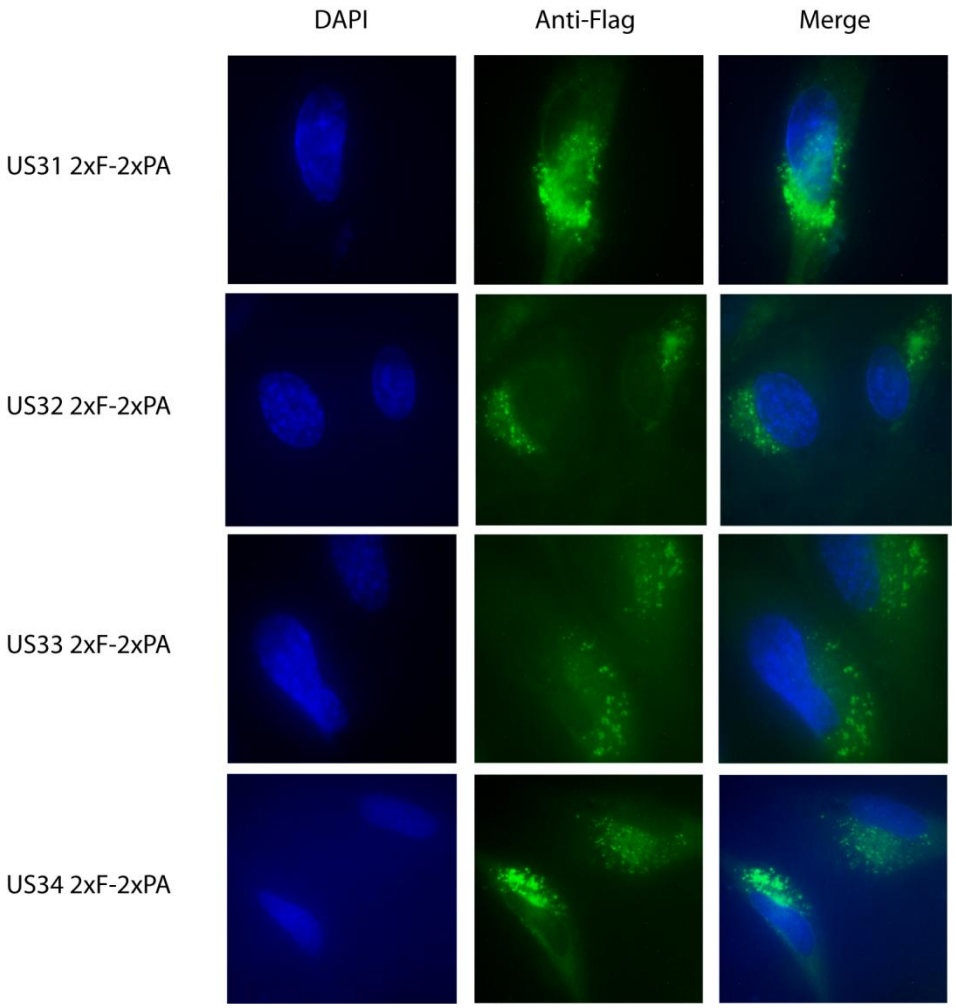


Figure 2.5. (*Continued*)

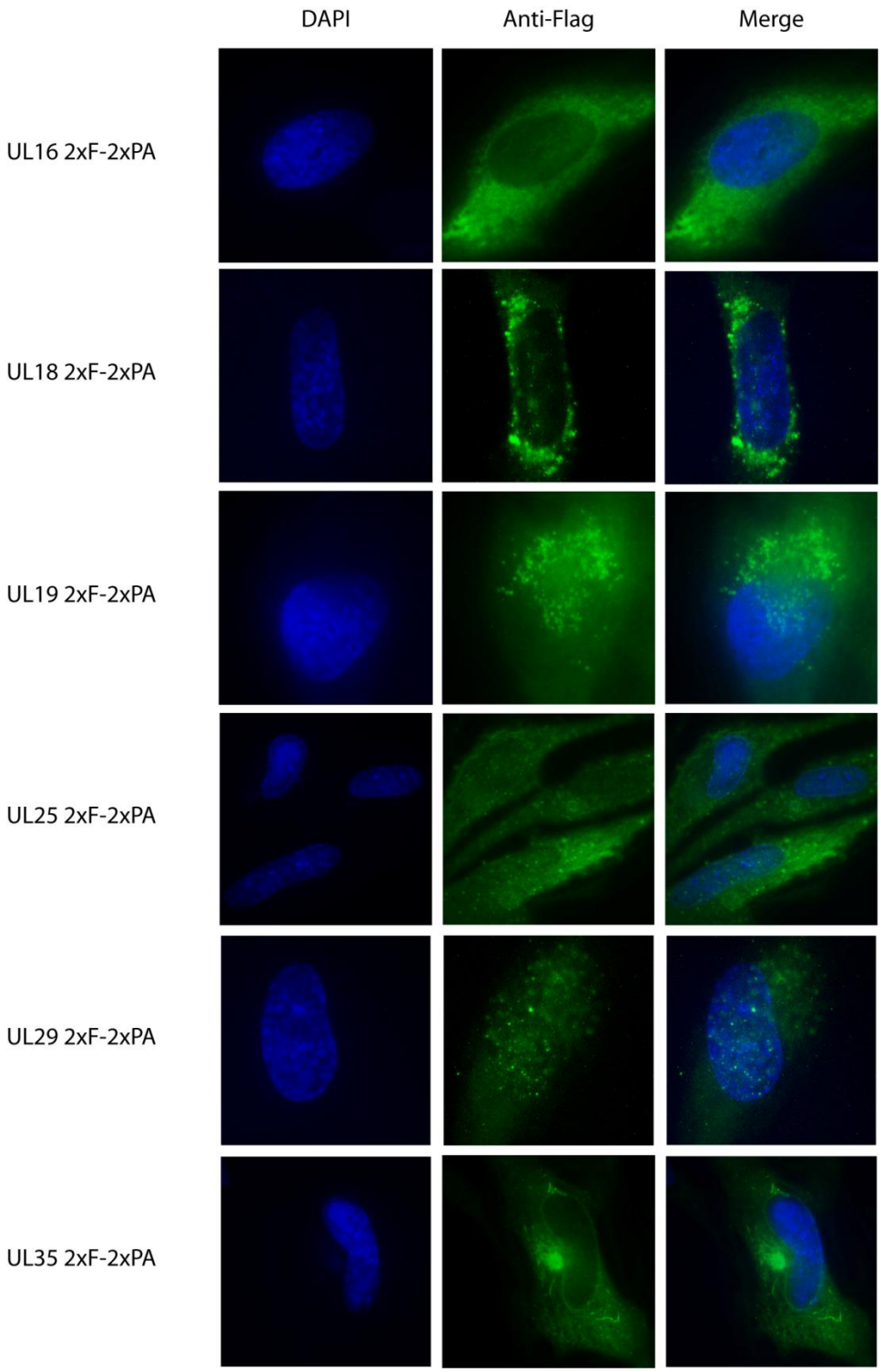


Figure 2.6. **HCMV ORFs with dispersive cytoplasmic immunofluorescence localization.** HFF cells were infected with recombinant virus for 72 hpi. Infected cells were fixed with methanol and stained with primary anti-FLAG antibody and secondary anti-mouse conjugated to FITC. Nucleus was stained with DAPI. Images were taken at 1000x magnification.

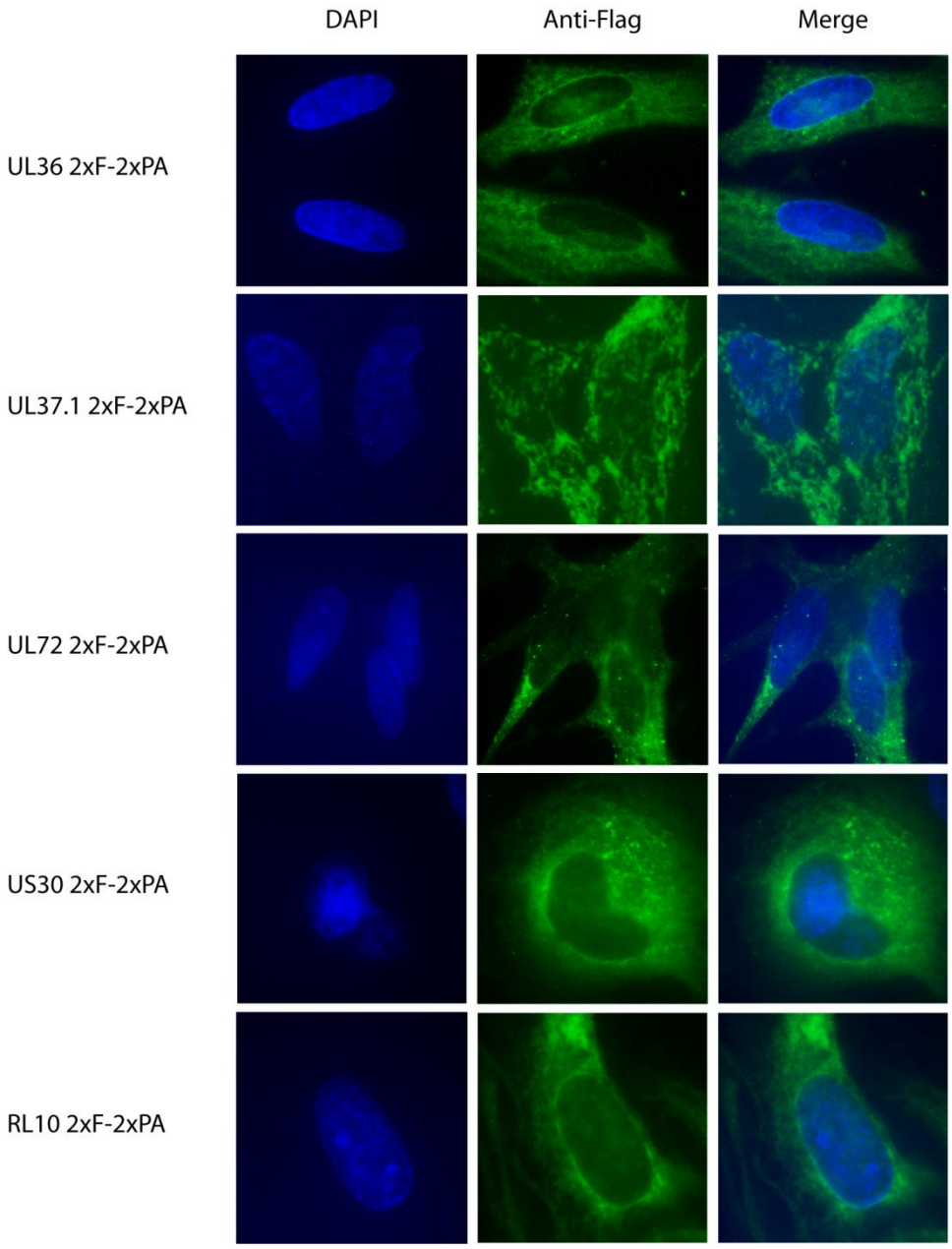


Figure 2.6 (*Continued*)

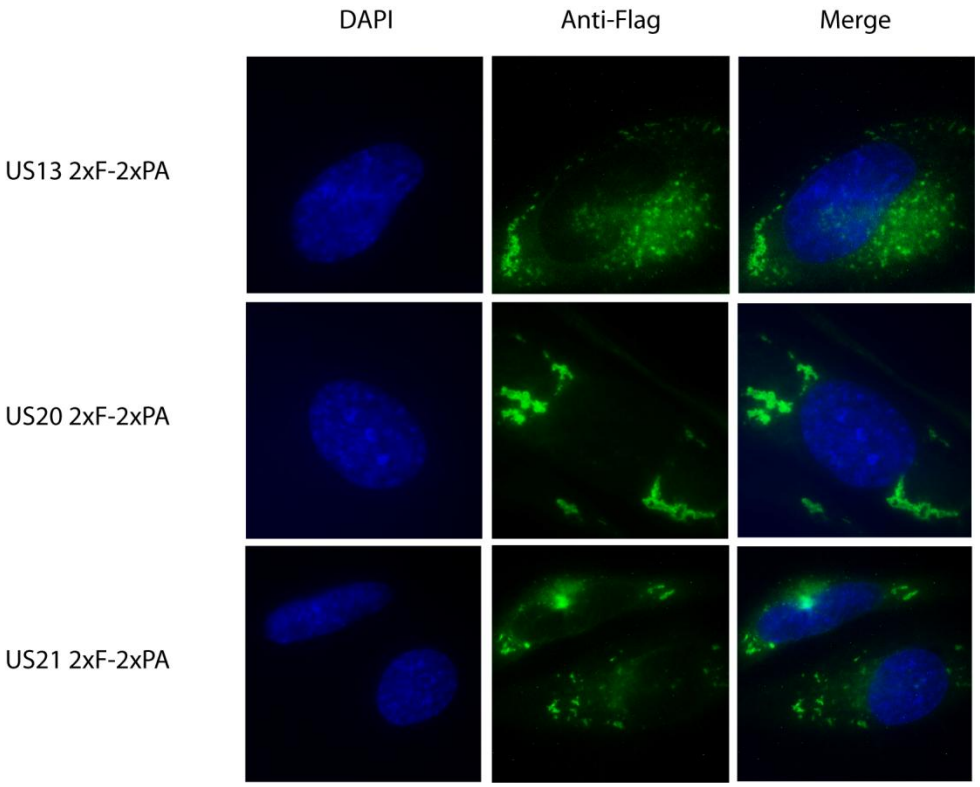


Figure 2.7. **HCMV ORFs with subcytoplasmic immunofluorescence localization.**
HFF cells were infected with recombinant virus for 72 hpi. Infected cells were fixed with methanol and stained with primary anti-FLAG antibody and secondary anti-mouse conjugated to FITC. Nucleus was stained with DAPI. Images were taken at 1000x magnification.

band or bands present and subtracted the 19 kilodaltons (kDa) associated with the tandem affinity tag we added to the C terminus of the viral protein. The results of the viral protein molecular mass profiling are indicated in Table 2.2**.**

Protein Expression kinetics

Our recombinant viruses were also suitable to determine when the viral protein was expressed during HCMV viral replication in HFFs. HCMV gene expression occurs in a cascade fashion with immediate early (IE) genes expressed from 0 to 6 hpi, early (E) genes from 6 to 24 hpi, and late (L) genes after 24 hpi [1]. Using immunoblotting, we found UL24 to be expressed at 4 hpi and UL71 to be expressed as early as 6 hpi, indicating that both proteins were expressed either during the immediate early or early phase of viral replication. We also found UL31, UL45, and UL79 to be expressed only at 72 hours post infection, similar to true HCMV late gene, pp28 [10] (Figure 2.8).

Discussion

In this study, we were able to successfully modify a mutagenesis system to construct 109 recombinant HCMV viruses, each expressing a different viral protein tagged at the carboxyl-terminus with a tandem affinity epitope tag. These recombinant viruses allowed us to determine the subcellular localization of the protein and its molecular weight under the context of infection in a human fibroblast cell culture model for 82 ORFs. By determining the localization of the viral protein, we were able to gain some insight into its function.

For the 17 viral genes that we found to express proteins with nucleus localization under the context of infection, our data was consistent with other labs for the following ORFs: UL28 [11], UL34 [12], UL44 [13], UL53 [14], UL69 [15], UL79 [16], UL83 [17], UL87 [18], UL84 [13], UL97 [19], UL112 [20], UL122 [21], and UL123 [22]. For UL77, although it has already been established that the HCMV gene is analogous in sequence and function to the well-studied Herpes Simplex Virus UL25 DNA packaging protein [23], we believe that we are the first to show the localization of HCMV UL77 to the nucleus. Furthermore, the nuclear staining of UL77, unsurprisingly, was very similar to HCMV UL44, which has been shown to localize to the viral replication compartment in the nucleus [13].

For ORF UL3 and UL31, we believe that we are the first thus far to show the nuclear localization of their proteins under the context of infection at 72 hours post infection. UL3 localized to specific locations in the nucleus similar to HCMV UL44, which suggests that UL3 may also be located in the viral replication compartment. UL3 has been found to be nonessential for viral replication in human fibroblasts [4] and our localization results were consistent with Salsman et al., who also found UL3 to localize to the nucleus using a plasmid overexpression system [24]. The protein expressed by UL31 had a dispersive, pan-nuclear localization and while it isn't essential for viral replication, the deletion of the gene has been shown to significantly reduce viral titers [4]. The function of UL31 has not yet been identified although we can hypothesize from its late gene expression that it may act as either a virion structure protein or possibly a late gene transactivator that enhances expression of other late genes. Our localization study of UL24 exhibited an intracellular localization that differed from Adair et al. While we found UL24 to localize to the nucleus, Adair et al. determined UL24 to localize to a juxtanuclear compartment [25]. This discrepancy can potentially be explained by the difference

ORF	Western blot band (+19kDa tag)	Normalized Peptide Size (-19kDa)	Predicted Size
RL 10	50	31	18804.48
UL 003	72	53	12255.02
UL 004	50	31	17362.69
UL 005	40,20	21,1	18876.12
UL 007	70	51	24434.25
UL 008	100	81	13750.6
UL 011	72,43	53,24	30790.25
UL 012	(no bands detected at 72 hpi)		8206.7
UL 013	95	76	54352.01
UL 016	60	41	26186.42
UL 017	46,36	27,17	12698.72
UL 018	95,35	76,16	41766.23
UL 019	36	17	11404.1
UL 024	55	36	34187.75
UL 025	>100, 26	>81, 5	73538.71
UL 028	smear till 95, 40,35	21, 16	30298.53
UL 029	60	41	40776.96
UL 031	100,95,72	81,76,53	73583.22
UL 034	75,70	56,51	45358.2
UL 035	100,40	81,21	72542.48
UL 036	smear,72,30	53,11	45084.59
UL 037.1	95	78	18968.25
UL 042	50	31	13598.63
UL 043	72,20	53, 9	47785.95
UL 044	70	51	46232.1
UL 045	170	151	101599.71
UL 048	>181		253266.88
UL 053	82	63	42322.97
UL 067	(no bands detected at 72 hpi)		13216.77
UL 069	mulitple bands		82596.17
UL 071	60	41	42180.73
UL 072	72,36	53,17	43491.68
UL 077	(no bands detected at 72 hpi)		70520.19
UL 078	95	76	47356.41
UL 079	50	31	33871.32
UL 082	multiple bands		62000.87
UL 083	multiple bands		62882.52
UL 084	95,34	76, 15	65413.43
UL 087	130	111	104762.82
UL 088	70	51	47638.62
UL 090	(no bands detected at 72 hpi)		7415.58

Table 2.2. **List of molecular weights of tagged HCMV proteins.**
Predicted sizes were determined by direct translation of the predicted ORF.

ORF	Western blot band (+19kDa tag)	Normalized Peptide Size (-19kDa)	Predicted Size
UL 094	55	36	38380.51
UL 095	(no bands detected at 72 hpi)		56958.59
UL 096	48	29	14464.27
UL 097	120	101	78247.12
UL 099	60, 50	41,31	20949.14
UL 100	70,60	51,41	42931.34
UL 103	50	31	28607.25
UL 109	(no bands detected at 72 hpi)		11956.11
UL 111a	40	21	8682.13
UL 112	120,70,50	101,51,31	43672.16
UL 115	36	17	30913.79
UL 116	95	78	34408.11
UL 117	50	31	45490.51
UL 119	45	26	15319.12
UL 121	36	17	20198.74
UL 122	95,72,55	76,53, 36	44772.69
UL 123	95, multi	76	45592.31
UL 124	50,38,35	31,19,16	15854.9
UL 127	(no bands detected at 72 hpi)		15246.53
UL 129	28	9	7181.79
UL 130	55	36	26189.19
UL 132	56	37	29971.25
US 13	45	26	29459.06
US 14	50	31	34238.66
US 17	(no bands detected at 72 hpi)		31934.07
US 18	30	11	30192.79
US 19	45	26	26421.87
US 20	42	23	28559.07
US 21	40	21	26956.29
US 22	smear		65029.84
US 23	95	76	68897.65
US 25	34,20	15,1	5832.68
US 26	95	78	70020.64
US 27	(no bands detected at 72 hpi)		42169.86
US 28	(no bands detected at 72 hpi)		41063.55
US 29	95,55,50	76, 36, 31	51003.02
US 30	72	53	39137.03
US 31	(no bands detected at 72 hpi)		18939.78
US 32	(no bands detected at 72 hpi)		22056.59
US 33	(no bands detected at 72 hpi)		15722.28
US 34	(no bands detected at 72 hpi)		17739.19

Table 2.2 (*Continued*)

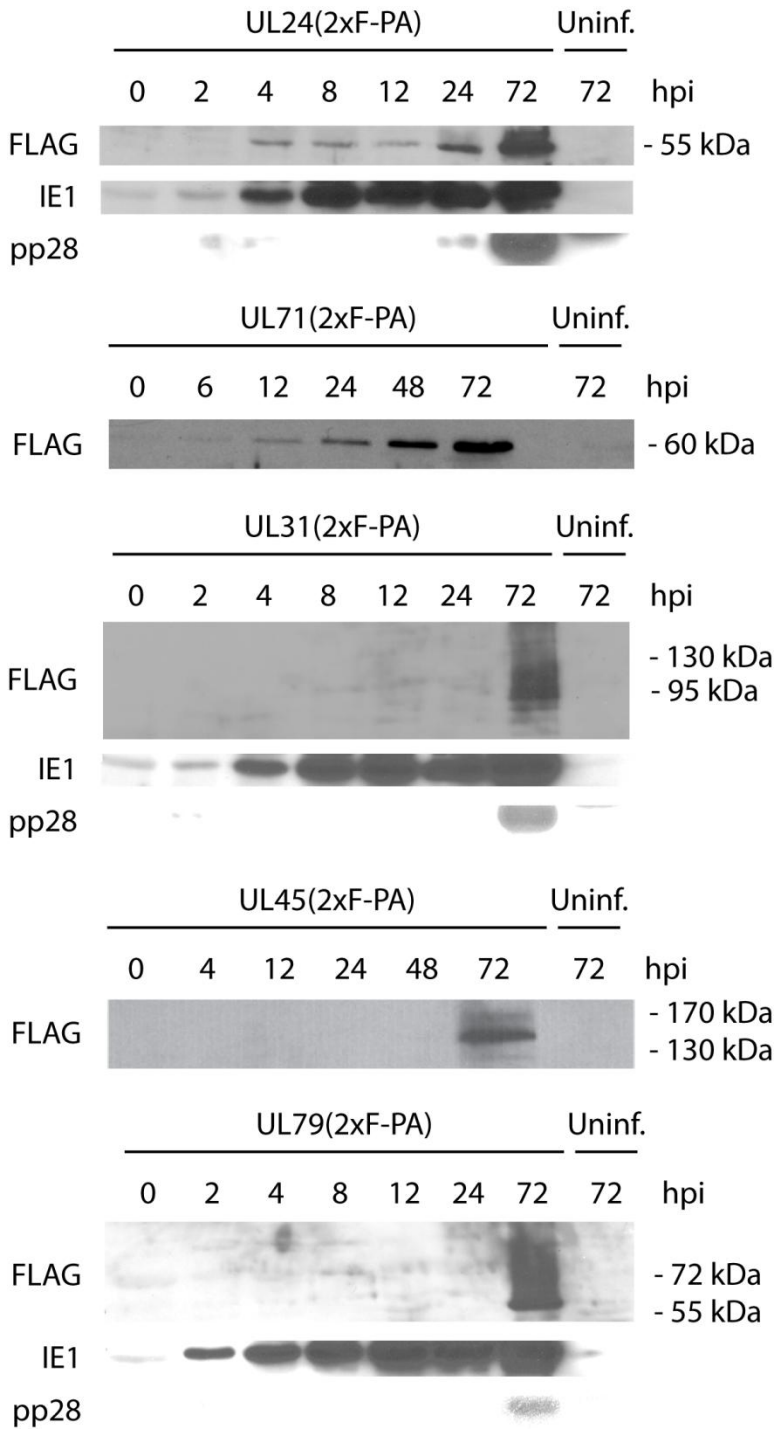


Figure 2.8. **Temporal protein expression of UL24, UL71, UL31, UL45 and UL79.** Human foreskin fibroblasts were infected at high MOI with recombinant virus and harvested at various time points. Lysates were separated by SDS-PAGE and immunoblotted with anti-FLAG antibodies to identify the viral protein. For some of the viruses, the membranes were also probed with IE1 antibody as a standard for immediate early expression and pp28 antibody as a standard for late protein expression.

in viral strains used, given that our HCMV BAC virus contains the Towne genome while their study used the AD169 strain.

We found 51 genes to encode proteins with juxtanuclear localization at 72 hpi. Of these 51 genes, previous labs have also shown similar intracellular localization for UL7 [24], UL12 [24], UL43 [25], UL48 [26], UL82 [27], UL94 [28], UL96 [29], UL99 [24], UL100 [30], UL103 [31], UL111a [24], UL124 [24], UL132 [32], US14 [24], US18 [33], US19 [24], US27 [24], and US28 [34]. Although no immunofluorescence localization studies have been published for UL4, UL71, UL115, UL119, and UL130, the function of their proteins have been characterized. Studies of proteins from UL45, UL71, UL115, UL119, and UL130 have all suggested that their activity takes place in the cytoplasm, which correlates with the results of this study [3, 36-38].

We believe that we are the first to show juxtanuclear localization at 72 hours post infection for the following 22 ORFs: UL5, UL8, UL11, UL13, UL17, UL42, UL67, UL78, UL88, UL90, UL109, UL116, UL121, UL127, UL129, US23, US25, US26, US29, US31, US32, and US34. More work will need to be done in order to determine which specific organelle these proteins are located in, although based upon their location, we can speculate that they may be in the endoplasmic reticulum (ER) or the Golgi apparatus.

Our localization results also differed from other groups for 4 ORFs. US17, US33, UL95, and UL117 were all found by other groups to localize to the nucleus at 72 hpi, while our results did not show a similar localization. For US17, Das et al. found that the nuclear localization was a dynamic process, thus it is possible that our results captured only the cytoplasmic localization. Furthermore, the difference could also be attributed to the human lung fibroblast model cell line used [39]. US33 was shown by Salsman et al. to localize to the nucleus using a plasmid overexpression system, which did not examine the protein under the context of infection. We can speculate that the US33 protein localization may be dependent on the temporal cascade, similar in fashion to UL82, which exhibited a biphasic nuclear localization during the early phase of infection and juxtanuclear localization after 72 hpi [27]. We are uncertain about why our results differed from Isomura et al. for UL95 and Zhikang et al. for UL117, it is possible that differences in the epitope tag or mutagenesis strategies used to construct our recombinant viruses may have contributed to the inconsistence [18, 40]. Lastly, our results for US22 also differed from Adair et al., which found the protein to localize in all cell compartments, rather than at the juxtanuclear site.

We found 11 ORFs to have a diffusive cytoplasmic localization that differed from the juxtanuclear localization. Our results for ORFs UL18 [41], UL 25 [42], UL36 [43], UL37.1 [44], and UL72 [45] were consistent with findings from other labs. We believe that we are the first to characterize the intracellular localization of UL19, UL29, US30, and RL10. When we compared our UL16 and UL35 localization results to the literature, UL16 was found by Dunn et al. to localize to the ER and Golgi [46] and UL35 was found to localize to the nucleus by Salsman et al. [47], while our results indicated a more dispersed localization throughout the entire cell. It is possible that the difference for UL16 could be the result of the model cell line used. In addition, the difference could be the result of the context of infection since Salsman et al. used an overexpression system.

The cytoplasmic localization for US13, US20, and US21 were sufficiently unique to be classified into their own category. All three genes are members of the US12 gene family, but considering that our localization results of other US12 family members (US13, US14, US17, US18, and US19) were different indicates that the unique localization is not a shared

characteristic of the gene family. It would be very interesting to identify the specific subcellular structure associated with these three ORFs.

For the 82 ORFs that we examined, no immunoreactive bands were seen by immunoblotting of lysates from HFFs infected with the recombinant virus for certain ORFs, while our IFA studies were able to detect a signal. Various possible explanations can be made to clarify this discrepancy, including issues with solubility of the protein when we harvested the lysate or difference in sensitivity between western blots and IFAs. To identify the molecular weight of those ORFs, the protein harvesting conditions may need to be further developed.

This project has proven that the ET recombination system can be used to successfully construct recombinant HCMV mutants with epitope tags at the C terminus, thus allowing us to study viral proteins in the context of infection. Our study was able to confirm the results from other groups and also characterize genes that have yet to be fully studied. Furthermore, the tandem affinity tag we used can open up avenues of research relating to immunoprecipitation experiments to study protein interactions. It is our hope that these results will become the foundation of new studies for these uncharacterized genes, which will further elucidate the pathogenesis of HCMV.

Materials and Methods

Cells and media

Human primary foreskin fibroblasts (HFF) (CC-2509, Clonetic, San Diego) were propagated using Dulbecco's Modified Eagle Medium (D-MEM) supplemented with 10% FBS, 1% Penstrep, and 0.2% Fungizone.

Construction and Propagation of Recombinant HCMV

A pZeo plasmid that was modified in our lab to have a tandem affinity tag containing two FLAG epitopes, a tobacco etch virus cleavage site, and two Protein A epitopes was used as the template for PCR to generate the tag cassette. Using the primers listed in Table 1 and the modified pZeo plasmid, we PCR amplified tag cassettes containing the tandem affinity tag and a zeocin resistance gene with 75 nucleotide arms that were homologous to the sequences flanking the insertion site. The resulting PCR products were then run on a 1% agarose gel and the 1.1kb band was gel purified using a gel extraction kit (Qiagen, Valencia, CA). The purified PCR product was then electroporated into an electroporation competent *E. Coli* EL350 strain, which harbored the wild-type Towne BAC. The electroporated bacteria were first incubated in a shaker at 30^0C to allow for recovery and lambda red recombination to occur and then plated on a Luria Bertani (LB) agar plate supplemented with 12.5µg/mL Chloramphenicol and 50µg/mL Zeocin. The bacteria were then returned to the 30^0C incubator for 48 hours. Afterwards, the positive clones were identified by colony PCR using pairs of primers (listed in Table 2) that flanked the ORF that we intended to tag. A successful recombination event was determined by a band shift of 1.1kb when the PCR products were run on a 1.5% gel. To obtain sufficient amount of the mutant tag BAC, we grew up large cultures and purified the BAC using the Nucleobond® PC500 kit (Clontech, Mountain View, CA). The purified mutant $HCMV_{Towne}$ BAC was then electroporated into human foreskin fibroblasts and the cells were then seeded in a T-25 Flask. The DMEM media was changed the day after electroporation and the cells were incubated at

37^0C and 5% CO_2 for 14 to 21 days till 100% cytopathic effect (CPE) was evident. CPE was determined by plaque formation and green fluorescent protein (GFP) expression from the modified BAC.

The virus was harvested by scraping the flasks with a plastic scraper. The infectious inoculum was then frozen in liquid nitrogen and thawed in a 37°C water bath. This freeze-thaw process was repeated three times and the infectious inoculum was then centrifuged for 30 minutes at 4000RPM, 4°C. The supernatant containing the infectious inoculum were then aliquoted and stored at -80°C.

Restriction Digest Profile and Southern Blot

The recombinant HCMV BAC DNA was harvested using lysing reagents from the QIAprep® Spin Miniprep Kit (Qiagen, Valencia, CA) and a modified protocol. The harvested BACs were digested with HindIII restriction enzyme (New England Biolabs, Ipswich,MA) for 16 hours at 37^0C. The digested products were then separated on a 0.8% agarose gel overnight at 4^0C. The resulting gel containing the restriction digest profile of the recombinant wild-type BAC was then placed into a capillary transfer apparatus overnight to allow for the DNA to transfer from the gel to the GeneScreen Plus® Hybridization Transfer Membrane (Perkin Elmer, Wellesley, MA). The membrane was then quickly rinsed in distilled water and the DNA was fixed by UV-crosslinking using a UV Stratalinker® 1800 (Agilent Technologies, Santa Clara, CA). The membrane was then incubated with prehybridization solution (20xSSC/Formamide/Non-Fat Dried Milk) containing salmon sperm DNA (ssDNA) for 3 hours at 65^0C. The template for the radioactive probe was prepared by PCR amplifying the Zeocin Resistance Gene from the modified pZeo plasmid, separating the PCR product from the plasmid template by gel electrophoresis and then excising the ~500 bp fragment from the agarose gel using the QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA). Using the excised fragment we generated the radioactive probe by "random primed" DNA labeling using the Random Primed DNA Labeling Kit (Roche, Mannheim, Germany) and alpha- ^{32}P dCTPs nucleotides (Perkin Elmer, Wellesley, MA). The membrane was then incubated with fresh hybridization solution, ssDNA, and the radiolabelled probe overnight at 65^0C. Following the overnight probing, the membrane was washed with 20xSSC-based solutions of various concentrations, exposed, and visualized using the phosophoimager, Storm 840 (GE Healthcare, Waukesha, WI).

SDS-PAGE and Western blotting

Human Foreskin Fibroblasts (HFF) seeded the day before on a 12-well plate with $1.5x10^5$ cells/well were infected with mutant tag viruses at a Multiplicity of Infection (MOI) of 1. After two hours of viral adsorption, the infectious inoculum was removed and replaced with fresh supplemented D-MEM. The infected cells were then incubated for 72 hours at 37^0C and 5% CO_2. Afterwards, the cells were washed with PBS and then lysed using M-PER (Thermo-scientific, Rockford, IL). The harvested lysates, mixed with beta-mercaptoethanol and non-reducing sample buffer, were run on a 5% Stacking, 10% Resolving SDS-PAGE gel. After gel electrophoresis, the proteins were then transferred from the SDS-PAGE gel to a nitrocellulose membrane (BioRad, Hercules,CA). Next, the membrane was treated with 5% Non Fat Dry Milk (NFDM) for two hours and then stained with primary mouse-antibody against the FLAG epitope for two hours (Sigma-Aldrich, St. Louis, MO). Afterwards, the membranes were washed three

times with PBS supplemented with 0.05% Tween20 and then stained with peroxidase horse anti-mouse IgG secondary antibody (Vector Labs, Burlingame, CA) for two hours. Again, the membrane was washed three times with PBS supplemented with 0.05% Tween20 and then stained with Amersham ECL™ Western Blotting Detection Reagent. The membranes were then exposed to film and the film was developed. For the temporal protein expression study, fibroblasts were infected in a similar matter and then harvested at specific time points. For some of the ORFs, we used anti-IE1 antibody (Millipore, Billerica, MA) and anti-pp28 antibody (Virusys Corp., Taneytown, MD) as standards for immediate early and late gene expression, respectively.

Immunofluorescence Assay

Human Foreskin Fibroblasts (HFF) were seeded on glass coverslips that were placed in a 12-well plate at $1.5x10^5$ cells/well and allowed to grow overnight at 37^0C and 5% CO_2. The cells were then infected with the mutant tag virus at a MOI of 1 and the virus was allowed to adsorb for 2 hours. After which, the infectious inoculum was removed and replaced with supplemented DMEM. The infected cells were then incubated for 72 hours at 37^0C and 5% CO_2, washed with PBS, and fixed with methanol for one hour at -20^0C. After fixation, the cells were treated with 1% Bovine Serum Albumin (BSA) in Phosphate Buffered Saline (PBS) for 30 minutes and then stained with primary mouse antibody against the FLAG epitope (Sigma-Aldrich, St. Louis, MO) for one hour at 37^0C. Next, the cells were washed three times with PBS and stained with anti-mouse secondary antibody conjugated to Fluorescein (Vector Labs, Burlingame, CA) for one hour at 37^0C. The cells were washed again three times with PBS and further stained with DAPI for 5 minutes at 37^0C. The coverslips were then mounted on glass slides using mounting solution and sealed with nail polish. The coverslips were imaged using the Nikon Eclipse TE300 Fluorescence at 1000x magnification.

References

1. Mocarksi, E. S., Shenk, T., and Pass, R. F. (2007). "Cytomegalovirus". In *Fields Virology* (P. M. H. David M. Knipe, Ed.), pp. 2701-2772. Lippincott Williams & Wilkins, New York, NY, USA.
2. Zhang, G., et al. (2007). Antisense transcription in the human cytomegalovirus transcriptome. *Journal of virology* 81: 11267-11281.
3. Varnum, S. M., et al. (2004). Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *Journal of virology* 78: 10960-10966.
4. Dunn, W., et al. (2003). Functional profiling of a human cytomegalovirus genome. *Proceedings of the National Academy of Sciences of the United States of America* 100: 14223-14228.
5. Yu, D., Silva, M. C., and Shenk, T. (2003). Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 12396-12401.
6. Lichty, J. J., Malecki, J. L., Agnew, H. D., Michelson-Horowitz, D. J., and Tan, S. (2005). Comparison of affinity tags for protein purification. *Protein expression and purification* 41: 98-105.

7. Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nature biotechnology* 17: 1030-1032.
8. Muyrers, J. P., Zhang, Y., Testa, G., and Stewart, A. F. (1999). Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic acids research* 27: 1555-1557.
9. Lesniewski, M., Das, S., Skomorovska-Prokvolit, Y., Wang, F. Z., and Pellett, P. E. (2006). Primate cytomegalovirus US12 gene family: a distinct and diverse clade of seven-transmembrane proteins. *Virology* 354: 286-298.
10. Kerry, J. A., et al. (1997). Translational regulation of the human cytomegalovirus pp28 (UL99) late gene. *Journal of virology* 71: 981-987.
11. Mitchell, D. P., Savaryn, J. P., Moorman, N. J., Shenk, T., and Terhune, S. S. (2009). Human cytomegalovirus UL28 and UL29 open reading frames encode a spliced mRNA and stimulate accumulation of immediate-early RNAs. *Journal of virology* 83: 10187-10197.
12. Biegalke, B. J., Lester, E., Branda, A., and Rana, R. (2004). Characterization of the human cytomegalovirus UL34 gene. *Journal of virology* 78: 9579-9583.
13. Xu, Y., Colletti, K. S., and Pari, G. S. (2002). Human cytomegalovirus UL84 localizes to the cell nucleus via a nuclear localization signal and is a component of viral replication compartments. *Journal of virology* 76: 8931-8938.
14. Dal Monte, P., et al. (2002). Analysis of intracellular and intraviral localization of the human cytomegalovirus UL53 protein. *The Journal of general virology* 83: 1005-1012.
15. Winkler, M., Rice, S. A., and Stamminger, T. (1994). UL69 of human cytomegalovirus, an open reading frame with homology to ICP27 of herpes simplex virus, encodes a transactivator of gene expression. *Journal of virology* 68: 3943-3954.
16. Perng, Y. C., Qian, Z., Fehr, A. R., Xuan, B., and Yu, D. (2011). The human cytomegalovirus gene UL79 is required for the accumulation of late viral transcripts. *Journal of virology* 85: 4841-4852.
17. Cristea, I. M., et al. (2010). Human cytomegalovirus pUL83 stimulates activity of the viral immediate-early promoter through its interaction with the cellular IFI16 protein. *Journal of virology* 84: 7803-7814.
18. Isomura, H., et al. (2011). The human cytomegalovirus gene products essential for late viral gene expression assemble into prereplication complexes before viral DNA replication. *Journal of virology* 85: 6629-6644.
19. Marschall, M., et al. (2005). Cellular p32 recruits cytomegalovirus kinase pUL97 to redistribute the nuclear lamina. *The Journal of biological chemistry* 280: 33357-33367.
20. Yamamoto, T., Suzuki, S., Radsak, K., and Hirai, K. (1998). The UL112/113 gene products of human cytomegalovirus which colocalize with viral DNA in infected cell nuclei are related to efficient viral DNA replication. *Virus research* 56: 107-114.
21. Sourvinos, G., Tavalai, N., Berndt, A., Spandidos, D. A., and Stamminger, T. (2007). Recruitment of human cytomegalovirus immediate-early 2 protein onto parental viral genomes in association with ND10 in live-infected cells. *Journal of virology* 81: 10123-10136.
22. Arcangeletti, M. C., et al. (2003). Human cytomegalovirus proteins PP65 and IEP72 are targeted to distinct compartments in nuclei and nuclear matrices of infected human embryo fibroblasts. *Journal of cellular biochemistry* 90: 1056-1067.
23. Isomura, H., et al. (2010). The human cytomegalovirus UL76 gene regulates the level of expression of the UL77 gene. *PloS one* 5: e11901.

24. Salsman, J., Zimmerman, N., Chen, T., Domagala, M., and Frappier, L. (2008). Genome-wide screen of three herpesviruses for protein subcellular localization and alteration of PML nuclear bodies. *PLoS pathogens* 4: e1000100.
25. Adair, R., et al. (2002). The products of human cytomegalovirus genes UL23, UL24, UL43 and US22 are tegument components. *The Journal of general virology* 83: 1315-1324.
26. Ogawa-Goto, K., et al. (2002). An endoplasmic reticulum protein, p180, is highly expressed in human cytomegalovirus-permissive cells and interacts with the tegument protein encoded by UL48. *Journal of virology* 76: 2350-2362.
27. Hensel, G. M., et al. (1996). Intracellular localization and expression of the human cytomegalovirus matrix phosphoprotein pp71 (ppUL82): evidence for its translocation into the nucleus. *The Journal of general virology* 77 (Pt 12): 3087-3097.
28. Liu, Y., et al. (2009). The tegument protein UL94 of human cytomegalovirus as a binding partner for tegument protein pp28 identified by intracellular imaging. *Virology* 388: 68-77.
29. Tandon, R., and Mocarski, E. S. (2011). Cytomegalovirus pUL96 is critical for the stability of pp150-associated nucleocapsids. *Journal of virology* 85: 7129-7141.
30. Krzyzaniak, M. A., Mach, M., and Britt, W. J. (2009). HCMV-encoded glycoprotein M (UL100) interacts with Rab11 effector protein FIP4. *Traffic (Copenhagen, Denmark)* 10: 1439-1457.
31. Ahlqvist, J., and Mocarski, E. (2011). Cytomegalovirus UL103 controls virion and dense body egress. *Journal of virology* 85: 5125-5135.
32. Spaderna, S., et al. (2005). Deletion of gpUL132, a structural component of human cytomegalovirus, results in impaired virus replication in fibroblasts. *Journal of virology* 79: 11837-11847.
33. Das, S., and Pellett, P. E. (2007). Members of the HCMV US12 family of predicted heptaspanning membrane proteins have unique intracellular distributions, including association with the cytoplasmic virion assembly complex. *Virology* 361: 263-273.
34. Fraile-Ramos, A., Kohout, T. A., Waldhoer, M., and Marsh, M. (2003). Endocytosis of the viral chemokine receptor US28 does not require beta-arrestins but is dependent on the clathrin-mediated pathway. *Traffic (Copenhagen, Denmark)* 4: 243-253.
35. Bubeck, A., et al. (2002). The glycoprotein gp48 of murine cytomegalovirusL proteasome-dependent cytosolic dislocation and degradation. *The Journal of biological chemistry* 277: 2216-2224.
36. Schauflinger, M., et al. (2011). The tegument protein UL71 of human cytomegalovirus is involved in late envelopment and affects multivesicular bodies. *Journal of virology* 85: 3821-3832.
37. Ryckman, B. J., Chase, M. C., and Johnson, D. C. (2010). Human cytomegalovirus TR strain glycoprotein O acts as a chaperone promoting gH/gL incorporation into virions but is not present in virions. *Journal of virology* 84: 2597-2609.
38. Ryckman, B. J., et al. (2008). Characterization of the human cytomegalovirus gH/gL/UL128-131 complex that mediates entry into epithelial and endothelial cells. *Journal of virology* 82: 60-70.
39. Das, S., Skomorovska-Prokvolit, Y., Wang, F. Z., and Pellett, P. E. (2006). Infection-dependent nuclear localization of US17, a member of the US12 family of human cytomegalovirus-encoded seven-transmembrane proteins. *Journal of virology* 80: 1191-1203.

40. Qian, Z., Xuan, B., Hong, T. T., and Yu, D. (2008). The full-length protein encoded by human cytomegalovirus gene UL117 is required for the proper maturation of viral replication compartments. *Journal of virology* 82: 3452-3465.
41. Maffei, M., et al. (2008). Human cytomegalovirus regulates surface expression of the viral protein UL18 by means of two motifs present in the cytoplasmic tail. *J Immunol* 180: 969-979.
42. Zini, N., et al. (1999). The novel structural protein of human cytomegalovirus, pUL25, is localized in the viral tegument. *Journal of virology* 73: 6073-6075.
43. Patterson, C. E., and Shenk, T. (1999). Human cytomegalovirus UL36 protein is dispensable for viral replication in cultured cells. *Journal of virology* 73: 7126-7131.
44. Williamson, C. D., and Colberg-Poley, A. M. (2011). Intracellular sorting signals for sequential trafficking of human cytomegalovirus UL37 proteins to the endoplasmic reticulum and mitochondria. *Journal of virology* 84: 6400-6409.
45. Caposio, P., Riera, L., Hahn, G., Landolfo, S., and Gribaudo, G. (2004). Evidence that the human cytomegalovirus 46-kDa UL72 protein is not an active dUTPase but a late protein dispensable for replication in fibroblasts. *Virology* 325: 264-276.
46. Dunn, C., et al. (2003). Human cytomegalovirus glycoprotein UL16 causes intracellular sequestration of NKG2D ligands, protecting against natural killer cell cytotoxicity. *The Journal of experimental medicine* 197: 1427-1439.
47. Salsman, J., Wang, X., and Frappier, L. (2011). Nuclear body formation and PML body remodeling by the human cytomegalovirus protein UL35. *Virology* 414: 119-129.

ORF	2xFLAG-2PA_Tag Forward Primer	2xFLAG-2PA_Tag Reverse Primer
UL 2	CTACGCGGCATATTTTTGATTACGCTCGTCATCTGGACCGTAGTGTGGCTCAAACTGCTTCGAGACGCTCTTTTACGGATGGACTACAAGGACGACGATGACAAGATG	AGGAAAGTGCAGGTTCCTGTTGCTATGTTACCACAGGAGATCACGGAACATAAATGTTTTCTGCGTATGTTTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 3	ATACGCTACATACCTGCCACAGTGGGCAGCTTTTACCATGCCAGAAAACTCGGTCGCGATCCTTATAGTCAAAGACGGATGGACTACAAGGACGACGATGACAAGATG	CCGCGTAGTAGTCGAAAGCCGTACTGAGCGTGCGAAGCGGGTAGGGTGCCGAACGACGGGTATGCGTCGTCGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 4	AGCAGGAATTACTATTTCAGAAGAGAAGATGCGAATTCCACCTTCTATTACTCTTGTTACAATCTGACCGTGTCCCGGATGGACTACAAGGACGACGATGACAAGATG	CTTAACATTTCAAAAGCAACGTAAACACAATAGCTACAGCTGCGCGGTTCTGTGGAACTTCACGTGCGATCTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 5	CGTTTCTGCATAGACCGTCTCCGGGACATCGCCCGTCCTCTGAAATACCGCTATCAACGTCTCGTCGCTACCGTGCGGATGGACTACAAGGACGACGATGACAAGATG	TTTCAGACTCTCTGACTGAAAACAAATATGCAAAAGCAAAACACAACAAACTATACACAGCTGGCTAACTAGCTAAGGCGGCCGCACTAGTGATAGATCT
UL 6	GATTTGGGTTTGCCTCAATCGGCTTGGCGGCGCTGGCGAATCCACGTGGACGATGAAGAACGTGGTCTGTTAACGCGGATGGACTACAAGGACGACGATGACAAGATG	CGCCGGAAACCATGATTTATTTACCACGGTACGATTTGGAAACAGTCACGCTCAAACAGTACTTTTTATTTCCTAAGGCGGCCGCACTAGTGATAGATCT
UL 7	GTTGTCGTGTTTTTGATTCTGATAGTTGTTTGTATTATGGGGTGGTGGAAATTGTTGTGTAGTAAATCAGAATTACGGATGGACTACAAGGACGACGATGACAAGATG	ATTTTCCGACGTAGATAGTCCCGGGCTAGTCATTGTGGGCACAAAACCTTCTCCCTGATAAAAAGCACATTACTAAGGCGGCCGCACTAGTGATAGATCT
UL 8	GCACCTCGAAAAGTTTACGATCGCTGGAAAGACAGTAAACAGTATGGACAAGTGTTCATGACGGACACAGAACTACGGATGGACTACAAGGACGACGATGACAAGATG	AGTTCCAATAATTGGATTTTCGTGTACAAAGAATGATTGTTATCCATCGAAGTAATAACGCGTACCGGAACATCAAGGCGGCCGCACTAGTGATAGATCT
UL 9	ATTTGTTTCAAATTTCCCCAGAAAGCTTGGAATAAATTCACACAATACCGATACAACAGTATGCTCGCCGCCACTCGGATGGACTACAAGGACGACGATGACAAGATG	ATGTTTACGGTATTCAAACAGTGAGCTGAAAAATAAACGTACATAGCTTTTAGTTTCCTCGACGGTGATTCTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 10	CTCCGAGAGAAATGGGACACCAAGGGATACCTTTACAAAGGGACCGATGGCCTGCCCACTACGGACTAATTATCGCGGATGGACTACAAGGACGACGATGACAAGATG	CATGTTCTCAGGTTCCGTAGGGACTGTCCCTATATTCAAACAGTGAGTTTGAAACCGGACATATCCGTCCGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 11	CGAATACCGCAAAAGCTGTGGCTGCTCTGGCAGCATGACAAGCACGGCATCGTGCTCATCCCCCAAACCGATCTGCGGATGGACTACAAGGACGACGATGACAAGATG	GTCCGCCGCGCCAGCTGGCACGGAGTTGGCGTTTCACAGTGATTTCATGCAATCATTTCCTACGCGACTTGCTTAAGGCGGCCGCACTAGTGATAGATCT
UL 12	GCACGGAGTTGGCGTTTCACAGTGATTTCATGCAATCATTTCCTACGCGACTTGCTTACAGATCGGTTTGGGGGACGGATGGACTACAAGGACGACGATGACAAGATG	TTATCATTTTCTACTTTCGAATACCGCAAAAGCTGTGGCTGCTCTGGCAGCATGACAAGCACGGCATCGTGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 13	CCGCATCGTGCCCTGTTTCGTCTATGCTTAGGACTGTGGGTCTCGAGCTACCTGGTTCGACGGCCCATGACGATTCGGATGGACTACAAGGACGACGATGACAAGATG	TCTTTATTGGTAGTGTTTGACACGCATAAGCGACCGGGGATGGGGGGAAATAAAGGAATGGCTCGGTGTATTCTAAGGCGGCCGCACTAGTGATAGATCT
UL 14	GCAGAGGAAACTGCCGCGGGAGAAGCCTCTGCGGTAGCGGCGGCGGCCGTCTCTGAGGAAGAGCAGCAGCGGGAGCGGATGGACTACAAGGACGACGATGACAAGATG	GCCGGCGCCCGTCATCTGACATTCCGTTTTCCTGCCGTGACTGCGAATCATCCGCTTCATGGCTCTCCTCGTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 15	GTGCCGTTCACCCTCCCAACCACAGTTTATACTGCGCAACCTGGAGGGATCGGGCGTCCACAGGTACCGCGTCTGCGGATGGACTACAAGGACGACGATGACAAGATG	ACGGCGCTGGTGGCCGACGTGCACGACCTGCGCCACTCGGACCGCTCCTGCGACCTAGCTTTCGGATCTCGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 16	CGGATCCCCCAGCGATTGTGTCAGCGGCTGCGCATTCGCCTGCCGCATCGATACCAGCGGTTACGCACCGAGGACCGGATGGACTACAAGGACGACGATGACAAGATG	AGACCCGTCCACGGGGGACATGGCCTTCTTATAGCAGCGTGAACGTTGCACGTGGCCTTTGCGGTTATCCGTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 17	CGACGACTTTTCTTTCCTCGGGAGGACTCGGAGCCACTGATGTCGGATCCGGCACGGTCTCCCGAAGAGGAGGAGCGGATGGACTACAAGGACGACGATGACAAGATG	TTATATACTTTCAGTCCGCGTTTTGACCCCTCCTATCTTCTTTGATGATGTATCCTCTTAGCCGTGTGTTGTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 18	CACTACTTTACCACCTTGAAACAATACCTACGTAACCTGGCCTTTGCGTGGCGCTATCGCAAGGTCCGGTCGTCACGGATGGACTACAAGGACGACGATGACAAGATG	TGAGAAACATGACCGTGTTGACGGGGGCACGCGGTAACCGACGTCGAAACAGCTCATACAGGGCGTTGATGGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 19	ACGCTGAAGCGGATTCGGTTACGAAATTCGTGGGTAGCGTCAGAAGACGAGCTAGATGTGAGTCGCAGGGACGCGCGGATGGACTACAAGGACGACGATGACAAGATG	CAACACGAAACCGCGTCACTCACGGACGTAGGTTATTTTGAAAACCTACGTTAATCCTGAACGCGTTTCGTGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 20	CCCCACTTAATAGAATTCCCGCCGTCTCCTCCGCCATCGCCTGGACCCATGCACATGGTGGTCTGCATGCCAGCACGGATGGACTACAAGGACGACGATGACAAGATG	GTGCTCAAGTTCAGTGTAGATTTATGGAAAATATGTAGTCCGTACCGCTTGGGGCTCAGAGTCCAAAGTCCGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 20a	CTTGCTGGGGGTGGTATTACCATCGGAGTTAGTGCTGGTACTGGGTTGTTCCACCGTCACGCTGCAAAAACACGACGGATGGACTACAAGGACGACGATGACAAGATG	GGCACCTTCTCAGAAATCGAAGCGCAGGTAAACAGGTAAGAAATACAAAAAATAACGTGATTGTGAACGCGGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 21	TGGTGCTACTGCTGGACGATGTTCCCCCCCATGTGCCCGGTTTTGCTCCCTACCGCGTACCTCGTCCCCATCCCACGGATGGACTACAAGGACGACGATGACAAGATG	CTTTTATATGATCCCTGTACAGATGTAAATAAAATGCTTTTATTTAAAACTGGTCCCAATGTTCTTCGGGAATCAAGGCGGCCGCACTAGTGATAGATCT
UL 23	GGAACCAGGAGCATGCGGAAGCTAAATAAGCCTACCTGTCCTTACGGAGTGGACCACCAACTTTTTGACGACGCGCGGATGGACTACAAGGACGACGATGACAAGATG	AAAAATAAAAACAGCTTTTTATTTACAGGAACGGGGAAAAAAAGGCACACGGTCCGTGGGAGACGCGGGTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 24	GAAAACCGACGTTTCAGCGGCCACAACGGCATCTACGATCGCGTGCCCGACTGCCCCAAAGGACGTCAGCACCGTCGGATGGACTACAAGGACGACGATGACAAGATG	CGCGAGTCAAACGGGAAGTAGCGTCGGTGTTTTATGCCCCAAGCAGCGTCGTCGTCACTCGTGGCGTCACAGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 25	CCGCGCGACTATCCTCAACGCGACGAGCGGGACCGTCACCGTCGCGACCGCCGCGACAGCGGGGAATACTGTTGCCGGATGGACTACAAGGACGACGATGACAAGATG	ATTAAAAAACAACACATACTTTCCTGTGACTTTTTATCATAAACCGTTCCGCCCTGCTGCTTCGTTCCACCATCAAGGCGGCCGCACTAGTGATAGATCT
UL 26	CGTGTGACGCCGGGCAAGCAAGAAATCACCGACGCGATGTTCGAAGCCGGCAACGTGCCATCAGCGCTGTTGCCGCGGATGGACTACAAGGACGACGATGACAAGATG	CGGTGTGTGGACGCTCGTCCTCTCTGTCTTCTCCGGGTTTTTTTTTTCATGTTTTTTTTTCTTCCTATTTTGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 27	CGGGCCCACATTCAAAAGTTTGAGCGTCTTCATGTACGCCGTTTTCGGCCTCACGAGGTCGGAGGTCACGCCACACGGATGGACTACAAGGACGACGATGACAAGATG	GCGCGCCTGCTCGTCATCGCCGCGCCTCGTCGGTCGATGACCCCACGGTGCTTATAACGCGCCGCCACGGCTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 28	CCCGCGTACGACGTCAGCGAGTACGTGTTTTCGGGCCGTAGCGTGCTGGATAGCGTGAGCGGCACGGGGGCGTCGCGGATGGACTACAAGGACGACGATGACAAGATG	GGCGAAGCGACGCGGAGCGACGGCGGTAAAGCCAAACACCGGCTATATAGCTAGTCATCACAGTCTCCTCCTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 29	AGTATCGGCCCCGTGGACCGCAGCTCGCTGTACGAGGCCAATCCCGAGCTGAGACTGCCGTTCAAAAAGCGTAGGCGGATGGACTACAAGGACGACGATGACAAGATG	GTGTCTCAAACACGCTACTTTCGGTTATAAAAACACCGTCGCCCTATTTCTGGGCGTGTGTACACTGATGACTCAAGGCGGCCGCACTAGTGATAGATCT
UL 30	TGTCTTGTGTTGCCCAACTGCTGTCAGGTCTCCGTAGATCGCTCCCGGGTGCCCGAGACGGGGGGCAGATGGCTACGGATGGACTACAAGGACGACGATGACAAGATG	GCGACCCGCTCCACCCCGTCCTCCACACGCTCAGCCGCGACTGAGCGCCGGGGCGCGCCGCTACTTGGGTTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 31	GCCGCCGGAGCAACCACGTTCTTTCTGGGCGACATGCAATTGCCCGCGGACAACTTTCTCACGTCTCCCCATCCCCGGATGGACTACAAGGACGACGATGACAAGATG	CACACACGCAGACGTACTTTTTAATGAAACCATCGGATAGTGACGTGTCGGGAAAGGAGGACGGACGGAGGGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 32	GGGGGCGCCAAACGCCGTCGGACGCCGTGCAGAACATCCTCCAAAAGATCGAGAAGATTAAGAACACGGAGGAACGGATGGACTACAAGGACGACGATGACAAGATG	CCGTCCTCCTTTCCCGACACGTCACTATCCGATGGTTTCATTAAAAAGTACGTCTGCGTGTGTGTTTATTAACTAAGGCGGCCGCACTAGTGATAGATCT

Appendix Table 2.1 **Tag Cassette Primers**

ORF	2xFLAG-2PA_Tag Forward Primer	2xFLAG-2PA_Tag Reverse Primer
UL 33	AAAACCGCGTCCAACGACACATCCACAAAAATCCCCCATCGACTCTCACAATCGCATCATAACCTCAGCGGGGTACGGATGGACTACAAGGACGACGATGACAAGATG	GGGGACGTGTACCGTATATGAGGGGAAATGGCGACGGGTTCTGGTGCTTTCTGAATAAAGTAACAGGAAAGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 34	GTCGCCGCGACCCCCTCACCTTCAGTCACCCCAGCCCTTACCCCCGTCACGTCCCCCATAACCCCGTTGTGTATTCGGATGGACTACAAGGACGACGATGACAAGATG	GCTCCCTTTGAAAACGAAACCAAAACCAGAGCGGAACTTGAGAAATCAACGCTTTATTGTTCTCCAGTGACGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 35	AGCATGCCGAGCACTAGCTACGGCAGCGACGTCGAAGACCCGCGGGACGATTTGGCCGAAAATCTACGGCATCTCCGGATGGACTACAAGGACGACGATGACAAGATG	AGCATGCCGAGCACTAGCTACGGCAGCGACGTCGAAGACCCGCGGGACGATTTGGCCGAAAATCTACGGCATCTCAGGCGGCCGCACTAGTGATAGATCT
UL 36	GCCGGTCGACGACACATTATTCCCGGCGCCGCCAACGGCATGCCGCCCCTCACCCCGCCACACGCCTACATGAACCGGATGGACTACAAGGACGACGATGACAAGATG	ATGCTGTGTTATATACAAAAACATGCACATAAACAAACGGGACCACCGTGCTCGTCATCCTCTCCTCAATCAGTTAGGCGGCCGCACTAGTGATAGATCT
UL 37.1	CTGCGGAGGCACCGGTTCCTGTGGCAACGGCGGCGACAAGCTCGGCTGCTGCAGCACAACGGCCCCAGCAGTCTCGGATGGACTACAAGGACGACGATGACAAGATG	CAGCAATAGCGGTAAAGTCCCTTCTTATACTATCCCGGAGTCTGTGGTTTTTTTGTTTACCCCTGCTTACTGGTGAGGCGGCCGCACTAGTGATAGATCT
UL 37.3	TTGCTTGAAGATTTCCGCTACATTTGTCGGGACGAGGGAAGTTCCAGCACAAAAACGACGTTCATCGGATCGTCCGGATGGACTACAAGGACGACGATGACAAGATG	AGACTCCTCCGTCCGCCTTCGGAGAACTACGGCGCGGCGGCACGGCCTTTATAGACACTATCAGCGTTGACGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 38	TTGCCGCCGGTCGTGGTGAAGGATGTGATGTACACGGCCGGGGAGGGCGACGTGGTACAGATGGTGGTCGTGGTCCGGATGGACTACAAGGACGACGATGACAAGATG	GTATAGCAACTATTAAAAAATACTACAAAAAAAAAAAGCTGAACATGGTCATCTAGCAGCAAAGTTCTCCTTCTAAGGCGGCCGCACTAGTGATAGATCT
UL 39	GTACATGGCCCAGGTCTCCAGACCGAACGTGGCGGGTCGCAGCTTCAGATGGTAGGCCACCCGCTCCGAGAGTTGCGGATGGACTACAAGGACGACGATGACAAGATG	GTGCACGTGCGCGGCGTAAACGAAAGCGCTTTTGGGCTCACCCATCTGCAGTCCTGTTGCCTGAACGAGCATTCAAGGCGGCCGCACTAGTGATAGATCT
UL 42	TTCTTGGCCGTGGTGTTCACCGTGGTGATTAACCGGGACAATTCCAATACAACGGGGACGGTGTCAACATCGGGACGGATGGACTACAAGGACGACGATGACAAGATG	GGCCAGGTCGTAGCGTAGTTATTATTCCGTAGCAGCAATGATGGTACAGTCAAGCACATGATCTATTTCCCGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 43	TACCGCGCCACCGGCTGTCCCAGAGAACTGTTTTGCCGTCGTTTGCGCAAAAAGAAGGGGCTCTTTGCTCGAAGGCGGATGGACTACAAGGACGACGATGACAAGATG	GGGCTCCATGTCGGTGGCAGTGACGGCGACGGTGGTAACTGTGGTGGAGACGGTACCGACGGCGTCCGCGGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 44	GCGAAAGAGGAGAGCGACAGCGAGGATTCTGTAACGTTCGAGTTCGTCCCTAACACCAAGAAGCAAAAGTGCGGCCGGATGGACTACAAGGACGACGATGACAAGATG	GTCGTCTCCGCCGGCGTTTTCTCCATCGCGGGACCGCGCCGTGCGCGCAGGCCGCGTGCCCGGGCACGCGCTCTAAGGCGGCCGCACTAGTGATAGATCT
UL 45	GACGTGCTTCTGGCCAGGATGCTCAAGTGGTACCACTGGCGCTTTAAGGTCGGAGTATATAAGTACTGTGCCTCTCGGATGGACTACAAGGACGACGATGACAAGATG	ACGACGACCGGGAACGCTGCGAAGACGCCCGGCGTCTAATAATACAGCCGCGCCGAGCCAGCGGGCCCCCGACTAAGGCGGCCGCACTAGTGATAGATCT
UL 46	CACAAAGAACTCAAACTGTCCGTCGTCCGTTTGGACAACTTTAGCGTCGAATTGGGAGACTTTCGAGAATTCGTCCGGATGGACTACAAGGACGACGATGACAAGATG	GTTTATAAAGTCTGTGTGCGATGTGTTTATTTTTTTCTTCTGTGTCTCCTCCCCGTATGCTGTCAGCGCCGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 47	GCGCAGCAGGACATCTTGTCTTCGATCGACCCGCCCGCCGGCTCGACATCGGTGTCCCTGCCGCCGGCCTCGCCACGGATGGACTACAAGGACGACGATGACAAGATG	AGACGCATTGATTGCCCGCTCGCGCTCCAAAGCGAGCGATGTCGCCCTGGTGGCAGCTGGCCTGCGTGACTTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 48	GTGCGATTGGACCTGCGACAACTGGCGCAATCCGTACAGGACACTATTCAACACATGCGGTTTCTCTATCTTTTGCGGATGGACTACAAGGACGACGATGACAAGATG	GATACGTTGTTTTGTAACATAAAAACGATAAAAAGCCTATTGTTTTTATTACCCGCTACTGTCAGTGTCGGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 48A	TCCAGCCTGTTTAATGACAAGTGCGCCTTTAAGCTGGACCTGTTGCGCATGGTAGCCGTGTCGCGCACCCGGCGCCGGATGGACTACAAGGACGACGATGACAAGATG	TTTTTTATGTTACAAAACAACGTATCACTTTTACGGTGATTTATTCTTGCTATTCTTTTTCCCCTTGGGCTGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 49	CACATTGTGCAGCAGGCGCGCGGCGGTTTAGCAGCGGCTTCCTCTTCGTCCTCCCCCCACGGCCTGCCCCATGTCCGGATGGACTACAAGGACGACGATGACAAGATG	TCCAAAACGACGTTGACGACGTGACGGTGTTTTTCGTCCCGCTTGTTGGCCACCGTGGGTCCCGGCGCGGTGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 50	CTCCTGCTGGTGCTGGCTCTGCTGTGGCGCGGTCATGGCCAATCGCCGTCTCCGACACGCTCCGCACACCGCGACCGGATGGACTACAAGGACGACGATGACAAGATG	AGCTGCACATGGAAATGGTCCTCGTGAACGCAGAGGGGCGCGAAGAGACGGCGCAGAGCCTGGTGGAACTCCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 51	TTAACGGCGATCGTCAGCAAGCTCATGGACATTAACGGAATCCTGGAGGGAAAAGACGAGCCGGCGCCGGGTAAACGGATGGACTACAAGGACGACGATGACAAGATG	ACGCCGCGTGGCCTGCACCAGATCCTGATGGAGAACCTTGTTCATCTCCATCGCACCGACGCCACCGCCGATTTAAGGCGGCCGCACTAGTGATAGATCT
UL 52	CGCGATTTCACGCAGCTGTTGGAGAGTTGCGACATCAAGCTGGTGGACCCCACGTACGTGATAGACAAGTATGTCCGGATGGACTACAAGGACGACGATGACAAGATG	CGTTGGCGGCGCTTGCGGAGCAGGGAGCGCAAGGCTGAGCGTCGTTCGCGCGGCGTGCGCACGCCGCTCACGCTAAGGCGGCCGCACTAGTGATAGATCT
UL 53	CCTCAGTCTCAGCAGCATTATCACCGTCCCCAGTCACCACCGCCGCCGCTGTTTCTCAACAGCATTCGTGCGCTCGGATGGACTACAAGGACGACGATGACAAGATG	GCAAGGACAAGTTTTTGCACATGGTGCTGCCGCGGCGCTTGCACTTGGAGCCGGCTTTTCTGCCGTACAGTGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 54	CACATGGTGCTGCCGCGGCGCTTGCACTTGGAGCCGGCTTTTCTGCCGTACAGTGTCAAGGCGCACGAATGCTGTCGGATGGACTACAAGGACGACGATGACAAGATG	ACTACTCTCACCACCAGCCTCAGTCTCAGCAGCATTATCACCGTCCCCAGTCACCACCGCCGCCGCTGTTTCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 55	AACCTGCTAGACCGACTGCGACACCGCAAAAACGGCTACCGACACTTGAAAGACTCCGACGAAGAAGAGAACGTCCGGATGGACTACAAGGACGACGATGACAAGATG	ATGTCCAGTTTAACCCCGTATATCACAAGTCTCTGTCACTTTTTTTGTCTAGTTTTTTTTCTCCTCTTGGTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 56	CGACGCGCCTCATCGCTGCTGGATTTGGCCCGCGACGAACATGGAATCCAGGATCTGGTGCCTGGTAGTCTGCGTCGGATGGACTACAAGGACGACGATGACAAGATG	TGACTGTGAGTAGCAGAAGTTCCACGAGTAGAAGATGAGGAAACCGCAGCACCCAGACAGACGATACACAAGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 57	GAGGCTGGCGGCGTCGGCGGGAGCAGCGGGGGCGGCGGGGGCAGCGGGTTATTGCCGGCCAAACGCAGCCGGTTGCGGATGGACTACAAGGACGACGATGACAAGATG	GGCTGCGACTGTTGCTGCTGTTGTTGCGATGCTCGTGGCTGCGGCGGCCGTTGTCGCGGCGTCTGCTGGCGGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 59	CACCGTTCCCCATCACGGTGTTTCATGGCAATGGGCGAAAGTTACTTTTTACAAGACGAAAAATGTGTTTTTTGGCGGATGGACTACAAGGACGACGATGACAAGATG	TTACGAGTCAATAAAACAGGAAAAGATCCGAACTTTAAAATTGTGTATTTTTATTTTCCCATCCCCTCTTTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 60	CTATATAGGGTATGCATCCTAGGGGTGGAAGCTAGGCTCCGCCCCTACGGGGCGTCGCCTAGCATCCGGTATACGCGGATGGACTACAAGGACGACGATGACAAGATG	GCGTGCGTCACCTCCGGCGGAAATCATCTCTGATGACGTAGCGAGCGAAGCGAGCTACGTCATCAGTCCGTTCTAAGGCGGCCGCACTAGTGATAGATCT
UL 62	CTGAACCAACGTTGGGTGAACTGCCGGGCCCATCGTGTGGGGACGGTGATTGTGGAGGGGCGTACTGACGGCAACTCGGATGGACTACAAGGACGACGATGACAAGATG	GCTGGCCGATGCGGCGGCGAGGTTGGGTGTGGCCGGAAGCGCTCGGGGTCGACGGTGGGCCGCCATGACACCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 64	TATGGAGGGATATCAGGTCATCATTGTGTATCAAAAGATGATTTGTACAACAGGGAAGGATACGGTTTTAAAGGTCGGATGGACTACAAGGACGACGATGACAAGATG	ATGCTCTACTTAGAACCTTTTTAATGACAGAATGAACTCCATGTTATACGCTCTTTATATAGTTTCTCTGCACTAAGGCGGCCGCACTAGTGATAGATCT
UL 67	TTTCACGATCGTTGTGCATCATACAACGACACTTTTTATCCTACAAATTTCACACCTATCTATATACACACAATACGGATGGACTACAAGGACGACGATGACAAGATG	CGCGAAATCAATTGCTAGAACATAAACGTTCTCAACAGGTATGAAATGAACAAACTAGATGATGCTATAACCTTAAGGCGGCCGCACTAGTGATAGATCT
UL 69	GACGACTCGGTGGTAATGCGCTGCCAGACGCCCGATTACGAGGACATGCTGTGTTACAGCGATGATATGGATGACCGGATGGACTACAAGGACGACGATGACAAGATG	CACCATTTTTTCTTTTCTTTGTAACGGGATAAGGGACAGCAATCATCACGCACAACACCCTTCACTCTCTTTTTAAGGCGGCCGCACTAGTGATAGATCT

Appendix Table 2.1 **Tag Cassette Primer** (*Continued*)

ORF	2xFLAG-2PA_Tag Forward Primer	2xFLAG-2PA_Tag Reverse Primer
UL 70	TCCAACGCCAAAAACGTCCACATTTCCATCAAAATCAGACCGCCCGACGCGCCCATGCCGCCGGCGACCGCCGTCCGGATGGACTACAAGGACGACGATGACAAGATG	CGGGCACGCTGATGTATATATAGATGAGACCGCTGCCGGGGGGCGGGTCACCGGCGCCGTGGAAAGTGAGGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 71	TTTGTGGGTTCCGCGCGTGAACTCAGGGGCGTGAAGAAAAAGAAACCCACGGCGGCGGCCTTGCTGTCCTCCGTGCGGATGGACTACAAGGACGACGATGACAAGATG	AGTACGTCTATGAGAAACCCCGTACTCGTCGTGTCCGTGATCACGTACGTTTTCCAAAACGTGCCAGGCTGTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 72	AGCTCCGACGCAGAAAGCCCTGTCAACACGAGTGACGTTATTTTTGACGTCGGCCCGAAACCTCCCCGTCATCCCCGGATGGACTACAAGGACGACGATGACAAGATG	GCACGTTTTGGAAAACGTACGTGATCACGGACACGACGAGTACGGGGTTTCTCATAGACGTACTTTATTAGGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 73	CTCATGGGAGTTTTTTGTATCGTACTACGACATTGCTGTTTCCAGAACTTTACTGCAACCACCACCAAAGGCTATCGGATGGACTACAAGGACGACGATGACAAGATG	TCAGCCTACAGTCGTCCACATATGGAAACCTTACCCCGCCGGAACACCGCCGGGCTGTGAACCTGTCCACCCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 74	GGAAACCTTACCCCGCCGGAACACCGCCGGGCTGTGAACCTGTCCACCCTCAATAGCCTTTGGTGGTGGTTGCAGCGGATGGACTACAAGGACGACGATGACAAGATG	GGACTATGCTTAATGCTCTCATTCTCATGGGAGTTTTTTGTATCGTACTACGACATTGCTGTTTCCAGAACTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 75	CTCCTCATGATGTCCGTCTACGCGCTATCGGCCATCATCGGCATCTATCTGCTCTACCGCATGCTCAAGACATGCCGGATGGACTACAAGGACGACGATGACAAGATG	GTTTTAGAAGATTTTGTGTGAATGTCTTTAACTTTCTCTGTCCCTTTTCTCATAAACTGTCAGGTTCTACAGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 76	ACCACCGTGTCGTCCGCGTTCGTCGTCCCCAGGGACCGGCGTGGCAGGAGCTTCCGCCGTCCCACACGGTCTTTACGGATGGACTACAAGGACGACGATGACAAGATG	GAAGCACCGGCGGCCGCTGACGCGGTCGCCGCGTCAGACGGGGTGGCGGGTCCCGTGATGGCATCGTGCCGACTAAGGCGGCCGCACTAGTGATAGATCT
UL 77	CTCTTTAACGTCAACGATATTTACGAACTGTTGTACTTTTTAGTGTTGGGTTTTCTTCCGAGCGTGGCGGTGTTGCGGATGGACTACAAGGACGACGATGACAAGATG	CGCTCTCGCCGAACGAATATACCCTCTCGGGGACGCTCGCCCTTTATGCAGCAAGCGACACGTGGTGGAAATTAAGGCGGCCGCACTAGTGATAGATCT
UL 78	TCGGCCGAGAGTTCCTCCGTCCTCTGCACCGACGGCGAAAACACCGTCGCGACGGACGCGACGGTAACGGCATTACGGATGGACTACAAGGACGACGATGACAAGATG	CTGGCTAACGACGTGTGAACCTGACGTAATTTATCTGCCACTTTTCTCCCCGCTGCCGTACAACGCCGCCGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 79	GCCTCGTCCTGTCTCAACCCAGAGGCCGTACTGGCCGCCCCGCGCCCTTCATATGCCGACGCTGGCTAACGACGTGCGGATGGACTACAAGGACGACGATGACAAGATG	ACGGTAACGGCATTATGAGCGGCGGCGTTGTACGGCAGCGGGGAGAAAAGTGGCAGATAAATTACGTCAGGTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 80	GCTTCCCAGAGCCCGCCCAAAGACATGGTGGATCTGAATCGGCGGATTTTTGTGGCTGCGCTCAATAAGCTCGAGCGGATGGACTACAAGGACGACGATGACAAGATG	CCGCGGTGTGCTTTATTAGGGTATCACGGTGTAGAAAAAAAAGAGAGGGAAGCCCTAAATATAGCGTCTCTCTTAAGGCGGCCGCACTAGTGATAGATCT
UL 82	TCCGTGCCCGCGCCACGACCCTCGCCCATCAGCACCGCTTCCACTTCCAGCACCCCACGCAGTCGACCCCGCATCCGGATGGACTACAAGGACGACGATGACAAGATG	CTGTTTTTATTATTTAGAAATAAATACAGGGAATGGGAAAAACACGCGGGGGGAAAACAAAGAAGTCTCTCTCTAAGGCGGCCGCACTAGTGATAGATCT
UL 83	CCCAAACGTCGCCGCCACCGGCAAGACGCCTTGCCCGGGCCATGCATCGCCTCGACGCCCAAAAGCACCGAGGTCGGATGGACTACAAGGACGACGATGACAAGATG	GGCCAAAAGTCGCGTGTCTGAGTGGACGTGGGTTTTTATAGAGTCGTCCTAAGCGCGTGCGCGGCGGGTGGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 84	CAAGACTGTAAGCTTACGGAAAATACGACAGAGAAAACGAGTCCTGTCACTTTAGCCATGGTCTGCGGCGATCTCCGGATGGACTACAAGGACGACGATGACAAGATG	ACCGATTTTTTTGTTTTAACCCCGTATCGCGCGGACGCCTAGTGTCCGTTTCCCATCACCAGGGTTCTCTGTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 85	ATCCGCGAACAGGTCGTCTTCACGGTGTGTGACGTTTCACCCGATAACAAATCCGCGACCTGCATATTTAAAGGCCGGATGGACTACAAGGACGACGATGACAAGATG	CGGCGGGCCCGATTCCGAAGGTTGGGGTCGGCGCGTGGCATGCTTGGTGTCTGCGGGCGCGAGAGGGCCGGCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 86	ACGTCCGAGACGCACTTTGGTAACTACGTCGTAGGCGAAATCATCCCGCTGCAGCAATCCATGTTATTTAACTCGCGGATGGACTACAAGGACGACGATGACAAGATG	CCGGGCCCGCGTTTGCGTCTAGGTGACTCGCCGCTAACCTGCGGTCGTCGCCGTCCTCCTCACCGGACGGCCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 87	CCTTCTGTCCCAACCTCCGTCAGCGTTGACGGCAGTTCTGAACCCACGTCGCCGCGAGCGCGGTTTGCATCACGACGGATGGACTACAAGGACGACGATGACAAGATG	CTGCGTCGTGCCAACCCGGCGTCGGACGCTCCTCCGGACGAAACGCCGCGGCGGCAGCGGCCGCGGCTTCCATCAAGGCGGCCGCACTAGTGATAGATCT
UL 88	CGCCTGGGTTACGATCGTCTCGTTTCTGCCGATGCCGGCGTCAGTCGCCGGCACCTGGTGGCTCTGCTGCGTGCCCGGATGGACTACAAGGACGACGATGACAAGATG	CGATCTTATCATAGCCGTCATCATGGCCACCTACCTGTGTGACGAGATACACGCCATCCGTTTCAGGGTCAGCTAAGGCGGCCGCACTAGTGATAGATCT
UL 89.1	AACAAGGACAACGTGATCAGCATCGACCACCGCGGCGCCAAGAGCACGGCGCTTTTTGCCAGTTGCTACAACACCCGGATGGACTACAAGGACGACGATGACAAGATG	AACAAAAAAAGGAGGACATGGACTCGCCACGGTTTGTGGCAAGGCGTATGTTATCATCATGGAGCTACTCACGTTAGGCGGCCGCACTAGTGATAGATCT
UL 89.2	GACGATCTTATCATAGCCGTCATCATGGCCACCTACCTGTGTGACGAGATACACGCCATCCGTTTCAGGGTCAGCCGGATGGACTACAAGGACGACGATGACAAGATG	CCTGGGTTACGATCGTCTCGTTTCTGCCGATGCCGGCGTCAGTCGCCGGCACCTGGTGGCTCTGCTGCGTGCCTAAGGCGGCCGCACTAGTGATAGATCT
UL 90	CGCCTCACCTGGCGGCTCACCTGGCTCACCTGTCCGGCCCCACCCCGACTCTGTAGGCTTCCAGCCGAGCGTGCGCGGATGGACTACAAGGACGACGATGACAAGATG	CACCAGCAGCAAGTGGAAGTTTTGTCCTCGGATGCTCTGTGTAGAGAGGAGACAGAAAAGGGACTCTTATGATTAAGGCGGCCGCACTAGTGATAGATCT
UL 91	TCTCCTGTTCCCTCCGCCCCCAAAACTGTCAGCGACGCTCAGACGTCATGTGCGACGTCTCGGGCGCCTGTGACACGGATGGACTACAAGGACGACGATGACAAGATG	TGCGAAAGCAGCGCGTGCACGCGTAGAGCGAGTGTAACTGGATCTCCTCGGTAAACGCGTTCTGGACGTGCCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 92	AAGAAGCGCGAGGACGCGCTTCTGAAACAGATGCGTTCCGAATATGGAAACGCACCTGTATTCGGATCTGGCGTTCGGATGGACTACAAGGACGACGATGACAAGATG	TCCTCGTTACTCAGCACCTCCTGGTCGAGCACCAGATGTAGAGGCAATTGCTCATCGTCAGCGAACCGCGCCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 93	ATGCATGTGGAAGCTGGTCGGCAAGAGTCGGAAACACCGCGAGTATCGGGCCGTCGCTTGCCGTTCGACGATCTTCGGATGGACTACAAGGACGACGATGACAAGATG	CAGATGATCCATTCGCCGTCGCGGTAGAGCAACAGCAACTGGCATAAGATACACGAGCTGTCGTCCTCCGGACTAAGGCGGCCGCACTAGTGATAGATCT
UL 94	GACCTGGTCAAGGTAGAGGAACCTGTGTCACGTATGATAGTGTGTTCCTGTCCGGTGCTTAAGAACCTAGTGCACCGGATGGACTACAAGGACGACGATGACAAGATG	AACCGTGGCGAGTCCATGTCCTCCTTTTTTGTTGTTTCTTGTTTCTTCTCCCCGTGAACTGTCAGACCCCGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 95	TTCCCCGTCTACCACGTCGGAAAGTTGCTGGACGCCCTCTCGCTGAACGACGCGGGTCTCATCACGTTGAATCTACGGATGGACTACAAGGACGACGATGACAAGATG	GCAGAAACTGATGCTGCTGTCGCTCAAGGTCGACGCGCATCACGTCCTTTAAGAGCTGTTTGTTGACCGACGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 96	GACGACGTGAGAGCCACCTTGGTGGACTCGGTTTCGGCGACGTGCGATTTGGACCTGGAGGTCGACGACGCCGTCCGGATGGACTACAAGGACGACGATGACAAGATG	CACTGTCGTTATGTTGTGTTTTTTTTTTTAATAAAATCTGAACAGAGGCGTGACGGGGATTGCTATACCTGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 97	AACGCCTTTCGGCGCACCACCAGCATAATCTGCGAGGAGGACCTTGACGGTGACTGCCGCCAACTGTTCCCCGAGCGGATGGACTACAAGGACGACGATGACAAGATG	TTTTCCCCGCCGGTGGGTTTGTACCTTCTCTGTCGCCTTTCCCCTCAGCAACCGTCACGTTCCGCGTCCCGGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 98	ACGAACCTACCGATATGGGTGCCGAACTCTGCAAACGAATATGTTGTGAGTTCGGTACCACGCCCGGTGAGCCCCCGGATGGACTACAAGGACGACGATGACAAGATG	CTTCGTCCGAGGAGGAAGTCGGAGGGATGTTGTCGTAGGAGCGTAGAGACACCTGGCGACCCAGAGCATCTTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 99	ACTCCGCGCCAAAAGAAGATTTCACAACGTCCACCCACCCCCGGGACAAAAAGCCCGCCGCCTCCTTGCCCTTTCGGATGGACTACAAGGACGACGATGACAAGATG	CACACGGAGAAACACCCACGGGTGTCCCATTCCCGACTCGCGAATCGTACGCGAGACCTGAAAGTTTATGGGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 100	CACCACGACAGCTTGGAAAGCCGTCGCCTCCGCGAAGAAGGACGACGACGATGATGAAGACTTCGAGGACGCTCGGATGGACTACAAGGACGACGATGACAAGATG	GCTCCCGACCCACGGTCTGCGTGTCTCTTTTTTATGTCCATGTCTCCAAGTCTGGTGCGGGTGGCGGCGGGGTCAAGGCGGCCGCACTAGTGATAGATCT

Appendix Table 2.1 **Tag Cassette Primer** (*Continued*)

ORF	2xFLAG-2PA_Tag Forward Primer	2xFLAG-2PA_Tag Reverse Primer
UL 102	TCGCTGGCGGTTGACGCGCAACACGCTGCTAAGCGGGTGGCTTCCGAGGGCCTGCGGTTTTTCCGGCTCAACGCTCGGATGGACTACAAGGACGACGATGACAAGATG	CTGAGAGCTTTGACGCACTTTTTTTTACTAGTATCCACGTCACTTACCCACGTAGTTCCCCTACGTGACTCGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 103	GTGCTCAACGCCAACTACCCCAATTTGCCCTCACCCCCCAAGCTGCCACCGCGCTGGGAACGGGGAGAGGAAGAGCGGATGGACTACAAGGACGACGATGACAAGATG	GTAAAAGGGACGTCAAAGCTCACTACGTGTTGCGTGTTTTTTTTTCTATGATATGCGTGTCTAGTTCGCTTCTCAAGGCGGCCGCACTAGTGATAGATCT
UL 104	CCAGAGCGTGTCGGCCGATTTCGAAATGGAGGCCCTGATGATCCGCGGCGTGCTGGAGGTCCATACGGATTTCACCGGATGGACTACAAGGACGACGATGACAAGATG	GTGTGCAGCCACAGCTTGTCGCCGCGCACCGTAAAGTCGAGCACTTGCGGCTCCATGATCATCACATTCTGTCTAAGGCGGCCGCACTAGTGATAGATCT
UL 105	CCCTATGAGAAGAACACCGCTATCACCCCCTATATCTGTCGCGCGCTCAAAGACAAACGCACCACGCTTATTTTTCGGATGGACTACAAGGACGACGATGACAAGATG	TAGACAACGTGGTTCTTGTACGTGGTTTTTACCCTGCTCAATAAAGTCACGTTTTCCTTACACGGTGTTGTGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 108	AAAGGACAGATATCAATCTTCTTAACTTATATCATGGTAAATTCAGAATCCTATGGTGTCTTATTATCTCTAAAGCGGATGGACTACAAGGACGACGATGACAAGATG	TCATGATAGTAGCCAGGTTATAAGGATCATATATATCTCGTCAGGGAAATACAAGTTAGACCATAATGTTGACTAAGGCGGCCGCACTAGTGATAGATCT
UL 109	CTTCGCGAGGGACATTATCGCCGCCGACCCGCGTGTGCCCTGCCTCACGGCTGGTCCGTGATGAACTCTTGCTCGCGGATGGACTACAAGGACGACGATGACAAGATG	GCACTACTATGTAACTAGGATATGGTCCTATAAGGTGTCTTCTATCACGGTGGCTTGTTTATCGCTTGGCGGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 110	TTTATCCTCTTTTGTTGTTCCAAAGAGAAAAGAAAAAAAAAACAAGCAGCCACTCTAACCATCACACTCACGAGTCGGATGGACTACAAGGACGACGATGACAAGATG	CGTAGTGACGTCGTTGCTGCTGCTCTTGCCTTCATGGCGGTATTTCTCTTCCTCCCCCCTAACCCCATATTAAGGCGGCCGCACTAGTGATAGATCT
UL 111a	CATCGAGTAAAACCTACGTTGGTAGGTCACGTAGGTACGGTTTATTGTGACGGTCTTTCTTTTCCGCGTGTCGGGCGGATGGACTACAAGGACGACGATGACAAGATG	CAGCCTTTGACCACCGTACCGTCGAGCCACACGGAGTAGTCGTCCTCACGTTGCTACAAGAGGAAAACTACGTCAAGGCGGCCGCACTAGTGATAGATCT
UL 112	GACGGACTCGAAATTATCGATCCCAGCGAGGAAGCTGCGATCGCCGCCGCCTCGATCGCGGCGTTTTTCGACGATCGGATGGACTACAAGGACGACGATGACAAGATG	ACCAAAGGGGGTGGAAACCCAAATGTCCAAGCGCGTCCTGTTTCATAATTTTTCCGGTCTCGGCTCGGTTTTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 114	AACGACCATTTTATACTGGCCAACGCGTATCTGGACACGCATTACCGAGAGACCATGGACTGGCGACTCTGTGGGCGGATGGACTACAAGGACGACGATGACAAGATG	ATTTTTTTTACAATAAAGAGATAGACATCACAGTTCACCACCTTGTCTCCCCGGTGTGTCTATTATCATCAATCAAGGCGGCCGCACTAGTGATAGATCT
UL 115	CCGCCCGAGCTGAAGCAGACGCGCGTCAACCTGCCGGCTCACTCGCGCTATGGCCCTCAAGCAGTGGATGCTCGCCGGATGGACTACAAGGACGACGATGACAAGATG	CAGTCGGCGCTCAGACAAAACACGCGAGCCTGCTCGTCGGGGGTTAACAGAGAGCCTTTATTATCAGCAATGTTAAGGCGGCCGCACTAGTGATAGATCT
UL 116	GGGATCGAAGACCTGAGCGCCAACTTTCGGCGCCAACTGGCTCCTTACCGTCACACTCTCATCGTGCCGCAGACTCGGATGGACTACAAGGACGACGATGACAAGATG	GCAGAAGGCAACACCACAGCAGTGCCACCGGTCCAGGTGAGAAAGAGAAGCCGCAATCCGGGCGGCGGCACATCAAGGCGGCCGCACTAGTGATAGATCT
UL 117	TTGACGTACCCTTCGATCGCGGCGCCCCACTTGCCGCTGTACAACGAATTCACCAGCTTTCGCCTGCCCACCTCACGGATGGACTACAAGGACGACGATGACAAGATG	CCACCGCTTCACACAGTAAGCAAAAGCACAGGGCCAGGAAAAGCAACCAGCCCCGCCATCGCCGCCGCCGCTTCAAGGCGGCCGCACTAGTGATAGATCT
UL 119	GCACGGACAGACTGCAAGGGTCCCTTCACTCAGGTGGGGTATCTTAGCGCTTTTGCGCCCGGTGATAAAGGTAAGCGGATGGACTACAAGGACGACGATGACAAGATG	GATAAAATTCATATGCTGTTTATTCTGTTAGCAGGTTGGTGGGGGAGGAAGGGGAATAGAACAGAGGCGGTATTAAGGCGGCCGCACTAGTGATAGATCT
UL 121	AGCATCGTCGCGGCGGCCCTGTGGAAGGTAGATTACGACAGATCCGTTGCGGTTGTGTCCAAAAGCTATAAATCACGGATGGACTACAAGGACGACGATGACAAGATG	GCGTCACCCCCGCACGGTACATAGTTACCCTCTCGACGTCGCCGGCTGTCAATGACGTGCCTGCGTCAGTGGCTAAGGCGGCCGCACTAGTGATAGATCT
UL 122	ACCCCTCCCGAAGACCTGGACACCCTGAGCCTGGCCATCGAGGCAGCCATCCAGGACCTGAGGAACAAGTCTCAGCGGATGGACTACAAGGACGACGATGACAAGATG	TCACCACACGTTAATACTGTCACGGGGAATCACTATGTACAAGAGTCCATGTCTCTCTTTCCAGTTTTTCACTTAAGGCGGCCGCACTAGTGATAGATCT
UL 123	GATGGTGCTGAGGAACCCACCGCCTCTGGAGGCAAGAGCACCCACCCTATGGTGACTAGAAGCAAGGCTGACCAGCGGATGGACTACAAGGACGACGATGACAAGATG	ATAGAGTATAATATAGAGTATACAATAGTGACGTGGGATCCATAACAGTAACTGATATATATATACAATAGTTTAAGGCGGCCGCACTAGTGATAGATCT
UL 124	GCTCCGCCGCTACCGGAGAAGAGCCCATGTCCGGAACACGTACCCGAGATTCCGCGTGGAGATCCCACGCTATGTTCGGATGGACTACAAGGACGACGATGACAAGATG	AGGATGGGGTCCCATTTATTATTTACAAATTCACATATACAACAACGCCGTCCCCCGTGCCCGCAGTTTTTATTAAGGCGGCCGCACTAGTGATAGATCT
UL 127	CGGCGACAGAAATCTCAAAACGCGTATTTCGGACAAACACACATTTTATTATTCACTGCAGCATATAGCCCATTTCGGATGGACTACAAGGACGACGATGACAAGATG	TAACCCGGATCGTGGGCCTGGATCAGTACCTGGAGAGCGTTAAGAAACACAAACGGCTGGATGTGTGCCGCGCTAAGGCGGCCGCACTAGTGATAGATCT
UL 129	CTGTGGTGCGTCTGGATCTGTCTCTCGACGTTTCTGATAGCCATGTTCCATCGACGATCCTCGGGAATGCCAGAGCGGATGGACTACAAGGACGACGATGACAAGATG	ATCTCAGCCGTTTTCTCGGGACTGTAGCAGACTTCGCCGTCCGGACACCGCAGCCTGTGGATTCATGAAAATCTAAGGCGGCCGCACTAGTGATAGATCT
UL 130	ACACCTTCTGCACCCATCCCAATCTCATCGTTTGAGCCCGTCGCGCGCGCAGGGAATTTTGAAAACCGCGCGTCACGGATGGACTACAAGGACGACGATGACAAGATG	GCACCCGCGGCACGCGGCTGTGACCCAATAGCAGCCACAACGCCGTCAAGAACGGCGTCAGGTTTTTGGGACTCAAGGCGGCCGCACTAGTGATAGATCT
UL 132	AGCACCAGCTGGCAGATTCCCAAACTAATGAAAGTTGCCATGCAACCCGTCTCGCTCAGAGATCCCGAGTACGACCGGATGGACTACAAGGACGACGATGACAAGATG	ACATCATACACATAGTATAGGCGAGGTGATGGGGCGGGGAAAGAGTTGGAACCGAAAGACAAAAAAAAAAGCCTAAGGCGGCCGCACTAGTGATAGATCT
US 13	AACCTCTGGCGCCTGCCCTGGACGACCGTCTTTGCCGCCTTCAGATCCTCGTATTGCGAAGGTGGCGGTGGCTCGCGGATGGACTACAAGGACGACGATGACAAGATG	GCTCGACGAACAGTCGTCGGGGCTTCAGGTACCCGGCAAGTTTTATAGAGAAAGGGGGACGATGGGTGGTGGCTAAGGCGGCCGCACTAGTGATAGATCT
US 14	CATTACCTGATTGACAACGGCACGCTCAGCGTCATCCTCAACACCACCACCGCGACGTTCCAGAGCAAGGCTGCCCGGATGGACTACAAGGACGACGATGACAAGATG	CCATGAGGCGGGTGATGCGCCCGAGTGAACGGGTGAGCGTCTCGGTGGAGTCTTCTTATAAACCAGCGGAGCTCAAGGCGGCCGCACTAGTGATAGATCT
US 15	TTCACTCTGCTGATGGTGCTGAGAATCATGACCCTGCGCACCTTTTTGCAAACCTACTTTTCCTCTGACAAGCTGCGGATGGACTACAAGGACGACGATGACAAGATG	TTTCCCGCTGCGTGGAAACTGTCTCCATGTCGGGACCGCAGCGCCCGGCGGCGTATCCGCAAGGTCTCGAAGCTAAGGCGGCCGCACTAGTGATAGATCT
US 16	TCCGACGCGATGACGTTAACGCCGAACGACCTGTCCCATGCTTCGTCGTCGACGCCGTTGTCCACCCTCTCGCCCCGGATGGACTACAAGGACGACGATGACAAGATG	TTTCTCTTCTCATGGTGCGCTGCGTTCTCTGGAAACGGCTGCTCTGTCCGAAAACCAGTTCCGAACGAAAATCTAAGGCGGCCGCACTAGTGATAGATCT
US 17	GCCACCAACGCGGCCTCCATTTCGGGCTGTGACCTTTTGCGGGAGGTACAGAGAAACCTCACGCGAACCATGGCGCGGATGGACTACAAGGACGACGATGACAAGATG	CCCGCGGTTCTAACAGGCTTGATTGGTGGAGACGGCCGGCGCGGCGGGTGGGGGAAACGACGAGTTTTTCCGTTAAGGCGGCCGCACTAGTGATAGATCT
US 18	GGCCACGTCTGGGTGCAGCAGTACGCCGAGAAACACGGCGGACGCATCGACGGCGTGAGTCTCCTCAGCTTGTTACGGATGGACTACAAGGACGACGATGACAAGATG	TTCATCGTCTGTCTCTGGAGGGAAGGTGGGGAACGTTTAAGCGAGCAGGAGCGTGTCATCTCCCCCATCTTTTTAAGGCGGCCGCACTAGTGATAGATCT
US 19	CTACGCCTCTATGTCGAAAATGTGGCTTTATTCATCGGCATGTACCATCTTCTGAGGCTCTGGTTGTGGAGCCCACGGATGGACTACAAGGACGACGATGACAAGATG	CCGGTTACCGCGGATTTGATTGACGTCACGAGTGTGGTCAAACCGTGGCGGCACCCTGTATCCGACCGGTCGTCAAGGCGGCCGCACTAGTGATAGATCT
US 20	CGTTGGATTAGTCTTTCGGACGGCGTGCCTTTGGACAACGGGACTTTGACAGCCGCCAGTACGACGGGGAAGTCCCGGATGGACTACAAGGACGACGATGACAAGATG	CATTCTAGCGGGACGACATGAAGCATGGCGACAAGCGCGGCTGCTGTGAAAACGGGCGCGGTTTTATAGGCATTAAGGCGGCCGCACTAGTGATAGATCT
US 21	AGCGCCCACACTGCTCAGACGACGGTCGCTACGACGGTCGCTGCCACAGCAGCGGCGTCGCCCCAGTTCGTCTCCCGGATGGACTACAAGGACGACGATGACAAGATG	GCACCTACCCTTCTCTTATACACAAGCGAGCGAGTGGGGCACGGTGACGTGGTCACGCCGCGGACACGTCGATTAAGGCGGCCGCACTAGTGATAGATCT

Appendix Table 2.1 **Tag Cassette Primer** (*Continued*)

ORF	2xFLAG-2PA_Tag Forward Primer	2xFLAG-2PA_Tag Reverse Primer
US 22	CGCGGACACACGCTGAAACAGGTCCCCATCCCCGAACCCGACCAGATGGGAGACGACGGACCAGACCCGGGTCCCCGGATGGACTACAAGGACGACGATGACAAGATG	ATTTTATTAAAAATTTATCTGTGCAGAATCATAAGTTTATGATGAATAAAAACGGGGAAAGGGAATCTGCTTTTAAGGCGGCCGCACTAGTGATAGATCT
US 23	GCGCGACGCATCAATCTGAATACGGATTTGTCTCCGGAGTGGGTCAAGAGTTTCGATTTTCGGGAGCACTTTGTGCGGATGGACTACAAGGACGACGATGACAAGATG	TGCGCGAAAGCTTAATATAGGAGCAGCGTCAGACGAATCGCGGCTGGTGGCCCTGGGGGTGGGACGCGCCGCCTAAGGCGGCCGCACTAGTGATAGATCT
US 24	CAGTCGCGGTTAGTGACGTACTCGGACCCTTTTCCCTTAAAGAGCCTGACGGGTGTGCGCGAGTACATCCAGATTCGGATGGACTACAAGGACGACGATGACAAGATG	CACCCCAGAAAAGAAGCAGAGAAAGAGACTTCTTTGTGCGACACGTTTTATTCCGCGTTCTCCGCTCGACGTTCAAGGCGGCCGCACTAGTGATAGATCT
US 25	CAGTCCCGACGTCCGTCGCCATCTGGCTCCCTGCTGTTGGCTACCTCTCGGATTTCAAAAAAGAGCACGTGCCGACGGATGGACTACAAGGACGACGATGACAAGATG	ACGAGTACAAAACAGGAAAGGAATACAAATAAAAAAAGACGCTGTGACACTTTGGCTCTTTCCTGTGCACCGTCAAGGCGGCCGCACTAGTGATAGATCT
US 26	GCCGCCGTGCTGGAGGGCCGCCCCACGCCCAAACCTACCATGCCTCATCCTGTTTCTTACTTGCCCTTTTGGATGCGGATGGACTACAAGGACGACGATGACAAGATG	CAAAGTGTCACAGCGTCTTTTTTTATTTGTATTCCTTTCCTGTTTTGTACTCGTAAACTGTTGACGTTGTTTTTAAGGCGGCCGCACTAGTGATAGATCT
US 27	GCCAAACACACCAAACGTACCCATTATGACAGAAAAAATGCACCTATGGAGTCCGGGGAGGAGGAATTTCTGTTGCGGATGGACTACAAGGACGACGATGACAAGATG	GCCGACGTGACCACGAAGAAACGTGCAATTAGCAAAAATAGATGTGCGGCGGACGCGTGAGAGAGGATCGAATTAAGGCGGCCGCACTAGTGATAGATCT
US 28	AGCTCGCCGAGTCGAAGAGAGACATCTTCCGACACGCTGTCCGACGAGGTGTGTCGCGTCTCACAAATTATACCGCGGATGGACTACAAGGACGACGATGACAAGATG	TGTATATCGGGCACGGGGAAAAGAGGGGCGGACACGGGGTTTGTATGAAAAGGCCGAGGTAGCGCTTTTTTATTAAGGCGGCCGCACTAGTGATAGATCT
US 29	GGAGATACGGTTCACCATAGCCTGGATGTGGTTTCCCAGTGTGCTGCTGATACTGGGGTTGTTGACGCCTCCGAGCGGATGGACTACAAGGACGACGATGACAAGATG	AACGTCGCACCGTCTCGAAGACGGCACTGTTCTCGAATGGACATGTTTCGTCCGACATCGACAGTGCAGCCGTTAAGGCGGCCGCACTAGTGATAGATCT
US 30	CCCCCTCCTGTAGCGTCAGCTCCTCCTGCAGTTCCCTCGCAGCCCCCCGAGTATTCGTCTGTGTTCCGCCTGTACGGATGGACTACAAGGACGACGATGACAAGATG	ACACCGACCGCACTCGAGAGTAAAGACAAATGAGACGCTGAAGGCCGCGATCAGCCTCCCGTCTCTTTATTTTTAAGGCGGCCGCACTAGTGATAGATCT
US 31	AGCGGCCAGAGTAGCGGTGACGAGAGTAACTGCAGTCTACGCACCCACGGCGTGTACACACGGGGTGAACAACACCGGATGGACTACAAGGACGACGATGACAAGATG	CAACGGGTCGTCTTTTTAAATCCCCGCAGATATCCGGTTGATGTAGCCAGTCGCCTACACGCGACTTATCGATTAAGGCGGCCGCACTAGTGATAGATCT
US 32	ACGCGGGCTGAGCCGGACCGCGTGGTGTTCCAGTTAGGGGGAGTGCCGCCTCGCCGTCACCGGGAAACTTACGTGCGGATGGACTACAAGGACGACGATGACAAGATG	ATGTTTCCTGAACGTCAATCACACACATCACACGGGGATTTACGCTATGTTGTTTATTGTCATGCCGTGTTCTTAAGGCGGCCGCACTAGTGATAGATCT
US 33	CTTAAGGCTCATGGTGCCTCTGTGCCAGAAACTCCTGTGGGAACCCGCTCACCACCCCGTTCTTATGCGATTCTCCGGATGGACTACAAGGACGACGATGACAAGATG	AGGAAACATGTCCCCATCATCAGCGTCACAACTGACGTGGGTTGGGTACTGACGTGCAGGATATTACGCGAGTCAAGGCGGCCGCACTAGTGATAGATCT
US 34	CGCTTCAATGTCGGTGATTTTCACGGAGCGTCATGGGAAAACGGTACCGCTCCCGATGGAGAGCCCGGGGTATGCCGGATGGACTACAAGGACGACGATGACAAGATG	CGTAGACGATGAACCGTACGAATCGCGGCACAACGACTGGACGACGTCGTTTACGTAATTTTAAGAAGAATTTCAAGGCGGCCGCACTAGTGATAGATCT
RL 1	GGGCACGACGCGGTGAACGTGGCGCGTGGATGCCGGCCGAAACATTTACATGTCCCAAGGATAAACGTCCCCGGCGGATGGACTACAAGGACGACGATGACAAGATG	TTATTGGATATCGGTGAAGCAGCGTGGCGGCGAAATACGCGATCCCTGGGCTGGTAGATCCCCCTACCCCGTCTAAGGCGGCCGCACTAGTGATAGATCT
RL 2	CAAAAAAAAAGGAACAAAATAATAGATTAGCAGAAGGAATAATCCGTGCGACCGAGCTTGTGCTTCTTTTGTTACGGATGGACTACAAGGACGACGATGACAAGATG	ATTTTATCTCAGCTTTTCTCCCAAACCTCGGTTTCTTCCTATTCTTAAGTTTTCCCTAGTATATTTGCCTCCTTAAGGCGGCCGCACTAGTGATAGATCT
RL 4	GTCAATCCGCAAGGGTGTTCCTGCCCCCTCGACGCGCCTGTCGCGATCCTCATGGCGAGGACCGCGATCTCCGTACGGATGGACTACAAGGACGACGATGACAAGATG	CTGCTGACCGACGCGTCAGGATGTGCAGGGCATGCGGGGGAATCAGGACCGGACACGGGATAATTTCATCTACCTAAGGCGGCCGCACTAGTGATAGATCT
RL 10	TCCAGACAGAAAGATAACTCCCCGTTGATCGAACCGACGGGCACAGACGACGAAGAGGACGAGGACGACGTCCGGATGGACTACAAGGACGACGATGACAAGATG	GACGTGACTATGACAAGCGTGAGGGGGGTGCTGTAGGTCTGCATGGTGCAAAACACGTTCTCGCCTTCCTTGTCAAGGCGGCCGCACTAGTGATAGATCT
RL 11	GAACAGGACGCTGAACGGTTGCGGACGGACCTAGATACGGAACCTCTGTTGTTGACGGTGGACGGGGATTTGGAGCGGATGGACTACAAGGACGACGATGACAAGATG	AAAAATAATAAACATGCATATTATTTGCGTATATGATGACTTGTTCCACCGTCGATGTTGTGTGCGCATCTTTTAAGGCGGCCGCACTAGTGATAGATCT

Appendix Table 2.1 **Tag Cassette Primer** (*Continued*)

ORF	Up2 Forward Primer	Dn2 Reverse Primer
UL 2	GATTCCAAACCGGATACGCTACATACCTGCCACAGTGGGCAGCTTTTACC	CTATGTTACCACAGGAGATCACGGAACATAAATGTTTTCTGCGTATGTTT
UL 3	TTTACCGTACCCAGACAACGGTGCTTTATAGACTCATCACTTAAGGCGGG	ACTGAGCGTGCGAAGCGGGTAGGGTGCCGAACGACGGGTATGCGTCGTCG
UL 4	CGTTTATCGTCACCACTAGTGCAATATTTTGATCGTAAGGCTGAAAGAGT	AAACACAATAGCTACAGCTGCGCGGTTCTGTGGAACTTCACGTGCGATCT
UL 5	TATTGTGTTTACGTTGCTTTTGAAATGTTAAGCGTCCCTACGGCGCTAAC	CAAATATGCAAAAGCAAAACACAACAAACTATACACAGCTGGCTAACTAG
UL 6	GAGAGTCTGAAACGGGGTGGGAGGGACTTTTGCGGGTAGTGCACGCTAAG	ACCACGGTACGATTTGGAAACAGTCACGCTCAAACAGTACTTTTTATTTC
UL 7	AGTACTGTTTGAGCGTGACTGTTTCCAAATCGTACCGTGGTAAATAAATC	GGGCTAGTCATTGTGGGCACAAAACCTTCTCCCTGATAAAAAGCACATTA
UL 8	CAGAATTATAGTAATGTGCTTTTTATCAGGGAGAAGGTTTTGTGCCCACA	TGTACAAAGAATGATTGTTATCCATCGAAGTAATAACGCGTACCGGAACA
UL 9	AAAGACAGTAAACAGTATGGACAAGTGTTCATGACGGACACAGAACTATG	GAGCTGAAAAATAAACGTACATAGCTTTTAGTTTCCTCGACGGTGATTCT
UL 10	GTAAACATAATGACGTACATATACGTGGTTATACAACAGGTGTTTGTGCT	ACTGTCCCTATATTCAAACAGTGAGTTTGAAACCGGACATATCCGTCCGC
UL 11	GGGACAGTCCCTACGGAACCTGAGAACATGTGGAAATCACCTGTGGTAGA	GAGTTGGCGTTTCACAGTGATTTCATGCAATCATTTCCTACGCGACTTGC
UL 12	GCAGCGGGAGCAGATGATAACGCAAGAAGCGACCGCAGTGGGCCCACAGC	ACCGCAAAAGCTGTGGCTGCTCTGGCAGCATGACAAGCACGGCATCGTGC
UL 13	CCCTGATTCCCTTCATAAAGCTGTTGACCGGCCCTAGAAAGACCAAGAGC	CGCATAAGCGACCGGGGATGGGGGGAAATAAAGGAATGGCTCGGTGTATT
UL 14	GTCAAACACTACCAATAAAGATAATCTGCCAATCGCACCTTATATATAAT	TCCGTTTTCCTGCCGTGACTGCGAATCATCCGCTTCATGGCTCTCCTCGT
UL 15	AAAGAACAAAAAACACCCATCCCAGCGGTACCGTACCTCGGCGACGCTCC	ACGACCTGCGCCACTCGGACCGCTCCTGCGACCTAGCTTTCGGATCTCGC
UL 16	AAGCCTGGGTACTTATGGGGAGCGCGCACAAAGGACCGTCAGGCGCCGGC	GCCTTCTTATAGCAGCGTGAACGTTGCACGTGGCCTTTGCGGTTATCCGT
UL 17	CCCGTGGACGGGTCTCTTTGACACGAGCGCGGCACGCCGTTGCCACGAGC	TTGACCCCTCCTATCTTCTTTGATGATGTATCCTCTTAGCCGTGTGTTGT
UL 18	CTGAAAGTATATAACGCCGATCATGTCCGAGGAACTGTTAATAAAACGCC	GGGGGCACGCGGTAACCGACGTCGAAACAGCTCATACAGGGCGTTGATGG
UL 19	TACCTACGTAACCTGGCCTTTGCGTGGCGCTATCGCAAGGTCCGGTCGTC	CGGACGTAGGTTATTTTGAAAACCTACGTTAATCCTGAACGCGTTTCGTG
UL 20	TGGAACGGTCTTTATATATACAAACGCCGTTATGCTCAGTGTCCGGCAAG	TATGGAAAATATGTAGTCCGTACCGCTTGGGGCTCAGAGTCCAAAGTCCG
UL 20a	TTTTTAGAGCAGAACCTTACAGCTTTTTAATAAAAACAAGATAGTCAAC	CGCAGGTAAACAGGTAAGAAATACAAAAAATAACGTGATTGTGAACGCGG
UL 21	GCGTCTGGCTGTGTGCCGTTAAATACCTTGGGTGACGACATCTCGAGGTC	ATGTAAATAAAATGCTTTTATTTAAAACTGGTCCCAATGTTCTTCGGGAA
UL 23	CACCGTTGACTGTGACGCCACGAGTGACGACGACGCTGCTTGGGGCATAA	TTTACAGGAACGGGGAAAAAAAGGCACACGGTCCGTGGGAGACGCGGGT
UL 24	GCGTGGTGTGCGACGGGCCCCGGGCTCGCCCACTGACTCGGCGCGTCAC	GTCGGTGTTTTATGCCCCAAGCAGCGTCGTCGTCACTCGTGGCGTCACAG
UL 25	CATGTATGCAGGTAAGCAACTGAGCCGAACGCACCTCAGCAGACGAGAGG	CCTGTGACTTTTTATCATAAACCGTTCCGCCCTGCTGCTTCGTTCCACCA
UL 26	TGGCGGCGCGTTATAAGCACCGTGGGGTCATCGACCGACGAGGCGCGGCG	TCTGTCTTCTCCGGGTTTTTTTTTTCATGTTTTTTTTCTTCCTATTTTG
UL 28	CCACGGCGGGTCGTTGGCTCCCGCTGTGCTGGCCGCCGCTGCACGGCATC	GCGGTAAAGCCAAACACCGGCTATATAGCTAGTCATCACAGTCTCCTCCT
UL 29	ACTGCTGCTTCTGCTTTTTTGTCTCCTGTGGATCGTCGCGGACTGCCGGC	GGTTATAAAAACACCGTCGCCCTATTTCTGGGCGTGTGTACACTGATGAC
UL 30	GTTGAAAACGCGCATGATCTCGCGGAGCCATCTACGCGCCTGTCAGGGAG	CCACACGCTCAGCCGCGACTGAGCGCCGGGGCGCCGCTACTTGGGTTT
UL 31	TCCATGGCGGTTTACGACGAGGAAACCATGCGGCAGTCGCAGGCGGTGCA	AATGAAACCATCGGATAGTGACGTGTCGGGAAAGGAGGACGGACGGAGGG
UL 32	CGGGCGCAGTCCGGGGCGACGACGCTTCCGGGTTCTGGAGAAAAGCCAGC	CACTATCCGATGGTTTCATTAAAAAGTACGTCTGCGTGTGTGTTTATTAA

ORF	Up2 Forward Primer	Dn2 Reverse Primer
UL 33	CCGCCTAGAACCGCAGTACCAGTACTCCGCATGTCAACAGTACCTGTAAC	GGGAAATGGCGACGGGTTCTGGTGCTTTCTGAATAAAGTAACAGGAAAGC
UL 34	GGCCCGCTCGCACGGACCTATACTATTACCGCCCCACCGCCGTCGTCGTC	AAACCAGAGCGGAACTTGAGAAATCAACGCTTTATTGTTCTCCAGTGACG
UL 35	TTCTCAAGTTCCGCTCTGGTTTTGGTTTCGTTTTCAAAGGGAGCCCCATC	ATCAACGTCTCGTCCTGAGACAGACACGTATAAAAAGAGGAAAACCGCGT
UL 36	GCTGTCAGGAATACCTGCACCCCTTTGGCTTCGTCGAGGGTCCGGGCTTT	ATGCACATAAACAAACGGGACCACCGTGCTCGTCATCCTCTCCTCAATCA
UL 37.1	TTTCAAGACGACGTGAGACCCACACGCGGGTTTCACTTCTTTCTTTAATT	TCTTATACTATCCCGGAGTCTGTGGTTTTTTGTTTACCCCTGCTTACTG
UL 37.3	GATGTACGTACCACGGTACGGACATTAACGTCACTTCCAACGCCACGAGT	AGAACTACGGCGCGGCGGCACGGCCTTTATAGACACTATCAGCGTTGACG
UL 38	AGAAGGGACTTTACCGCTATTGCTGCTATTCATAGAGAAGGATAGAAAGG	CTACAAAAAAAAAAAGCTGAACATGGTCATCTAGCAGCAAAGTTCTCCTT
UL 39	ACACACACCACACGTCACGACACCGATCGATTTTCTTTATTCTTAGTGTG	AAAGCGCTTTTGGGCTCACCCATCTGCAGTCCTGTTGCCTGAACGAGCAT
UL 42	TCGGTACCGTCTCCACCACAGTTACCACCGTCGCCGTCACTGCCACCGAC	TATTCCGTAGCAGCAATGATGGTACAGTCAAGCACATGATCTATTTCCCG
UL 43	CTAGAGCGCGTGCCCGGGCACGCGGCCTGCGCGCACGGCGCGGTCCGCGG	ACGGCGACGGTGGTAACTGTGGTGGAGACGGTACCGACGGCGTCCGCGGC
UL 44	TTTCGCTCGCTCGCGCCCGCTCCTTAGTCGAGACTTGCACGCTGTCCGGG	CCATCGCGGGACCGCGCCGTGCGCGCAGGCCGCGTGCCCGGGCACGCGCT
UL 45	GTCTTTGTCCATCACGTAGTCAAGTACTCGATTATGGCCGACAGCGTGTG	AGACGCCCGGCGTCTAATAATACAGCCGCGCCGAGCCAGCGGGCCCCCGA
UL 46	CCCGCAGCTGCTCTATCAACTTTTTGAAATCTACCGTGCGCCTCGCCATC	GTGTTTATTTTTTTCTTCTGTGTCTCCTCCCCGTATGCTGTCAGCGCCGC
UL 47	TGTCCTCGTCGGCCGGGTCGCGCGGCCGTTTGGCCACCGCGCGCGCGTCC	CGCTCCAAAGCGAGCGATGTCGCCCTGGTGGCAGCTGGCCTGCGTGACTT
UL 48	GAAATGGTGACCGAGGGGTACACGCAACCGCAGCTCCCCACCGGAGAGCC	AAAACGATAAAAAGCCTATTGTTTTTATTACCCGCTACTGTCAGTGTCGG
UL 48A	CGGCGGTTTAGCAGCGGCTTCCTCTTCGTCCTCCCCCCACGGCCTGCCCC	TATCACTTTTACGGTGATTTATTCTTGCTATTCTTTTTCCCCTTGGGCTG
UL 49	CGGCGGTTTAGCAGCGGCTTCCTCTTCGTCCTCCCCCCACGGCCTGCCCC	TATCACTTTTACGGTGATTTATTCTTGCTATTCTTTTTCCCCTTGGGCTG
UL 50	CCCGCTGGTGCTGGCTCTCCTGCTGGTGCTGGCTCTGCTGTGGCGCGGTC	CGTGAACGCAGAGGGGCGCGAAGAGACGGCGCAGAGCCTGGTGGAACTCC
UL 51	TACTCGGATTCATGGCGATCGGCGCCGCTGATTGAGGACGCGGAAAAAGA	TCCTGATGGAGAACCTTGTTCATCTCCATCGCACCGACGCCACCGCCGAT
UL 53	GGAGAGTTGCGACATCAAGCTGGTGGACCCCACGTACGTGATAGACAAGT	GGTGCTGCCGCGGCGCTTGCACTTGGAGCCGGCTTTTCTGCCGTACAGTG
UL 54	ATTTGCTTTGTGATTTTGCTTCGTAAGCTGTCAGCCTCTCACGGTCCGCT	GTCTCAGCAGCATTATCACCGTCCCCAGTCACCACCGCCGCCGCTGTTTC
UL 55	CAAGAGACCACGACGCGCCTCATCGCTGCTGGATTTGGCCCGCGACGAAC	TCACAAGTCTCTGTCACTTTTTTTGTCTAGTTTTTTTTCTCCTCTTGGT
UL 56	GGCCGGCGCCAGACCGGACGACAGCGTCTCGTACGTGAGCGAGTCGAGTC	CACGAGTAGAAGATGAGGAAACCGCAGCACCCAGACAGACGATACACAAG
UL 57	CTGGATGTGGCGCCAGACTGTGTGCTGAGTTACGTGGAAAGCCGTTTCCA	GTTGCGATGCTCGTGGCTGCGGCGGCCGTTGTCGCGGCGTCTGCTGGCGG
UL 59	ACCGCCTCGCCGGCCACGGGGTTGATTCCTGTTCTTATGCCGACACCAG	AAGATCCGAACTTTAAAATTGTGTATTTTTATTTTCCCATCCCCCTCTTT
UL 60	AGAGGCCCCGCCTAGGTGGGCGGAGCGGTAATTTTCCACCGCCGCGGCCC	ATCATCTCTGATGACGTAGCGAGCGAAGCGAGCTACGTCATCAGTCCGTT
UL 62	CTCGGCCAGGGGGTACCGAGGCGGTGCCCGCGACTCGCCCCTCCTCCAAG	TTGGGTGTGGCCGGAAGCGCTCGGGGTCGACGGTGGGCCGCCATGACACC
UL 64	GAGCCTGAGATGATGATGATGGCTACGAAGGACGGGCGGACGGGCAAACG	AATGACAGAATGAACTCCATGTTATACGCTCTTTATATAGTTTCTCTGCA
UL 67	GCCGCCGCCGCGGTTGCTACTACTTTCTTAAGTGATGCGAATTGGTGGCT	TAAACGTTCTCAACAGGTATGAAATGAACAAACTAGATGATGCTATAACC
UL 69	CTATATATACATCAGCGTGCCCGAACGTGACCTTCCTAGCGACGACGGCC	AACGGGATAAGGGACAGCAATCATCACGCACAACACCCTTCACTCTCTTT

Appendix Table 2.2. **ORF Flanking Primers**

ORF	Up2 Forward Primer	Dn2 Reverse Primer
UL 70	ACAGAGGAAGACGAAGACGGGAAAGAGAATAAAAATAACGTCCCGGATAA	GATGAGACCGCTGCCGGGGGGCGGGTCACCGGCGCCGTGGAAAGTGAGGC
UL 71	GGGAAGACGCAGTGATCCGTCGGTGTCTGCGAGAGTACGTTGGCGACTAT	TACTCGTCGTGTCCGTGATCACGTACGTTTTCCAAAACGTGCCAGGCTGT
UL 72	ACCAAAAGACCTAGTACTAGTGTGTTCCACTCCATCACACCACCGCCACG	GATCACGGACACGACGAGTACGGGGTTTCTCATAGACGTACTTTATTAGG
UL 73	TGAGAGTCGATTCGATCGGTAAACATCGTAAGCATCGTGGCGGTGGTGTG	TGGAAACCTTACCCCGCCGGAACACCGCCGGGCTGTGAACCTGTCCACCC
UL 74	GCAAGCCAAACCACAAGGCAGACGGACGGTGCGGGGTCTCCTCCTCTGTC	TCTCATGGGAGTTTTTTGTATCGTACTACGACATTGCTGTTTCCAGAACT
UL 75	CTGCTGTATCCCACGGCCGTGGACCTGGCCAAACGAGCCCTTTGGACACC	TGTCTTTAACTTTCTCTGTCCCTTTTCTCATAAACTGTCAGGTTCTACAG
UL 76	CGATGTCATTGGCCGCTGCGAAGGGAGAAGAGGGGACACGCGAGTAAGTC	CGGTCGCCGCGTCAGACGGGGTGGCGGGTCCCGTGATGGCATCGTGCCGA
UL 77	TTTCGACCTGTGTACCGATTCTGTTCTGGACTATCTGGGACGGCGTCAGG	CCTCTCCGGGGACGCTCGCCCTTTATGCAGCAAGCGACACGTGGTGGAAA
UL 78	TCCCCGGAGAGGGTATATTCGTTCGGCGAGAGCGGGCGGCGGTGGTGGGT	GACGTAATTTATCTGCCACTTTTCTCCCCGCTGCCGTACAACGCCGCCGC
UL 79	TCGTCCATCGTCATTGTCGTCACCGTCGCTACCCGCTCACCGAGCGAACG	CGGCGTTGTACGGCAGCGGGGAGAAAAGTGGCAGATAAATTACGTCAGGT
UL 80	TCGCGGGCCATCATCGTTCGCTCGGTGAGCGGGTAGCGACGGTGACGACA	ATCACGGTGTAGAAAAAAAGAGAGGGAAGCCCTAAATATAGCGTCTCTC
UL 82	CCGTTCACCTTTGCGCATCCCCTGACCCCCCCCTCATCCCGCCTTCGCG	AATACAGGGAATGGGAAAAACACGCGGGGGGAAAACAAAGAAGTCTCTCT
UL 83	GCCGCAGAGGGCGCGCCGCTCAGTCGCCTACACCCGTACGCGCAGGCAGC	TGGACGTGGGTTTTTATAGAGTCGTCCTAAGCGCGTGCGCGGCGGGTGGC
UL 84	CTGCATATTTAAAGGCTGAGCCGGCCCTCTCGCGCCCGCAGACACCAAGC	CGTATCGCGCGGACGCCTAGTGTCCGTTTCCCATCACCAGGGTTCTCTGT
UL 86	CTGCAGACCAAGGCGCACATCCACCCGGGATTCGCACTCACGGCCGTACG	GTGACTCGCCGCTAACCTGCGGTCGTCGCCGTCCTCCTCACCGGACGGCC
UL 87	ACCCACAGTTTTGTGATGTACAAGCACACCTTCAAGGAGCCCGCTTGCAC	CGGACGCTCCTCCGGACGAAACGCCGCGGCGGCAGCGGCCGCGGCTTCCA
UL 88	GTTGACGGCAGTTCTGAACCCACGTCGCCGCGAGCGCGGTTTGCATCACG	ATGGCCACCTACCTGTGTGACGAGATACACGCCATCCGTTTCAGGGTCAG
UL 89.1	CCTCTCGCCGCTGCCGCCTAACCTCCGCTCGCACCACCGCCGCCGCCATC	CTCGCCACGGTTTGTGGCAAGGCGTATGTTATCATCATGGAGCTACTCAC
UL 89.2	CCGAGCGTGCGTAATCATAAGAGTCCCTTTTCTGTCTCCTCTCTACACAG	TCTGCCGATGCCGGCGTCAGTCGCCGGCACCTGGTGGCTCTGCTGCGTGC
UL 90	TAGTGGCGTGCGCGACCCCCAGTCGGTTGAGTTCCGCCAGCAACGAGTTC	TGTCCTCGGATGCTCTGTGTAGAGAGGAGACAGAAAAGGGACTCTTATGA
UL 91	GCGCGCCCATAAAAACGAAAGTGTCGTCGTCGCGACCCGCCACAGCCGCC	GTAGAGCGAGTGTAACTGGATCTCCTCGGTAAACGCGTTCTGGACGTGCC
UL 92	CCTTCTCCTGTTCCCTCCGCCCCCAAAACTGTCAGCGACGCTCAGACGTC	GGTCGAGCACCAGATGTAGAGGCAATTGCTCATCGTCAGCGAACCGCGCC
UL 93	TTACTGGGTGCTGCCGGGCGGCTTTGCTGTGTTCTCGCGCGTCACTCTTC	GGTAGAGCAACAGCAACTGGCATAAGATACACGAGCTGTCGTCCTCCGGA
UL 94	TTACTGGGTGCTGCCGGGCGGCTTTGCTGTGTTCTCGCGCGTCACTCTTC	CCTTTTTTTGTTGTTTCTTGTTTCTTCTCCCCGTGAACTGTCAGACCCCG
UL 95	GCAGCTCCGCGTAGCGCTCCTGGATCTTGGCGGCCGAGTCTCCGCGCAAC	CTCAAGGTCGACGCGCATCACGTCCTTTAAGAGCTGTTTGTTGACCGACG
UL 96	TTGCTGGACGCCCTCTCGCTGAACGACGCGGGTCTCATCACGTTGAATCT	TTTTTTTTAATAAAATCTGAACAGAGGCGTGACGGGGATTGCTATACCTG
UL 97	TCGGTGTGGTAGCTAGTGCAGCTCTAGGAACAGGGAAGACTGTCGCCACT	ACCTTCTCTGTCGCCTTTCCCCTCAGCAACCGTCACGTTCCGCGTCCCGG
UL 99	GTGCTAGACCGTTGGAGTCGCGACCTGTCCCGCAAGACGAACCTACCGAT	TGTCCCATTCCCGACTCGCGAATCGTACGCGAGACCTGAAAGTTTATGGG
UL 100	CATCGGCGCGCCCCCATCGCCTCCCGAGCGAGCGGGCCGCCGCTATCGCC	GTCTCTTTTTTATGTCCATGTCTCCAAGTCTGGTGCGGGTGGCGGCGGGG
UL 102	GTCGGCGAAAAAAGACCCCGCGGGCCTTCGCGACTCTCTTCTGTCCGAGG	TTTTACTAGTATCCACGTCACTTACCCACGTAGTTCCCCTACGTGACTCG

ORF	Up2 Forward Primer	Dn2 Reverse Primer
UL 103	CACGGCAGACGAGGAGCGGCGCGGCCCAGAGCGTGTCGGCCGATTTCGAA	CTACGTGTTGCGTGTTTTTTTTTCTATGATATGCGTGTCTAGTTCGCTTC
UL 104	CACGGCAGACGAGGAGCGGCGCGGCCCAGAGCGTGTCGGCCGATTTCGAA	CTACGTGTTGCGTGTTTTTTTTTCTATGATATGCGTGTCTAGTTCGCTTC
UL 105	TTCAGAACGAGGTGCTCATCAACTACTGCGACATCGCCGACAACTGGGTC	TGGTTTTTACCCTGCTCAATAAAGTCACGTTTTCCTTACACGGTGTTGTG
UL 108	TCTTTTTTCTTTTTAGTCGATGGAACTTTTCTTCGGTACGGGTTCTTGTT	AGGATCATATATATCTCGTCAGGGAAATACAAGTTAGACCATAATGTTGA
UL 109	CCGCCATGAAGGCAAGAGCAGCAGCAGCAACGACGTCACTACGATGATTG	TGGTCCTATAAGGTGTCTTCTATCACGGTGGCTTGTTTATCGCTTGGCGG
UL 110	CCTTCGTCCCAGACGGACGGCTATCGGTTCGCGCGCTCGTCTTTTCTTCG	CTGCTCTTGCCTTCATGGCGGTATTTCTCTTCCTCCCCCCTAACCCCATA
UL 111a	CTACCTGGGACGCGCAGTTGGGCGGCGGACTGGGGCGGCATGCTGCGGTG	CGAGCCACACGGAGTAGTCGTCCTCACGTTGCTACAAGAGGAAAACTACG
UL 112	TGGATGATAACCATGTGCTTTTCGTGCGCTATGTGTCGTCCCGTCTGTA	ATGTCCAAGCGCGTCCTGTTTCATAATTTTTCCGGTCTCGGCTCGGTTTT
UL 114	GCCGCCCGAGCTGAAGCAGACGCGCGTCAACCTGCCGGCTCACTCGCGCT	AGACATCACAGTTCACCACCTTGTCTCCCCGGTGTGTCTATTATCATCAA
UL 115	CGGCGCCAACTGGCTCCTTACCGTCACACTCTCATCGTGCCGCAGACTTG	CGCGAGCCTGCTCGTCGGGGGTTAACAGAGAGCCTTTATTATCAGCAATG
UL 116	CCCCACTTGCCGCTGTACAACGAATTCACCAGCTTTCGCCTGCCCACCTC	TGCCACCGGTCCAGGTGAGAAAGAGAAGCCGCAATCCGGGCGGCGGCACA
UL 117	CCCCACTTGCCGCTGTACAACGAATTCACCAGCTTTCGCCTGCCCACCTC	TGCCACCGGTCCAGGTGAGAAAGAGAAGCCGCAATCCGGGCGGCGGCACA
UL 119	GCCTCCTGCCACATATAGCTCGTCCACACTCCGTCTCGTCACACGGCAAC	TTCTGTTAGCAGGTTGGTGGGGGAGGAAGGGGAATAGAACAGAGGCGGTA
UL 121	CTGTTTAATAAAAGTAGCTTTTTTTATACATCTCCGTCTCTGGTCTCGTG	AGTTACCCTCTCGACGTCGCCGGCTGTCAATGACGTGCCTGCGTCAGTGG
UL 122	CGCGTCCTTTCAAGGTGATTATTAAACCGCCCGTGCCTCCCGCGCCTATC	CGGGGAATCACTATGTACAAGAGTCCATGTCTCTCTTTCCAGTTTTTCAC
UL 123	AAATGAGTCAGGAGGACGGATACTTATATGTGTTGTTATCCTCCTCTACA	CAATAGTGACGTGGGATCCATAACAGTAACTGATATATATATACAATAGT
UL 124	CTTGGCAGAGGACTCCATCGTGTCAAGGACGGTGACTGCAGAAAAGACCC	TTACAAATTCACATATACAACAACGCCGTCCCCCGTGCCCGCAGTTTTTA
UL 127	ATGGCTAATTGCCAATATTGATTCAATGTATAGATCGATATGCATTGGCC	TCAGTACCTGGAGAGCGTTAAGAAACACAAACGGCTGGATGTGTGCCGCG
UL 129	TTGTGGCTGCTATTGGGTCACAGCCGCGTGCCGCGGGTGCGCGCAGAAGA	TGTAGCAGACTTCGCCGTCCGGACACCGCAGCCTGTGGATTCATGAAAAT
UL 130	GAGTTCAGCGTGCGGCTCTTTGCCAACTAGCCTGCGTCACGGGAAATAAT	ACCCAATAGCAGCCACAACGCCGTCAAGAACGGCGTCAGGTTTTTGGGAC
UL 132	ACACGTCGTTCGCGGACATAACGAGAAATCCACGTCGCCACGTCTCAAGA	GAGGTGATGGGGCGGGGAAAGAGTTGGAACCGAAAGACAAAAAAAAAGC
US 13	AAGACTCCACCGAGACGCTCACCCGTTCACTCGGGCGCATCACCCGCCTC	CTTCAGGTACCCGGCAAGTTTTATAGAGAAAGGGGGACGATGGGTGGTGG
US 14	CTGTAGCTTCGAGACCTTGCGGATACGCCGCCGGGCGCTGCGGTCCCGAC	GAGTGAACGGGTGAGCGTCTCGGTGGAGTCTTCTTATAAACCAGCGGAGC
US 16	CCCCACGGATCTCGCGCCTTAGACGCACGGTCATATAGCCTCCGGCTGTC	CGTTCTCTGGAAACGGCTGCTCTGTCCGAAAACCAGTTCCGAACGAAAAT
US 17	ACACTCTATAAACGGTTTTTCATACGCGCCTTTTGATCGCCACCGCCGTC	TTGGTGGAGACGGCCGGCGCGGCGGGTGGGGGAAACGACGAGTTTTTCCG
US 18	ACTGTTTCATCGACGCCTACCTTAGACCGACAGCGGTCGTAAGCGGCAGC	AAGGTGGGGAACGTTTAAGCGAGCAGGAGCGTGTCATCTCCCCCATCTTT
US 19	CTAATGCCTATAAAACCGCGCCCGTTTTCACAGCAGCCGGCGCTTGTCGCC	ACGTCACGAGTGTGGTCAAACCGTGGCGGCACCCTGTATCCGACCCGTCG
US 20	GGCGGTAAGTGAGATCCTGGTGATGATAGCCGGTTTGGCCGTCATCGTGA	GCATGGCGACAAGCGCGGCTGCTGTGAAAACGGGCGCGGTTTTATAGGCA
US 21	CGCTTTGTGTATTTAGACGAATCTCGGCGATAACCGCCGGCGTTGCCGCC	CAAGCGAGCGAGTGGGGCACGGTGACGTGGTCACGCCGCGGACACGTCGA
US 22	TTTCGCGCAGCGCGCTTTATCCGACTCGCTGTCGAGACGGCTCCGCCGGC	GCAGAATCATAAGTTTATGATGAATAAAACGGGGAAAGGGAATCTGCTT

Appendix Table 2.2. **ORF Flanking Primers** (*Continued*)

ORF	2xFLAG-2PA_Tag Forward Primer	2xFLAG-2PA_Tag Reverse Primer
US 23	CTCTTTCTCTGCTTCTTTTCTGGGGTGTCTAGCTGGCGGCCTCTTTTGAC	GCAGCGTCAGACGAATCGCGGCTGGTGGCCCTGGGGGTGGGACGCGCCGC
US 24	GGCTCGGCTTCTGTCAGTGACGCCGGTTGTCCGGCTGAGAGTGTCGCGCA	AAGAGACTTCTTTGTGCGACACGTTTTATTCCGCGTTCTCCGCTCGACGT
US 25	TCCAACGCGCCTGTGGAGGGCCAATCGGACCGCGGGAGCTCTCCAAGTGG	ATACAAATAAAAAAAGACGCTGTGACACTTTGGCTCTTTCCTGTGCACCG
US 26	AAGAAACTCCATAAAATAGGCTGCCAAGTGCCGCTCCACGCCGCGGCACC	TTATTTGTATTCCTTTCCTGTTTTGTACTCGTAAACTGTTGACGTTGTTT
US 27	TGTAATGCTTTTTACAGGACCGTTCAACAGGTGATACTACCTGCAAGGTA	GTGCAATTAGCAAAAATAGATGTGCGGCGGACGCGTGAGAGAGGATCGAA
US 28	GTGCGTGGACCAGACGGCGTCCATGCACCGAGGGCAGAACTGGTGCTATC	GAGGGGCGGACACGGGGTTTGTATGAAAAGGCCGAGGTAGCGCTTTTTTA
US 29	CGGAAAAGTTTATGGGGAAAAAGACGTAGGAAAGGATCATGTAGAAAAAC	GGCACTGTTCTCGAATGGACATGTTTCGTCCGACATCGACAGTGCAGCCG
US 30	CGACATCGGTGACACAGCTTCAGAAACAACGTGTGTGGCGCACGCTACTT	GGCACTGTTCTCGAATGGACATGTTTCGTCCGACATCGACAGTGCAGCCG
US 31	GTCGGTGTCTCGTCGGTGAGACGAGGCCGCCGCCCGACAAGTTCGATCTC	CCCGCAGATATCCGGTTGATGTAGCCAGTCGCCTACACGCGACTTATCGA
US 32	CGTTGTCATCCGGCTTAGAGCAAACCGTCCTTTTATCATCTTCCGTCGCC	ACACATCACACGGGGATTTACGCTATGTTGTTTATTGTCATGCCGTGTTC
US 33	CGAGCGGCAATCCAGACGAGCAGACCAAGGACGTTGATGAGTTGCTCTAA	GCGTCACAACTGACGTGGGTTGGGTACTGACGTGCAGGATATTACGCGAG
US 34	CGCCGTCGGCACTTGGCTTCAGAGCAGCGCCTCGGGGCGATGCGACGGCG	TCGCGGCACAACGACTGGACGACGTCGTTTACGTAATTTTAAGAAGAATT
RL 10	GTGCGGGGAGGATCGACGTGTGCGGTGCTTGTGGAACACGGTGTTTTAAT	GGGGGGTGCTGTAGGTCTGCATGGTGCAAAACACGTTCTCGCCTTCCTTG

Appendix Table 2.2. **ORF Flanking Primers** (*Continued*)

Chapter 3

Functional Characterization of HCMV Open Reading Frame US20

Abstract

Human Cytomegalovirus (HCMV) is a beta-herpesvirus that has been shown to cause high morbidity and mortality in neonates and immunocompromised individuals, such as organ transplant recipients and AIDS patients. Given the high-protein-coding capacity of HCMV, substantial work still needs to be done in order to understand this complex virus and its associated diseases. In this study, we characterized the US20 open reading frame (ORF), a member of the US12 gene family, and its protein product (pUS20). Using a recombinant HCMV virus expressing US20 protein tagged at the C terminus with a tandem affinity tag, we confirmed the *in silico* predictions that pUS20 is a membrane protein that can form homo-dimers. Furthermore, we determined pUS20 to be expressed during the early phase of replication, with a unique specific cytoplasmic localization that occurred only during the late phase of replication. Co-localization results suggested pUS20 did transiently localize to early endosome, but did not localize to either the endoplasmic reticulum (ER) or the trans-Golgi network (TGN) at 72 hours post infection. This study also found that the US20 ORF itself to be dispensable for viral replication in cultured human foreskin fibroblasts, retinal pigment epithelial cells, human microvascular endothelial cells, and differentiate monocytes. Using recombinant viruses expressing truncated forms of pUS20, we dissected the roles of the transmembrane domains on pUS20's expression levels, subcellular localization, and dimerization. Lastly, using an immunoprecipitation and mass-spectrometry approach with 293T cells transiently expressing pUS20, we were able to identify multiple potential protein binding partners. Currently, we have been able to confirm the association of pUS20 with valosin containing protein (VCP) under the context of infection. VCP is implicated in various multiple cellular processes, including cell cycle regulation, nuclear envelope formation, and the ubiquitin proteasome system. Mutations in VCP have also been shown to be correlated with certain neurodegenerative diseases, which presents the hypothesis that US20 and VCP may be a mechanism that contributes to brain damage caused by HCMV infection in neonates and children.

Introduction

Human Cytomegalovirus (HCMV), a ubiquitous beta-herpesvirus, is the largest virus known to infect human with a genome size of roughly 250 kilobases. While HCMV infection is life-long, a vast majority of healthy children and adults will remain generally asymptomatic. Serious clinical manifestations can occur following congenital infection of children or reactivation of the virus in immunocompromised individuals, such as organ transplant recipients or HIV infected patients [1]. While significant genetic and molecular work has been done to characterize many of HCMV's 160 or more genes, the function of a majority of the genes is still not well understood.

The US12 family is a distinct gene family, containing 10 contiguous genes (US10 through US21), located in the unique short (US) region of the HCMV genome (Figure 3.1) [2]. As of yet, homologs of the US12 family have only been found present in cytomegalovirus (CMV) strains capable of infecting higher primates, including chimpanzee CMV and rhesus CMV [3]. The US12 family is present in both clinical and lab-adapted strains of HCMV, such as Towne and AD169. Furthermore, all members of this family share a common characteristic of having multiple hydrophobic domains, which has led to comparisons of the US12 family with the seven-transmembrane (7TM) super group of proteins [3]. Global gene deletion studies have determined individual US12 family members to be nonessential for viral replication in cultured human fibroblasts [4, 5]. In addition, proteomic studies of HCMV particles using mass spectrometry did not find any US12 family member in the virion [6]. Currently, the function of the gene family and its genes has not been determined, however, based upon sequence motif analysis, roles in apoptosis and cell proliferation have been suggested [3].

In this study we present a characterization of the US20 open reading frame (ORF) and its membrane protein product, pUS20. Using a recombinant HCMV virus expressing a C terminus tagged pUS20, we examined the protein's expression level and subcellular localization during the viral replication cycle. In addition we found, through co-immunoprecipitation and immunoblotting, that pUS20 was able to form homo-dimers during infection. We dissected the seven predicted transmembrane domains to determine their roles in pUS20's protein levels, subcellular localization and dimerization. Using viral growth curves, we found US20 ORF to be dispensable for viral replication in multiple cell types, including fibroblasts, epithelial, endothelial, and monocytes/macrophages. Lastly, through immunoprecipitation and mass-spectrometry and co-immunoprecipitation, we were able to confirm the association of pUS20 and valosin containing protein (VCP) during infection.

Results

Bioinformatic Analysis of US20 peptide

To identify the potential transmembrane domains of US20, we submitted the translated nucleotide sequence of the predicted US20 ORF to several online transmembrane domain prediction websites, including HMMTOP (http://www.enzim.hu/hmmtop/), Sosui (http://bp.nuap.nagoya-u.ac.jp/sosui/), TMHMM (http://www.cbs.dtu.dk/services/TMHMM-2.0/), and TopPred (http://mobyle.pasteur.fr/cgi-bin/portal.py?#forms::toppred). All of the online algorithms predicted pUS20, the protein product of US20 ORF, to have seven transmembrane domains (Figure 3.2A), with each algorithm predicting slightly different lengths for each domain.

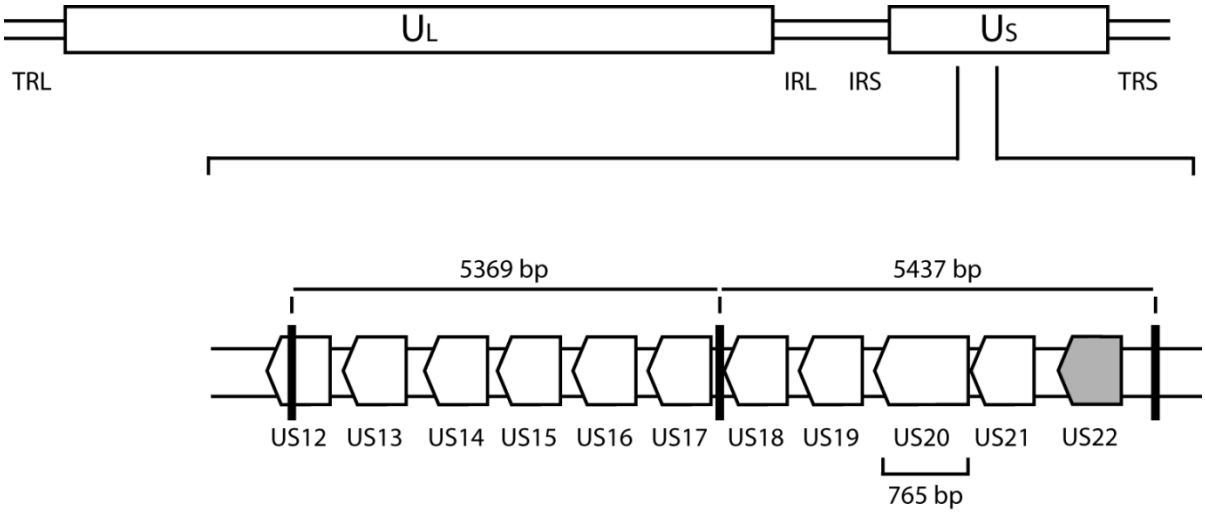
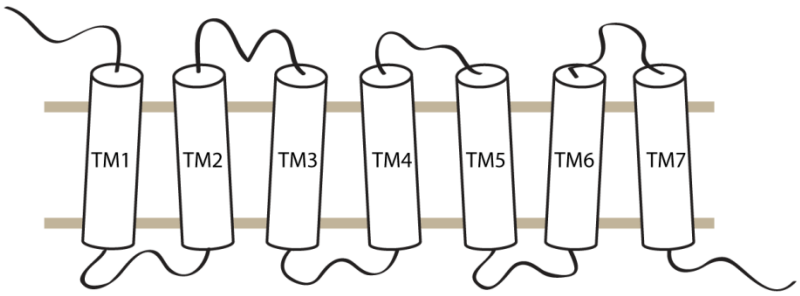


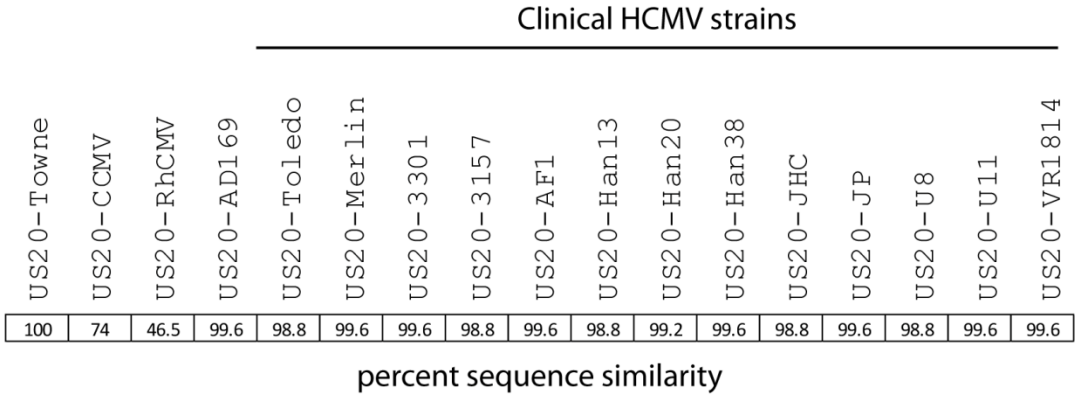
Figure 3.1. **Genomic Map of US12 gene family of the HCMV Towne strain.**
The thick bands denote the Hind III restriction enzyme sites and the predicted lengths between the sites. Genes indicated in white are members of the US12 gene family.

A



	TM1	TM2	TM3	TM4	TM5	TM6	TM7
HMMTOP	31-49	62-80	89-106	119-136	145-163	176-199	208-226
Sosui	28-50	61-83	90-112	114-135	144-166	175-197	207-229
TMHMM	29-51	61-83	90-109	114-136	143-162	177-199	211-233
Mobyle	32-52	62-82	89-109	117-137	144-164	175-191	208-228
Predicted	28-52	61-83	89-112	114-137	143-166	175-199	207-233

B



C

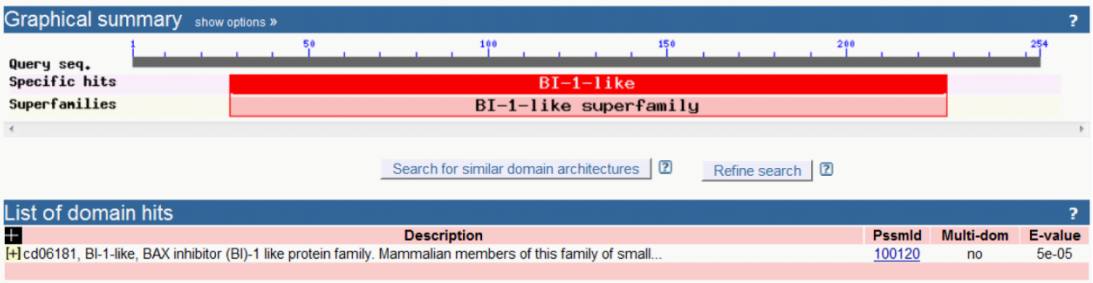


Figure 3.2. **Bioinformatic predictions of HCMV pUS20**.
(A) The predicted nucleotide sequence of US20 was translated and submitted to various transmembrane prediction algorithms. All algorithms predicted pUS20 to have seven transmembrane domains. We consolidated the results and generated our prediction of US20's transmembrane domains in the listed row marked "Prediction". (B) The amino acid sequence of pUS20 among the viral strains AD169, various HCMV clinical strains, chimpanzee CMV and rhesus CMV were compared using the ClustalW method. (C) The US20 translated sequence was submitted to BLAST search and predicted to have a Bax-inhibitor-1 domain.

For our experiments, we consolidated the results by assuming the size of the transmembrane domain to be as large as possible. A transmembrane domain map was then constructed, which we annotated as "Predicted" (Figure 3.2A).

Given that we intended to study US20 using the Towne strain, which has been characterized as a lab-adapted strain [7], we compared the amino acid sequence of pUS20 of the Towne strain with other lab adapted and clinical strains of HCMV to ensure that there was no major deletions, mutations or significant variability in the Towne homolog (Figure 3.2B). We found that the pUS20 homolog in the Towne strain had over 98% sequence similarity to both lab adapted (ie. AD169) and clinical strains (ie. Toledo and Merlin). Thus, we were confident that the US20 gene in our lab-adapted strain would function in a manner similar to its clinical strain counterpart. We also compared the pUS20 sequence from the Towne strain to its homolog in chimpanzee CMV (CCMV) and rhesus CMV (RhCMV) and found a sequence similarity of 74.4% and 41.9%, respectively.

Lastly, we also submitted the translated nucleotide sequence of US20 to the NCBI Basic Local Alignment Search Tool (BLAST) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) and found the gene to have a Bax-inhibitor-1 like domain (Figure 3.2C), which was consistent with previously predictions [3].

US20 ORF encodes a transmembrane peptide

To validate the prediction that pUS20 was a transmembrane protein, we examined the ability of pUS20 to aggregate at high temperatures during sample preparation for western blot assay, a phenomenon referred to as "thermal aggregation" [8]. For our studies, we used a bacterial artificial chromosome (BAC)-infectious clone containing the genome of HCMV Towne, which has been shown to replicate at levels similar to the HCMV Towne strain in human foreskin fibroblasts (HFF) [4]. Applying the mutagenesis strategy previously used by our lab for deletion studies [4], we constructed the recombinant virus, US20-3xF-His, by inserting a tandem affinity tag, consisting of a 3xFLAG tag epitope and 6x Histidine epitope linked with a tobacco etch virus (TEV) cleavage site, at the 3' terminus of the US20 gene (Figure 3.3A). We infected HFFs with US20-3xF-His virus at a high multiplicity of infection (MOI) and harvested the cells at 72 hours post infection. The lysate was treated with reducing agent beta-mercaptoethanol, incubated either at 37°C for 45 minutes or 95°C for 10 minutes, separated by SDS-PAGE, and then immunoblotted with anti-FLAG antibodies. The 95°C incubated sample only had a band greater 170 kilodaltons (kDa) and appeared to not have entered the resolution portion of the SDS-PAGE gel. This result suggested that the sudden denaturation by the high temperature caused pUS20 to aggregate. However, the 37°C incubated sample had a prominent band present at 28kDa, which corresponded with the predicted size of pUS20 (Figure 3.3B). The "thermal aggregation" verifies the presence of hydrophobic domains and supports the prediction that pUS20 is a transmembrane protein.

HCMV pUS20 localization is conserved in various human cell types

The subcellular localization of pUS20 was examined by immunofluorescence assay using anti-FLAG antibodies. HFFs, retinal pigment epithelial (RPE) cells, and human microvascular endothelial (HMVEC) cells were infected with US20-3xF-His virus. After 72 hours, the cells

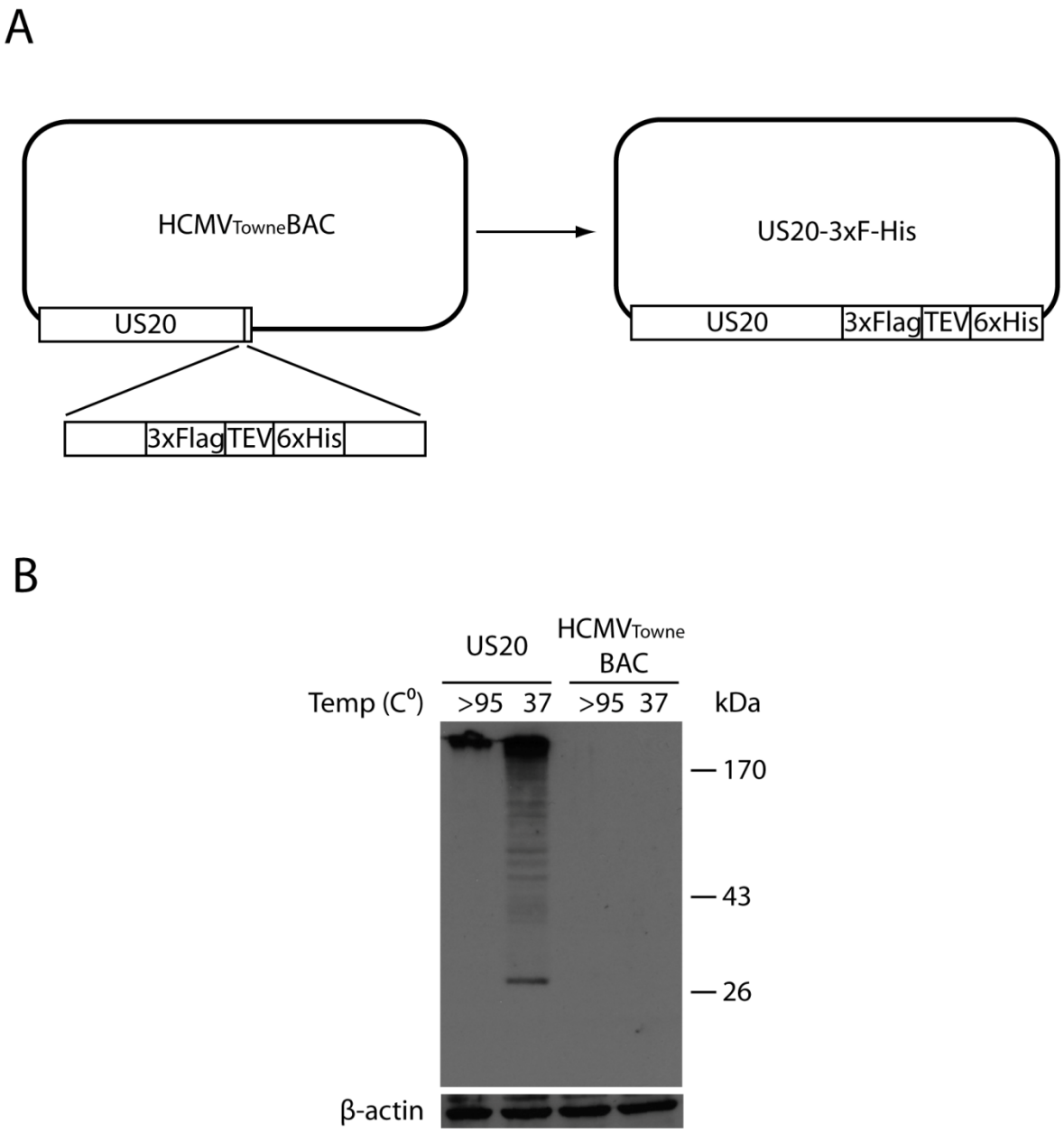


Figure 3.3. **Thermal Aggregation of HCMV pUS20**.
(A) Construction strategy of US20-3xF-His recombinant virus. (B) Human foreskin fibroblasts (HFF) were infected with either the US20-3xF-His (US20) virus or the wild-type Towne BAC ($HCMV_{Towne}$ BAC) at a high MOI. Lysate was harvested at 72 hours post infection. The samples were then treated with beta-mercaptoethanol and either incubated at 95^0C for 10 minutes or 37^0C for 45 minutes. The samples were run on an SDS-PAGE gel and immunoblotted using anti-FLAG antibody to detect tagged pUS20 or β-actin antibody (loading control).

were fixed and probed with anti-FLAG antibodies. We observed pUS20 a specific cytoplasmic localization that was consistent in HFFs, RPEs, and HMVECs (Figure 3.4).

HCMV pUS20 transiently localizes to early endosomes, while it does not localize to the ER or TGN at 72 hpi

Using previous localization studies which examined US12 family genes in the secretory pathway as a guide [9], we performed immunofluorescence assays using organelle markers to determine pUS20's cytoplasmic localization in relation to the endoplasmic reticulum (anti-Bip), early endosomes (anti-EEA), and the trans-Golgi network (anti-golgin245) (Figure 3.5) under the context of infection. In our studies we did not find any co-localization of tagged pUS20 with either Bip or Golgin245, which suggests that the pUS20 does not localize to the ER or the trans-Golgi network at 72hpi. We did find a small number of cells in which there was co-localization of pUS20 and EEA1, which suggests that pUS20 does traffic through but does not remain in endosomes.

HCMV pUS20 is expressed as early as 4 hours post infection

To determine when pUS20 was expressed during the viral replication cycle, we used the recombinant HCMV virus, US20-2xF-PA, which was previously constructed by our group. Similar to the US20-3xF-His virus, the US20-2xF-PA virus was constructed to express pUS20 tagged at the C terminus, however, we used a different tandem affinity tag containing two FLAG tag epitopes and two Protein A epitopes linked with a tobacco etch virus (TEV) cleavage site. We infected HFFs with US20-2xF-PA virus at a high MOI and then harvested the cells at 0, 4, 8, 12, 24, and 72 hours post infection. pUS20 expression in the harvested lysates was determined by immunoblotting using anti-FLAG antibodies. We identified an immunoreactive band of 47 kDa in size, which correlated with the predicted 28 kDa of US20 and the 19 kDa tandem affinity tag. This 47 kDa band representing pUS20 was present in all lanes except for 0 hpi, indicating that pUS20 is expressed as early as 4 hours post infection (Figure 3.6), suggesting that pUS20 is expressed during the early phase of protein expression.

Unique HCMV pUS20 cytoplasmic localization is prominent as early as 36 hours post infection in human foreskin fibroblasts

We followed the western blot time course assay with an immunofluorescence time course assay in HFF cells to determine when during viral replication did pUS20 localize to the specific cytoplasmic structure. Briefly, HFF cells were infected with US20-2xF-PA virus at a high MOI and then probed with anti-FLAG antibodies. To allow us to use the FITC fluorophore, we removed the fluorescence caused by the green fluorescent protein (GFP) expressed by the BAC by adding methanol, which not only fixed the cells, but also denatured the GFP to quench the fluorescence. We found pUS20 to localize to the cytoplasmic structure as early as 36 hours post infection (Figure 3.7). This observation suggests that pUS20's localization to the cytoplasmic structure occurs during the late phase of replication, which occurs between 24 to 36 hours post infection [1].

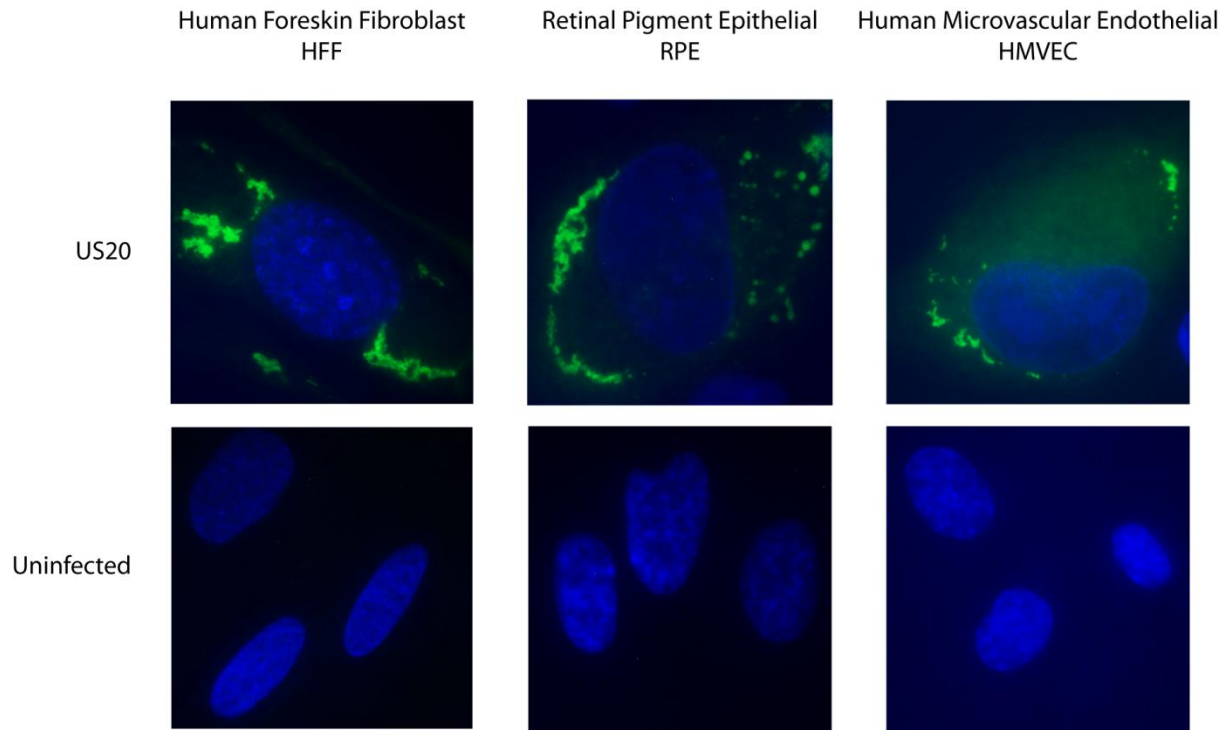


Figure 3.4. **Immunofluorescence assay shows that cytoplasmic localization of pUS20 is consistent in various cell types**.
HFFs, RPEs and HMVECs were infected at a high MOI with the US20-3xF-His virus. Cells were then fixed with methanol at 72hpi and probed with anti-FLAG antibody. Secondary anti-mouse conjugated with fluorescein was then added for visualization. DAPI staining was performed to visualize the nucleus. Images presented are merged images.

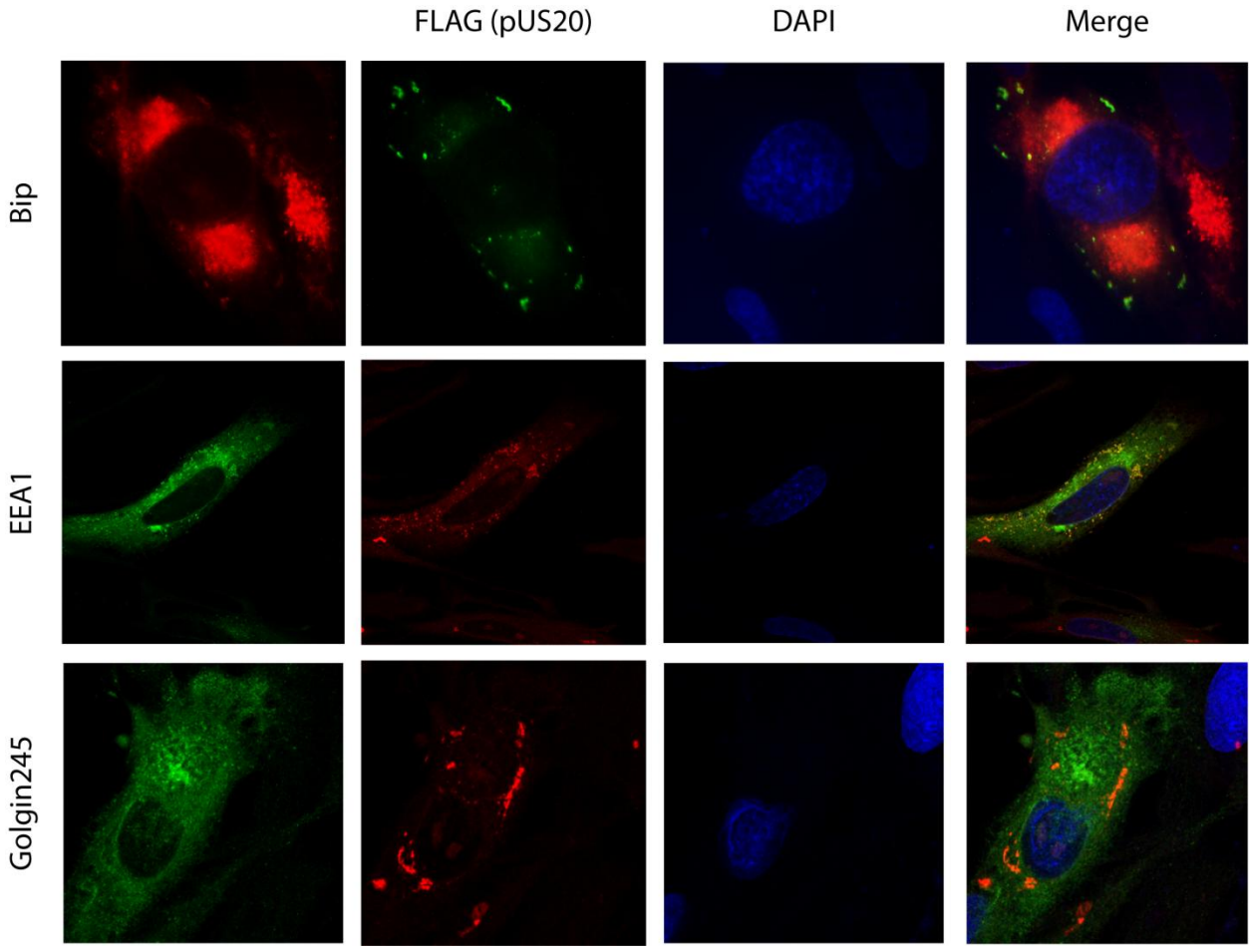


Figure 3.5. **Immunofluorescence assay shows co-localization of pUS20 and EEA1.** HFFs were infected with US20-3xF-His virus at high MOI for 72 hpi. Cells were fixed with methanol and then incubated with primary anti-FLAG; DAPI; and either anti-Bip, anti-EEA1, or anti-Golgin245. Anti-FLAG antibody was visualized with secondary anti-mouse conjugated with Alexa Fluor® 647 or FITC, while anti-EEA1 and anti-Golgin245 were visualized with secondary anti-rabbit conjugated with Alexa Fluor® 488. Anti-Bip was visualized with secondary anti-rabbit conjugated with Texas Red.

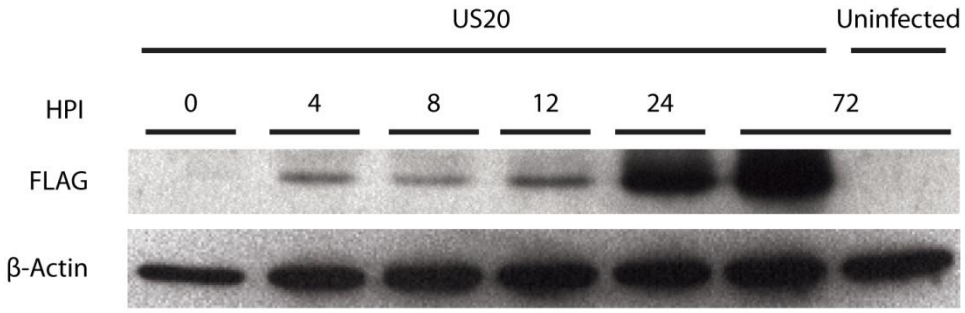


Figure 3.6. **Western Blot determines that pUS20 is expressed as early as 4 hpi.**
HFF cells were infected at a high MOI with the US20-2xF-PA virus, which expresses pUS20 tagged with a FLAG and Protein A epitope at the C terminus. Infected cellular lysates were harvested at the indicated time points and separated on a 10% SDS-PAGE gel. The resulting membrane was immunoblotted with anti-FLAG (FLAG) antibody to identify tagged pUS20 and anti-β actin as a loading control.

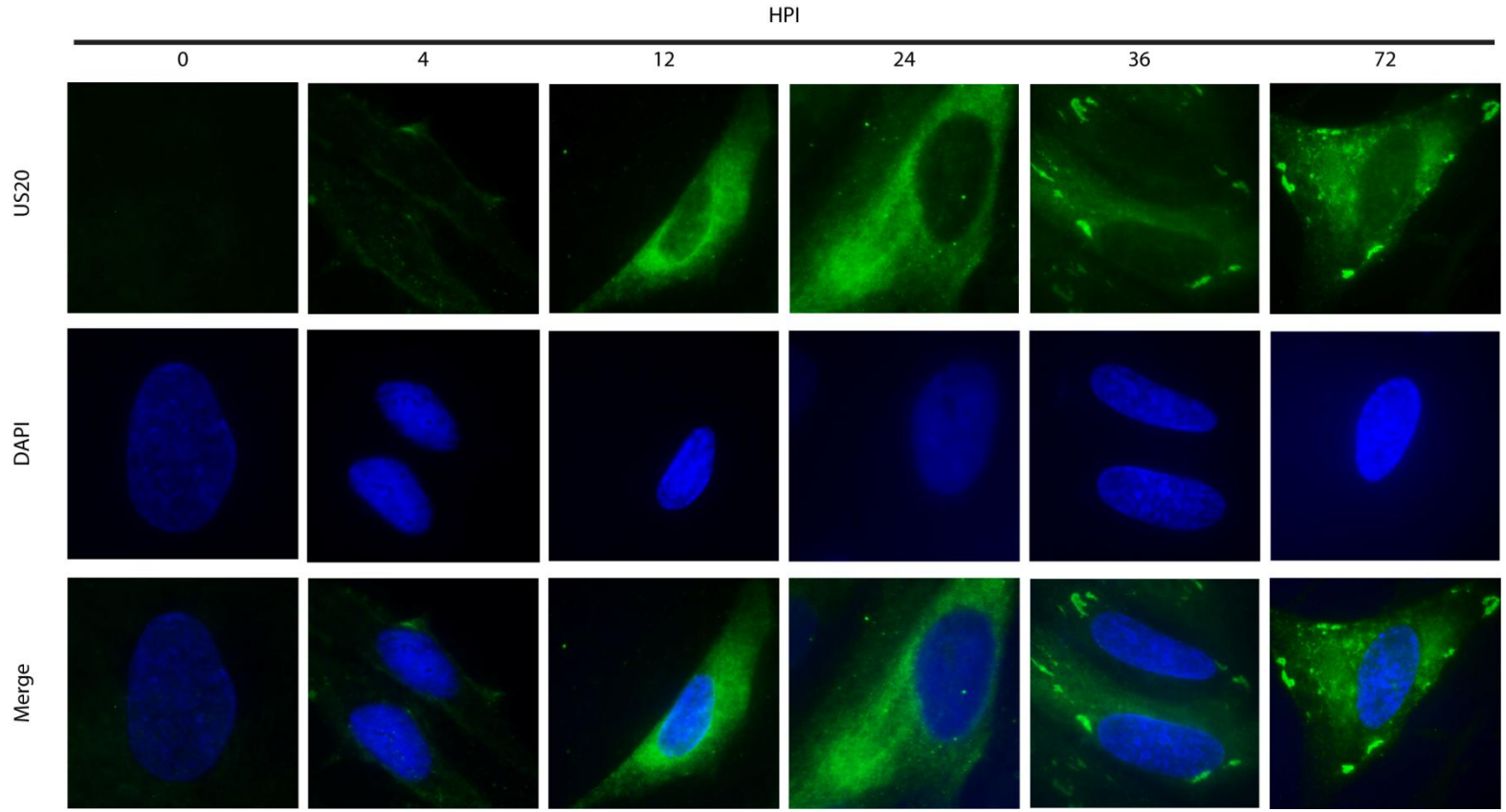


Figure 3.7**. Immunofluorescence shows that pUS20's specific cytoplasmic localization occurs as early as 36 hpi**.
HFF cells were infected at a high MOI with the US20-2xF-PA virus, which expressed the US20 protein tagged with a FLAG and Protein A epitope at the C terminus. Cells were fixed at the indicated time points with methanol and an immunofluorescence assay was performed using anti-FLAG antibodies. The primary antibody was visualized with secondary anti-mouse antibodies conjugated with fluorescein(Green). DAPI staining was then performed to visualize the nucleus.

Subcellular localization of HCMV pUS20 is not dependent on late gene expression

Given that we found pUS20 to be expressed as early as 4 hours post infection, we sought to determine if the subcellular location of pUS20 during the late phase of viral replication was dependent on other viral proteins expressed after 24 hours post infection. To block late gene expression during infection, we initially pretreated HFFs with phosphonoacetic acid (PAA), which has been previously shown to be a successful inhibitor of viral DNA synthesis. This inhibition results in the blocking of late gene expression [10]. The cells were then infected with US20-3xF-His virus and the media was then replaced after 2 hours with fresh media supplemented with PAA. The cells were fixed after 72 hpi and subjected to an immunofluorescence assay using anti-FLAG antibody. We found that pUS20 was still able to localize to the cytoplasmic structure regardless of the presence or absence of PAA in the media (Figure 3.8). As a control for the effectiveness of PAA, a similar immunofluorescence assay was performed on PAA treated, infected cells using anti-pp28 antibody, which has been commonly used as a marker for late gene expression [11, 12]. The significant reduction of pp28 expression we observed in PAA treated cells confirmed that late gene expression was significantly inhibited by the drug (Figure 3.8). This result further provides evidence that pUS20 is expressed before the late phase of replication and that late gene expression is not essential for pUS20's cytoplasmic localization.

pUS20's C terminus tail does not face the extra-cellular space

Since pUS20 was predicted to have seven transmembrane domains, we hypothesized that pUS20 could share characteristics with other multi-transmembrane spanning proteins, including the superfamily of G-protein-coupled receptors (GPCRs). GPCRs are expressed on the surface of the cell and commonly have their N terminus facing the extracellular space and the C terminus facing the intracellular space [13]. However, there are also multi-spanning transmembrane proteins that have their C terminus tail facing the extracellular space [14]. To examine if the pUS20 cytoplasmic localization we observed was at the cell surface, we performed a selective-permeabilization immunofluorescence assay [15, 16]. Briefly, we infected HFFs with US20-2xF-PA virus for 72hpi and then treated the cells with only paraformaldehyde, which fixed but did not permeabilize the cells, thus preventing any antibodies from entering the cell during the immunostaining process. As a negative control, we also infected HFFs with UL123-2xF-PA virus, which had the UL123 gene tagged with our FLAG and Protein A epitope at the C terminus. UL123, also known as IE1, is a transactivator protein that has been shown to localize to the nucleus [17].

We observed no fluorescence signal of pUS20 in cells that were only fixed and not permeabilized (Figure 3.9), while the control cells that were fixed and permeabilized with Triton-X did show pUS20 cytoplasmic localization. Similar results were also observed with UL123, which was expected considering that UL123 conventionally localizes to the nucleus. This result indicated that the C-terminus tail of pUS20 did not face the extracellular surface. However, if the topology and orientation of pUS20 was similar to other GPCRs, this result was expected since the C terminus tail for GPCRs typically faces the cytoplasm. To determine if the orientation of pUS20 was similar to GPCR, we examined the staining of US20 (1-206) virus (Figure 3.13), which expressed a truncated form of pUS20 that not only had no transmembrane (TM) domain 7 but was also tagged at the end of the peptide loop between TM6 and TM7, which

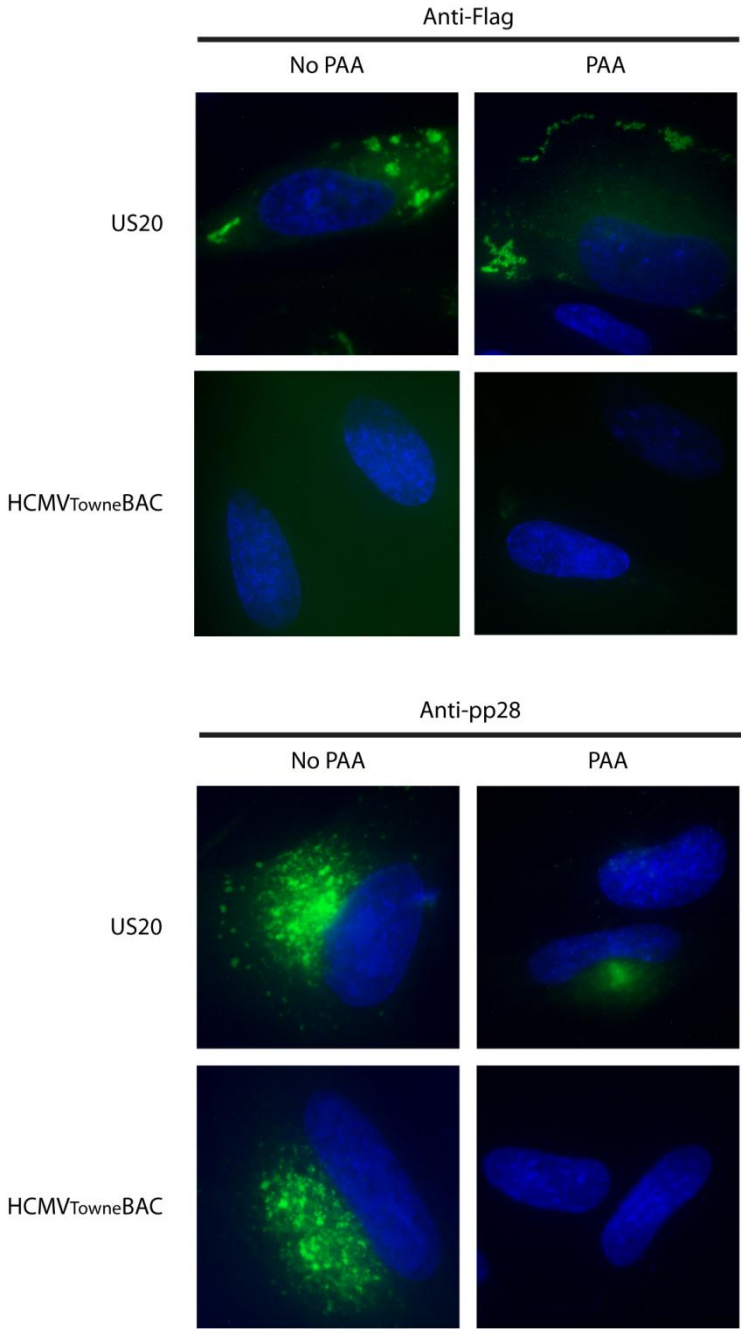


Figure 3.8. **Immunofluorescence shows that PAA does not inhibit pUS20's cytoplasmic localization.**
HFFs, pretreated with 500 ug/mL phosphonoacetic acid (PAA), were infected with either the US20-3xF-His (US20) or wild-type Towne BAC (HCMVTowne BAC) virus at an MOI of 1. After 2 hours, the infectious inoculum was then replaced with fresh media supplemented with PAA. Infected cells were fixed with methanol at 72hpi and then incubated with either anti-FLAG or anti-pp28 antibodies. Secondary anti-mouse conjugated with fluorescein was then added for visualization. DAPI staining was performed to visualize the nucleus. The images presented are merged.

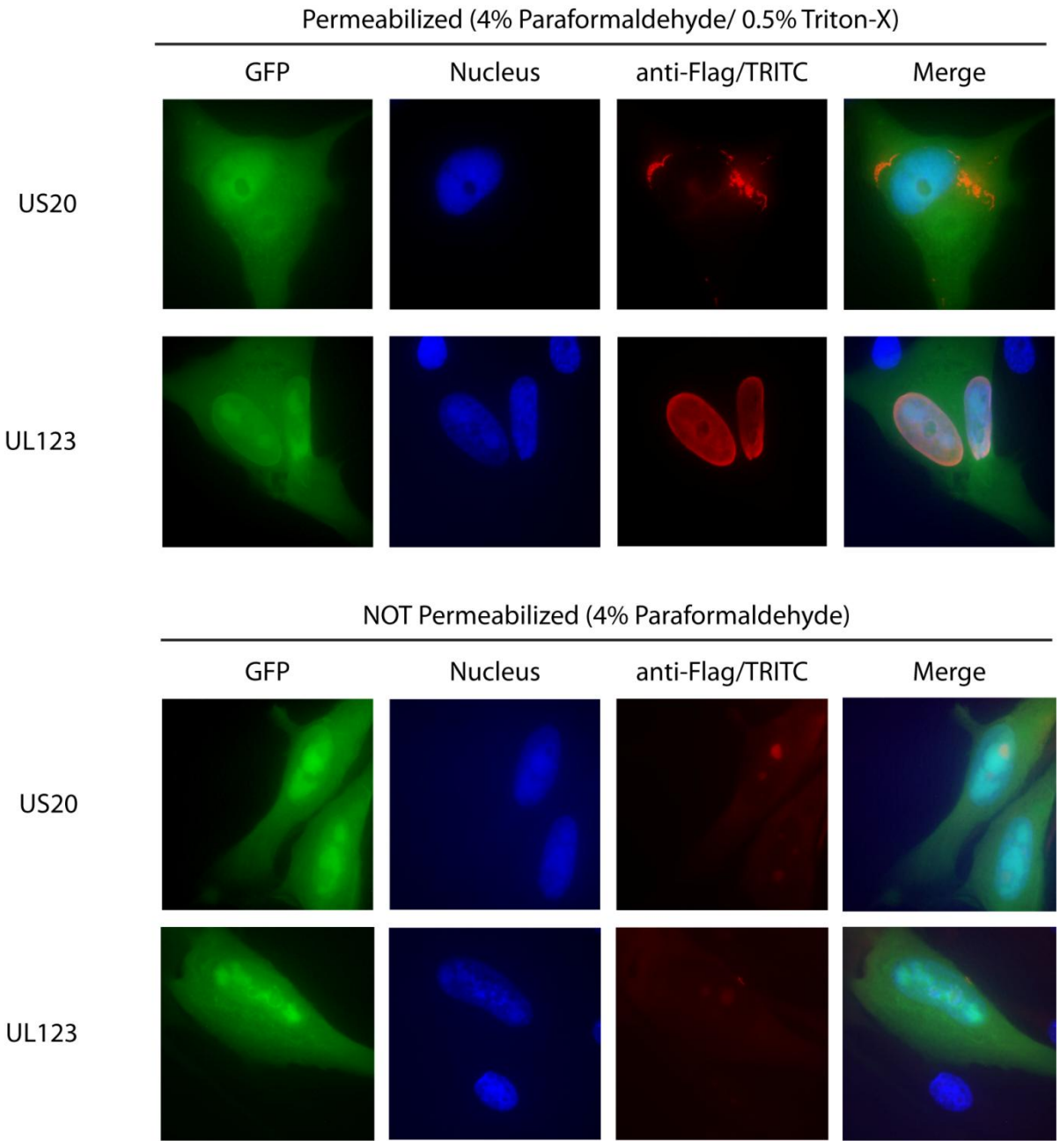


Figure 3.9. **Immunofluorescence shows no US20 staining in unpermeabilized cells infected with US20-2xF-PA virus.**
HFF cells were infected with either US20- or UL123- tagged virus (C terminus tagged with FLAG and Protein A epitopes) at a high MOI. Cells were then fixed at 72hpi with 4% paraformaldehyde and either permeabilized with 0.5% Triton-X or mock permeabilized (PBS). Primary anti-FLAG antibody was added and then visualized using the Histostain SP kit (Invitrogen) and streptavidin-TRITC. DAPI staining was then performed to visualize the nucleus.

would be expected to face the extracellular space. We found that nonpermeabilized cells infected with US20 (1-206) virus also did not exhibit cytoplasmic staining (data not shown). From this data, we concluded that the C-terminus of pUS20 does not face the extracellular space and that the specific cytoplasmic localization we observed is not at the surface of the cell. These results show that pUS20 remains internalized inside of the cell and does not traffic to the cell surface.

HCMV US20 gene is not essential for viral replication in HFFs , RPEs, HMVECs, and differentiated THP-1 macrophages.

To study the role of the US20 gene in the context of viral replication, we constructed a deletion mutant virus which replaced the US20 gene with a Zeocin antibiotic resistance marker (Figure 3.10A). The PCR results and restriction digest profile of the recombinant BAC confirmed the replacement of US20 and no nonspecific genomic recombination in the BAC (Figure 3.10B, C**)**. The recombinant BAC was transfected into HFFs and the resulting virus, US20ko, was harvested after two weeks. To further confirm the deletion of the US20 gene, we harvested RNA from HFFs infected with either the US20ko virus or the wild-type HCMV Towne BAC virus at 72 hours post infection and performed a northern blot using a radiolabelled probe specific to the last 300 nucleotides of the US20 gene. We found two bands present in the wild-type virus lane, indicating that the US20 gene expressed a two kilobases transcript and a six kilobases transcript (Figure 3.10D**)**. As expected, the bands were not present in the lane containing RNA harvested from the cells infected with the US20ko virus.

To examine the role of the US20 gene in HCMV replication *in vitro*, we conducted multi-step growth curves of the US20ko virus in HFFs, retinal pigment endothelials (RPE), and human microvascular endothelial cells (HMVECs). While HFFs are not usually infected with HCMV [18], both RPEs [19] and HMVECs [1] have been shown to be natural reservoirs for lytic HCMV infection. The deletion of US20 did not change replication levels of HCMV in HFFs (Figure 3.11A**)**, which correlated with previous results from our lab [4] and others [5]. Because of HCMV's large genome and coding potential, there is significant data suggesting that the virus encodes tropism factors that influence replication in specific cell types [1, 4] and temperance factors, which cause the virus to become hypervirulent upon deletion [4]. Previous data in our lab indicated that the deletion of either US16 or US19 ORFs caused their respective mutant knock out virus to growth better in HMVECs relative to the wild-type virus [4]. Given that US16 and US19 are part of the US12 family of genes, which US20 is also a member, we thought the US20ko virus would also show a similar phenotype. However, our multi-step growth curve results showed that the US20ko virus grew at wild-type levels indicating that the US20 ORF did not influence viral replication levels in either RPEs (Figure 3.11B) or HMVECs (Figure 3.11C).

We also sought to determine if US20 played a role in HCMV replication in macrophages, which, depending on its differentiation state, is a major site of latent and lytic infection [1]. Using THP-1 myelomonocytic cells differentiated with 12-O-tetradecanoylphorbol-13-acetate (TPA), which has been shown to be permissive to infection by the Towne strain [20], we conducted a multi-step growth curve study. While THP-1 cells have been shown to be permissive to infection by HCMV within 24 hours of treatment with TPA [21], a recent study showed that the monocytes go through multiple differentiation states after treatment evident by cellular morphological changes [22]. Furthermore, a longer differentiation time of THP-1 by TPA was found to increase titers [22]. For our experiment, we pretreated undifferentiated THP-1

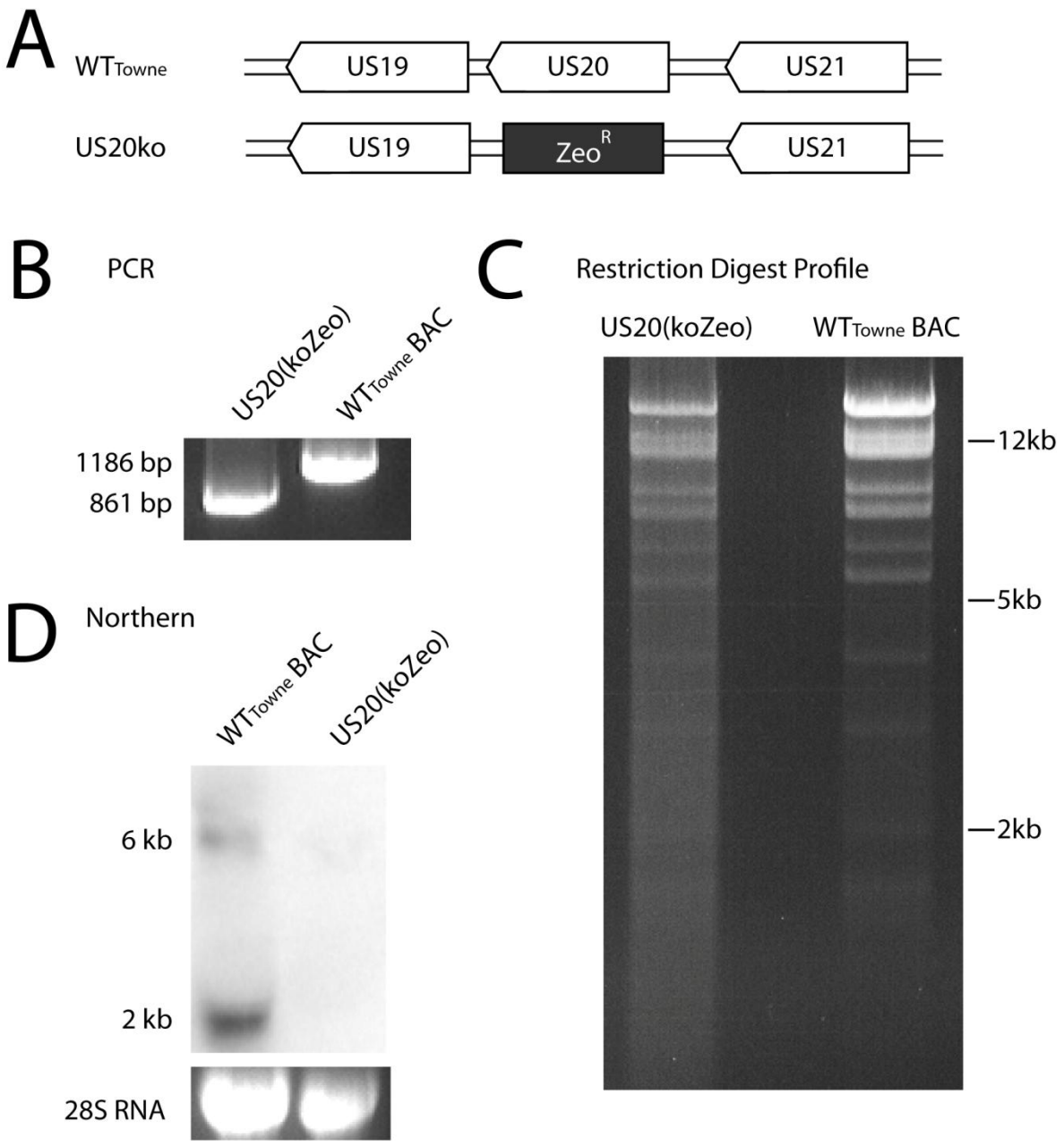


Figure 3.10. **Construction of US20ko virus**.
(A) Schematic of the deletion of the US20 ORF in the wild-type Towne BAC (WT_{Towne} BAC) and replacement with a Zeocin resistance gene (ZeocinR) by insertion deletion mutagenesis. (B) PCR using primers flanking US20 ORF. (C) Restriction digest was performed on the recombinant and wild-type Towne BAC using Hind III enzyme. (D) RNA from HFF cells infected with either the US20 knock out (US20(koZeo)) or wild-type Towne BAC virus after 72 hpi was harvested and separated on an agarose gel. The subsequent membrane was probed with a radiolabelled oligomer specific to the last 300 nucleotides of US20 and imaged using a phosophoimager.

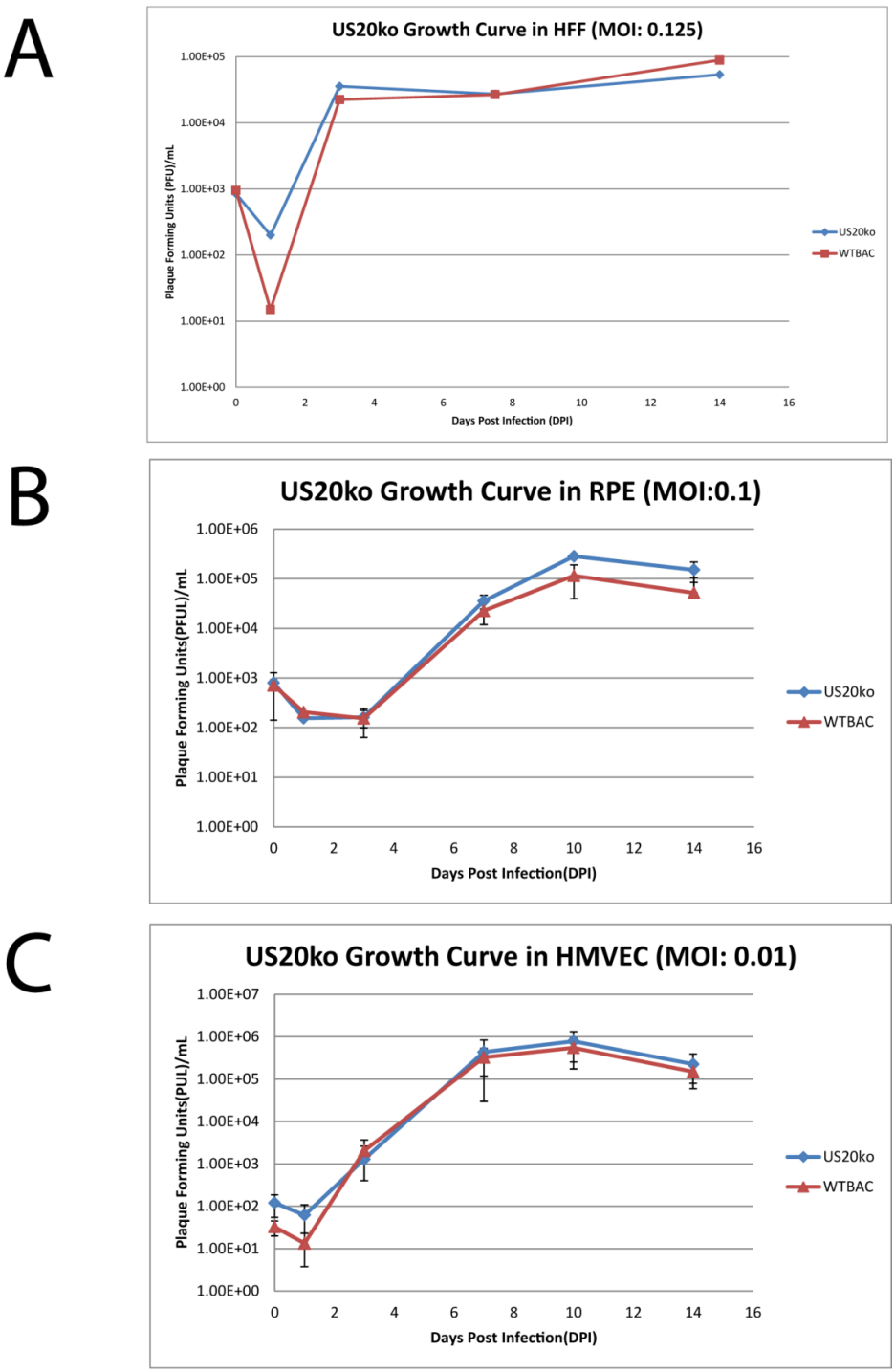


Figure 3.11. **Viral growth curves show that the deletion of US20 ORF does not impact fitness of the virus in various cell types**.
(A)HFF, (B) RPE, (and C) HMVEC were infected with either US20 deletion virus (US20ko) or wild-type Towne BAC (WTBAC) at their respective MOI. Infected cells were harvested at various time points and the lysate was titrated in HFF cells. GFP expressing plaques were counted using a fluorescence microscope. Each experiment was performed in triplicate.

cells with TPA for either 1, 3, or 7 days and then infected them with either the US20ko or the wild-type virus at a MOI of 0.1. For all three differentiation time points, we found that the US20ko viral titers were at levels similar to the wild type virus (Figure 3.12), which indicated that the deletion of the US20 ORF did not impact viral replication at various stages of monocyte differentiation. We also noticed that the titers at 7 days post infection for THP-1 cells that were differentiated for 72 hours had titers that were 10 times less than the infected THP-1 cells were differentiated for only 24 hour, which was unexpected.

Deletion of predicted pUS20 transmembrane domains alters protein levels of pUS20 and destabilizes the specific cytoplasmic localization

To study the role of the predicted transmembrane domains in the cytoplasmic localization of pUS20, we constructed several viruses that expressed truncated forms of pUS20 using the previously mentioned homologous recombination techniques. Using the predictions (Figure 3.2A) as a guide, we systematically deleted each of the predicted transmembrane (TM) domains (Figure 3.13A). In addition to deleting the predicted domains, we also inserted a 3xFlag and polyhistidine tandem affinity tag (3xFlag-6xHis-KanR) at the C terminus end to allow us to track the truncated protein. PCR and restriction digestion of the modified BACs confirmed that the tag inserted in the US20 gene and that no unwanted genomic recombination took place (Figure 3.13B). The recombinant BACs were also sequenced, which not only confirmed the deletion of the predicted transmembrane domains but also insured that the tag was inserted in frame. Based upon visualization of CPE and GFP expression in infected HFFs, we determined that the truncated mutants grew at levels similar to the wild-type (Figure 3.13C).

In our immunoblotting experiments of our truncation mutants, we observed that the protein levels of the truncated pUS20 in infected HFFs decreased for some of the recombinant viruses. We quantified the band intensities using ImageQuant, using IE72 (IE1) as a loading control and to normalize for possible variation in virion concentration of the inoculum. We found that the protein levels of truncated pUS20 in the viruses, US20 (1-142), US20(1-113), US20 (1-88), US20(1-60), and US20(1-27) to be less than 50% of the protein levels of the full length pUS20 of the US20(3xF-His) virus. While the pUS20 levels of US20 (1-206) and US20 (1-174) were roughly equivalent to US20-3xF-His (Figure 3.14A). These results indicated that the lack of predicted TM5, TM6, and TM7 reduced protein levels of pUS20.

To determine if the change in protein levels was occurring at the RNA stage or the protein stage, we performed qRT-PCR to measure levels of US20 mRNA in RNA samples taken from HFFs infected with US20(1-60), US20-3xF-His, and wild-type Towne BAC. The results showed that the US20 mRNA levels were the same for both US20(1-60) and US20-3xF-His (Figure 3.14B), suggesting that the change in protein level was not occurring during transcription. A growth curve study of US20(1-60), US20-3xF-His and wild-type Towne BAC in HFF quantitatively showed that truncating pUS20 did not impact the fitness of the virus (Figure 3.14C). This confirmed that the reduction in protein levels was not because of a reduction in viral replication.

For the immunofluorescence assay, we infected HFFs with the recombinant viruses expressing the truncated forms of pUS20 and fixed the cells at 72 hpi (Figure 3.15). We found that the truncated pUS20 expressed by the US20(1-88), US20(1-113), US20(1-143), US20(1-174), and US20(1-206) viruses all exhibited a distinctive cytoplasmic localization similar to the phenotype that we observed with the nontruncated pUS20 expressed by the US20-3xFlag-His

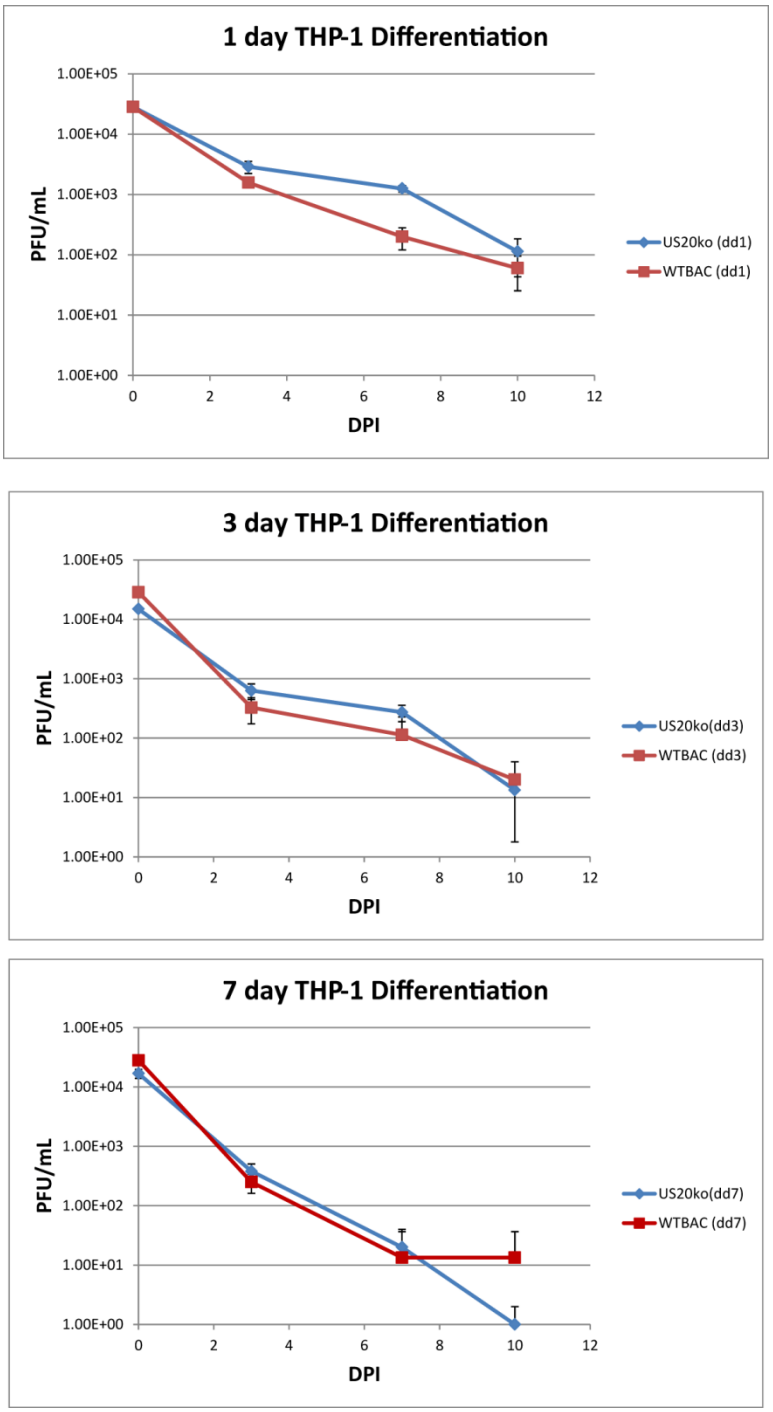


Figure 3.12. **Viral growth curve shows that deletion of US20 ORF does not impact infection and replication in differentiated THP-1 cells**.
THP-1 cells were initially differentiated for either 1, 3, or 7 days with TPA and then infected with either US20 deletion virus (US20ko) or wildtype Towne BAC (WTBAC) virus at an MOI of 0.1. The supernatant was harvested and titered on HFFs. GFP expressing plaques were counted using a fluorescent microscope. All experiments were performed in triplicate.

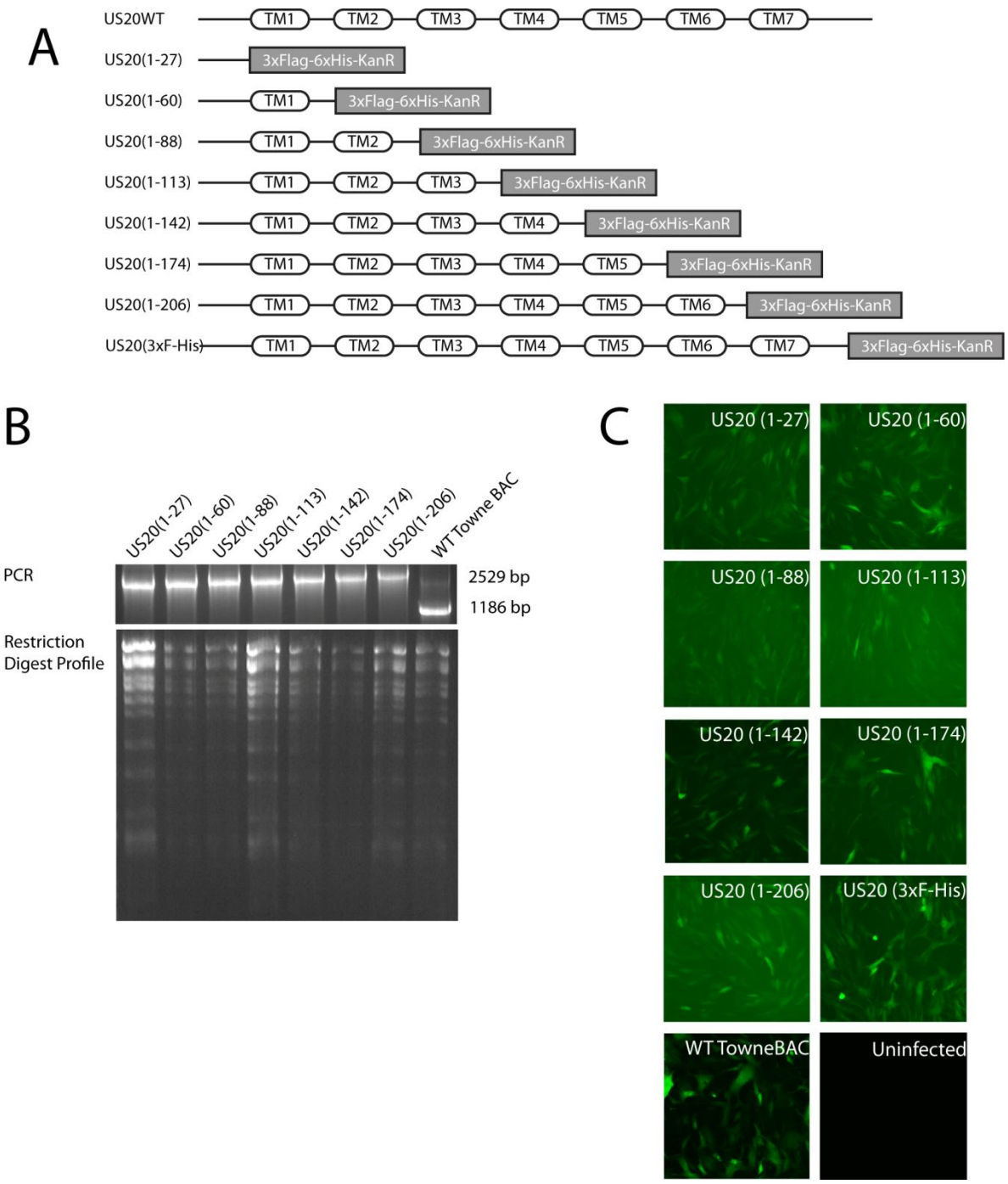


Figure 3.13. **Construction of US20 Truncation mutants.**
(A) Schematic of truncation mutants. Insertion deletion mutagenesis was used to delete the predicted transmembrane domains and insert an affinity tag containing a 3xFLAG and a poly-histidine epitope at the C terminus. (B) PCR using primers flanking US20 confirmed the deletion and insertion of the tag. Restriction digests were performed on the recombinant BACs using Hind III restriction enzymes. (C) HFFs were infected with the truncation mutants for 72 hpi and viral fitness was determined by visualization of GFP expressing cells.

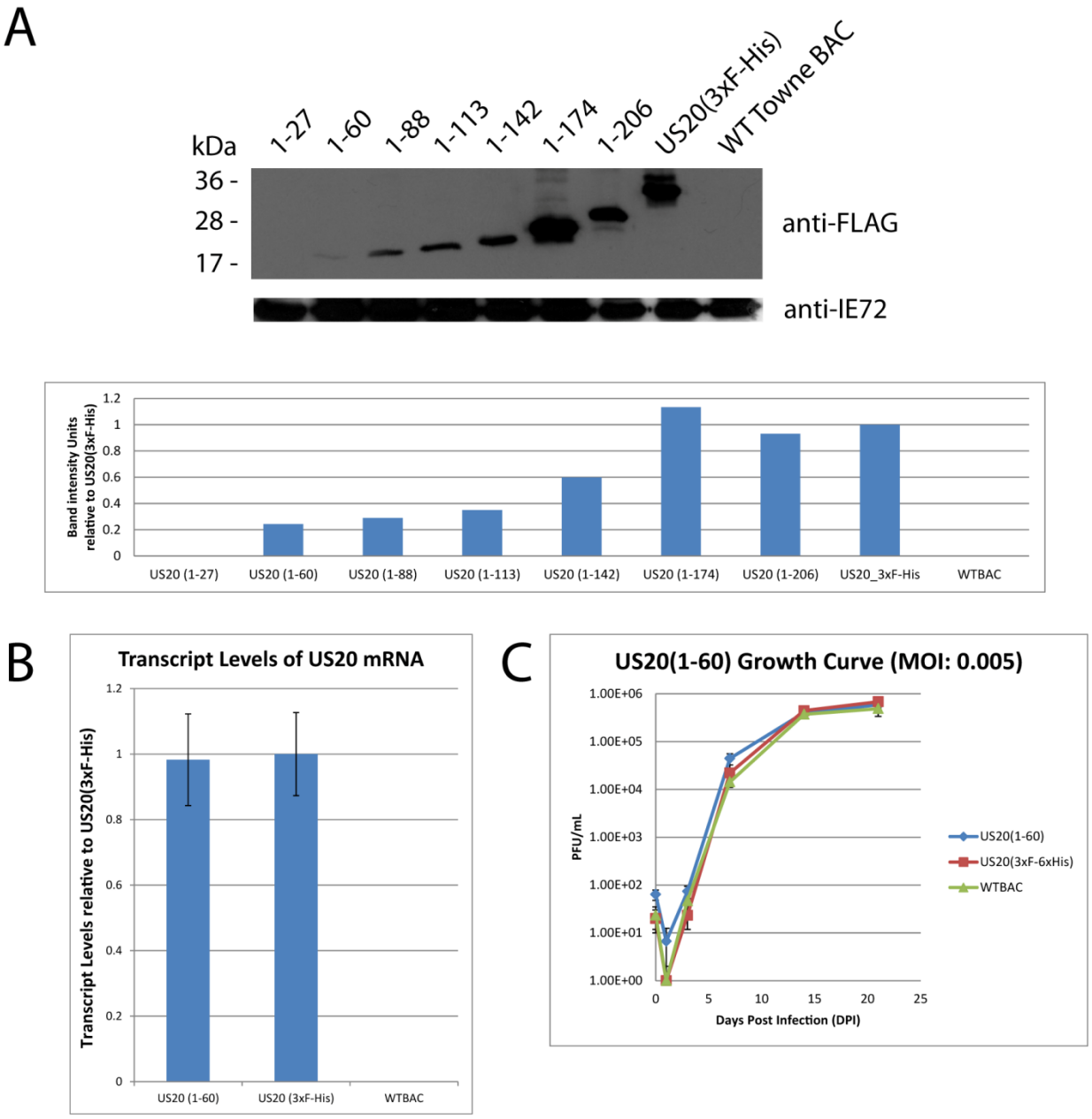


Figure 3.14. **Truncation of pUS20 results in reduction in protein levels.**
(A) HFF cells were infected at an MOI of 1 with the US20 truncation mutants, the tagged US20 virus (US20-3xF-6xHis) and the wild-type Towne BAC (WTBAC). The lysates were harvested and immunoblotted with either anti-FLAG to identify pUS20 or anti-IE72 to control for viral infection. ImageQuant was used to quantify the expression, using IE72 for normalization. The results were the normalized in reference to US20-3xF-His. (B) qRT-PCR results measuring US20 mRNA levels in lysates harvested from US20(1-60), US20(3xF-His), and WTBAC virus. CT values were first normalized to IE1 transcript levels and then normalized to US20(3xF-His). (C) Viral growth curve of US20(1-60) in HFFs infected at MOI of 0.005.

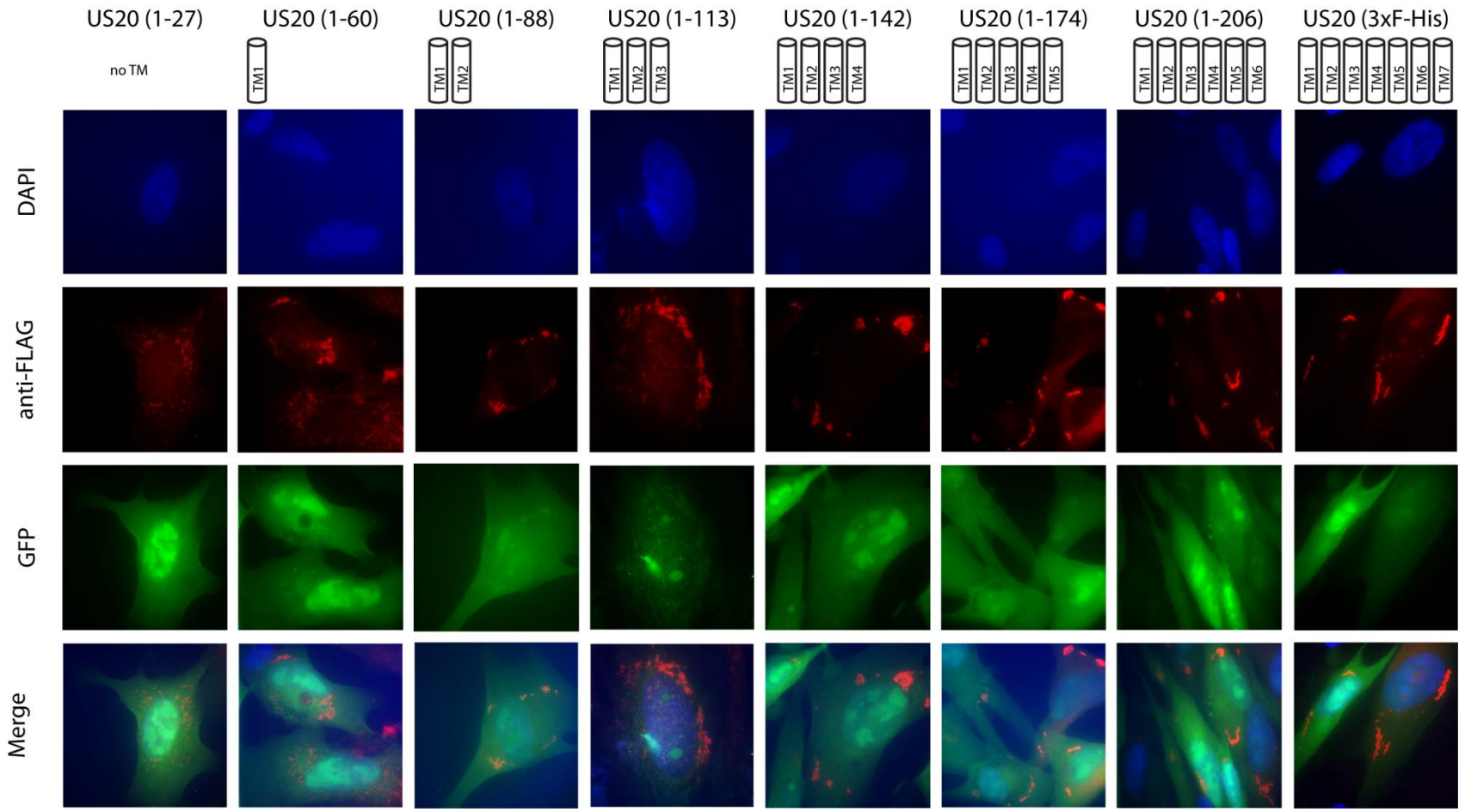


Figure 3.15. **Immunofluorescence assay of truncated pUS20 shows that deletion of TM2-TM7 does not impact subcellular localization**.
HFF cells were infected with the truncation mutant virus at a high MOI. After 72 hours, infected cells were fixed with 4% paraformaldehyde and permeabilized with 0.5% Triton-X. US20 truncated proteins were probed using anti-FLAG antibodies and visualized using the Histostain SP kit (Invitrogen) and TRITC-Streptavidin. DAPI staining was performed to visualize the nucleus. GFP expression indicates viral infection.

virus. We also observed that US20(1-60) virus, which expressed the truncated pUS20 with only TM1, showed both the distinctive and a dispersive cytoplasmic localization, while US20(1-27), which expressed the truncated pUS20 with no TM domains, showed only the dispersive cytoplasmic localization. This data suggests that the transmembrane (TM) domains 2, 3, 4, 5, 6, and 7 are not essential for the specific cytoplasmic localization of pUS20.

pUS20 forms oligomers

It has become generally accepted that GPCRs can form homo- and hetero- dimers [23] and that this dimerization can play a role in transport to the plasma membrane [24], ligand-promoted regulation [25], signal transduction [26], and receptor internalization [27]. Since both GPCRs and pUS20 share a common trait of having seven transmembrane domains, we hypothesized that pUS20 could also form dimers and oligomers. Briefly, samples were prepared by infecting HFFs with the US20-3xF-His virus, harvesting the proteins at 72 hpi, selectively adding dithiothreitol (DTT), and then incubating the samples at 37°C. DTT is a reducing agent that has been used by other groups in the past to show homo-dimerization [28]. For the US20-3xF-His samples, we were able to see the 28-kDa -immunoreactive band, representing the tandem affinity tagged pUS20 monomer, under reducing (DTT) and non-reducing (minus DTT) conditions (Figure 3.16). We also observed an immunoreactive band that migrated at a size predicted for a tandem affinity tagged pUS20 dimer (56 kDa), which showed an increase in immunostaining under non-reducing conditions.

To confirm the homo-dimerization of pUS20, we used a co-immunoprecipitation approach, which has not only been used to demonstrate the existence of homo-dimers, but also hetero-dimers of GPCRs [29]. Prior to the co-transfection experiment, we wanted to see if the cloned and tagged pUS20 behaved similarly to the pUS20 expressed during viral infection. We repeated the "thermal aggregation" experiment (Figure 3.2A) using lysates from 293T cells either transiently expressing cMyc-US20 or US20-FLAG. In both cases, using the respective antibodies, we found an immunoreactive band at 25kDa in samples incubated at 37°C, which was slightly smaller than the size of the pUS20 with the 3xF-6xHis tag (Figure 3.17A). For the samples that were incubated at 97°C, we saw an immunoreactive band greater than 170kDa and no bands present at 25 kDa in cells transiently expressing cMyc-US20 and several large bands greater than 55 kDa in cells transiently expressing US20-FLAG. Even though we saw a 25 kDa band in both the 37°C and 97°C incubated samples from cells transiently expressing US20-FLAG, the band intensity was less for the 97°C incubated sample. In addition, we determined that both cMyc-US20 and US20-FLAG could form homo-dimers and homo-oligomers through immunoblotting using identical sample preparation conditions from the infection experiments (Figure 3.17B). From these results, we concluded that both transiently expressed tagged pUS20 in 293T cells behaved similarly to the tagged pUS20 expressed during HCMV infection.

For the co-immunoprecipitation, 293T cells were co-transfected with plasmids expressing N-terminus c-Myc tagged pUS20 (cMyc-US20) and C-terminus FLAG tagged pUS20 (US20-FLAG). The protein lysates were harvested and US20-FLAG complexes were immunoprecipitated with anti-FLAG antibodies. The elutions were then separated under non-reducing conditions (no DTT) and the resulting membrane was immunoblotted with either anti-FLAG or anti-cMyc antibodies. From our eluted US20-FLAG complexes from cells transiently expressing both plasmids, we had immunoreactive bands of 50 kDa and 75 kDa in size using anti-FLAG antibodies for staining (Figure 3.18A) and immunoreactive bands of 25 kDa, 50 kDa,

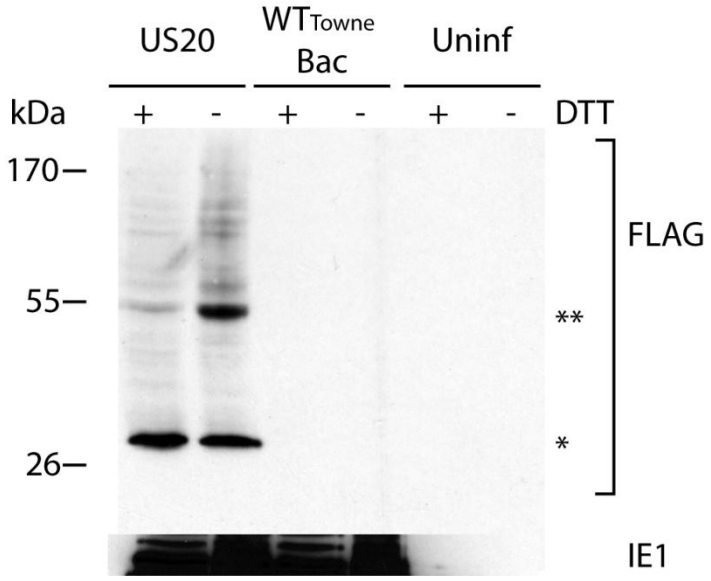


Figure 3.16. **Immunoblot of pUS20 shows homo-dimerization**.
HFF cells were infected with either US20-tagged virus (C terminus tagged with 3xFLAG and poly histidine epitopes) or wildtype Towne BAC (WTTowne BAC virus. Cells were harvested at 72hpi and protein samples treated with either 0.1M dithiothreitol (DTT) (reduced) or water (non-reduced), incubated at 37°C for 30 minutes and separated on a 10%SDS-PAGE gel. The * indicates the monomer, while the ** represents the dimer.

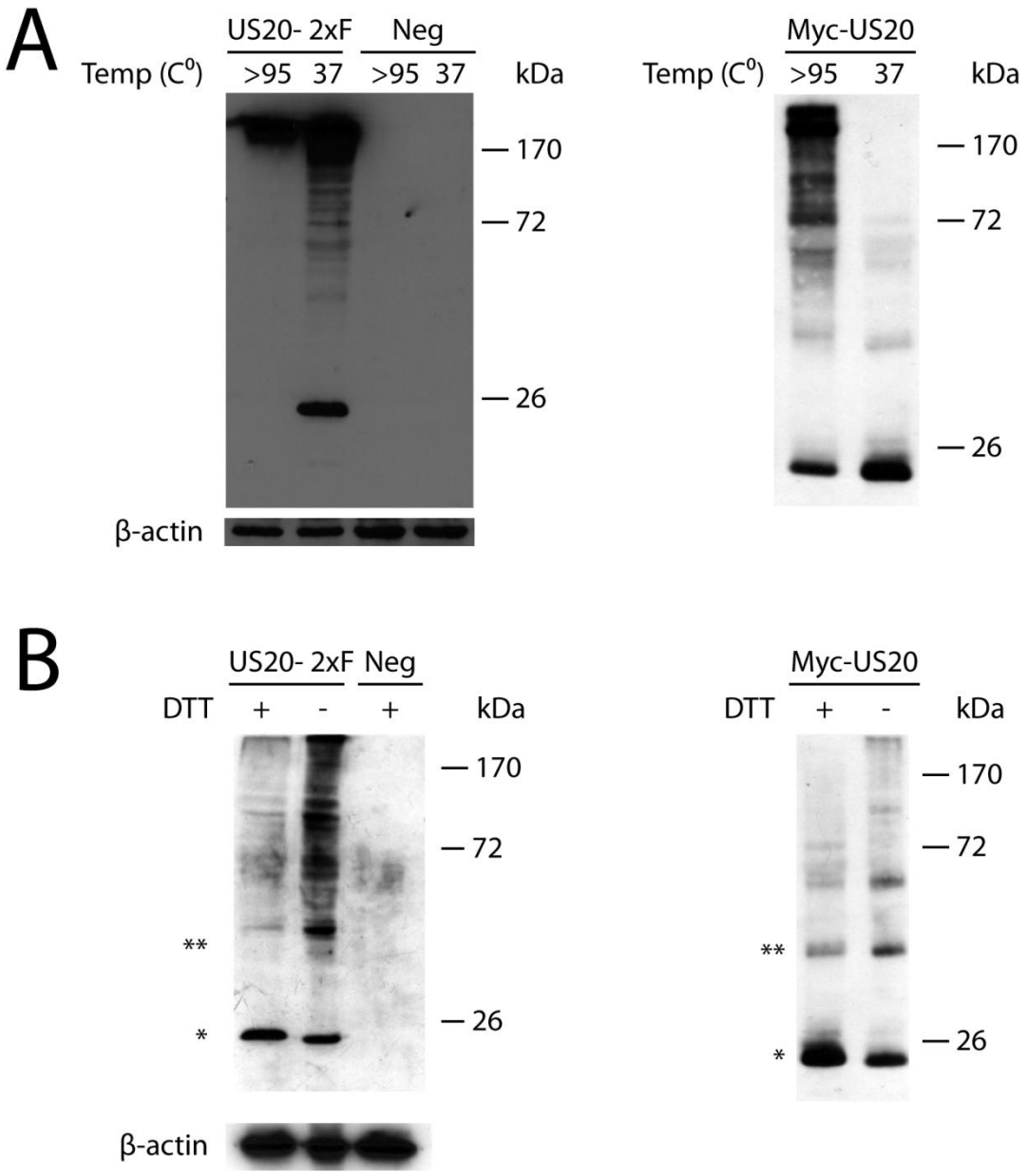


Figure 3.17. **Immunoblotting of cMyc-tagged pUS20 and FLAG-tagged pUS20 shows dimerization.**
293T cells were transfected with either pcDNA plasmid expressing US20 tagged at the C terminus with FLAG epitopes (US20-2xF), pCMV-Myc plasmid expressing US20 tagged at the N terminus with a cMyc epitope (myc-US20) or an empty plasmid (Neg). Proteins were then harvested after 48 hours. (A) Samples were treated with beta-mercaptoethanol and heated at either >95° or 37° Celsius. (B) Samples were either treated with DTT (+) or water (-) and then heated at 37° Celsius. All samples were run on 10% SDS-PAGE gel. The resulting membranes were immunoblotted with either anti-FLAG or anti-cMyc antibody to identify pUS20. Anti-beta actin immunoblotting served as a loading control. * indicates the US20 monomer, while ** represents the dimer.

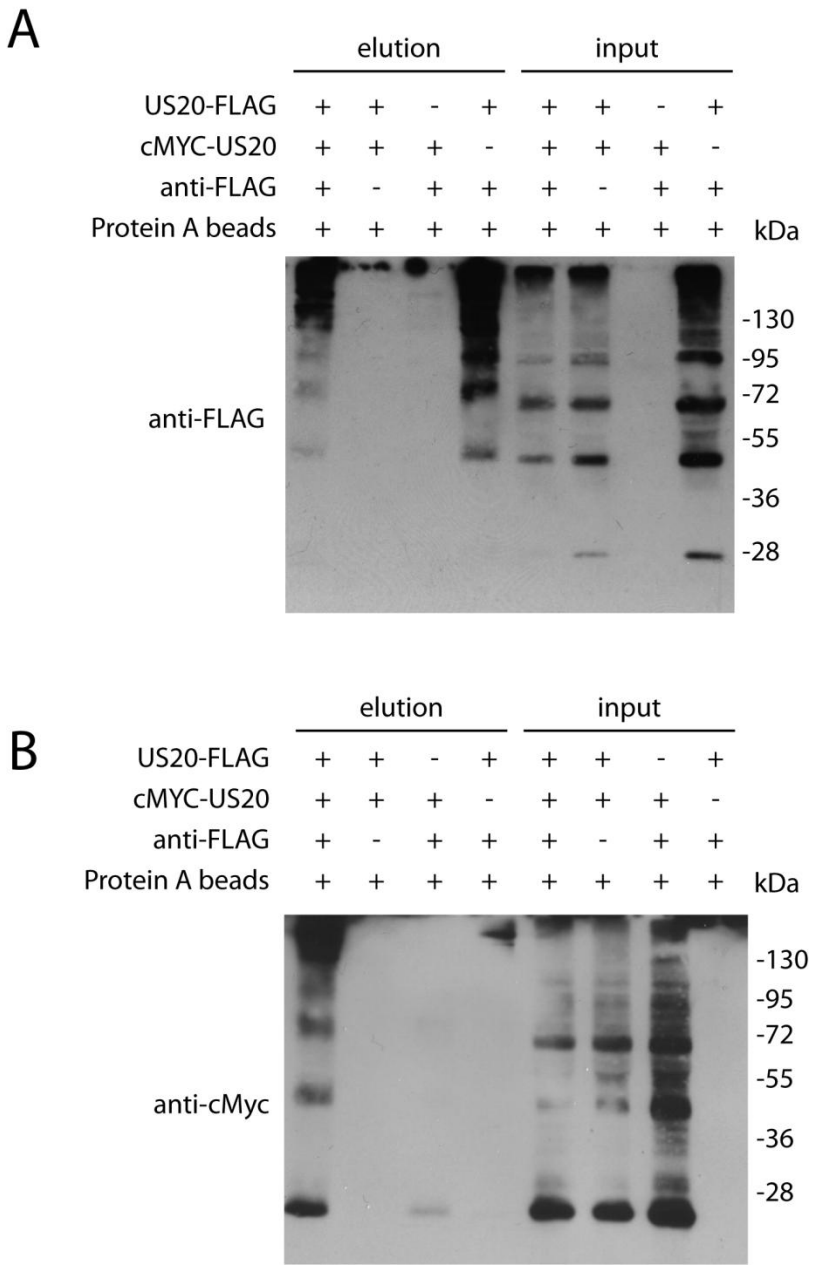


Figure 3.18. **Co-immunoprecipitation shows dimerization of transiently expressed cMyc-US20 and US20-Flag**.
293T cells were transfected with either a plasmid expressing a N- terminus cMyc-tagged pUS20 (cMYC-US20), a plasmid expressing C-terminus FLAG-tagged pUS20 (US20-FLAG), or both plasmids. Protein lysates were harvested after 48 hours and then subjected to immunoprecipitation (IP) using anti-FLAG antibody and Protein A/G beads. Proteins were eluted from the beads with 3xFLAG peptides (150 µg/uL) and separated on a 10% SDS-PAGE gel. Membranes were then immunoblotted with either (A) anti-FLAG or (B) anti-cMyc antibodies.

and 75 kDa in size when we probed with anti-cMyc antibody (Figure 3.18B). This result showed that both US20-FLAG and cMyc-US20 peptides did associate with each other strengthening the hypothesis that pUS20 can form homo-dimers and higher order oligomers.

To identify specific regions of pUS20 important for its homo-dimerization, we examined the ability of our truncated pUS20 peptides expressed by our recombinant viruses to form dimers or oligomers. Following the approach we took with the full-length tagged pUS20 expressed by US20-3xF-His, we infected HFFs with the recombinant viruses and harvested the lysates after 72 hpi. The lysates were then selectively treated with DTT and then separated on a SDS-PAGE gel. We found that lanes containing lysates from US20(1-142), US20(1-174), and US20(1-206) had two distinct immunoreactive bands, the smaller band representing the truncated monomer and the second band representing the dimer (Figure 3.19). While we did identify an immunoreactive band representing the monomer for US20(1-27), US20(1-60), US20(1-88), and US20(1-113), there was no "dimer" band visible suggesting that the truncated pUS20 expressed by these viruses were unable to form dimers. From these results, we believe that pUS20 does not need TM5, TM6, and TM7 in order to form dimers.

HCMV US20 ORF gene does not modulate Caspase 3 activation after staurosporine treatment

Because our BLAST search predicted a Bax-inhibitor-1-like domain in pUS20, we hypothesized that pUS20 may have functions similar to the Bax-inhibitor-1 protein. Originally discovered in yeast cells [30], Bax-inhibitor-1 (BI-1) and its orthologs have been shown to inhibit BAX-induced apoptosis in yeast, mammalian, and plant cells [30-32]. The topology of BI-1 is still not fully understood with bioinformatical analysis suggesting six to seven transmembrane domains, with the seventh domain being less hydrophobic than the first six domains [33]. The shared characteristic of having multiple transmembrane domain further strengthened the possibility that pUS20 could function similarly to BI-1.

We decided to examine US20 ORF's ability to inhibit BAX-induced apoptosis using staurosporine (STS), a potent inhibitor of protein kinases produced by *Streptomyces staurosporeus* [34], as the apoptosis inducer. STS has been used successfully by various groups to trigger Bax-induced apoptosis [35, 36]. Furthermore, overexpression of BI-1 in 293T cells has been successfully shown to inhibit apoptosis induced by STS [30]. We first sought to examine if the overexpression of US20-FLAG would prevent be able to prevent apoptosis induced by STS. Briefly, 293T cells were transfected with either the US20-FLAG expressing plasmid or empty plasmid and then treated with STS for six hours. The cells were then assayed for apoptosis by measuring Caspase-3 activity, one of the main apoptosis executioners, with the expectation that inducing apoptosis would cause an increase in Caspase-3 cleavage activity [37, 38]. We found that there was no significant difference in Caspase-3 activity between cells over expressing US20-FLAG and the control, suggesting that pUS20 is unable to inhibit apoptosis induced by STS (Figure 3.20A). We believe that the higher levels of Caspase-3 activity in 293T cells treated with STS that were not transfected, relative to the levels in transfected cells was the result of cell death caused by transfection process. The protein lysates from the experiment were also harvested and immunostained with anti-FLAG antibodies to indicate transient expression of US20-FLAG (Figure 3.20B).

To see if the deletion of the US20 ORF would impact STS-induced apoptosis during HCMV infection, we infected HFF cells with either the US20ko or wild-type Towne BAC virus for 72 hours and then treated the cells with STS for 6 hours. The protein lysates were then

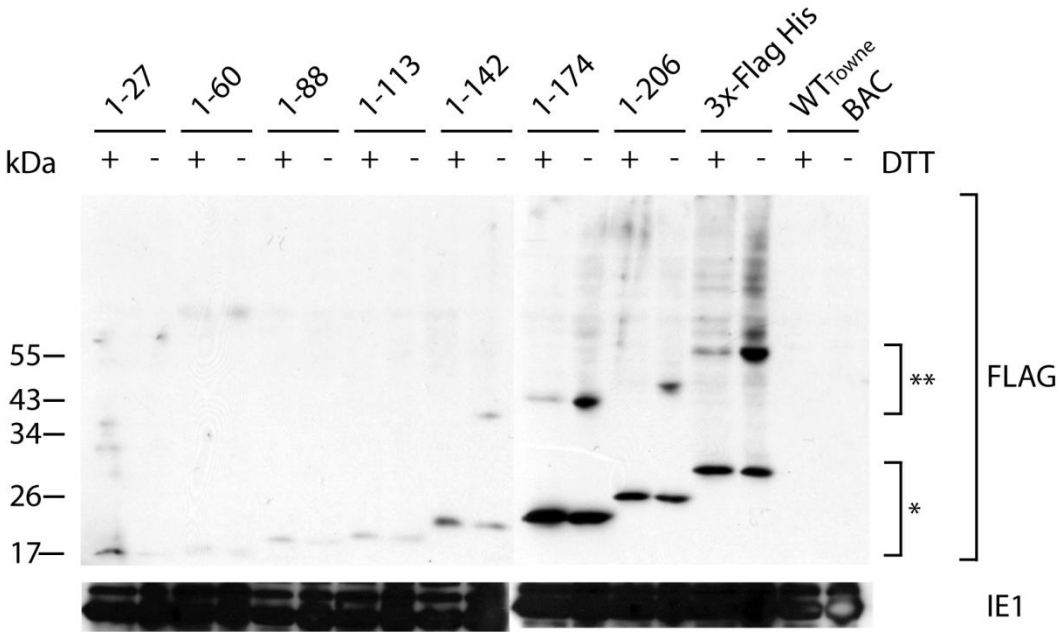


Figure 3.19. **Immunoblot of pUS20 truncated mutants suggests that TM5-TM7 are not necessary for dimerization.**
HFF were infected with recombinant viruses expressing truncated forms of pUS20, with each truncated peptide tagged at the C terminus with a 3xFLAG and poly histidine epitope. In addition, we also infected cells with a virus expressing the full length-tagged US20 (3x-Flag His) or wildtype Towne BAC virus (WTTowne BAC) as controls. Infected cells were harvested at 72hpi, lysates were treated with either 0.1M dithiothreitol(DTT) (reduced) or water (non-reduced) and then incubated at 37°C for 30 minutes. Samples were run on a 10%SDS-PAGE gel and immunblotted with either anti-FLAG antibody or anti-HCMV IE1 antibody. The * indicates the monomer, while the ** represents the dimer.

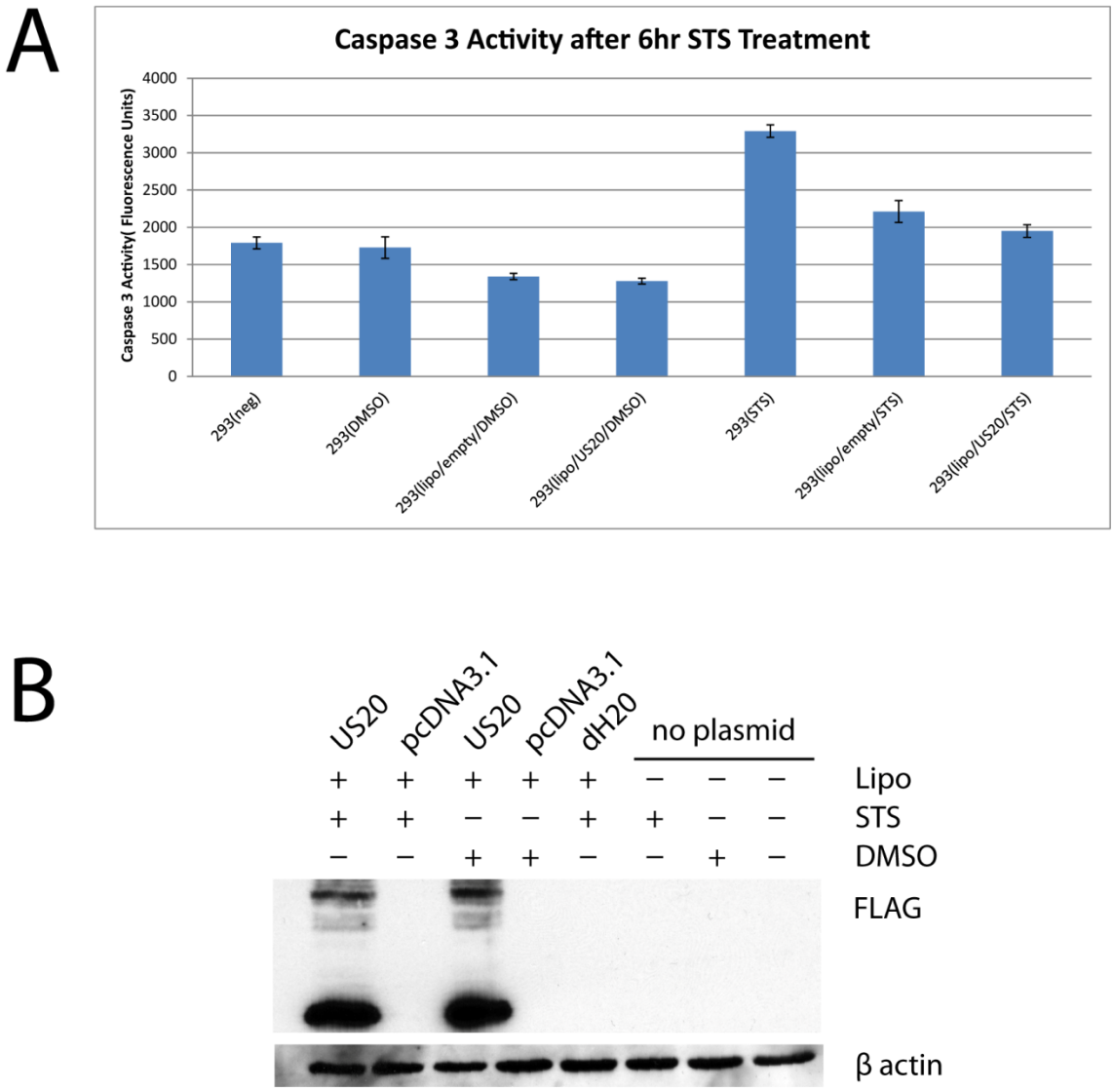


Figure 3.20. **Over expression of US20-FLAG does not change Caspase-3 activity induced by STS treatment.**
(A) 293T cells transiently expressing US20 tagged at the C terminus with FLAG epitopes (US20) were treated with 1uM of STS or mock treated (DMSO) for six hours. Cells were then lysed and Caspase 3 activity was measured using the Caspase-3 Assay Kit(BD Pharmingen). Resulting fluorescence was measured using a luminometer. Experiments were repeated in triplicate. (B) Cellular lysates from figure A were also run on 10% SDS-PAGE gel, transferred and immunoblotted with either anti-FLAG or anti-beta actin.

harvested and then assayed for Caspase-3 activity. As expected, cells that were treated with STS had higher levels of Caspase-3 activity, however, we did not see any difference in levels between samples from cells infected with either US20ko virus or wild-type Towne BAC virus indicating that the deletion had no impact on apoptosis (Figure 3.21). We did notice that uninfected HFFs that were treated with STS had higher levels of Caspase-3 activity than the infected cells, which we believe to be the result of HCMV's termination of cell cycle progression in human fibroblast cells [39]. This termination would prevent cellular replication in infected cells, resulting in fewer cells relative to the uninfected at the 72 hours harvest time point. The results of this experiment suggest that US20 ORF may not play a role in apoptosis despite having the Bax-inhibitor-1 domain.

HCMV pUS20 associates with valosin containing protein (VCP)

To help elucidate the function of US20 ORF, we used immunoprecipitation (IP) to identify proteins that interacted with pUS20. Briefly, we transiently expressed cMyc-US20 in 293T cells and performed an IP using anti-cMyc antibodies and Protein A/G beads. The eluted proteins were separated on a 10% SDS-PAGE gel and then stained with coomassie blue. We identified one band of 25 kDa, two bands of 72 kDa and two bands of 95 kDa in size present in the lane containing lysates from the cMyc-US20 expressing cells but were absent in the negative control (empty plasmid) (Figure 3.22A). The two bands of 95 kDa were excised as one group, the two bands of 72 kDa were also excised as another, and the 25kDa band was excised on its own. After excision from the gel, the bands were submitted for identification by mass spectrometry. We identified the bands to be Sodium Potassium ATPase alpha-1 subunit, Calnexin, transitional endoplasmic reticulum ATPase (VCP), succinate dehydrogenase flavoprotein subunit, and three heat shock proteins (HSP-90, HSP-71, HSP 70) (Figure 3.22B).

As of yet, we have been able to confirm the association of pUS20 and transitional endoplasmic reticulum ATPase (also referred to as VCP) under the context of infection by co-immunoprecipitation (co-IP). Briefly, we infected HFFs at high MOI with the US20-3xF-His or wild-type Towne BAC virus and harvested the lysate after 72 hpi. We incubated the lysate with anti-FLAG antibodies and Protein A/G beads, washed the beads, and eluted the attached proteins. The proteins were then separated on a 10% SDS-PAGE gel and then either immunoblotted with anti-FLAG or anti-VCP. An immunoreactive band at 95 kDa representing VCP was present in both the input lanes and the lane containing the elution from the co-IP of the US20-3xF-His infected cells (Figure 3.23A). A reciprocal co-IP was also performed using anti-VCP and Protein A/G beads, but the presence of a nonspecific band roughly the same size as pUS20 caused the results to be inconclusive (data not shown). In addition to the co-IP, we also found VCP to co-localize with certain aggregates of pUS20 in the context of infection using confocal fluorescence microscopy (Figure 3.23B). These results confirm that pUS20 does associate with VCP under the context of HCMV infection.

Discussion

In this study we extensively profiled the HCMV US20 ORF using a recombinant HCMV virus expressing an epitope tagged US20 protein. We were able to show that US20 ORF encodes a single protein of 28 kDa in size, which we designated pUS20. Furthermore, our kinetic studies found pUS20 to be expressed during the early phase of HCMV infection, which

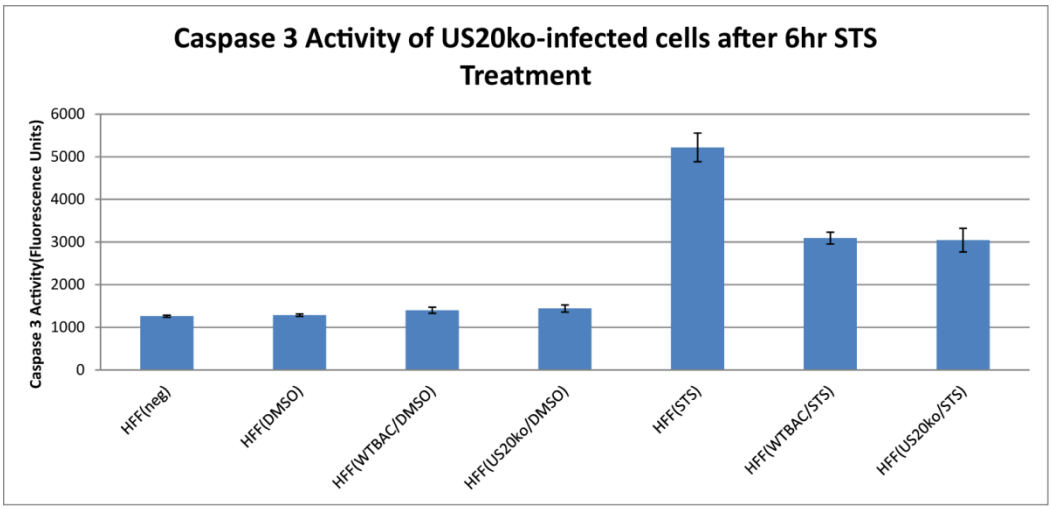
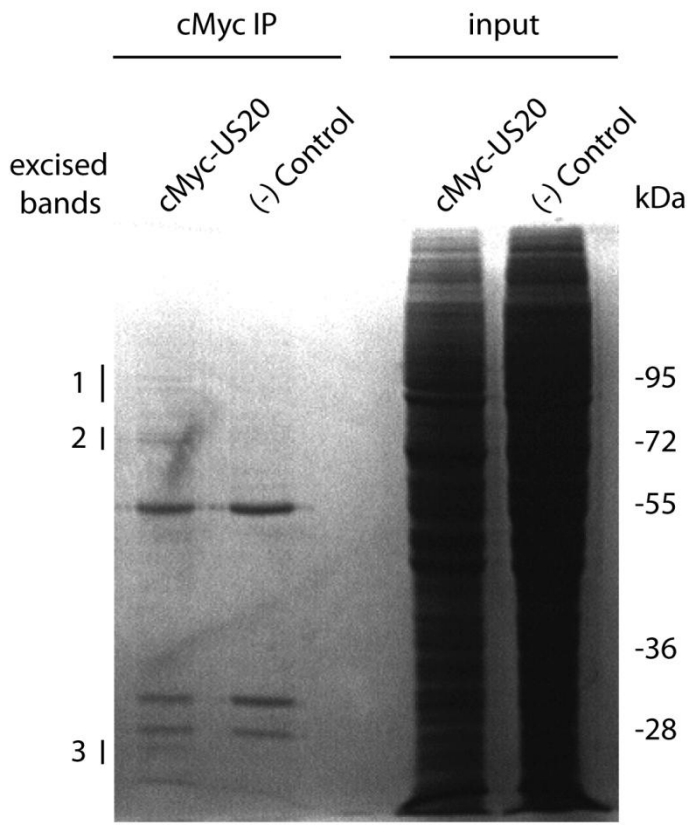


Figure 3.21. **The presence of US20 ORF in HCMV does not impact Caspase-3 activity in infected HFF cells treated with STS.**
HFF cells were infected with either HCMV US20 deletion virus (US20ko) or wild-type Towne BAC (WTBAC) virus. After 72 hours, the infected cells were treated with either 1uM of staurosporine (STS) or mock treated (DMSO) for six hours. Cells were then lysed and Caspase-3 activity was measured using the Caspase-3 Assay kit (BD Pharmingen). Resulting fluorescence was measured using a luminometer. Experiments were repeated in triplicate.

A



B

Excised Bands	Identified Protein	Predicted molecular weight (kDa)	SwissProt Accession #
1	Na(+)/K(+) ATPase alpha-1 subunit	112.9 (Large isoform)/ 74.1 (Small isoform)	P05023
	Calnexin	67.6	P27824
	Transitional endoplasmic reticulum ATPase (VCP)	89.3	P55072
	HSP-90-beta	83.3	P08238
2	Heat shock cognate 71 kDa protein	70.8	P11142
	Heat shock 70 kDa protein 1A/1B	70.1	P08107
	Succinate dehydrogenase [ubiquinone] flavoprotein subunit	72.7	P31040
3	cMyc-US20		

Figure 3.22. **Identifying pUS20's interacting partners using immunoprecipitation.** Protein lysates from 293T cells transiently expressing cMyc-US20 or empty plasmids ((-) control) were immunoprecipiated using anti-cMyc antibodies and Protein A/G beads. Eluted proteins were separated on a 10% SDS-PAGE gel and stained with coomassie blue (A). We identified five distinct bands that were not present in the negative control, which we submitted for identification by mass spectrometry (B).

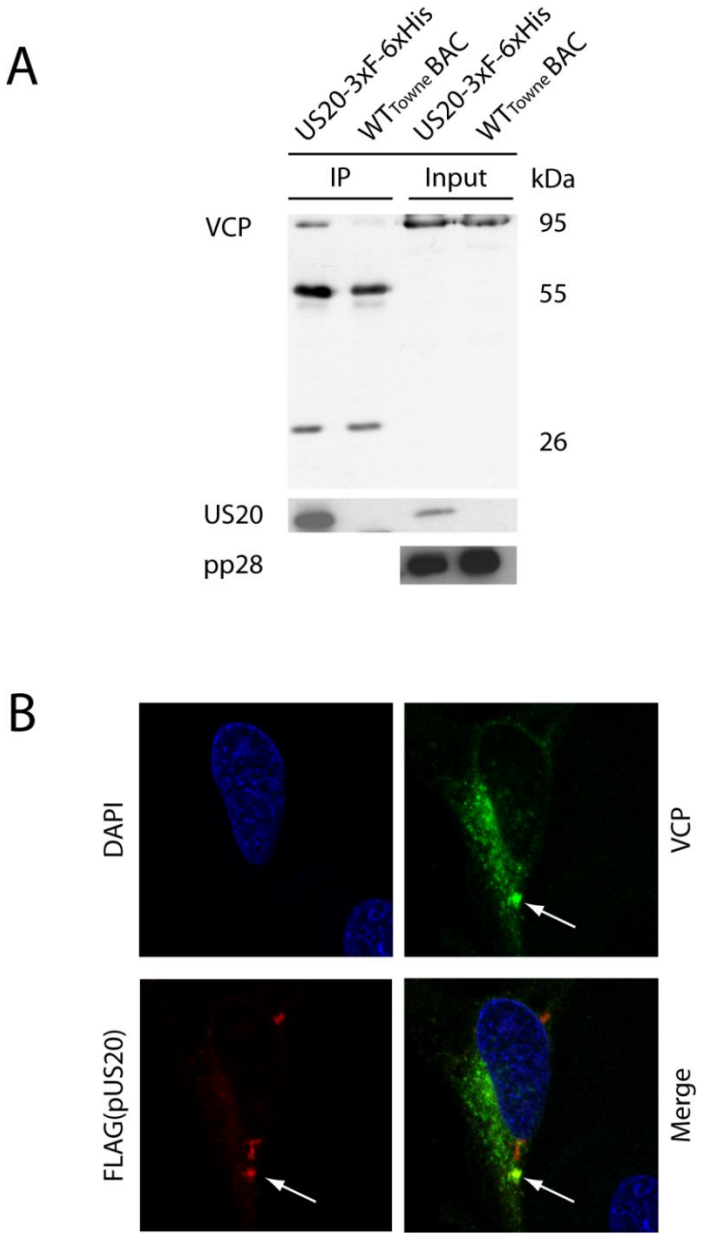


Figure 3.23. **Confirmation of pUS20 and VCP association by co-immunoprecipitation and IFA**.
To confirm the localization of pUS20 and VCP, we performed a co-immunoprecipitation using anti-FLAG and Protein A/G beads on lysates harvested from HFF cells either infected with US20-3xF-6xHis or wild-type Towne BAC (WTTowne BAC). Eluted proteins were separated on a 10% SDS-PAGE gel and immunoblotted using anti-VCP, anti-FLAG, and anti-pp28 (A). The bands present at 26 and 55 kDa represent the light and heavy chain of the anti-FLAG antibody, respectively. An IFA was performed on HFFs infected with US20-3x-6xHis virus for 72 hpi. Cells were fixed with methanol and then stained with anti-VCP, anti-FLAG, and DAPI. Anti-VCP was visualized with secondary anti-rabbit conjugated to AlexaFluor®488, while anti-FLAG was visualized with secondary anti-mouse conjugated to AlexaFluor®647. The arrow indicates co-localization of the two stained proteins.

was consistent with Chambers et al. who also determined the US20 transcript to be expressed during the early phase of replication by DNA microarray [40]. Through immunofluorescence, we found pUS20 to localize to a specific cytoplasmic structures after 36 hpi, which we found to not be a part of the ER or the TGN using Bip and Golgin245 antibodies, respectively, as markers. We did find a small population of cells exhibiting co-localization of pUS20 and EEA1, suggesting that while pUS20 may be present in endosomes, the localization is transient. In addition, we found that the cytoplasmic localization was not dependent on other viral genes expressed during the last phase of HCMV replication.

Through topology prediction software, we identified seven hydrophobic regions, which indicated that pUS20 was a membrane protein with seven transmembrane domains. This prediction was supported by our evidence that tagged pUS20 showed thermal aggregation properties, which is a common phenomenon for heat-denatured membrane proteins [8]. Given the prediction of seven transmembrane domains, we considered that pUS20 would localize to the surface of the cell in a fashion similar to G-protein coupled receptors (GPCRs) [41]. However, our IFA studies involving selective permeabilization of the cellular membrane did not detect pUS20 at the surface of the cell, indicating that pUS20 remains internalized in the cell.

Our growth curve studies of the recombinant virus with the US20 gene knocked out concluded that the gene is not essential for infection and replication in HFF, RPE, HMVEC and differentiated THP-1. Despite the fact that US20 is dispensable for viral growth in the cell lines we studied, we still believe US20 to be essential for *in vivo* infection, since the virus would be under evolutionary pressure to remove genes that did not contribute to infection and replication of its host. An example would be the ORF m155 in murine cytomegalovirus (MCMV). While the m155 knock out virus showed no growth defect in cell culture, a dramatic attenuation of viral virulence and growth was observed in the animal model [42]. These results suggest that we may not be able to observe a replication phenotype of the US20ko virus in a cell culture model and that a more complex model may be necessary to observe the virus' impact on replication and infection.

One property that pUS20 does appear to share with some GPCRs, is its ability to form homo-dimers or homo-oligomers, which we showed through co-immunoprecipitation. While we are still not sure what functional role this property plays in HCMV infection, we can hypothesize based upon existing literature. Homo-dimerization has been shown to allow GPCRs to escape the ER quality-control system [23], so the dimerization of pUS20 may allow the protein to escape degradation. Homo-dimerization has also been found to be important for internalization of receptors [43], which could explain why pUS20 is not expressed on the surface of the cell. In addition to homo-oligomers, GPCRs have also been found to form hetero-dimers and there is evidence to support its role in pharmacological diversity [26] and internalization [44]. Furthermore, BILF1, a GPCR encoded by herpesvirus Epstein-Barr virus (EBV), was shown to hetero-dimerize with a subset of human chemokine receptors, although the consequence of the "hijacking" was not explored [45]. Given that both EBV and CMV are herpesviruses, it is possible that pUS20 may also hetero-dimerize with other seven transmembrane proteins in a similar fashion.

The recombinant viruses expressing truncated forms of pUS20 allowed us to further decipher the impact of the seven predicted transmembrane (TM) domains on the properties we observed. We found that the deletion of the TM5 through TM7 reduced the protein levels of pUS20, with additional TM deletions resulting in further decreases in protein levels. This suggests that TM1 through TM5 may be important for maintaining high protein level. The lack

of difference in US20 RNA transcripts in HFFs infected with either US20(1-60) and US20-3xF-His suggests that the reduced protein level phenotype occurred after transcription. In addition, we found that deleting TM 2 through 7 resulted in a destabilization of pUS20's cytoplasmic localization. Lastly, we determined that the deletion of TM 4 through TM7 resulted in the inability of pUS20 to form a homo-dimer. Given these results, we believe that there may be a correlation between the stability of pUS20 protein levels and its ability to dimerize, which would support the theory that pUS20 needs to dimerize in order to avoid the ER quality-control system. We can also infer from these results that the unique cytoplasmic localization of pUS20 may not relate to its ability to dimerize.

Our initial hypothesis of US20 ORF was that it played a role in apoptosis, given the bax-inhibitor-1 domain that we predicted via BLAST. However, our overexpression of pUS20 in 293T cells was unable to inhibit apoptosis induced by staurosporine and we observed no difference in apoptosis resistance for the HFFs infected with US20ko virus. These results suggest that the US20 ORF does not play a major role in apoptosis.

Through a combination of immunoprecipitation and mass spectrometry, we were able to identify multiple potential proteins that could associate with pUS20. So far, we have been able to confirm the association of Valosin Containing Protein(VCP) aka Translational Endoplasmic Reticulum ATPase with pUS20. VCP is a ubiquitously expressed protein belonging to the AAA+ (ATPase associated with various activities) protein family, that has been implicated in many cellular processes including cell cycle regulation, nuclear envelope formation, Golgi biogenesis, autophagy, and the ubiquitin proteasome system [46]. In addition to cellular process, VCP has also been shown to play a role in viral infection, specifically the HBx protein of Hepatitis B [47]. The protein has also been found to be associated with neurodegenerative disease, which links VCP to brain function [48]. Given the abnormalities in brain development of fetuses with congenital HCMV infection [49], the interaction of pUS20 and VCP could be a factor in HCMV infection of the brain.

While the associations have not yet been confirmed by co-immunoprecipitation, the pull down of Sodium Potassium ATPase alpha-1 subunit and succinate dehydrogenase flavoprotein presents new possible hypotheses for the function of US20. Sodium Potassium ATPase alpha-1 subunit is a component of the membrane-associated enzyme that catalyzes the transfer of Sodium and Potassium ions across the surface of the plasma membrane [50]. Previous studies of HCMV in fibroblasts have shown that HCMV infection causes an increase in host cell plasma membrane potential and this may be the result of viral-stimulation of Na+/K+ ATPase [51]. Succinate dehydrogenase flavoprotein (SDHA) is a key enzyme in the tricarboxylic acid cycle (TCA) and can also contribute to the electron transport chain [52]. Previous studies have shown that HCMV infection increases the transcript levels of SDHA, along with other metabolic components in an effort to provide sufficient resources for viral replication [53]. The potential association of pUS20 with these two proteins suggests that US20 may modify these host proteins to benefit the virus during infection.

In this study we have examined the innate characteristics of the US20 ORF and its protein product. The identification of pUS20 and VCP's interaction is also exciting since it may potentially be a mechanism explaining how HCMV causes brain damage in infants. Despite the fact that the function of US20 ORF still remains uncertain at this point, this study provides a foundation and direction for future studies to further explain US20's role in HCMV infection.

Materials and Methods

Cells and media

Human primary foreskin fibroblasts (HFF) (CC-2509, Clonetics, San Diego, CA) and human embryonic kidney 293 cells (HEK293T) (ATCC, Manassas, VA) were propagated using Dulbecco's Modified Eagle Medium (D-MEM) (Invitrogen, Carlsbad, CA) supplemented with 10% FBS, 1% Penstrep, and 0.2% Fungizone. Retinal pigment epithelials (RPE) (C4000-1, Clontech, Mountain View, CA) that were immortalized with human telomerase reverse transcriptase (hTERT) were propagated with DMEM/F-12 media (Invitrogen, Carlsbad, CA) supplemented with 10% FBS, 17.3mL Sodium Bicarbonate, 1% Penstrep, 0.2%Fungizone, and 5mL L-glutamine. Human microvascular endothelial cells (HMVEC) (CC-2527, Clonetics, San Diego, CA) were propagated with the EGM-2 MV BulletKit (Lonza, Basel, Switzerland). Human acute monocytic leukemia cell line (THP-1) (ATCC, Manassas, VA) were propagated in RPMI Medium 1640 (Invitrogen, Carlsbad, CA) supplemented with 10% FBS, 1mM Sodium Pyruvate, 0.05mM β-Mercaptoethanol, 10mM HEPES, 1% Penstrep, and 0.2% Fungizone.

Plasmids

The US20-FLAG plasmid was constructed using the pcDNA3.1(+) plasmid (Invitrogen, Carlsbad, CA) as the backbone. Using the primers US20-2xFLAG-KPN1up and 2xFLAG-NOT1dn, we PCR amplified US20 and its C terminus FLAG affinity tag using the US20-2xF-PA BAC, which we previously constructed, as the template. During the amplification, we also inserted a KPN1 cut site upstream of the cloned US20 and three UAA stop codons and a NOT1 cut site downstream. Both the PCR product and pcDNA3.1(+) were digested with KPN1 and NOT1 restriction enzymes (New England Biolabs, Ipswich, MA), separated on 10% agarose gel, gel excised and purified using a gel extraction kit (Qiagen, Valencia, CA). The digested plasmid and PCR product were then ligated together using ligase (Invitrogen, Carlsbad, CA) and then transfected into chemically competent *E. Coli* Top10 cells (Invitrogen, Carlsbad, CA). Cells were allowed to recover for one hour at 37°C and then plated on LB plates supplemented with ampicillin antibiotics. Colonies were picked after 16 hours, propagated and the purified DNA was submitted for sequencing at the UC Berkeley DNA Sequencing Facility (Berkeley, CA). The cMyc-US20 plasmid was constructed in a similar fashion using the pCMV-Myc plasmid (Clontech, Mountain View, CA) as the backbone and the PCR primers, cMyc-US20-Sal1up and cMyc-US20-Kpn1dn.

Antibodies

Anti-FLAG antibodies, used for identifying pUS20 expressed from US20-FLAG plasmid, US20-2xF-PA virus and US20-3xF-His virus, were purchased from Sigma Aldrich (Saint Louis, MO). Antibodies against EEA1 (made in rabbit), Golgin45 (made in rabbit), Bip/GRP78 (made in rabbit), and β-actin (made in mouse) were purchased from Santa Cruz Biotechnologies, Inc. (Santa Cruz, CA). HCMV Immediate Early 1 (IE1 aka IE72) antibody (made in mouse) was purchased from Millipore, Inc. (Billerica, MA). HCMV pp28 antibody (made in mouse) was purchased from Virusys Corporation (Taneytown, MD). C-Myc antibody (made in mouse) was purchased from Clontech (Mountain View, CA). Anti-VCP (made in rabbit) was purchased from Cell Signaling Technology ® (Beverly, MA). Secondary antibodies for immunoblotting

(peroxidase anti-mouse IgG) and immunofluorescence (fluorescein anti-mouse IgG, texas red anti-mouse IgG, texas red anti-rabbit IgG) were purchased from Vector Laboratories (Burlingame, CA). Lastly, secondary antibodies for immunofluorescence, anti-mouse Alexa Fluor® 647 and anti-rabbit Alexa Fluor® 488 were purchased from Invitrogen (Carlsbad, CA).

Bioinformatic Analysis

The sequence of the US20 ORF in the Towne strain was obtained from our own sequencing data of the wild-type Towne BAC [4]. The predicted peptide sequence of pUS20 was obtained by using the translation software in Lasergene EditSeq (DNASTAR, Madison, WI). To obtain prediction topology data of pUS20, we submitted the peptide sequence to several online algorithms listed on the ExPASy website. Prediction of domains in pUS20 was performed by submitting the peptide sequence to the BLAST website and using the protein blast program (http://blast.ncbi.nlm.nih.gov/Blast.cgi). Sequence similarity analysis of pUS20 homologs was performed by obtaining the US20 DNA sequences of different viral strains from the nucleotide database at NCBI (http://www.ncbi.nlm.nih.gov/genbank/). The *in silico* translated sequences were then aligned and compared using Lasergene MegAlign (DNASTAR, Madison, WI).

Construction and Propagation of Recombinant HCMV virus

The US20-3xF-His virus, along with the recombinant viruses expressing truncated forms of pUS20, was constructed by allele exchange mutagenesis [4]. Using a pUCK19 plasmid as our template which was modified in our lab to be a tandem affinity tag containing: a 3xFlag epitope, a tobacco etch virus cleavage site, and a poly histidine epitope, we generated the targeting cassette by PCR. Using primers 5-13, listed in the supplementary primer table, and the template, we PCR amplified tag cassettes containing the tandem affinity tag and a kanamycin resistance gene with 70-75 nucleotide arms that were homologous to the sequences flanking the insertion site. The resulting PCR products were then run on a 1% agarose gel and the 1.6 kilobase band was gel purified using a gel extraction kit (Qiagen, Valencia, CA). The purified PCR product was then electroporated into an electroporation competent *E. Coli* EL350 strain, which harbored the wild-type Towne BAC. The electroporated bacteria were incubated in a shaker at 30^0C to allow for recovery and lambda red recombination to occur and then plated on a Luria Bertani (LB) agar plate supplemented with 12.5µg/mL Chloramphenicol and 100µg/mL Kanamycin. The bacteria were then returned to the 30^0C incubator for 48 hours and positive clones were identified by colony PCR using primers 14 and 15 that flanked the predicted US20 ORF. A successful recombination event was determined by a band shift of 1.5kb when the PCR products were run on a 1.5% gel. To obtain sufficient amount of the recombinant BAC, we grew up large cultures and purified the BACs using the Nucleobond® PC500 kit (Clontech, Mountain View, CA). PCR amplicons of the insertion site in the recombinant BACs were submitted to UC Berkeley DNA Sequencing Facility (Berkeley, CA) to confirm insertion was in frame. Purified recombinant BACs was then electroporated into human foreskin fibroblasts and the cells were then seeded in a T-25 Flask. The DMEM media was changed the day after electroporation and the cells were incubated at 37^0C and 5% CO_2 for 14 to 21 days till 100% cytopathic effect (CPE) was evident. CPE was determined by plaque formation and green fluorescent protein (GFP) visualization using a Nikon Eclipse TE300 Fluorescence microscope (Nikon, Tokyo, Japan).

The construction of the US20ko virus was performed in a manner similar to the approach taken above. We used primers 16 and 17 and the pZeo plasmid as a template to generate our deletion cassette which contained the Zeocin resistance gene flanked by arms homology targeting the US20 ORF. Steps identical to the ones explained above were taken to generate the recombinant BAC with the replacement of US20 ORF with the Zeocin resistance gene, except for the substitution of 50 µg/mL of Zeocin for kanamycin antibiotics.

All viruses generated from the recombinant BACs were harvested by scraping the flasks with a plastic scraper. The infectious inoculum was then frozen in liquid nitrogen and thawed in a 37°C water bath. This freeze-thaw process was repeated three times and the infectious inoculum was then centrifuged for 30 minutes at 4000RPM, 4°C to remove cellular debris. The supernatant containing the infectious inoculum were then aliquot and stored at -80°C.

Restriction Digest Profile and Southern Blot

The recombinant HCMV BAC DNA was harvested using lysing reagents from the QIAprep® Spin Miniprep Kit (Qiagen, Valencia, CA) and a modified protocol. The harvested BACs were digested with HindIII restriction enzyme (New England Biolabs, Ipswich, MA) for 16 hours at 37^0C. The digested products were then separated on a 0.8% agarose gel overnight at 4^0C. The resulting gel containing the restriction digest profile of the recombinant BACs was then placed into a capillary transfer apparatus overnight to allow for the DNA to transfer from the gel to the GeneScreen Plus® Hybridization Transfer Membrane (Perkin Elmer, Wellesley, MA). The membrane was then quickly rinsed in distilled water and the DNA was fixed by UV-crosslinking using a UV Stratalinker® 1800 (Agilent Technologies, Santa Clara, CA).

The membrane was then incubated with prehybridization solution (20xSSC/Formamide/ Non-Fat Dried Milk) containing salmon sperm DNA (ssDNA) for three hours at 65^0C. The template for the radioactive probe was prepared by PCR amplifying the Zeocin Resistance Gene from the modified pZeo plasmid, separating the PCR product from the plasmid template by gel electrophoresis and then excising the ~500 bp fragment from the agarose gel using the QIAquick® Gel Extraction Kit (Qiagen, Valencia, CA). Using the excised fragment, we generated the radioactive probe by "random primed" DNA labeling using the Random Primed DNA Labeling Kit (Roche, Mannheim, Germany) and alpha- ^{32}P dCTPs nucleotides (Perkin Elmer, Wellesley, MA). The membrane was then incubated with fresh hybridization solution, ssDNA, and the radiolabelled probe overnight at 65^0C. Following the overnight probing, the membrane was washed with 20xSSC-based solutions of various concentrations, exposed, and visualized using the phosophoimager, Storm 840 (GE Healthcare, Waukesha, WI).

Thermal Aggregation

Protein lysates from either HFFs infected with US20-3xF-His after 72 hours or 293T cells transiently expressing either cMyc-US20 or US20-FLAG were harvested using M-PER lysing reagent (Thermo-scientific, Rockford, IL). Loading dye supplemented with 2-mercaptoethanol was then added to the samples. Thermal aggregation was induced by boiling the samples at >95°C for 10 minutes, while the control samples were incubated at 37°C for 30 minutes. The samples were then separated on a 5% stacking and 10% resolving SDS-PAGE gel and the proteins were electrotransferred to a nitrocellulose membrane (BioRad, Hercules,CA). Membranes were blocked with 5% Nonfat Dry Milk (NFDM) for two hours and then stained

with either anti-FLAG or anti-cMyc antibodies to probe for pUS20 for two hours. Anti-β-actin antibody was also added to measure protein loading levels. Following the primary antibody incubation, the membranes were washed three times with PBS supplemented with 0.05% Tween20 and then stained with peroxidase horse anti-mouse IgG secondary antibody for two hours. Membranes were again washed three times with PBS supplemented with 0.05% Tween20 and then treated with Western Lightning® plus-ECL (Perkin Elmer, Wellesley, MA). The treated membranes were then exposed to film.

Immunofluorescence Assay

HFFs, RPEs, and HMVECs were seeded on glass coverslips that were placed in a 12-well plate at $1.5x10^5$ cells/well and allowed to grow overnight at 37^0C and 5% CO_2. The cells were then infected with US20-2xF-PA at a multiplicity of infection (MOI) of 1 and the virus was allowed to adsorb for two hours. After which, the infectious inoculum was removed and replaced with supplemented DMEM. The infected cells were then incubated for 72 hours at 37^0C and 5% CO_2, washed with phosphate buffered saline (PBS) and fixed with methanol for one hour at -20^0C. After fixation, the cells were treated with 1% bovine serum albumin (BSA) in PBS for 30 minutes and then stained with anti-FLAG for one hour at 37^0C. Next, the cells were washed three times with PBS and stained with anti-mouse secondary antibody conjugated to Fluorescein (Vector Labs, Burlingame, CA) for one hour at 37^0C. The cells were the washed again three times with PBS and further stained with DAPI (Invitrogen, Carlsbad, CA) for five minutes at 37^0C. The coverslips were then mounted on glass slides using mounting solution and sealed with nail polish. The coverslips were imaged using a Eclipse TE300 Fluorescence microscope (Nikon, Tokyo, Japan) at 1000x magnification. Images were processed and merged using Adobe Photoshop.

For the immunofluorescence studies involving the secretory pathway, HFFs were infected with US20-3xF-His virus and fixed with methanol. Cells were then incubated with primary antibodies against FLAG, Bip, EEA1, and Golgin245, followed by secondary antibodies, anti-mouse Alexa Fluor® 647 and anti-rabbit Alexa Fluor® 488. Visualization and co-localization studies were carried out at the CNR Biological Imaging Facility (Berkeley, CA) using the LSM710 Confocal Microscope (Nikon, Tokyo, Japan).

Protein Kinetic Studies

For the immunoblotting component of the study, HFFs were infected with the US20-2xF-PA virus and cells were harvested at 0, 4, 8, 12, 24, and 72 hours post infection. The lysates were harvested and sample dye supplemented with β-mercaptoethanol was added. The samples were then separated on a 10% SDS-PAGE gel and then electrotransferred to a nitrocellulose membrane. Western blotting using anti-FLAG antibodies to visualize pUS20 was performed as describe above. For the immunofluorescence (IFA) part of the study, HFFs infected with US20-2xF-PA were fixed with methanol at 0, 4, 12, 24, 36, and 72 hpi. The IFA was competed as previously described.

For the phosphonoacetic acid (PAA) study, HFFs were pretreated with 500 µg/mL of PAA for an hour prior to infection and the infected with US20-3xF-His or wild-type Towne BAC virus. After two hours, the media was replaced with fresh DMEM supplemented with PAA. At 72 hpi, the cells were fixed with methanol and then probed with either primary antibody anti-

FLAG or anti-pp28. Secondary anti-mouse conjugated to fluorescein was added for visualization.

Selective permeabilization immunofluorescence assay

HFFs were infected with either US20-2xF-PA or UL123-2xF-PA, which were constructed from a previous study. After 72 hours, all cells were fixed with 4% paraformaldehyde and then either treated with 0.5% Triton-X (permeabilized) or PBS (not-permeabilized). Primary anti-FLAG antibody was added followed by visualization with Histostain SP kit (Invitrogen, Carlsbad,CA) and streptavidin-TRITC (Vector Labs, Burlingame, CA).

Northern Blot

Human foreskin fibroblasts (HFF) were infected with either recombinant US20 knock out or wildtype Towne BAC virus at a MOI greater than 1, and then harvested after 72hpi. Total RNAs were isolated from the cells using Trizol (Invitrogen, Carlsbad, CA) following the manufacturer's recommendation and then separated on a 1% agarose gel supplemented with 37% formaldehyde, 20X MOPS and ethidium bromide. Stability of the RNA was confirmed by the presence of the ribosomal RNA. Using a capillary transfer apparatus, we transferred the RNA from the gel to the GeneScreen Plus® Hybridization Transfer Membrane (Perkin Elmer, Wellesley, MA). Using primers 18 and 19, we generated a 300 nucleotide PCR product that was complementary to the 3' end of the predicted US20. Using the PCR product, we generated a radioactive probe by "random primed" DNA labeling using the Random Primed DNA Labeling Kit (Roche, Mannheim, Germany) and alpha- ^{32}P dCTPs nucleotides (Perkin Elmer, Wellesley, MA). The membrane was initially blocked with prehybridization solution containing ssDNA and then incubated with fresh hybridization solution, ssDNA, and the radiolabelled probe overnight at 65^0C. Following the overnight probing, the membrane was washed with SSC-based solutions of various concentrations, exposed, and visualized using the Storm 840 phosphorimager (GE Healthcare, Waukesha, WI).

Plaque-based growth assays

HFF, RPE, and HMVEC were infected at a low MOI with either recombinant US20 knock out or wild-type Towne BAC derived virus. THP-1 cells were pretreated with 12-O-tetradecanoylphorbol-13-acetate (TPA) for either 1, 2, or 7 days prior to infection with the viruses. Viral progeny were harvested from infected cells at various time points starting from 0 up to 21 days post infection by either physical scraping and repeated freeze-thawing cycles or harvesting the supernatant.

The harvested infectious lysate was titrated by first serially diluting the lysates in 10 fold increments and then adding the dilutions to HFFs. The virus was allowed to adsorb to the cells for two hours and then an agarose overlay, consisting of a 1:1 mixture of 2% Type VII agarose (Sigma-Aldrich, St. Louis, MO) and 2x DMEM, was added over the cells. The cells were stored in a 37°C incubator at 5% CO_2 for 14 days, after which, the number of infectious particles was determined by counting the number of plaques present in the well using a Nikon Eclipse TE300 Fluorescence microscope (Nikon, Tokyo, Japan).

Quantification of protein expression

HFF were infected with recombinant viruses expressing truncated forms of pUS20 at an MOI of 1 for 72 hours. Loading dye supplemented with DTT was added to the protein lysates harvested from the cells. SDS-PAGE and Western blot techniques were used as described above using anti-FLAG and anti-IE72 antibodies. The resulting film was scanned and the bands were quantified using ImageQuant™ TL software (GE Healthcare Life Sciences, Piscataway, NJ). Arbitrary intensity values were then obtained for all bands. Values for bands reactive to anti-FLAG were normalized to its corresponding band reactive to anti-IE72. The data was then presented as fold differences relative to the full length tagged pUS20 (US20-3xF-His).

Qualitative real-time PCR

RNA from HFF cells infected with recombinant viruses for 72 hpi was harvested using Trizol (Invitrogen, Carlsbad, CA) following the manufacturer's recommendation. Reverse transcriptase PCR was performed on the harvested RNA to obtain the cDNA. qRT-PCR was performed on the cDNA using primers 20 and 21, which amplified the 3xF-His epitope tag, to quantify the US20 transcript. We also used primers 22 and 23 to quantify IE1 to be used as a loading control. The resulting Ct values were first normalized to IE1 and then normalized to the Ct values of the US20-3xF-His virus.

Dimerization and co-immunoprecipitation

For all studies examining dimerization under the context of infection, we harvested the HFF cells after 72 hours. Protein sample buffer was added to protein lysates supplemented with either DTT or water. The samples were then incubated at 37°C for 30 minutes and then separated on a 10% SDS-PAGE gel. A western blot was then performed using anti-FLAG or anti-IE1.

For studies examining transient expressing of pUS20, 293T cells were first seeded on a 12 well plate at $1.5x10^5$ cells/well. After 12 hours, the cells were transfected with 2µg of either cMyc-US20 or US20-FLAG plasmid using 2 uL of Lipofectamine 2000 (Invitrogen, Carlsbad, CA). The protein lysates were then harvested after 48 hours and also selectively incubated with sample buffer containing DTT. SDS-PAGE and western blotting were then performed using anti-FLAG and anti-cMyc antibodies. β-actin immunoblotting was also used as a loading control.

For the co-immunoprecipitation study, we co-transfected 293T cells using Lipofectamine 2000 with 1µg of cMyc-US20 and 16 µg of US20-FLAG plasmid. Cells were lysed with M-PER supplemented with Complete Protease Inhibitor Cocktail (Roche, Basel, Switzerland) for 30 minutes on ice. Non-soluble debris was pelleted by centrifugation at 14,000 RPM for 30 minutes at 4°C and the supernatant was transferred to a fresh tube. The supernatant was then pre-cleared with Protein A/G PLUS-agarose beads (Santa Cruz Biotechnology, Inc, Santa Cruz, CA) for one hour. After centrifugation and transfer of supernatant, the pre-cleared lysate was then incubated with anti-FLAG antibodies for two hours in a tumbler at 4°C. Protein A/G PLUS-agarose beads were then added to the supernatant and returned to the tumbler for one hour at 4°C. The beads were then washed four times with pre-chilled PBS and then transferred to a spin column (Thermo Fisher Scientific, Rockford, IL) The proteins were then eluted with 150 µg/uL 3xFLAG peptides (Sigma Aldrich, St. Louis, MO). Sample buffer with no reducing agent was added to the eluted proteins. After a 30 minute incubation at 37°C, the proteins were then

separated on a 10% SDS-PAGE gel, transferred to a membrane and then immunoblotted using anti-FLAG or anti-cMyc antibodies.

Caspase 3 Assay

293T cells were seeded at $1.5x10^5$ in a 12 well plate and then transfected with US20-FLAG plasmid. After 48 hours, the cells were treated with 1µM Staurosporine (STS) (EMD, Gibbstown, NJ) for six hours. The cells were then washed with PBS and then lysed using the Cell Lysis Buffer provided in the Caspase-3 Assay Kit (BD Biosciences, Franklin Lakes, NJ). Capase-3 activity levels were then quantified by following the manufacturer's protocol and then detecting the fluorescence at the 420 nm wavelength using a Spectra Max® M2 plate reader (Molecular Devices, Sunnyvale, CA).

For the experiment involving recombinant viruses, we infected HFF with either the US20ko or the wild-type Towne BAC virus at a MOI of 1. After 72 hours, we then treated the infected cells with 1µM STS. Caspase 3 activity was measured using the approach described above.

Immunoprecipitation and mass spectrometry

For the immunoprecipitation, we seeded a T-75 with $8x10^6$ cells and then transfected the cells with 24µg of cMyc-US20 plasmid with 60uL of Lipofectamine 2000. After 48 hours, the cells were washed with PBS and removed from the surface of the flask with a cell scrapper. The cells were then pelleted by centrifuging at 300xg for five minutes and lysed on ice for 30 minutes using M-PER supplemented with Complete Protease Inhibitor Cocktail. The sample was centrifuged at 14,000 rpm for 30 minutes at 4°C to pellet the insoluble debris. The supernatant was then pre-cleared with Protein A/G PLUS-agarose beads for 30 minutes and then briefly centrifuged to pellet the beads. The supernatant was then incubated with anti-cMyc antibodies for two hours in a tumbler at 4°C followed by an one hour incubation after adding Protein A/G PLUS-agarose beads. The beads were then washed four times with ice-cold PBS and then transferred to a spin column. Protein sample buffer supplemented with 0.1M DTT was added to the column and then incubated for 30 minutes at 37°C. The eluted proteins were collected by centrifugation and then separated on a 10% SDS-PAGE gel. The gel was then stained with coomassie blue for 30 minutes and then destained overnight using an acetic acid methanol solution. Bands were visualized and submitted to the HHMI Mass Spectrometry Laboratory (Berkeley, CA) for processing and identification by mass spectrometry.

For the co-immunoprecipitation study, $3.0 x10^5$ HFFs were infected with US20-3xF-His virus. After 72 hours, the cells were harvested and lysed using the approach describe above. We performed an IP using anti-FLAG antibodies and the eluted proteins were separated on a 10% SDS-PAGE gel. The proteins were then transferred to a nitrocellulose membrane and then immunoblotted with either anti-FLAG or anti-VCP antibodies.

Acknowledgements

We would like to thank David King and Sharleen Zhou of the HHMI Mass Spectrometry Laboratory for their assistance with the identification of the proteins from the immunoprecipitation of pUS20. We would also like to thank Steven Ruzin and Denise

Schichnes of the CNR Biological Imaging Facility for their help with the confocal microscopy work.

References

1. Edward S. Mocarksi, T. S., Robert F. Pass (2007). "Cytomegalovirus". In *Fields Virology* (P. M. H. David M. Knipe, Ed.), pp. 2701-2772. Lippincott Williams & Wilkins, New York, NY, USA.
2. Weston, K., and Barrell, B. G. (1986). Sequence of the short unique region, short repeats, and part of the long repeats of human cytomegalovirus. *Journal of molecular biology* 192: 177-208.
3. Lesniewski, M., Das, S., Skomorovska-Prokvolit, Y., Wang, F. Z., and Pellett, P. E. (2006). Primate cytomegalovirus US12 gene family: a distinct and diverse clade of seven-transmembrane proteins. *Virology* 354: 286-298.
4. Dunn, W., et al. (2003). Functional profiling of a human cytomegalovirus genome. *Proceedings of the National Academy of Sciences of the United States of America* 100: 14223-14228.
5. Yu, D., Silva, M. C., and Shenk, T. (2003). Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 12396-12401.
6. Varnum, S. M., et al. (2004). Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *Journal of virology* 78: 10960-10966.
7. Gerna, G., Percivalle, E., Sarasini, A., Baldanti, F., and Revello, M. G. (2002). The attenuated Towne strain of human cytomegalovirus may revert to both endothelial cell tropism and leuko- (neutrophil- and monocyte-) tropism in vitro. *The Journal of general virology* 83: 1993-2000.
8. Okada, N., et al. (2011). Identification of TMEM45B as a protein clearly showing thermal aggregation in SDS-PAGE gels and dissection of its amino acid sequence responsible for this aggregation. *Protein expression and purification* 77: 118-123.
9. Das, S., and Pellett, P. E. (2007). Members of the HCMV US12 family of predicted heptaspanning membrane proteins have unique intracellular distributions, including association with the cytoplasmic virion assembly complex. *Virology* 361: 263-273.
10. Stinski, M. F. (1978). Sequence of protein synthesis in cells infected by human cytomegalovirus: early and late virus-induced polypeptides. *Journal of virology* 26: 686-701.
11. Kerry, J. A., et al. (1997). Translational regulation of the human cytomegalovirus pp28 (UL99) late gene. *Journal of virology* 71: 981-987.
12. Martinez, J., Lahijani, R. S., and St Jeor, S. C. (1989). Analysis of a region of the human cytomegalovirus (AD169) genome coding for a 25-kilodalton virion protein. *Journal of virology* 63: 233-241.
13. Dorsam, R. T., and Gutkind, J. S. (2007). G-protein-coupled receptors and cancer. *Nature reviews* 7: 79-94.
14. Tan, S., Tan, H. T., and Chung, M. C. (2008). Membrane proteins and membrane proteomics. *Proteomics* 8: 3924-3932.
15. Hoffmeister, H., Gallagher, A. R., Rascle, A., and Witzgall, R. (2011). The human polycystin-2 protein represents an integral membrane protein with six membrane-spanning domains and intracellular N- and C-termini. *The Biochemical journal* 433: 285-294.

16. Wang, J., Chu, B. B., Ge, L., Li, B. L., Yan, Y., and Song, B. L. (2009). Membrane topology of human NPC1L1, a key protein in enterohepatic cholesterol absorption. *Journal of lipid research* 50: 1653-1662.
17. Ahn, J. H., and Hayward, G. S. (1997). The major immediate-early proteins IE1 and IE2 of human cytomegalovirus colocalize with and disrupt PML-associated nuclear bodies at very early times in infected permissive cells. *Journal of virology* 71: 4599-4613.
18. Detrick, B., Rhame, J., Wang, Y., Nagineni, C. N., and Hooks, J. J. (1996). Cytomegalovirus replication in human retinal pigment epithelial cells. Altered expression of viral early proteins. *Investigative ophthalmology & visual science* 37: 814-825.
19. Miceli, M. V., Newsome, D. A., Novak, L. C., and Beuerman, R. W. (1989). Cytomegalovirus replication in cultured human retinal pigment epithelial cells. *Current eye research* 8: 835-839.
20. Turtinen, L. W., and Seufzer, B. J. (1994). Selective permissiveness of TPA differentiated THP-1 myelomonocytic cells for human cytomegalovirus strains AD169 and Towne. *Microbial pathogenesis* 16: 373-378.
21. Lee, C. H., Lee, G. C., Chan, Y. J., Chiou, C. J., Ahn, J. H., and Hayward, G. S. (1999). Factors affecting human cytomegalovirus gene expression in human monocyte cell lines. *Molecules and cells* 9: 37-44.
22. McCormick, A. L., Roback, L., Livingston-Rosanoff, D., and St Clair, C. (2010). The human cytomegalovirus UL36 gene controls caspase-dependent and -independent cell death programs activated by infection of monocytes differentiating to macrophages. *Journal of virology* 84: 5108-5123.
23. Terrillon, S., and Bouvier, M. (2004). Roles of G-protein-coupled receptor dimerization. *EMBO reports* 5: 30-34.
24. Margeta-Mitrovic, M., Jan, Y. N., and Jan, L. Y. (2000). A trafficking checkpoint controls GABA(B) receptor heterodimerization. *Neuron* 27: 97-106.
25. Roess, D. A., and Smith, S. M. (2003). Self-association and raft localization of functional luteinizing hormone receptors. *Biology of reproduction* 69: 1765-1770.
26. Galvez, T., et al. (2001). Allosteric interactions between GB1 and GB2 subunits are required for optimal GABA(B) receptor function. *The EMBO journal* 20: 2152-2159.
27. Rocheville, M., Lange, D. C., Kumar, U., Sasi, R., Patel, R. C., and Patel, Y. C. (2000). Subtypes of the somatostatin receptor assemble as functional homo- and heterodimers. *The Journal of biological chemistry* 275: 7862-7869.
28. Salim, K., et al. (2002). Oligomerization of G-protein-coupled receptors shown by selective co-immunoprecipitation. *The Journal of biological chemistry* 277: 15482-15485.
29. Persani, L., Calebiro, D., and Bonomi, M. (2007). Technology Insight: modern methods to monitor protein-protein interactions reveal functional TSH receptor oligomerization. *Nature clinical practice* 3: 180-190.
30. Xu, Q., and Reed, J. C. (1998). Bax inhibitor-1, a mammalian apoptosis suppressor identified by functional screening in yeast. *Molecular cell* 1: 337-346.
31. Perfettini, J. L., Reed, J. C., Israel, N., Martinou, J. C., Dautry-Varsat, A., and Ojcius, D. M. (2002). Role of Bcl-2 family members in caspase-independent apoptosis during Chlamydia infection. *Infection and immunity* 70: 55-61.
32. Kawai, M., Pan, L., Reed, J. C., and Uchimiya, H. (1999). Evolutionally conserved plant homologue of the Bax inhibitor-1 (BI-1) gene capable of suppressing Bax-induced cell death in yeast(1). *FEBS letters* 464: 143-147.

33. Henke, N., Lisak, D. A., Schneider, L., Habicht, J., Pergande, M., and Methner, A. (2011). The ancient cell death suppressor BAX inhibitor-1. *Cell calcium* 50: 251-260.
34. Tamaoki, T., and Nakano, H. (1990). Potent and specific inhibitors of protein kinase C of microbial origin. *Bio/technology (Nature Publishing Company)* 8: 732-735.
35. Morales, A. P., et al. (2011). Endoplasmic reticulum calcium release engages Bax translocation in cortical astrocytes. *Neurochemical research* 36: 829-838.
36. Zhang, X. D., Gillespie, S. K., and Hersey, P. (2004). Staurosporine induces apoptosis of melanoma by both caspase-dependent and -independent apoptotic pathways. *Molecular cancer therapeutics* 3: 187-197.
37. Hemrajani, C., Berger, C. N., Robinson, K. S., Marches, O., Mousnier, A., and Frankel, G. (2009). NleH effectors interact with Bax inhibitor-1 to block apoptosis during enteropathogenic Escherichia coli infection. *Proceedings of the National Academy of Sciences of the United States of America* 107: 3129-3134.
38. Belmokhtar, C. A., Hillion, J., and Segal-Bendirdjian, E. (2001). Staurosporine induces apoptosis through both caspase-dependent and caspase-independent mechanisms. *Oncogene* 20: 3354-3362.
39. Song, Y. J., and Stinski, M. F. (2005). Inhibition of cell division by the human cytomegalovirus IE86 protein: role of the p53 pathway or cyclin-dependent kinase 1/cyclin B1. *Journal of virology* 79: 2597-2603.
40. Chambers, J., et al. (1999). DNA microarrays of the complex human cytomegalovirus genome: profiling kinetic class with drug sensitivity of viral gene expression. *Journal of virology* 73: 5757-5766.
41. Jean-Alphonse, F., and Hanyaloglu, A. C. (2011). Regulation of GPCR signal networks via membrane trafficking. *Molecular and cellular endocrinology* 331: 205-214.
42. Abenes, G., et al. (2004). Murine cytomegalovirus with a transposon insertional mutation at open reading frame m155 is deficient in growth and virulence in mice. *Journal of virology* 78: 6891-6899.
43. Cvejic, S., and Devi, L. A. (1997). Dimerization of the delta opioid receptor: implication for a role in receptor internalization. *The Journal of biological chemistry* 272: 26959-26964.
44. Xu, J., He, J., Castleberry, A. M., Balasubramanian, S., Lau, A. G., and Hall, R. A. (2003). Heterodimerization of alpha 2A- and beta 1-adrenergic receptors. *The Journal of biological chemistry* 278: 10770-10777.
45. Vischer, H. F., Nijmeijer, S., Smit, M. J., and Leurs, R. (2008). Viral hijacking of human receptors through heterodimerization. *Biochemical and biophysical research communications* 377: 93-97.
46. Ju, J. S., et al. (2009). Valosin-containing protein (VCP) is required for autophagy and is disrupted in VCP disease. *The Journal of cell biology* 187: 875-888.
47. Jiao, B. Y., Lin, W. S., She, F. F., Chen, W. N., and Lin, X. (2011). Hepatitis B virus X protein enhances activation of nuclear factor kappaB through interaction with valosin-containing protein. *Archives of virology*.
48. Watts, G. D., et al. (2004). Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein. *Nature genetics* 36: 377-381.
49. Cinque, P., Marenzi, R., and Ceresa, D. (1997). Cytomegalovirus infections of the nervous system. *Intervirology* 40: 85-97.

50. Canfield, V. A., Norbeck, L., and Levenson, R. (1996). Localization of cytoplasmic and extracellular domains of Na,K-ATPase by epitope tag insertion. *Biochemistry* 35: 14165-14172.
51. Rugolo, M., Baldassarri, B., and Landini, M. P. (1986). Changes in membrane permeability in cells infected by human cytomegalovirus. *Archives of virology* 89: 203-212.
52. Rutter, J., Winge, D. R., and Schiffman, J. D. (2010). Succinate dehydrogenase - Assembly, regulation and role in human disease. *Mitochondrion* 10: 393-401.
53. Munger, J., Bajad, S. U., Coller, H. A., Shenk, T., and Rabinowitz, J. D. (2006). Dynamics of the cellular metabolome during human cytomegalovirus infection. *PLoS pathogens* 2: e132.

Appendix 1. Primer Table

Primer #	Primer Name	Sequence
1	US20-2xFLAG-KPN1up	CGGGGTACCGAAatgcaggcgcaggaggctaacgcgctgctgctctcccgcatggaggctctcgagtggttcaaaaagttca
2	2xFLAG-NOT1dn	ATAAGAATGCGGCCGCTTATTATTACTTGTCATCGTCGTCCTTGTAGTCCATCTTGTCATCGTCGTCCTTGTAGTCCATCCG
3	cMyc-US20-Sal1up	TATAGTCGACCcaggcgcaggaggctaacgc
4	cMyc-US20-Kpn1dn	TATAGGTACCttaggacttccccgtcgtac
5	US20upt3xflaghis75	CGTTGGATTAGTCTTTCGGACGGCGTGCCTTTGGACAACGGGACTTTGACAGCCGCCAGTACGACGGGGAAGTCCGACTACAAAGACCATGACGG
6	US20dntflaghis75	CATTCTAGCGGGACGACATGAAGCATGGCGACAAGCGCGGCTGCTGTGAAAACGGGCGCGGTTTTATAGGCATTACATATGAATATCCTCCTTAGTT
7	US20upt_1-27	gcgcaggaggctaacgcgctgctgctctcccgcatggaggctctcgagtggttcaaaaagttcaccgtatggctgGACTACAAAGACCATGACGG
8	US20upt_1-60	cagctggctttcagcttcggtttgggaagcgtttttggttggggttcccacaaaaccgcaacttttgcgtcgagGACTACAAAGACCATGACGG
9	US20upt_1-88	ttctttctcaccgtgctcgtgcccatcgtctgcatgttcatcacgtacacgttgggcaacgaacaccctagtaacGACTACAAAGACCATGACGG
10	US20upt_1-113	gccacggtgctttcatctatctgttggccaacagcctgacggcggccatcttccagatgtgctctgaaagccgcGACTACAAAGACCATGACGG
11	US20upt_1-142	tcctacgtgatgaccctggcgttgtttatctcctttacggggctggcgtttctaggtggccgtgaccgacgtcgcGACTACAAAGACCATGACGG
12	US20upt_1-174	TACGTGGTGATGCTGCTTTCGTTCCTCACGCTCGCTCTGCTAAGCGACGCCGATTGGCTGCAGAAGATAGTGGTGGACTACAAAGACCATGACGG
13	US20upt_1-206	AGCTTCTTTTTGGGTATTCTGGCCTACGACAGTCTCATGGTCATCTTTTTCTGCCCACCTAACCAATGCATCCGTGACTACAAAGACCATGACGG
14	US20checkup	GGCGGTAAGTGAGATCCTGGTGATGATAGCCGGTTTGGCCGTCATCGTGA
15	US20checkdn	GCATGGCGACAAGCGCGGCTGCTGTGAAAACGGGCGCGGTTTTATAGGCA
16	US20ko-up	AGAGAAGGGTAGGTGCGCCGCAGCGGCTTTGTGCCGAGACCGTCGCCACCTAACTTCGTATAATGTATGCTATACGAAGTTATCAAGTTTCGAGGTCGAGTGTCAGTC
17	US20ko-dn	GCATGGCGACAAGCGCGGCTGCTGTGAAAACGGGCGCGGTTTTATAGGCATAACTTCGTATAATGTATGCTATACGAAGTTATGGAACGGACCGTGTTGACAATTAAT
18	US20northup300	tcgttcctcacgctcgctct
19	US20northdn	ttaggacttccccgtcgtac
20	RTPCR-US20up	TACTACAAAGACCATGACGG
21	RTPCR-US20dn	AAGCAGCTCCAGCCTACACA
22	RTPCR-IE1up	cgagtgaccgaggattgcaa
23	RTPCR-IE2dn	caccatgtccactcgaacctt

Chapter 4

Predicted SUMOylation site of HCMV ORF UL44 is not essential for viral replication

Abstract

SUMOylation is a recently discovered post-translation modification that has been shown to be utilized by viruses to regulate protein function. Two proteins of Human Cytomegalovirus (HCMV), IE1 and IE2, have already been identified to be targets of SUMOylation. Initial studies by our collaborators were able to identify an interaction between the SUMOylation enzyme, UBC9, and UL44 by yeast two-hybrid and co-immunoprecipitation studies, leading to the hypothesis that UL44 is a SUMOylated. In this study, we used a recombinant HCMV virus containing a lysine to alanine mutation at the predicted SUMOylation ψKxE motif near the C terminus of UL44 to study the importance of the motif in the context of infection. We found that the mutation of the SUMOylation target site in UL44 did not affect DNA replication or viral replication in human foreskin fibroblasts. We were also able to observe through radiolabeling and immunoprecipitation experiments that the mutation caused a loss in expression of a high molecular weight form of UL44. These results indicate that the predicted SUMOylation site in UL44 is not essential for viral replication in tissue culture.

Introduction

Viral proteins expressed by Human Cytomegalovirus (HCMV) can undergo various types of post-translational modifications, examples include the glycosylation of viral glycoproteins [1], phosphorylation of pUL69 [2], myristoylation of pp28 [3]. SUMOylation is a post-translational modification in which an 11kDa SUMO moiety is covalently attached to a lysine residue in the consensus sequence, ψKxE (where ψ corresponds to a large hydrophobic amino acid, K is a lysine residue, x is any amino acid and E is a glutamic acid residue), of the target protein [4]. There are a wide range of functions that have been found to be regulated by the SUMOylation system, including cell cycle regulation, transcription, cellular localization, degradation, and chromatin organization [5]. Given that viruses can utilize or hijack cellular host systems to aid in propagation, it was not surprising when, two years after the initial discovery of SUMO in 1997, HCMV immediate-early 1 protein (IE1) was discovered to be SUMOylated [6]. Subsequent work by various groups showed that , while SUMOylation was not essential for IE1 function, the SUMOylated form of IE1 enhanced the expression of immediate early 2 protein (IE2) [7], which was also found to be a target of sumoylation by Lee and Ahn [8]. In addition to HCMV, SUMO has been found to play a role the pathogenesis of various human viral pathogens including herpes simplex, papillomavirus, HIV, SARS, and poxvirus vaccinia [4].

HCMV open reading frame (ORF) UL44 is a processivity factor, a component of the virally encoded dimeric DNA polymerase, that has been well established to be essential for viral replication *in vitro* [1, 9]. Previous studies of UL44 established that UL44 forms a C clamp holodimer structure that wraps around DNA, similar to PCNA and Herpes Simplex Virus UL42 [10]. While the N terminus has been found to be sufficient for all of the biochemical functions of UL44 in vitro, including dimerization, interaction with UL54, binding to dsDNA and stimulation of long-chain DNA synthesis [1], recombinant HCMV genomes lacking the C terminus residues of UL44 were found to inhibit viral replication [11]. This suggests that the C terminus may have some form of functional domain.

Preliminary studies from a collaborating lab identified, through yeast two-hybrid screening, an interaction between UL44 and SUMO-conjugating enzyme UBC9. This interaction was then validated by overexpression and co-immunoprecipitation experiments. Furthermore, using SUMOylation prediction algorithms, we were able to identify a ψKxE motif near the end of the C terminus of UL44, providing additional evidence that UL44 can be SUMOylated (Figure 4.1). In this study we found that the mutation of the lysine residue of the ψKxE motif in UL44, which is the substrate used by UBC9 to covalently link SUMO to the protein [12], was not essential for DNA replication or viral fitness in human foreskin fibroblasts.

Results

Recombinant HCMV with UL44K410A mutation does not significantly impact viral replication.

Using the SUMOsp 2.0 Online Server Program (http://sumosp.biocuckoo.org/online.php), which predicts SUMOylation sites based upon sequence, we identified a single ψKxE motif starting at amino acid 409 of UL44 (Figure 4.1**)**. To determine the relevance of this predicted motif for viral replication, we constructed the recombinant virus, UL44K410A, containing a double point mutation that changed the codon AAA at nucleotide position 1228 to GCA of UL44 (Figure 4.2A**)**. This mutation resulted in the substitution mutation of the lysine residue at

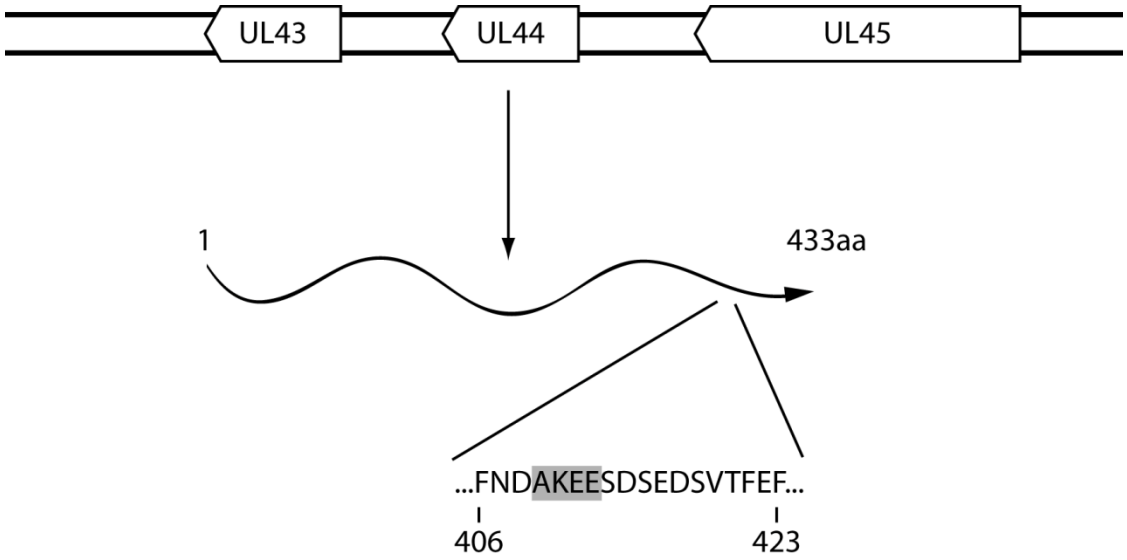


Figure 4.1. **Map of UL44 and predicted SUMOylation site**.
HCMV open reading frame (ORF) UL44 is located on the complementary strand between UL43 and UL45. The 1302 bp gene expresses a peptide of 433 amino acids in length, which contains a predicted ψKxE motif (shaded in grey) near the C terminus end the of the peptide, starting at amino acid 409.

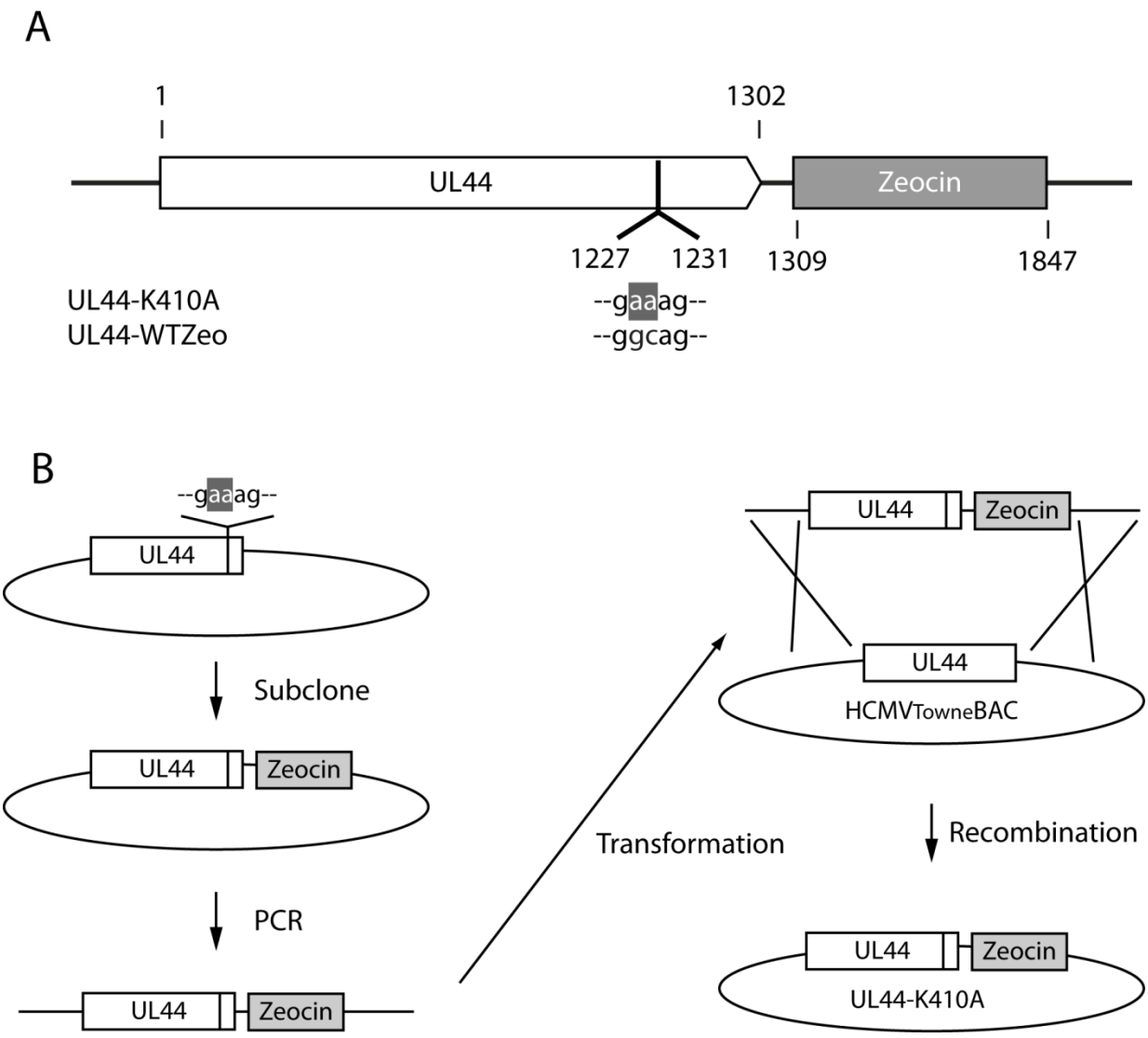


Figure 4.2. **Construction of UL44-K410A recombinant virus**.
(A) The UL44-K410A virus contained a point mutation, illustrated in grey, resulting in the replacement of the lysine residue at position 410 with an alanine. Both the UL44-K410A and the UL44-WTZeo virus had a Zeocin Resistance Gene (ZeoR) inserted downstream of the UL44 ORF in the viral genome. (B) The mutant UL44 ORF was subcloned upstream of a ZeoR gene of a pZeo plasmid. A target cassette consisting of both the mutant UL44 ORF and ZeoR flanked by arms of homology was constructed by PCR. The cassette was transformed into E. Coli containing the wild-type Towne BAC, where by homologous recombination generated the UL44-K410A construct.

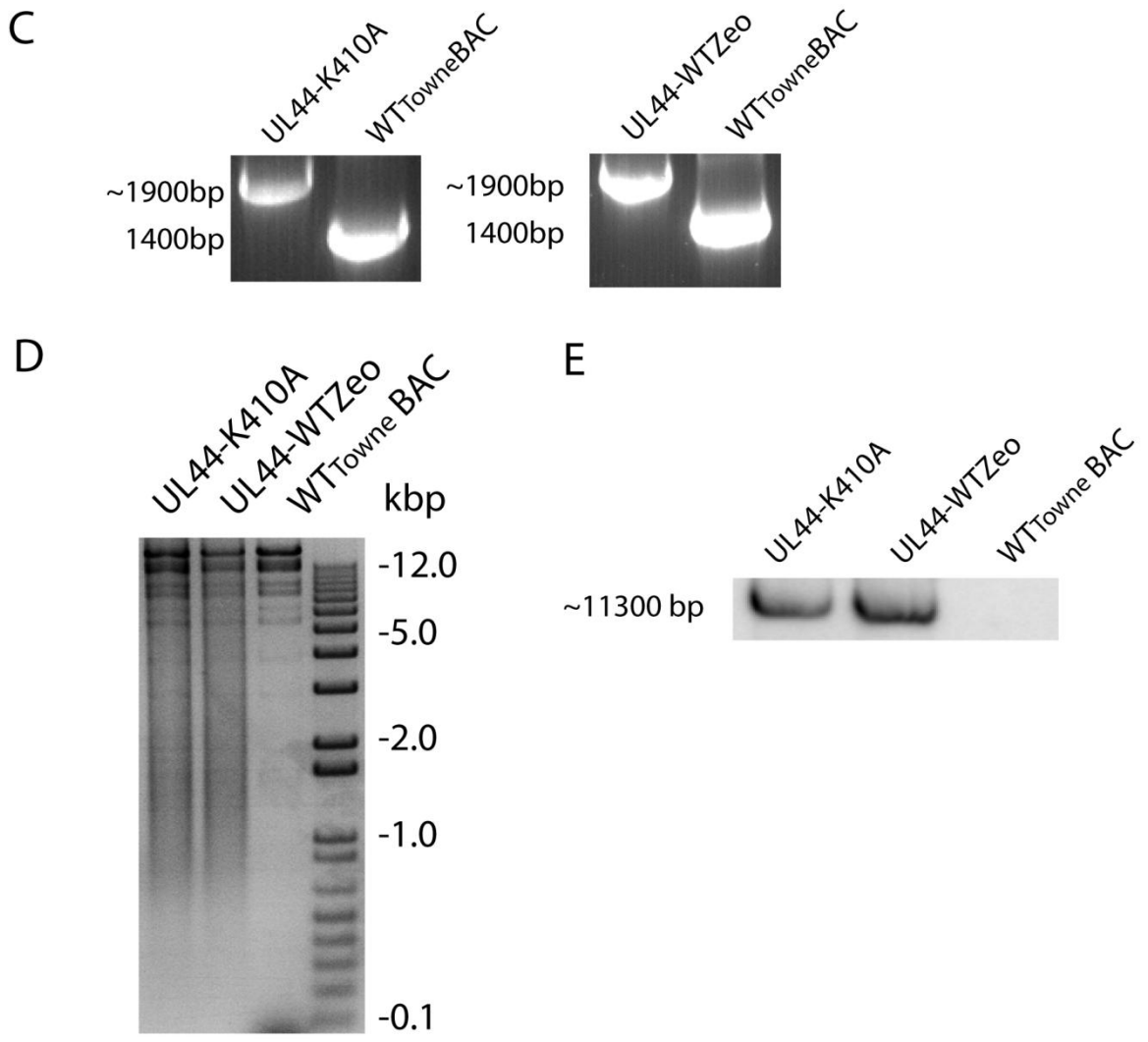


Figure 4.2. **Construction of UL44-K410A recombinant virus (continued)**.
(C) PCR of the UL44 ORF in the viruses UL44-K410A, UL44-WTZeo and wild-type Towne BAC(WTTowneBAC) using primers flanking the ORF. (D) The constructs were digested with HindIII restriction enzyme and separated on a 1% agarose gel. (E) The digested DNA in the gel was then transferred to a membrane and probed with a radiolabelled probe specific to the Zeocin Resistance Gene.

position 410 of the ψKxE motif to an alanine residue. This mutation has been shown by previous groups to successfully prevent SUMOylation of the target protein at that specific site [13, 14].

In addition to the UL44-K410A virus, we also constructed the control recombinant virus, UL44-WTZeo in which no mutation was present at the ^{410}lysine, but contained the Zeocin Antibiotic Resistance (ZeoR) gene in order to control for any phenotypes caused by the additional 7 nucleotides and the antibiotic resistance gene that were inserted into the genome of both the UL44-K410A and UL44-WTZeo virus. Briefly, using a plasmid constructed by our collaborators that contained a cloned UL44 with the point mutations, we subcloned the mutant UL44 ORF into a pZeo plasmid with a ZeoR gene seven nucleotides downstream of the cloning site. The mutant UL44 ORF and ZeoR insertion cassette was constructed with flanking arms of homology by PCR and the product was then transformed into an EL350 *E.Coli* strain harboring the wild-type Towne BAC. After recombination and selection, virus was then generated from the recombinant BAC (Figure 4.2B**).**

To confirm and validate the ET recombination between the cassette and the BAC, the recombinant BACs were checked by PCR using primers flanking the UL44 ORF. The shift in size from 1.4 kilobase pairs (kbp) to 1.9 kbp corresponded with the addition of the 500 nucleotide in length ZeoR gene, thus indicating that the cassette inserted into the correct location (Figure 4.2C). Furthermore, we digested the recombinant BACs with HindIII restriction enzyme and the resulting profiles of constructs UL44-K410A and UL44-WTZeo were identical to the wild-type Towne BAC, showing that no unwanted recombination occurred (Figure 4.2D). To confirm that the cassette only inserted once into the genome, a southern blot was performed on the restriction digest profile using a nucleotide probe specific for the ZeoR gene. The results identified a single band of 11.3 kbp in length, which corresponded to the expected size of the HindIII digested fragment containing the UL44 ORF (Figure 4.2E). Lastly, the mutant BACs were sequenced to confirm that the point mutations were present in the genome.

A multi-step, low growth curve study was performed using the UL44-K410A, UL44-WTZeo, and wild-type Towne BAC to examine if the predicted SUMOylation site at ^{410}lysine impacted viral replication in primary human foreskin fibroblasts (HFF) cells (Figure 4.3). We infected the HFFs at a MOI of 0.1 and harvested the cells at 0, 1, 3, 7, 10, and 14 days post infection (dpi). While all three viruses had similar titers at three dpi, both the UL44-K410A and UL44-WTZeo virus had titers 10 times greater than the wild-type Towne BAC virus at 10 dpi. We also observed that the UL44-K410A virus had titers 2 times greater than the UL-44WTZeo. The results from this growth curve study indicated that the ^{410}lysine site is not essential for viral replication in HFFs.

Mutation K410A does not affect DNA replication in tissue culture

Since UL44 has been established to be a processivity factor, we sought to examine if the predicted SUMOylation site at ^{410}lysine played a role in DNA replication. Using DNA harvested 3 dpi from HFFs infected with either UL44-K410A, UL44-WTZeo, or wild-type Towne BAC, we performed a real-time quantitative PCR (qPCR) experiment to quantitatively measure the viral genome copies using primers specific to the Major Immediate Early Protein (MIEP). The MIEP gene has been used by other several groups as a means to quantify HCMV genome copy number [15]. Genome copy numbers of UL44-K410A were found to be similar to levels of the

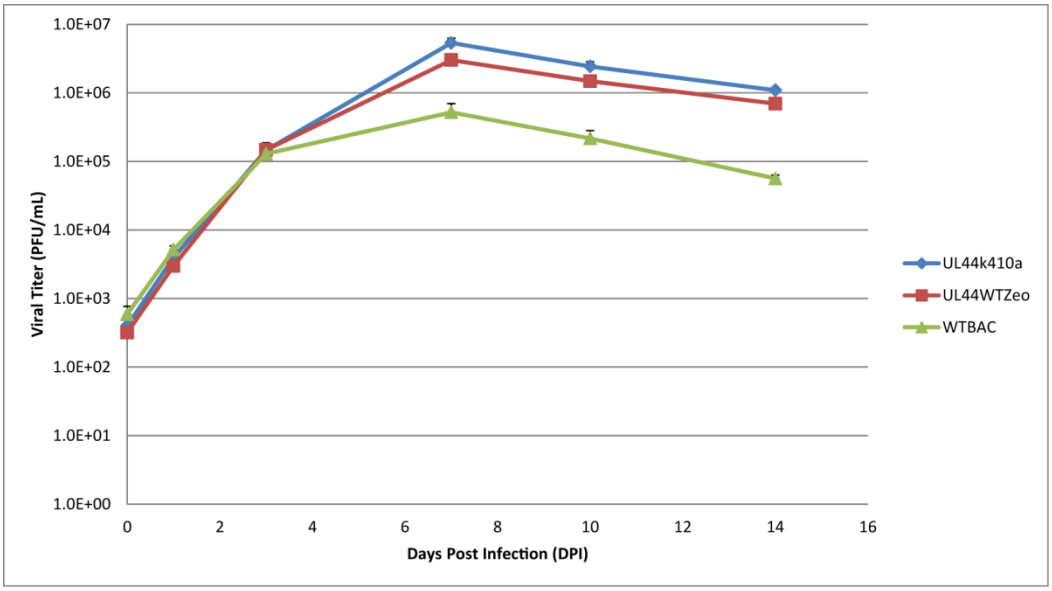


Figure 4.3. **Growth Curve of UL44-K410A recombinant virus in HFFs**.
HFFs were infected at a multiplicity of infection (MOI) of 0.1 with UL44-K410A, UL44-WTZeo and wild-type Towne BAC virus (WTBAC). Infected cells were harvested at 0, 1, 3, 7, 10 and 14 days post infection (dpi) and titered on HFF cells. GFP expressing plaques were counted by eye after 14 days using a fluorescence microscope. The experiment was performed in triplicate.

wild-type Towne BAC (Figure 4.4), thus the predicted SUMOylation site at ^{410}lysine does not impact DNA replication efficiency of HCMV in tissue culture.

UL44 K410A mutation weakens expression of a potential high molecular weight form of UL44

Given that SUMOylation increases the molecular weight of the target protein by 11 kDa or more, depending on the number of SUMO moieties attached, we hypothesized that the mutation of the lysine residue would prevent SUMOylation at that site resulting in a loss of a high molecular weight form of UL44. To determine if this was the case, we performed an immunoprecipitation(IP) on lysates labeled with 35S Methionine from 72 hours infected HFFs using anti-UL44 antibodies. We found three distinct bands that had a relative estimated molecular weight of 50, 60, and 70 kilodaltons (kDa) for both the wild-type Towne BAC and UL44-WTZeo virus (Figure 4.5**)**. However, only the 50 and the 60 kDa bands were present in the UL44-K410A lane. We found that the level of expression of the 50 kDa form to be the same for all three viruses and the molecular weight correlated with the literature [16]. This result suggests the possibility that both the 60 and 70 kDa bands were higher molecular weight forms of UL44 and the deletion of SUMOylation site may have inhibited the expression of the 70kDa form.

Discussion

In this study, we found that the mutation of the ^{410}lysine residue of the predicted ψKxE motif to be not essential for HCMV replication in HFFs in tissue culture. This result, coupled with our collaborators' data of the interaction between UL44 and UBC9, suggests that SUMOylation at ^{410}lysine is not essential for HCMV DNA replication or virion production in tissue culture. These results suggest that SUMOylation of UL44 may only occur under specific conditions that were not present in the HFF infection model that we used for this study. It is also a possibility that SUMOylation and its downstream effects may be viral strain specific, which was observed for HCMV Immediate Early 2 IE2, a transactivator gene essential for viral replication in tissue culture [9, 17]. A study performed by Lee and Ahn found that mutations of the SUMOylation sites of IE2 did not have an effect on the replication of the HCMV Towne strain in fibroblasts[8], while mutations of the same SUMOylation sites were found by Berndt et al. to impair viral replication of the AD169 and VR1814 strains[18]. Thus, it is possible that the lack of a replication defect by our UL44 mutant is specific to the Towne Strain.

Our immunoprecipitation using anti-UL44 antibodies identified three distinct bands for the wild-type Towne BAC virus, which we hypothesized to be various forms of UL44 that may have been post-translational modified. Some evidence to support this hypothesis was seen by Strang et al., who was able to identify a high molecular weight form of UL44 by IP and mass spectrometry [16]. In our studies, we observed that the K410A mutation of UL44 caused a reduction in expression of the largest band in the lane, suggesting that the largest band was the SUMOylated form of UL44. However, without having done additional experiments, we cannot rule out the possibility that the largest band is not UL44, but rather a completely different protein.

The function of SUMOylation in the context of HCMV infection is still unknown. Studies of SUMOylation of HCMV IE1 and IE2 indicated that SUMO may serve as a method to enhance transactivation of genes [7, 19], in an effort to increase viral replication. The

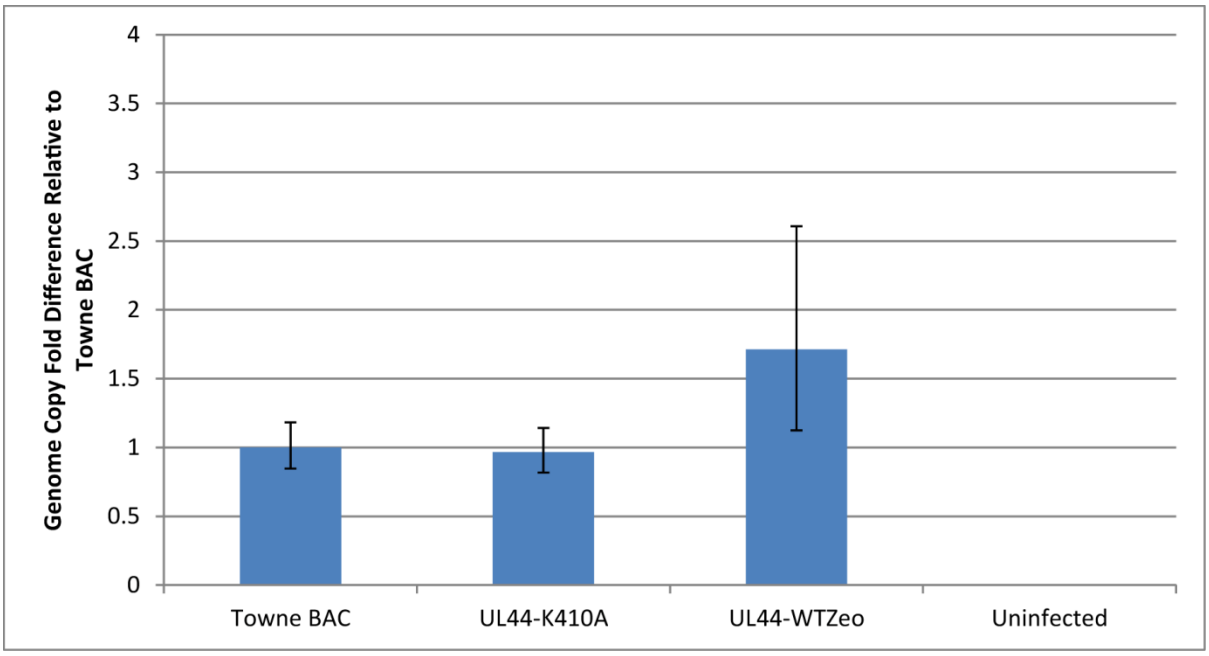


Figure 4.4. **Real Time Quantitative PCR of MIEP gene of UL44-K410A recombinant virus at 72 hours post infection**.
HFFs were infected with recombinant (UL44-K410A and UL44-WTZeo) or wild-type Towne BAC virus at a MOI of 0.05. Real time quantitative PCR (qPCR), using primers for the major immediate early protein (MIEP) gene, was used to determine the relative genome copy number from DNA samples harvested from the infected cells at 72 hpi. Results were normalized against GAPDH, the internal control, analyzed by the ΔΔCT method, and presented as fold change relative to the wild-type Towne BAC. The results presented are an average of quadruplicate experiments.

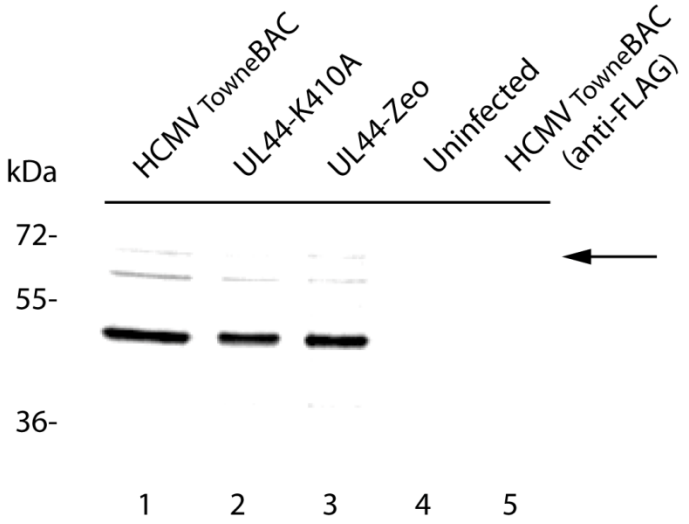


Figure 4.5. **Immunoprecipitation of UL44 in cells labeled with 35S-Methionine**. HFFs were infected with either recombinant (UL44-K410A, UL44-WTZeo), wild-type Towne BAC virus (HCMVTowneBAC), or mock infected (Uninfected). Infected cells were then labeled with 35S Methionine and harvested after 72hpi. IP was performed using anti-ICP36(UL44) antibodies or anti-FLAG as a control and IP'd proteins were run on a 10% SDS-PAGE gel. Gel was then fixed, dried, and then imaged with a phosphoimager. Arrow indicates a band present in lanes 1 and 3, but not as prevelant in lane 2.

SUMOylation of UL44 may also be a mechanism to enhance viral DNA replication, although this regulation may only be apparent in an *in-vivo* setting.

Materials and Methods

Cells and media

Human primary foreskin fibroblasts (HFF) (CC-2509, Clonetics, San Diego, CA) were propagated using Dulbecco's Modified Eagle Medium (D-MEM) (Invitrogen, Carlsbad, CA) supplemented with 10% FBS, 1% Penstrep, and 0.2% Fungizone.

Plasmids

The pCMV plasmid expressing the cloned UL44 with the K410A mutation was a gift from a collaborating lab (Ao Shen).

Construction and propagation of recombinant virus

To construct the HCMV Towne UL44-K410A mutant virus (designated UL44-K410A) and the control virus, HCMV Towne wild-type with Zeocin insertion (designated UL44-WTZeo), we first subcloned the entire UL44 mutant ORF, containing the point mutation or the wild-type UL44 into a pZeo plasmid, respectively. Using the pCMV-HA-UL44-K410A plasmid and the wild-type Towne BAC as the template, we amplified their respective UL44 ORFs and attached on flanking SphI and BamHI cut site, by PCR using primers, 44k410aSphIup and 44k10aBamH1dn . The PCR products were gel purified, digested with SphI and BamHI (New England Biolabs, Ipswich, MA), and ligated into the pZeo plasmid, which was also digested with the same restriction enzymes. The ligated products were then transformed into chemically competent *E. Coli* Top10 cells (Invitrogen, Carlsbad, CA) and plated on LSLB agar plates for Zeocin antibiotic selection (Invitrogen). The Top10 cells containing either the plasmid with UL44K410A (pZeo-UL44K410A) or UL44WT (pZeo-UL44WT) were propagated and the plasmids were purified and sequenced to confirm the ligation. Using the pZeo-UL44K410A as the template, we generated an insertion cassette by PCR using the following primers: 44k410aZeoup and 44k410aZeodn. The resulting PCR product contained the entire UL44ORF and downstream Zeocin resistance gene flanked by two targeting arms that had 46 nucleotides of homology to the nucleotides immediately flanking the UL44 ORF in the wild-type Towne BAC.

This insertion cassette was gel purified and then electroporated into *E. Coli* EL350 containing the wild-type Towne BAC for lambda RED recombination between the cassette and BAC. The electroporated cells were then plated on LSLB plates supplemented with chloramphenicol and zeocin antibiotics. To confirm the deletion of the wildtype UL44 and replacement with the insertion cassette, we performed colony PCR using primers: UL44up2 and UL44dn3. The PCR product was also sequenced to confirm that the point mutation was maintained in the mutant BAC, designated UL44K410A-BAC. The process was repeated using the pZeo-UL44WT as the template to generate the UL44WTZeo-BAC. To confirm that no undesired recombination occurred, we digested both mutant BACs and the wildtype $HCMV_{Towne}$ BAC with HindIII and ran the products on a 1% agarose gel to generate a restriction digest profile. To confirm that our insertion cassettes only inserted a single time into the BAC, the

restriction digest profile was transferred to a GeneScreenPlus® Hybridization Transfer Membrane (Perkin Elmer) and then subjected to a southern blot using a radiolabelled P^{32} probe with homology to the Zeocin resistance gene cassette. The UL44-K410A and the UL44-WTZeo mutant viruses were generated by electroporating the UL44K410A-BAC and UL44WTZeo-BAC, respectively, into HFFs. The wild-type Towne BAC virus was also produced in a similar manner. The electroporated cells were then incubated at 37°C and 5% CO_2 for three weeks and harvested when 100% cytopathetic effect was apparent in the cells.

Multi-step Viral Growth Curve

The replication and spread abilities of our mutant viruses were determined by multi-step virus growth curves in HFFs. The appropriate virus were added to a HFF monolayer at 95% confluency at a MOI of 0.1. The inoculum was removed after two hours and each well was then washed with PBS. Supplemented DMEM was added to the cells and then the plate was incubated at 37° C and 5% CO_2. Cells were collected by scraping the plate with a micropipette tip at the indicated time points post infection: 0, 1, 3, 5, 7, 10, and 14 days. To release virus from infected cells, collected cells was frozen at -80°C and then thawed in a 37°C water bath three times. Infectious virus at each time point was titrated on HFF cells seeded at 95% confluency. The inoculum was serial diluted in 10-fold increments with DMEM supplemented with serum. The diluted innoculum was then added to HFF cells and allowed to adsorb for two hours. The viral inoculum was then removed and an overlay containing DMEM supplemented with serum and 2% Type VII Agarose (Sigma Aldrich, St. Louis, MO) was added to each well. The infections were allowed to proceed for 14 days, and the fluorescing plaques were counted using a Eclipse TE300 Fluorescence microscope (Nikon, Tokyo, Japan). Titrations of each time point were performed in triplicate.

Real-Time Qualitative PCR

HFFs were infected with either UL44-K410A, UL44-WTZeo, or the wild-type Towne BAC virus at a MOI of 0.05. DNA from infected cells was harvested after 72 hours point infection using the QIAamp® DNA blood Mini Kit (Qiagen, Valencia, CA). PCR reactions were then prepared using Dynamo HS SYBR Green (Thermo Scientific, Waltham, MA), harvested DNA, and primers to amplify either the MIEP or GAPDH genes. The reactions were run in an iCycler thermal cycler (BioRad, Hercules, CA) for 40 cycles. CT Values were extrapolated from the iCycler software into Microsoft Excel and relative fold differences were calculated using the ΔΔCT method comparing the recombinant viruses to the wild-type.

Immunoprecipitation of 35S Methionine labeled UL44

HFFs were seeded on a 6 well plate ($3x10^5$) and then infected with either UL44-K410A, UL44-WTZeo, or the wild-type Towne BAC virus at a MOI of 1. After 72 hours, the cells were washed with PBS twice and then starved by incubating the cells with DMEM deficient in methionine for one hour. The media was then replaced with DMEM supplemented with one mCi of 35[S] methionine (Perkin Elmer, Waltham, MA) and further incubated for two hours at 37°C, 5% CO_2 to allow for incorporation. Briefly, the cells were washed twice with PBS and the radiolabeled proteins were harvested using M-PER® (Thermo Scientific, Rockford, IL) as a

lysing agent and then scraped with a plastic scraper. The resulting lysate was then incubated on ice for 30 minutes and then centrifuged at16,000 x g for 20 mins to remove cellular non-soluble debris. Preclearing was performed by adding Protein A/G Plus-Agarose beads (Santa Cruz Biotechnology, Santa Cruz, CA) and incubating for 30 minutes in a tumbler at 4°C. The precleared lysate with the addition of anti-UL44 (ICP36) (Virusys Corp., Taneytown, MD) or anti-FLAG (Sigma Aldrich, St. Louis, MO) was then incubated for two hours at 4°C in a tumbler. Protein A/G Plus-Agarose beads were then added to the mixture and then further incubated for an addition hour. The beads were repeatedly washed with PBS, eluted with reduced sample buffer, and boiled at 95°C for 10 minutes. Eluted IP'd proteins were separated on a 10% SDS-PAGE gel that was then fixed with 40% methanol/10% acetic acid and dried with a Slab Gel Dryer SQD2000 (Savant). The dried gel was then imaged using a Storm 840 phosphoimager (GE Healthcare, Waukesha, WI).

References

1. Edward S. Mocarksi, T. S., Robert F. Pass (2007). "Cytomegalovirus". In *Fields Virology* (P. M. H. David M. Knipe, Ed.), pp. 2701-2772. Lippincott Williams & Wilkins, New York, NY, USA.
2. Rechter, S., et al. (2009). Cyclin-dependent Kinases Phosphorylate the Cytomegalovirus RNA Export Protein pUL69 and Modulate Its Nuclear Localization and Activity. *The Journal of biological chemistry* 284: 8605-8613.
3. Sanchez, V., Sztul, E., and Britt, W. J. (2000). Human cytomegalovirus pp28 (UL99) localizes to a cytoplasmic compartment which overlaps the endoplasmic reticulum-golgi-intermediate compartment. *Journal of virology* 74: 3842-3851.
4. Boggio, R., and Chiocca, S. (2006). Viruses and sumoylation: recent highlights. *Current opinion in microbiology* 9: 430-436.
5. Hannoun, Z., Greenhough, S., Jaffray, E., Hay, R. T., and Hay, D. C. Post-translational modification by SUMO. *Toxicology* 278: 288-293.
6. Muller, S., and Dejean, A. (1999). Viral immediate-early proteins abrogate the modification by SUMO-1 of PML and Sp100 proteins, correlating with nuclear body disruption. *Journal of virology* 73: 5137-5143.
7. Nevels, M., Brune, W., and Shenk, T. (2004). SUMOylation of the human cytomegalovirus 72-kilodalton IE1 protein facilitates expression of the 86-kilodalton IE2 protein and promotes viral replication. *Journal of virology* 78: 7803-7812.
8. Lee, H. R., and Ahn, J. H. (2004). Sumoylation of the major immediate-early IE2 protein of human cytomegalovirus Towne strain is not required for virus growth in cultured human fibroblasts. *The Journal of general virology* 85: 2149-2154.
9. Dunn, W., et al. (2003). Functional profiling of a human cytomegalovirus genome. *Proceedings of the National Academy of Sciences of the United States of America* 100: 14223-14228.
10. Appleton, B. A., Loregian, A., Filman, D. J., Coen, D. M., and Hogle, J. M. (2004). The cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer. *Molecular cell* 15: 233-244.
11. Silva, L. A., Loregian, A., Pari, G. S., Strang, B. L., and Coen, D. M. The carboxy-terminal segment of the human cytomegalovirus DNA polymerase accessory subunit UL44 is crucial for viral replication. *Journal of virology* 84: 11563-11568.

12. Hay, R. T. (2005). SUMO: a history of modification. *Molecular cell* 18: 1-12.
13. Perdomo, J., Verger, A., Turner, J., and Crossley, M. (2005). Role for SUMO modification in facilitating transcriptional repression by BKLF. *Molecular and cellular biology* 25: 1549-1559.
14. Palacios, S., et al. (2005). Quantitative SUMO-1 modification of a vaccinia virus protein is required for its specific localization and prevents its self-association. *Molecular biology of the cell* 16: 2822-2835.
15. Groves, I. J., Reeves, M. B., and Sinclair, J. H. (2009). Lytic infection of permissive cells with human cytomegalovirus is regulated by an intrinsic 'pre-immediate-early' repression of viral gene expression mediated by histone post-translational modification. *The Journal of general virology* 90: 2364-2374.
16. Strang, B. L., Sinigalia, E., Silva, L. A., Coen, D. M., and Loregian, A. (2009). Analysis of the association of the human cytomegalovirus DNA polymerase subunit UL44 with the viral DNA replication factor UL84. *Journal of virology* 83: 7581-7589.
17. Yu, D., Silva, M. C., and Shenk, T. (2003). Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proceedings of the National Academy of Sciences of the United States of America* 100: 12396-12401.
18. Berndt, A., Hofmann-Winkler, H., Tavalai, N., Hahn, G., and Stamminger, T. (2009). Importance of covalent and noncovalent SUMO interactions with the major human cytomegalovirus transactivator IE2p86 for viral infection. *Journal of virology* 83: 12881-12894.
19. Kim, E. T., Kim, Y. E., Huh, Y. H., and Ahn, J. H. Role of noncovalent SUMO binding by the human cytomegalovirus IE2 transactivator in lytic growth. *Journal of virology* 84: 8111-8123.