January 2020

Biological And Computational Studies Of The Structure And Function Of Pul103, A Human Cytomegalovirus Tegument Protein

Ashley N. Anderson
Wayne State University

BIOLOGICAL AND COMPUTATIONAL STUDIES OF THE STRUCTURE AND FUNCTION OF PUL103, A HUMAN CYTOMEGALOVIRUS TEGUMENT PROTEIN

by

ASHLEY NICOLE ANDERSON

DISSERTATION

Submitted to the Graduate School

of Wayne State University,

Detroit, Michigan

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

2020

MAJOR: IMMUNOLOGY & MICROBIOLOGY

Approved By:

_____

Advisor Date

_____

_____

_____

DEDICATION

I dedicate this dissertation first to God because I am able to do all things through Him who strengthens me. Secondly, I dedicate this work to my mother, Vanessa Anderson, because I would not have made it to this point in my life without her. Thirdly, I dedicate this work to my immediate family: my brother, father, grandmother, and grandfather for always believing in me and providing encouragement.

ACKNOWLEDGEMENTS

I would like to thank members of the Kovari laboratory, past and present, for assisting me with the computational studies, Dr. Daniel Ortiz for his mentorship and for making the viruses used in my experiments, and other members of the Pellett laboratory, past and present, for their knowledge, friendship, and various contributions to my life in the lab.

TABLE OF CONTENTS

DEDICATION ..... **ii**

ACKNOWLEDGEMENTS ..... **iii**

LIST OF TABLES ..... **vi**

LIST OF FIGURES ..... **vii**

CHAPTER 1: INTRODUCTION **1**

Section I: Overview of herpesviruses 1

Section II: Overview of HCMV 5

Section III: Overview of pUL103 homologs and protein interaction partners 15

CHAPTER 2: PROTEIN SEQUENCE ANALYSIS AND STRUCTURAL PREDICTION OF PUL103 HOMOLOGS **40**

Introduction 40

Methodology 42

Results 44

Discussion 50

CHAPTER 3: IDENTIFICATION OF THE TEMPORAL WINDOW WHEN PUL103 IS REQUIRED FOR MAXIMAL CYTOPLASMIC VIRION ASSEMBLY COMPARTMENT ABUNDANCE **68**

Introduction 68

Materials and Methods 72

Results 75

Discussion 79

CHAPTER 4: FINAL CONCLUSIONS AND DISCUSSION 88

REFEERENCES 93

ABSTRACT 108

AUTOBIOGRAPHICAL STATEMENT 110

LIST OF TABLES

Table 1-1 Characteristics of the three subfamilies of the *Herpesviridae*..... 29

Table 1-2 Betaherpesvirus genes conserved across the *Herpesviridae* 31

Table 1-3 Characteristics of HCMV pUL103, HCMV pUL71 and their homologs HSV-1 pUL7, HSV-1 pUL51, EBV BBRF2, BSRF1, KSHV ORF42, and KSHV ORF55 37

Table 3-1 Names and properties of viruses used to study HCMV pUL103..... 81

LIST OF FIGURES

FIG 1-1 Diagram of a herpesvirus virion 28

FIG 1-2 General overview of the herpesvirus replication cycle 30

FIG 1-3 Three- dimensional reconstructions from cryoelectron microscopy images of herpes simplex virus type 1 pentons and hexons 33

FIG 1-4 Three- dimensional reconstructions of cryoelectron microscopy images of HCMV nuclear capsids 34

FIG 1-5 Protein-protein interactions within the tegument 35

FIG 1-6 Immunofluorescence assay (IFA) image of the cVAC in an HCMV-infected cell . 36

FIG 1-7 The N terminus (residues 1-125) of pUL103 is more conserved than the C terminus (residues 126-249)..... 38

FIG 1-8 Diagram of the pUL103 amino acid sequence with confirmed and putative sequence domains..... 39

FIG 2-1 Human herpesvirus pUL103 homologs share limited sequence identities..... 53

FIG 2-2 List of pUL103 homologs used for analysis..... 54

FIG 2-3 Phylogenetic analysis reveals relationships among pUL103 homologs 55

FIG 2-4 MUSCLE Alignment shows conserved protein chemistry across the 14 pUL103 homologs 56

FIG 2-5 MUSCLE Alignment shows potentially conserved cytomegalovirus motifs..... 57

FIG 2-6 MUSCLE alignment produced in this study correlates with similarity plot from Dr. Richard Roller 58

FIG 2-7 The N terminus (residues 1-125) of pUL103 is more conserved than the C terminus (residues 126-249)..... 59

FIG 2-8 I-TASSER is a useful tool that provided predicted structures for the pUL103 homologs whose structures had not been crystallized..... 60

FIG 2-9 There are differences in the predicted structures of the pUL103 homologs..... 61

FIG 2-10 Heatmap showing alternating regions of movement within the predicted pUL103 structure..... 62

FIG 2-11 Nucleotidyl transferases make up the list of the top ten structural analogs of the predicted pUL103 structure..... 63

FIG 2-12 The predicted structure of pUL103 aligned well with the structure of the catalytic domain of the human terminal uridylyl transferase 7 (TUT7) protein 64

FIG 2-13 pUL103 homologs share protein chemistry with several nucleotidyl transferases 65

FIG 2-14 The predicted structure of pUL103 aligned well with the structure of the non-catalytic domain of the human terminal uridylyl transferase 4 (TUT4) protein..... 66

FIG 2-15 There are differences between the predicted pUL7 structure and the actual crystallized protein..... 67

FIG 3-1 Time course of HCMV protein expression in cells infected with the UL103-FKBP-V5 virus in the presence and absence of Shield-1 82

FIG 3-2 Immunofluorescence signals of viral nuclear marker IE2 and cellular cytoplasmic Golgi marker GM130 do not overlap in HCMV-infected cells..... 83

FIG 3-3 A Golgi ring gradient: a tool created to assist with discernment of mature cVACs 84

Changes in the stability of pUL103-FKBP-V5 affect cVAC abundance 85

FIG 3-5 pUL103-FKBP-V5 is needed for maximal cVAC abundance between 72 and 84 hpi 87

FIG 4-1 Various images of HCMV-infected cells with characteristic nuclear inclusions with halos (owl's eye nuclei) 92

CHAPTER 1: INTRODUCTION

Section I: Overview of herpesviruses

Herpesvirales. The Order *Herpesvirales* consists of three virus families: the *Alloherpesviridae*: herpesviruses of fish and amphibians, *the Herpesviridae*: herpesviruses of mammals, birds, and reptiles, and the *Malacoherpesviridae*: herpesviruses of mollusks. Herpesviruses are linear, single-segment, dsDNA viruses. Classification as a herpesvirus is dependent on the structure of the virus particle (virion). Herpesvirus virions consist of an envelope, tegument, icosahedral capsid, and viral DNA core (Fig. 1-1). These viruses obtain their envelope, a lipid bilayer embedded with glycoproteins, from their hosts during secondary envelopment. The glycoproteins interact with cell surface markers to initiate entry into target cells (74). Directly beneath the envelope lies a layer of proteins and RNA, the tegument, that is only present in herpesviruses. These molecules immediately begin modulating the cellular environment to facilitate virion production upon entry (65, 94). Within the tegument lies the capsid, a protein shell that surrounds the viral DNA. The capsid primarily consists of the major capsid protein (MCP) organized into an icosahedral (20 faces) shape, with T=16 symmetry. Surrounded by the capsid is the viral DNA core. The DNA lacks chromatin and is packed into the shape of a ring called a torus. The DNA is packed so tightly into the capsid that it is pressurized, which is believed to assist with injection of the DNA into the nuclear pore of infected cells (74).

Disease manifestations. Herpesviruses produce new virions through a lytic replication cycle that results in the death of the host cell. These viruses also establish life-long infections within their hosts through latency: a state in which lytic replication is repressed and no new virions are produced. From latency the viruses can return to a lytic replication cycle through reactivation which usually occurs due to cellular stressors such as changes in nutrients, damage, etc. Across

the various hosts, symptoms of herpesvirus infections may include fever, lesions, inflammation, cancer, or death primarily in immunocompromised or immunodeficient populations. Disease in immunocompetent individuals usually ranges from none to mild cold or flu symptoms, may include some of the symptoms mentioned above, and rarely leads to death. However, there are exceptions. During organ transplantations, seemingly healthy, immunocompetent individuals can experience severe disease and/or death if the organs are infected with human cytomegalovirus (HCMV), a member of the *Herpesviridae* (48, 58, 74, 91, 105).

Lytic replication cycle. The lytic replication cycles of all herpesviruses are similar, consisting of the same basic steps (Fig. 1-2) (21, 74).

Attachment and entry. Herpesvirus infection begins with attachment of virion envelope glycoproteins (typically gB and others), to cell surface receptors heparan sulfate proteoglycans on permissive cells (Fig. 1-2, steps 1 and 2). Upon the receptor/ligand interactions, the virion enters the cell by either viral envelope fusion with the cellular membrane or endocytosis (Fig. 1-2, step 3). After membrane fusion the capsid and tegument are released into the cytoplasm (Fig. 1-2, step 4). During endocytosis, the capsid and tegument are not released into the cell until the viral envelope fuses with the membrane of the intracellular vesicle. Once in the cell, the capsid is trafficked along microtubules to the nucleus (Fig. 1-2, step 5a). Meanwhile, the tegument proteins begin modulating the cellular environment to facilitate production of new virions (Fig. 1-2, steps 5b and 6). At the nucleus, the capsid is docked to a nuclear pore by unknown mechanisms. With the capsid portal facing the nuclear pore, the viral DNA is propelled into the nucleus, possibly due to the potential energy of the capsid during DNA packaging (Fig. 1-2, step 7) (74).

DNA replication and viral cascade of gene expression. Once the viral DNA is in the nucleus, the viral cascade of gene expression begins starting with the immediate early (alpha)

genes. Herpesvirus immediate early genes require no de novo protein synthesis and are immediately transcribed in the nucleus and the transcripts translocated to the cytoplasm for translation (Fig. 1-2, steps 9 and 10). The products of immediate early genes are translocated to the nucleus to initiate transcription of early (beta) genes (Fig. 1-2, step 10). Once translated, some early gene products remain in the cytoplasm, while others are transported to the nucleus to initiate viral DNA replication and transcription of late (gamma) genes (Fig. 1-2, steps 11-14). Some late gene products are translocated to the nucleus for capsid assembly and egress, membrane-associated proteins are translated in the endoplasmic reticulum, and others such as tegument proteins are translated in the cytoplasm (Fig. 1-2, steps 15-17) (74).

Assembly, tegumentation, and egress. Capsid assembly begins in the nucleus and is also where the inner tegument consisting of pUL32 (pp150, "pp" denotes phosphoprotein), is acquired (Fig. 1-2, step 18). Capsids then traverse the nuclear envelope using the nuclear egress complex (NEC), viral proteins that destabilize the nuclear membrane (Fig. 1-2, steps 19-20). Once in the cytoplasm, the partially tegumented capsid acquires its outer tegument by interacting with large aggregates of tegument proteins. These large aggregates are usually near membranes of various sizes and Golgi origin (trans Golgi networks, TGN) (Fig. 1-2, step 21). After acquiring its full tegument, the capsid acquires its envelope by budding into TGN membranes, becoming a fully infectious virion (Fig. 1-2, step 22). The vesicle containing the virion, an exocytic vesicle, can carry one or multiple virions at a time. Within the vesicles, virions are trafficked to the cellular membrane for egress into the extracellular environment (Fig. 1-2, steps 23-24) (74, 75).

***Herpesviridae*.** The *Herpesviridae*, herpesviruses of mammals, birds, and reptiles, consists of three subfamilies: the *Alpha-*, *Beta-*, and *Gammaherpesvirinae*. These subfamilies are grouped on the basis of their host ranges, length of reproductive cycles, ability to cause infection in cell

culture, destruction of infected cells, and cells in which they establish primary infections and latency (Table 1-1). Individual viruses within the *Alphaherpesvirinae* tend to have a broad host range, relatively short reproductive cycles, infections spread rapidly in cell culture, infected cells are efficiently destroyed, they infect primarily mucocutaneous tissue, epithelial cells, and fibroblasts, and establish latency primarily within the dorsal root ganglia of neurons (15, 74). Individual viruses within the *Betaherpesvirinae* have a narrow host range, long reproductive cycles-for some over seven days, infections progress slowly in cell culture, infected cells are frequently enlarged (cytomegalia), they primarily infect fibroblasts, T, endothelial, epithelial cells, etc., and establish latency primarily within lymphoid tissues, hematopoietic progenitor cells, and other tissues (22, 64, 74, 99). Individual viruses within the *Gammaherpesvirinae* have a host range that is restricted to the family or order of the natural host, in cell culture they infect lymphoblastoid cells, primarily infect and establish latency within B, T, epithelial, and endothelial cells, among others, infections hardly damage cells except during reactivation from latency, and they can induce lymphoproliferation and cancers in infected cells. Unlike other herpesviruses, latency is the default infection strategy for most gammaherpesviruses, so information regarding their lytic replication cycles is limited (18, 55, 74, 101).

Human herpesviruses and disease manifestations. There are nine herpesviruses of humans across the three subfamilies. The human alphaherpesviruses include herpes simplex virus 1 (HSV-1), herpes simplex virus 2 (HSV-2), and varicella zoster virus (VZV). The human betaherpesviruses include human cytomegalovirus (HCMV), human herpesvirus 6A, 6B, and 7 (HHV-6A, -6B, -7). The human gammaherpesviruses include Epstein-Barr virus (EBV) and Kaposi's sarcoma-associated herpesvirus (KSHV). HSV-1 and HSV-2 are closely related viruses that cause oral and genital lesions, along with other more serious diseases such as encephalitis and

meningitis. VZV causes chicken pox in children and upon reactivation in older individuals causes shingles. Chicken pox is an itchy rash that spreads all over the body. Shingles is a painful rash that occurs in localized areas of the body. HCMV is an opportunistic pathogen that causes a range of symptoms from none to life-threatening. The most severe symptoms include organ-specific and systemic inflammation. Human herpesvirus 6A, 6B, and 7, which are more closely related than the simplex viruses, cause a range of symptoms as well, from asymptomatic infection to febrile illnesses, with some associated with neurologic complications, seizures, and encephalitis (50, 67, 100). EBV causes infectious mononucleosis, lymphoproliferative diseases, endemic Burkitt lymphoma, Hodgkin disease, nasopharyngeal carcinoma, as well as other B cell cancers in individuals with immunodeficiencies. KSHV causes Kaposi sarcoma, primary effusion lymphoma, and various lymphoproliferative diseases (67).

Section II: Overview of HCMV

HCMV significance. HCMV is a global pathogen, with the up to 85% of adults infected by the age of 40. It causes organ-specific and systemic inflammation that leads to severe disease in immunocompromised (transplant recipients, HIV patients) and immunodeficient (fetuses, infants) individuals. The immunocompromised populations tend to suffer from hepatitis, pancreatitis, pneumonitis, enteritis, retinitis meningitis and encephalitis. The disease usually manifests in immunodeficient populations as liver, lung, and spleen problems, petechiae, hearing loss, and neurodevelopmental disabilities (64, 67). HCMV is the most common viral congenital infection, as well as the leading cause of birth defects and childhood disabilities in the United States (49, 67). In fact, more children suffer from CMV-related diseases than from more well-known congenital conditions such as Down syndrome, fetal alcohol syndrome, or spina bifida (16). The virus is also the most common, non-hereditary cause of hearing loss in children (43).

HCMV virion characteristics. The virion is ~2000 Å in diameter; the tegument consists of at least 38 different proteins (including some of cellular origin) present in various quantities; the capsid has a diameter of ~1300 Å, and surrounds a genome ~236 kb that encodes at least 167 proteins, four large noncoding RNAs, two oriLyt RNAs, and at least 23 microRNAs. HCMV is the largest virus that infects humans, in terms of the size of its genome (73). HCMV utilizes its various gene products to manipulate the cellular environment into one most favorable for new virion production (20, 64).

HCMV replication cycle. The HCMV lytic replication cycle resembles that of other herpesviruses, but with some unique aspects (64).

HCMV cellular tropism and receptors. HCMV infects a variety of cells including fibroblasts, monocytes/macrophages, neutrophils, hepatocytes, endothelial, epithelial, smooth muscle, stromal, and neuronal cells. This broad cellular tropism is why HCMV causes such severe disease and is hard to treat. This broad tropism also suggests that HCMV either possesses one broad common receptor, multiple cell-specific receptors, or a receptor complex that consists of a combination of both cell-specific and broadly expressed cellular receptors. Several receptors, in addition to heparan sulfate, have been investigated and none have been shown to conclusively be involved in making cells permissive for HCMV infection (22, 64).

Attachment and entry. HCMV infection begins with attachment of virion envelope glycoproteins B, H, L, O, M, and N, as well as proteins pUL128, pUL130, and pUL131A in some instances, to cell surface receptors heparan sulfate proteoglycans on permissive cells. Upon the receptor/ligand interactions, the virion enters the cell by either viral envelope fusion with the cellular membrane or endocytosis. After membrane fusion the capsid and tegument are released

into the cytoplasm. During endocytosis, the capsid and tegument are not released into the cell until the viral envelope fuses with the membrane of the intracellular vesicle (22, 64).

I had primary responsibility for writing the following sections (pages 7-15), which were published as sections 9.3.1-9.3.4 and 9.4.3 in Close et al. 2018 Advances in Experimental Medicine and Biology (20).

Tegument proteins in action. Once in the cell, tegument proteins begin modulating the cellular environment to facilitate production of new virions. As mentioned earlier, HCMV virions contain at least 38 different tegument proteins in varying quantities. Tegument proteins are important for many processes throughout infection, including disassembly of virions, transcriptional regulation, modulation of cellular responses, and virion maturation. Many, if not most tegument proteins, perform multiple, distinct functions. Although it lacks a well-defined structure, there are clear elements of structural order throughout the tegument. Tegument protein pUL83 (pp65) is the most abundant protein in HCMV virions and dense bodies. While not required for replication, pp65 is important for the initiation of infection by enhancing activation of the major immediate-early promoter and is important for recruiting proteins into the virion during assembly. pUL48 (large tegument protein, LTP,) is the largest and second most abundant HCMV tegument protein. It is essential for replication and is involved in the release of viral DNA from the capsid. pUL32 (pp150) is the third most abundant tegument protein. It is essential for viral replication and plays a role in stabilizing DNA-containing capsids during virion maturation by forming the inner tegument. pUL47 (LTP-binding protein, LTPbp) is an essential protein involved in uncoating the capsid. pUL103 is a low-abundance tegument protein that plays important roles late in HCMV infection, including cVAC biogenesis, cell-to-cell spread, and virion maturation.

pUL103 may have roles early in the HCMV replication cycle due to its presence in nuclei and interactions with several proteins of the innate immune system (20).

Capsid trafficking and uncoating. The capsid is trafficked along microtubules to the nucleus. At the nucleus, the capsid is docked to a nuclear pore by unknown mechanisms. With the capsid portal, consisting of pUL104, facing the nuclear pore, the viral DNA is propelled into the nucleus, possibly due to the potential energy of the capsid during DNA packaging. These processes occur with assistance from pUL48, pUL47, and the smallest capsid protein (SCP) pUL48A (64).

DNA replication and viral cascade of gene expression. Once the viral DNA is in the nucleus, the viral cascade of gene expression begins starting with the immediate early (alpha) genes. HCMV immediate early (alpha) genes require no de novo protein synthesis and are transcribed and translated immediately upon entering the nucleus. The products of immediate early (alpha) genes are translocated to the nucleus to initiate transcription of beta genes. Once translated the early (beta) gene products are transported to the nucleus to initiate viral DNA replication and transcription of late (gamma) genes. Some late gene products are translocated to the nucleus for capsid assembly and egress, membrane-associated proteins are translated in the endoplasmic reticulum, and others such as tegument proteins are translated in the cytoplasm (74).

Capsid components and assembly. During capsid assembly, a protein scaffold, consisting of the HCMV assembly protease (PR-pUL80a, Table 1-2) and assembly protein precursor, (pAP-pUL80.5, Table 1-2) forms. The protease and assembly protein precursor are encoded by the same gene and share the same carboxyl terminus, which contains domains necessary for interaction with the major capsid protein (MCP-pUL86, Table 1-2). These proteins also have conserved amino terminal domains important for self-assembly into a scaffold, around

which the capsid proteins assemble. Capsid proteins are the building blocks of capsids with the major capsid protein (MCP, Table 1-2) as the major component. During assembly, MCP forms both pentameric and hexameric capsomeres, the major subunits of capsids (Fig. 1-3). Capsomeres are cylindrical and have a pore that runs along their length, which is ~160 Å. Capsomeres self-assemble onto the scaffold to form a capsid with T = 16 icosahedral symmetry. In this arrangement, 150 hexons form the triangular faces of the capsid and 11 pentons make up the vertices of those triangles (Fig. 1-4). Given that there are 6 molecules of MCP in each of the 150 hexons and 5 molecules of MCP in each of the 11 pentons, there are a total of 950 molecules of MCP per capsid. The twelfth pentonal position is occupied by a complex formed by 12 copies of the portal protein (pUL104 or PORT, Table 1-2), which provides the channel necessary for packaging the viral genome into the capsid during assembly and for its release into the nucleus at the initiation of infection.

During DNA encapsidation, the portal complex interacts with the terminase complex, which consists of ATPase (pUL89 or TER 1, Table 1-2) and DNA recognition (pUL56 or TER 2, Table 1-2) subunits. The small capsid protein (SCP-pUL48.5, Table 1-2) is the second most abundant protein in the HCMV capsid with 900 molecules per capsid. SCP is only found associated with the hexons, decorating their tips. In the absence of SCP, nascent capsids are devoid of viral DNA and the capsid-associated tegument protein pUL32 (pp150; “pp” denotes phosphoprotein). In contrast, SCP is dispensable for HSV-1 viral growth, demonstrating virus-specific structural and functional roles for this protein. Between the hexons and pentons reside small triplexes consisting of a dimer of the minor capsid protein (MnCP-pUL85, Table 1-2) and a monomer of the minor capsid-binding protein (MnCP-bp-pUL46, Table 1-2). These triplexes are important for stabilizing the nucleocapsid. The portal capping protein (PCP-UL77, Table 1-2) and the capsid

transport tegument protein (CTTP-pUL93) form the capsid vertex-capping (CVC) complex, also known as the capsid vertex-specific component (CVSC), which decorates pentons. The terminase-binding protein (TERbp-pUL51, Table 1-2) and the capsid transport nuclear protein (CTNP-pUL52, Table 1-2) associate with this complex as well. The role(s) of this complex and the associated proteins have not been determined, although they are thought to be involved with nucleocapsid stability by helping the capsid to withstand internal pressure from the DNA during encapsidation. They are also thought to be involved in the release of viral DNA into the nucleus during initiation of infection. In addition, the CVC and its associated proteins have been shown to be important for cleavage of concatemeric viral DNA into unit length genomes (20).

Encapsidation of virus genomes. Single copies of the virus genome are packed into newly formed capsids. The major events during genome encapsidation are assembly and then degradation of the scaffold, packaging of the virus genome into the rigidly constrained space inside the capsid, cleavage of the genome precisely at its termini, and closure (or corking) at the portal of the highly pressurized capsid. The virus accomplishes the process of DNA encapsidation rapidly, within fractions of a second. As the protease disassembles the scaffold, a virus genome is threaded into the capsid, short terminus first, through the channel formed by the portal complex. The ATPase subunit of the terminase is the motor for translocating viral DNA into capsids. For HCMV, the DNA recognition subunit of the terminase recognizes the intact packaging and cleavage sequence that is formed when pac-1 and pac-2 sequences at the genomic termini are juxtaposed in covalently circularized or concatemeric genomes. Precise cleavage by the terminase results in packaging one complete virus genome per capsid. Other viral proteins implicated in this process are pUL56 and pUL89. pUL56 binds to AT-rich sequences within the pac sequences and

has nuclease activity. pUL89 also has DNA cleavage activity; in addition to the terminase, it may be responsible for cleavage of the virus genome during encapsidation.

Once the capsid is assembled, it becomes partially tegumented in the nucleus. Proteins pUL32 and pUL48 form the inner tegument, an organized netlike layer that encloses the capsid shell. The inner tegument proteins tightly associate with the capsid in a manner that is resistant to treatment with various detergents. pUL32 interacts with the outside of the capsid in a manner that adds stability to the capsid, helping it to withstand the high internal pressure associated with the tightly packed genome (20).

Nuclear egress. Before capsids can enter the cytoplasm, they must pass through the nuclear envelope, which consists of the inner nuclear membrane (INM), the perinuclear space that is contiguous with the endoplasmic reticulum (ER), and the outer nuclear membrane (ONM). In uninfected cells, the distance across the nuclear envelope, from inner membrane to outer membrane, is typically ~50 nm. Nuclear envelope integrity is maintained by the linker of nucleoskeleton and cytoskeleton (LINC) complex, a multiprotein complex that directly connects the nuclear skeleton to the cytoskeleton. The complex includes Sad1p, UNC-84 (SUN) and Klarsicht, ANC-1, Syne homology (KASH) domain proteins. Interprotein interactions via SUN and KASH domains are important for nuclear membrane stability and for maintaining proper spacing between the inner and outer nuclear membranes.

The nuclear envelope is further stabilized by the nuclear lamina, a proteinaceous network that lines the inner nuclear membrane. Lamin proteins A/C and B associate with each other to form intermediate filaments that provide structural support to the cell. Lamin B also interacts with integral membrane proteins such as the lamin B receptor, which helps to tether the nuclear lamina to the inner nuclear membrane. To get past the nuclear lamina, herpesviruses employ conserved

proteins that make up the nuclear egress complex (NEC, pUL50-nuclear egress membrane protein (NEMP), pUL53- nuclear egress lamina protein (NELP), Table 1-2). In HCMV- infected cells, these proteins localize on the interior of the nucleus on the side that faces the cytoplasmic virion assembly complex (cVAC). The NEC competes with lamin proteins A/C and B for binding to each other, disrupting the lamina's fibrillary network. In addition, the NEC disrupts the binding of lamins to the nuclear envelope protein emerin, which links the nuclear lamina to the inner nuclear membrane. pUL50 also interacts with the HCMV-encoded nuclear rim- associated cytomegaloviral protein (RASCAL) that is thought to be important for NEC-dependent degradation of the nuclear lamina.

Additionally, the NEC recruits other viral and cellular proteins such as the viral protein kinase (pUL97) and cellular protein kinase C (PKC), both of which phosphorylate nuclear lamins, leading to destabilization and reorganization of the nuclear lamina. pUL97 has also been implicated in the phosphorylation of p32, a nuclear lamina component that interacts with the lamin B receptor, further contributing to nuclear lamina destabilization. These changes in the nuclear lamina facilitate egress of the capsids by allowing capsids to bud from the nucleus into the perinuclear space during primary envelopment. In HSV-1, the NEC forms a hexagonal lattice along the INM leading to formation of invaginations in the membrane and budding of capsids into the perinuclear space. X-ray crystallography showed that the HCMV NEC also forms hexameric rings, suggesting conservation of this invagination and budding activity across the herpesviruses. After enveloped capsids enter the perinuclear space, their membranes fuse with the ONM, and the capsids are released into the cytoplasm (20).

Tegumentation. Once in the cytoplasm, the partially tegumented capsid acquires its outer tegument by interacting with large aggregates of tegument proteins. The precise order of addition

of tegument proteins to maturing virions is unknown, but some information is available. Acquisition of the outer tegument occurs through several mechanisms (Fig. 1-5). Some tegument proteins aggregate in the cytoplasmic milieu and form complexes prior to incorporation in nascent virions. Others accumulate on membranes of various sizes and Golgi origin (trans Golgi networks, TGN). The marriage of partially tegumented capsids and membrane- associated tegument proteins is consummated during secondary envelopment, which occurs throughout the cytoplasmic virion assembly compartment (cVAC) (20).

Cytoplasmic virion assembly compartment (cVAC). One of the staples of HCMV infection, that appears to be involved in virion assembly and egress, is the presence of the cytoplasmic virion assembly compartment (cVAC). The cVAC is a juxtanuclear structure induced by HCMV and consists of cellular endosecretory machinery, primarily Golgi and early/recycling endosomes (Fig. 1-6). HCMV organizes the Golgi into a ring-like structure with the endosomes primarily in the center of the ring. Other secretory machinery such as late endosomes and the endoplasmic reticulum (ER) are located at the outermost edges of the cVAC. The cVAC is positioned at the center of the microtubule organizing center (mTOC) and induces the characteristic kidney bean shape of the nucleus. Within the cVAC, nascent capsids acquire their outer tegument, envelopes, and are trafficked to the cellular membrane for release into the extracellular milieu, indicating its significance in the HCMV replication cycle. Each infected cell has one cVAC, even cells with multiple nuclei due to cell fusion (syncytia). cVAC biogenesis is dependent on expression of several miRNAs and late proteins: pUL48, pUL94, pUL103, and gpUL132 (Table 1-2) (20, 106). The cVAC is present in various cells in vitro, and unpublished data suggests this structure exists in tissues as well (17, 23, 28, 53, 87, 89).

Egress. After acquiring its full tegument, the capsid acquires its envelope by budding into TGN membranes, becoming a fully infectious virion. The vesicle containing the virion, an exocytic vesicle, can carry one or multiple virions at a time. Within the vesicles, virions are trafficked to the cellular membrane for egress into the extracellular environment (74, 75).

Current treatments. There are five main medications for the treatment and prevention of HCMV infections: ganciclovir (GCV), valganciclovir (VGCV), cidofovir (CDV), foscarnet (FOS), and letermovir (LTV). GCV, VGCV, CDV, and FOS all target viral DNA replication. LTV targets viral DNA packaging. GCV was the first treatment approved for HCMV infections and is still the therapy of choice today. GCV is used in the prevention of HCMV disease and for treatment in immunocompromised hosts as well congenitally infected infants. GCV is a nucleoside analog. After undergoing phosphorylation by the viral kinase pUL97 and several cellular kinases, GCV is able to inhibit viral DNA synthesis by competitive inhibition of deoxyguanosine triphosphate. GCV has low bioavailability and necessitated the development of a proform of the drug, VGCV. Both drugs can exhibit toxicity with symptoms including hematological abnormalities such as anemia, neutropenia, and thrombocytopenia. CDV and FOS are both approved for the treatment of retinitis in AIDS patients caused by HCMV, and for use in cases where GCV resistance is present. Both also are limited in their use by their nephrotoxicity. The toxicity of these drugs is partially due to drug interactions between these treatments and other medications taken by the patients (46). Possible off-target effects on host cell DNA replication by these medications could also explain their toxic effects. LTV is the newest treatment available for treating and preventing HCMV infections in transplant recipients. This drug has been shown to be effective against several clinical and GCV-resistant strains and is well tolerated (10). The approval of this drug is a move in the right direction of developing more specific, diverse, and less

toxic treatments for HCMV infections. However, more work needs to be done in this area. In the Pellett laboratory, we believe that the protein HCMV pUL103 (p stands for protein, referred to henceforth as pUL103) has the potential to be an effective antiviral target.

Section III: Overview of pUL103 homologs and protein interaction partners

Introduction. pUL103 and its alpha- and gammaherpesvirus homologs (HSV-1 pUL7, EBV BBRF2, KSHV ORF42) have been studied and shown to be important for herpesvirus replication (Table 1-3); the same is true for the pUL103 interaction partner HCMV pUL71 and its homologs (HSV-1 pUL51, EBV BSRF1, KSHV ORF55) (Table 1-3). In the absence of these proteins, the viruses are attenuated. These proteins have various roles throughout herpesvirus infections, and we believe with continued study these proteins have the potential to be antiviral candidates.

Characteristics of HCMV pUL103 and pUL71

Genome information, expression kinetics, conservation, and sequence motifs/domains of pUL103. The UL103 gene is located in the unique long region of the genome. UL103 is a late gene, meaning it requires the expression of the immediate early and early genes for it to be transcribed and translated (64). The UL103 gene product pUL103 is a small (249 amino acids (aa), ~27.39 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (72, 74). When the protein sequences of pUL103 and its homologs were analyzed, it was shown that the average conservation score of the N terminus (residues 1-125), 3.8, is greater than that of the C terminus (residues 126-249), which is 2.3. Thus, the N terminus is more conserved than the C terminus, indicating the importance of this region to the function of the protein (Fig. 1-7). Within all pUL103 homologs is a conserved herpesvirus domain with the sequence G-F-x(8)-E-D-x-V-x(12)-R. This domain is also present in the protein topoisomerase

III of fission yeast, suggesting that these proteins may interact with nucleic acids (41). In addition to the conserved herpesvirus domain, the pUL103 protein sequence may contain a DEA(D/H) box and late "L" domains (Fig. 1-8). DEA(D/H) box domains are important for allowing proteins to interact with and unwind DNA and RNA (42). Late domains are motifs primarily found in the Gag proteins of retroviruses that allow for interaction with the endosomal sorting complexes required for transport (ESCRT) and their associated proteins, helping to facilitate membrane scission processes. The putative late domains present in pUL103 appear to follow the motif LYPXnL that allows for binding to the ESCRT-associated protein ALIX (63, 72, 78). More work needs to be done to validate the function of the putative domains.

Localization of pUL103. As a tegument protein, pUL103 is present in virions and dense bodies, enveloped tegument proteins. Within these particles pUL103 makes up a very small fraction of the total amount of protein present, around 0.1% (102). Thus, at the beginning of an infection there is very little pUL103 present, below the threshold of detection for techniques such as immunofluorescence. Being a gamma gene product, expression of pUL103 ramps up later in infection. At 120 hours post infection, the majority of pUL103 is present in the cytoplasm within the cVAC, with little in the nucleus (72).

Roles of pUL103 in HCMV infection. pUL103 is involved in virion and dense body egress, cVAC formation, cell-to-cell spread, and virion maturation. In pUL103-knockout experiments, viral titer decreases between 100 and 10,000-fold (1, 35, 110). In studies where viruses were used where protein expression of pUL103 was able to be regulated, knockdown of pUL103 results in decreases in cVAC formation, plaque sizes indicative of reduced cell-to-cell spread, and secondary envelopment, hampering virion maturation (26).

Protein interaction partners of pUL103. A study from our laboratory where a dual proteomics approach was used identified several cellular and viral protein partners of pUL103. The cellular proteins that were identified were primarily those involved in immune system signaling e.g. ISG15, IFIT1, STAT1, and IFI16 and intracellular trafficking e.g. DYNLL1, COPA, COPB1, ALIX, and MYO1C. The viral proteins that were identified were primarily those involved in associating with DNA e.g. pUL112/113, pUL34, pUL123 (IE2), pUL44, and pUL84 and other tegument proteins pUL56, pUL47, pUL45, and pUL71 (72).

The interaction between pUL103 and pUL71 was previously demonstrated to occur outside the context of infection, with data showing that pUL71 helps pUL103 to localize to Golgi membranes. During infection, the two proteins colocalize and are within close proximity to one another (38). Studies of the HSV-1 and EBV homologs of these proteins (pUL7 and pUL51; BBRF2 and BSRF1) have shown these proteins interact during infection, providing further evidence that the HCMV proteins might as well (3, 80, 108).

Genome information, expression kinetics, conservation, and sequence motifs/domains of pUL71. pUL71 is encoded by the UL71 gene located in the unique long region of the genome (64). UL71 is an early gene, meaning it requires the expression of immediate early genes for it to be transcribed and translated. The UL71 gene product pUL71 is a small (361 aa, ~39.71 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (74, 88). pUL71 possesses five sequence motifs that are necessary for its functions: an N-terminal palmitoylation site, N-terminal tyrosine-based trafficking motif, a leucine zipper-like motif, C-terminal tetra-lysine motif, and Vps4-interacting motif. The palmitoylation site is conserved across pUL71 homologs and is important for the localization of pUL71 to the cVAC, proper cVAC formation, and pUL71 stability. The tyrosine-based trafficking motif consists of four amino acids

in the following configuration: YXX (X stands for any amino acid; stands for an amino acid with a bulky hydrophobic side chain). The tyrosine-based trafficking motif is also conserved across pUL71 homologs and important for the localization of pUL71 to the cVAC, proper cVAC formation, as well as cell-to-cell spread, secondary envelopment, and virion egress. The leuzine zipper-like motif is important for proper cVAC formation. The C-terminal tetra-lysine motif is important for secondary envelopment (32, 62, 87). The importance of the Vps4-interacting motif is unknown (38).

Localization of pUL71. As a tegument protein, pUL71 is present in virions and dense bodies. Within these particles pUL71 makes up a very small fraction of the total amount of protein present, around 0.1%. Thus, at the beginning of an infection there is very little pUL71 present (102). Being a beta gene product, expression of pUL71 ramps up later in infection. At 120 hours post infection, the majority of pUL71 is present within the cVAC, localized at Golgi and trans-Golgi network (TGN) membranes (38, 87, 88).

Roles of pUL71 in HCMV infection. pUL71 is involved in virion egress, cVAC formation, cell-to-cell spread, and virion maturation. In one set of pUL71-knockout experiments the virus failed to grow indicating that pUL71 is essential for virus replication (35). In another set of pUL71-knockout and pUL71 mutant virus experiments, intracellular and extracellular virus output decreases between 10 and 10,000-fold. In addition there are decreases in cVAC formation, plaque sizes indicative of reduced cell-to-cell spread, and secondary envelopment, hampering virion maturation (88, 104, 110).

Roles of pUL103 and pUL71 in HCMV infection. In dual pUL71/pUL103-knockout experiments, extracellular viral titers and plaque sizes were reduced compared to the wild type virus. However, there were no differences in these biological endpoints between the dual

pUL71/pUL103 knockout virus and the single pUL71- and pUL103-knockout viruses. This suggests that pUL71 and pUL103 work together as a complex, within the same pathway or have redundant functions during HCMV infection. In other studies, pUL71 and pUL103 were shown to interact outside the context of and during infection, evidence that further supports the two proteins forming a complex (14, 38, 72)

Protein interaction partners of pUL71. Studies have shown that pUL71 interacts with the ESCRT protein Vps4 during and outside the context of HCMV infection (38). Studies from the Pellett and von Einem laboratories have shown that pUL71 interacts with pUL103 (38, 72).

Characteristics of HSV pUL7 and pUL51

Genome information, expression kinetics, conservation, and sequence motifs/domains of HSV-1 pUL7. The UL7 gene is located in the unique long region of the genome. UL7 is a late gene, meaning it requires the expression of the immediate early and early genes for it to be transcribed and translated. The UL7 gene product pUL7 is a small (296 aa, ~32.56 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (74, 79). As mentioned above, the N termini of pUL7 homologs are more conserved than their C termini (Fig. 1-7) and they contain a conserved herpesvirus domain (Fig. 1-8) with the sequence G-F-x(8)-E-D-x-V-x(12)-R (71).

Localization of pUL7. pUL7 associates with capsids and is present in virions (71, 80). Being a gamma gene product, expression of pUL7 ramps up later in infection. At 12 hours post infection, the majority of pUL7 is present in the cytoplasm in the vicinity of the TGN, with some in the nucleus (14, 71).

Roles of pUL7 in HSV infection. pUL7 is important for virion egress, cell-to-cell spread, secondary envelopment, stabilization of pUL51, and maintenance of focal adhesions. When UL7

deletion mutant viruses are constructed, viral titers decrease 10-1,000-fold (3, 98). There are also decreases in the numbers of cytoplasmic enveloped capsids, indicating defects with secondary envelopment. When pUL7 is deleted, pUL51 expression decreases substantially compared to wild type. In addition, in the absence of pUL7 focal adhesions failed to form, which seemed to correlate with changes in cell morphology and attachment. Focal adhesions are contact sites between the extracellular matrix and the cytoplasm. They are complex and dynamic structures that play roles in regulating cellular processes in response to extracellular stimuli and cell attachment and movement (3). These structures also seem to be unique to alphaherpesviruses, as they are not present in the human beta- and gammaherpesviruses (14).

Protein interaction partners of pUL7. Several studies have shown that pUL7 interacts with pUL51 (3, 14, 80). One study showed that pUL7 interacts with the adenine nucleotide translocator 2 (ANT 2) (98). Another study, which developed a platform for identifying novel protein interactions between HSV-1 proteins, identified several tegument (pUL55, pUL14, pUL16) and an envelope (pUL45) proteins as potential interaction partners of pUL7 (7).

Genome information, expression kinetics, conservation, and sequence motifs/domains of HSV-1 pUL51. The UL51 gene is located in the unique long region of the genome. UL51 is a late gene, meaning it requires the expression of the immediate early and early genes for it to be transcribed and translated. The UL51 gene product pUL51 is a small (244 aa, ~26.84 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (34, 74, 79). The N termini of pUL51 homologs are more conserved than their C termini and they contain a conserved N-terminal cysteine residue (palmitoylation site) and tyrosine-based trafficking motif (YXX) (80, 81).

Localization of pUL51. pUL51 is present in virions and dense bodies. Being a gamma gene product, expression of pUL51 ramps up later in infection. Between 6 and 24-hours post infection, the majority of pUL51 is present in the cytoplasm, where it colocalizes with several Golgi marker proteins such as Golgi-58K, GM130, and COP. The cysteine at position 9, the palmitoylation site, is important for the translocation of pUL51 to Golgi membranes and other membranous structures within the cell (14, 70, 79).

Roles of pUL51 in HSV infection. pUL51 is important for virion egress, cell-to-cell spread, secondary envelopment, stabilization and localization of pUL7, maintenance of focal adhesions, and formation of viral protein dense areas in neuronal cells. When UL51 deletion mutant viruses are constructed, viral titers decrease ~ 10-300-fold. Plaque sizes also decreased by ~4.5-100-fold, indicating defects in cell-to-cell spread. The tyrosine at position number 19, the leading residue of the conserved tyrosine-based trafficking motif, is necessary and sufficient for cell-to-cell spread in certain cell types. In addition, lack pUL51 leads to decreases in the numbers of cytoplasmic enveloped capsids, indicating defects with secondary envelopment. When pUL51 is deleted, pUL7 expression decreases below the threshold of detection and is inefficiently translocated into virions. In addition, in the absence of pUL51 focal adhesions failed to form, which seemed to correlate with changes in cell morphology and attachment. Focal adhesions are contact sites between the extracellular matrix and the cytoplasm. They are complex and dynamic structures that play roles in regulating cellular processes in response to extracellular stimuli and cell attachment and movement. These structures also seem to be unique to alphaherpesviruses, as they are not present in the human beta- and gammaherpesviruses (3, 80, 81). Mutation of pUL51 tyrosine 19, as mentioned above, leads to loss of single areas of high viral protein concentration in

cancerous neuronal cells (103). These structures are reminiscent of the cVACs present in HCMV-infected cells, suggesting this maybe a common mechanism shared between the viruses.

Roles of pUL7 and pUL51 in HSV infection. In dual pUL7/pUL51-knockout experiments, extracellular viral titers and plaque sizes were reduced compared to the wild type virus. However, there were no differences in these biological endpoints between the dual pUL7/pUL51-knockout virus and the single pUL7- and pUL51-knockout viruses. This suggests that pUL7 and pUL51 work together as a complex within the same pathway and/or have redundant functions during HCMV infection. Outside of and during the context of infection, pUL7 and pUL51 have been shown to interact with each other, solidifying the fact that they do form a complex (3, 14).

Protein interaction partners of pUL51. Several studies have shown that pUL51 interacts with pUL7 (3, 14, 80). One study showed that pUL51 may also interact with itself and pUL35, the small capsomere-interacting protein (SCP). The SCP decorates the outer layer of the hexons of the capsid, seemingly as a bridge between the capsid and tegument (7, 74, 79, 111).

Characteristics of EBV BBRF2 and BSRF1

Genome information, expression kinetics, conservation, and sequence motifs/domains of BBRF2. The *BBRF2* gene is located in the unique long region of the genome (56). *BBRF2* is a late gene, meaning it requires the expression of the immediate early and early genes for it to be transcribed and translated. The *BBRF2* gene product BBRF2 is a small (278 aa, ~30.58 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (56, 60, 66, 74). As mentioned above, the N termini of BBRF2 homologs are more conserved than their C termini (Fig. 1-7) and they contain a conserved herpesvirus domain (Fig. 1-8) with the sequence G-F-x(8)-E-D-x-V-x(12)-R (71).

Localization of BBRF2. BBRF2 is present in very low abundance in virions (47). Thus, at the beginning of an infection there is very little BBRF2 present. Being a gamma gene product, expression of BBRF2 ramps up later in infection. Outside the context of infection, BBRF2 localizes to the nucleus and juxtanuclear regions in the cytoplasm in a punctate pattern. In the presence of its binding partner BSRF1, BBRF2 localizes only to the cytoplasm, in the vicinity of Golgi, in transfected cells (60, 108).

Roles of BBRF2 in EBV infection. BBRF2 is important for cell-to-cell spread and stabilization of BSRF1. When BBRF2-knockout and deletion mutant viruses were constructed, permissive cells infected, supernatant collected, and new permissive cells infected with collected supernatant, the number of cells infected is less than that for the wild type virus. under the exact same conditions. This suggests a defect in infectivity in the absence of BBRF2. In the absence of BBRF2, BSRF1 protein levels decrease. Given that BSRF1 is able to be ubiquitinated and its degradation is inhibited in the presence of MG132 (proteasome inhibitor), BBRF2 may protect BSRF1 from ubiquitination and subsequent degradation by the proteasome. (60).

Protein interaction partners of BBRF2. Two studies have shown that BBRF2 interacts with BSRF1outside the context of and during infection. BBRF2 has also been shown to interact with BGLF2, homolog of HSV-1 pUL16. BGLF2 is a tegument protein important for virion maturation, egress, and virus infectivity (60, 108). BBRF2 has also been shown to interact with several host proteins: C1orf50, CEP76, GOLGA2, OPTN, PNMA1, PTBP1, RINT1, TFCP2, and TTC12 (82).

Genome information, expression kinetics, conservation, and sequence motifs/domains of BSRF1. The *BSRF1* gene is located in the unique long region of the genome (56). *BSRF1* is a late gene, meaning it requires the expression of the immediate early and early genes for it to be

transcribed and translated (61, 66, 74). The *BSRF1* gene product BSRF1 is a small (218 aa, ~23.98 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (47, 66, 74). The N termini of BSRF1 homologs are more conserved than their C termini and they contain a conserved N-terminal cysteine residue (palmitoylation site) and tyrosine-based trafficking motif (YXX) (80, 81)

Localization of BSRF1. BSRF1 is present in low abundance in virions (47). Thus, at the beginning of an infection there is very little BSRF1 present. Being a gamma gene product, expression of BSRF1 ramps up later in infection. During infection, BSRF1 localizes to the cytoplasm, in the vicinity of Golgi membranes (108).

Roles of BSRF1 in EBV infection. BSRF1 is important for cell-to-cell spread in the marmoset B cell line B95-8. In the presence of siRNAs against the *BSRF1* gene, new virions produced under these conditions fail to infect as many cells as virions produced in the presence of a control siRNA. These defective virions also contain less viral DNA than virions produced under the control condition (108).

Protein interaction partners of BSRF1. Two studies have shown that BSRF1 interacts with BBRF2 outside the context of and during infection (60, 108). BSRF1 has also been shown to interact with BGLF3.5 (homolog of HSV1 UL14, encapsidation chaperone protein (ECP)) and BALF1 (one of two vBcl2s) (74, 108). BSRF1 has been shown to interact with several cellular proteins: S100A6, SLC16A3, TBXAS1, and ZDHHC17 (82).

Characteristics of KSHV ORF42 and ORF55

Genome information, expression kinetics, conservation, and sequence motifs/domains of ORF42. The *ORF42* gene is located in the central unique region of the genome (25). *ORF42* is a late gene, meaning it requires the expression of the immediate early and early genes for it to

be transcribed and translated. The *ORF42* gene product ORF42 is a small (278 aa, ~30.58 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (13, 66, 74). As mentioned above, the N termini of ORF42 homologs are more conserved than their C termini (Fig. 1-7) and they contain a conserved herpesvirus domain (Fig. 1-8) with the sequence G-F-x(8)-E-D-x-V-x(12)-R (71).

Abundance and localization of ORF42. ORF42 is present in virions. Being a gamma gene product, expression of ORF42 ramps up later in infection. At 96 hours post infection, the majority of ORF42 is present in the cytoplasm (13).

Roles of ORF42 in KSHV infection. ORF42 is important for virion egress and late protein expression. When ORF42 deletion mutant viruses are constructed, viral titers decrease 10-100-fold. In addition, in the absence of full-length ORF42, protein levels of several early and late proteins decreased. Further studies indicated that global synthesis of new proteins was decreased in the absence of full-length ORF42 (13).

Protein interaction partners of ORF42. Several studies have shown that ORF42 interacts with ORF55 outside the context of infection (14, 60) In another study ORF42 is predicted to interact with the protein NME2 (29).

Genome information, expression kinetics, conservation, and sequence motifs/domains of ORF55. The *ORF55* gene is located in the central unique region of the genome (101). *ORF55* is predicted to be an early or late gene based on the expression kinetics of its homologs in HCMV (pUL71, early gene), HSV-1 (pUL51, late gene), and EBV (BSRF1, late gene). The *ORF55* gene product ORF55 is a small (227 aa, ~24.97 kDa), conserved tegument protein, with homologs in all viruses within the *Herpesviridae* (66, 74). The N termini of ORF55 homologs are more conserved

than their C termini and they contain a conserved N-terminal cysteine residue (palmitoylation site) and tyrosine-based trafficking motif (YXX) (80, 81).

Localization of ORF55. ORF55 is present in virions (52). Outside the context of infection, ORF55 has been shown to accumulate almost exclusively in the cytoplasm, colocalizing with Golgi membranes (14, 86). During infection, ORF55 primarily localizes to the cytoplasm, with little protein present in the nucleus (52).

Roles of ORF55 in KSHV infection. ORF55 is important for cell-to-cell spread and virion assembly. When permissive cells are incubated with a vBcl2 peptide that inhibits the interaction between ORF55 and vBcl2, lytic infection induced, supernatant collected, and new permissive cells infected with virions from supernatant, the number of cells infected is less than the number produced in the presence of the control peptide. Also, when the ORF55/vBcl2 interaction is disrupted less intracellular virions are produced (52).

Protein interaction partners of ORF55. Studies have shown that ORF55 interacts with ORF42 outside the context of infection (14, 60) In other experiments ORF55 has been shown to interact with the cellular proteins GOSR1 and STING (component of the cGAS/STING DNA sensing pathway) and the viral protein vBcl2 (ORF16) (29, 40, 52).

This work

Summary. Data from the Pellett laboratory supports the hypothesis that pUL103 has multiple roles throughout HCMV lytic infection and its various activities depend upon its structural domains and presence at specific times during infection. My objectives were to determine the structure-function relationship of pUL103 in HCMV infection and to identify the temporal window when pUL103 was required for cVAC biogenesis and maintenance for focused mechanistic studies. From my structural studies, I found that the predicted structure of pUL103 most closely

resembles that of the non-catalytic LIM domain of human terminal uridylyl transferase 4 (TUT4), a nucleotidyl transferase domain that lacks catalytic residues. Based on its predicted structure, pUL103 may thus function to bind RNA species for catalysis. This data provides the foundation for studies into whether pUL103 is capable of binding nucleic acids. During my biological studies to determine the period when pUL103 is required for cVAC biogenesis and maintenance, I found that late in infection, pUL103 needs to be present by 72 hours for normal cVAC biogenesis to occur. In addition, pUL103 needs to remain present longer than 84 hpi for the percentage of cVACs to be maintained during infection. This data helps us in achieving our long-term goal of identifying and characterizing viral proteins as potential targets for novel, diverse, and specific HCMV antivirals.

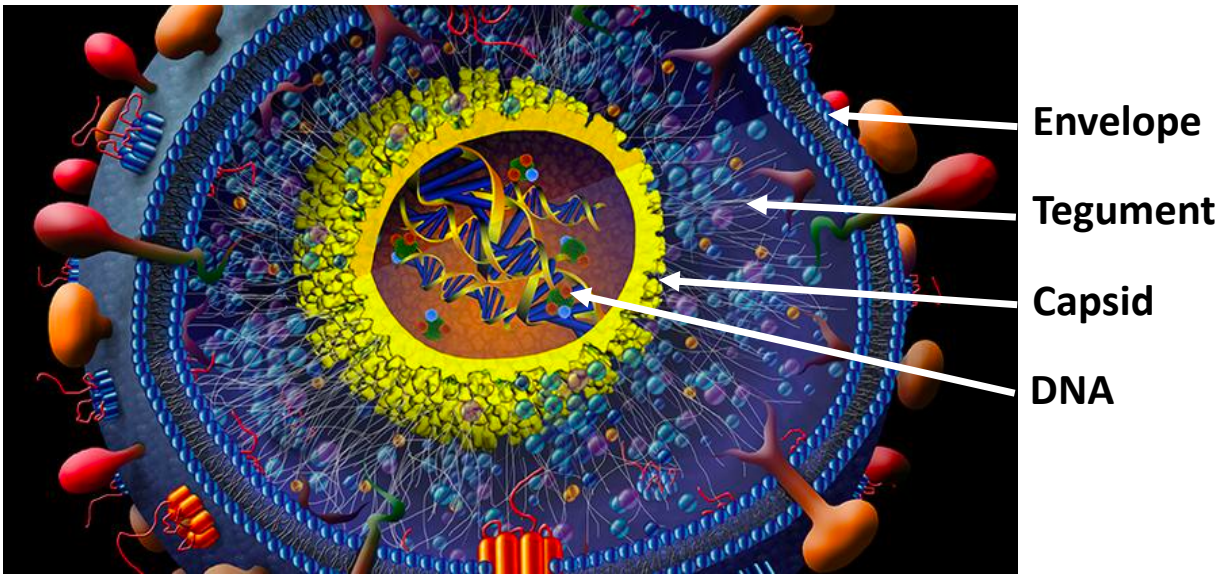

FIG 1-1 Diagram of a herpesvirus virion. Illustration showing the four main components (envelope, tegument, capsid, DNA) of an infectious herpesvirus particle. Image modified from that in reference (97).

Table 1-1 Characteristics of the three subfamilies of the *Herpesviridae*

Herpesvirus Subfamily	Host ranges	Reproductive cycles (length)	Infection in cell culture	Efficient destruction of cells	Source of primary infection	Latency reservoirs
Alphaherpesvirinae	Broad	Short	Spreads rapidly	Yes	Mucocutaneous tissue, epithelial cells, and fibroblasts	Neuronal dorsal root ganglia
Betaherpesvirinae	Narrow	Long	Progresses slowly, enlarges cells	No	Fibroblasts, T, endothelial, and epithelial cells	Lymphoid tissues, hematopoietic progenitor cells
Gammaherpesvirinae	Restricted	?	Induces lymphoproliferation and cancers	No	B, T, epithelial, and endothelial cells	B, T, epithelial, and endothelial cells

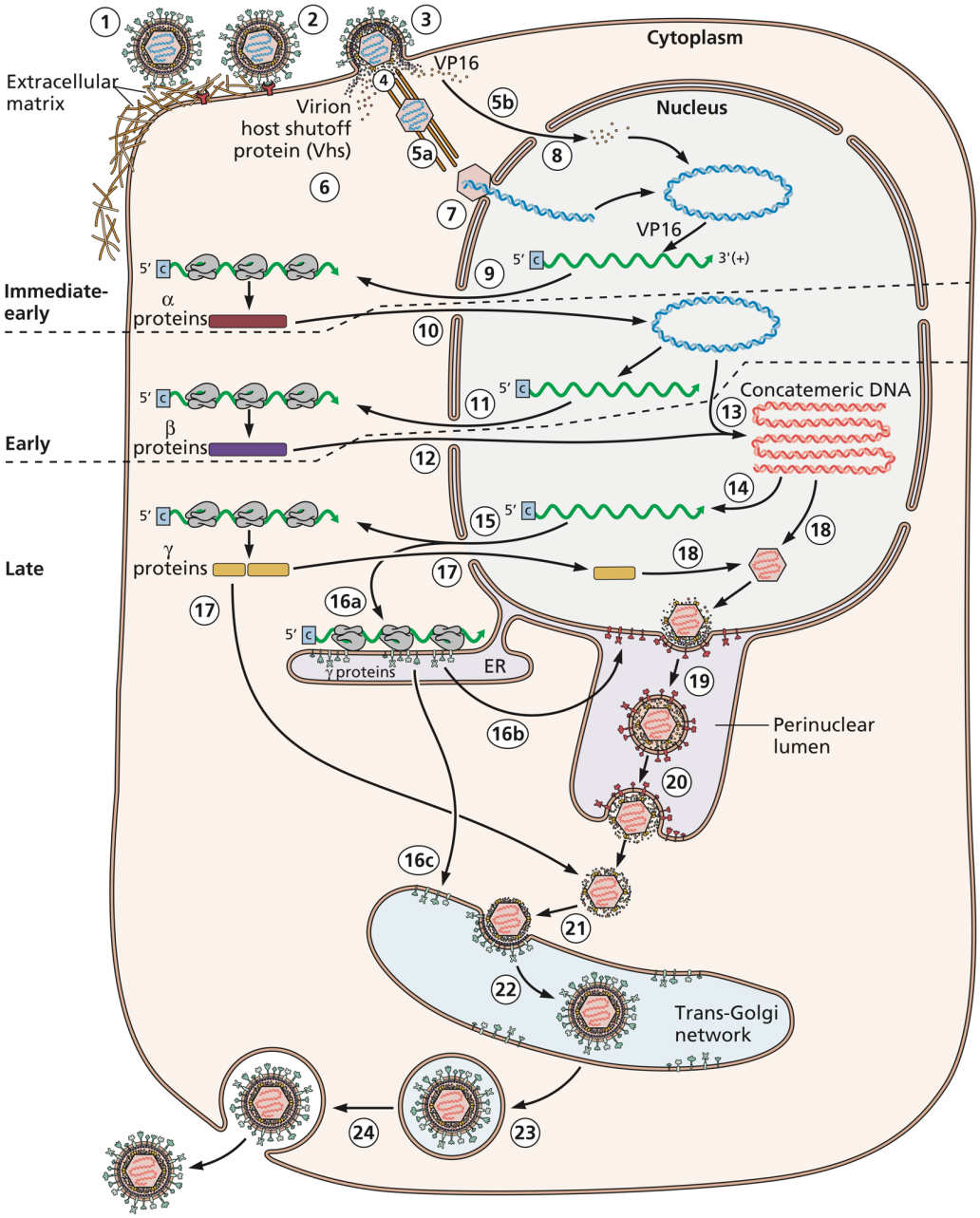

FIG 1-2 General overview of the herpesvirus replication cycle. Viral infection begins with attachment, followed by fusion of the cell membrane and viral envelope, entry of the capsid and tegument into the cytoplasm (steps 1-4). The capsid is trafficked to the nuclear pore and the DNA deposited into the nucleus (steps 5a-7). The viral cascade of gene expression begins with transcription and translation of the immediate early genes, followed by the early and late genes (steps 8-18). Capsids are assembled in the nucleus, acquire their inner teguments, undergo primary envelopment within the inner nuclear membrane, and are translocated into the perinuclear space (steps 18-19). The capsids undergo de-envelopment and are transported into the cytoplasm by fusion of their envelope with the outer nuclear membrane (step 20). In the cytoplasm, the capsids acquire their outer tegument, undergo secondary envelopment, and are translocated to the cell surface for egress (steps 21-24). Image is from reference (39).

Table 1-2 Betaherpesvirus genes conserved across the *Herpesviridae*

Function (gene name)	HCMV	HHV-6A, HHV-6B, and HHV-7	HSV
Gene regulation			
Multifunctional regulator of expression (MRE)	UL69	U42	UL54
Nucleotide metabolism			
Ribonucleotide reductase, large subunit (RR1)	UL45	U28	UL39
Uracil-DNA glycosylase (UNG)	UL114	U81	UL2
Deoxyuridine triphosphatase (dUTPASE)	UL72	U45	UL50
DNA replication			
Helicase/primase complex			
ATPase subunit (HP1)	UL105	U77	UL5
RNA pol subunit (HP2)	UL70	U43	UL52
subunit C (HP3)	UL102	U74	UL8
DNA polymerase (POL)	UL54	U38	UL30
ssDNA-binding protein (SSB)	UL57	U41	UL29
DNA polymerase processivity subunit (PPS)	UL44	U27	UL42
Virion			
Nonstructural; roles in virion maturation			
Alkaline exonuclease (NUC)	UL98	U70	UL12
Capsid transport nuclear protein (CTNP)	UL52	U36	UL32
Terminase-binding protein (TERbp)	UL51	U35	UL33
Terminase (TER)			
TER ATPase subunit (TER1)	UL89	U66	UL15
TER DNA recognition subunit (TER2)	UL56	U40	UL28
Assembly protease (PR)	UL80a	U53	UL26
Assembly protein precursor (pAP)	UL80.5	U53a	UL26.5
Capsid nuclear egress complex			
Nuclear egress membrane protein (NEMP)	UL50	U34	UL34
Nuclear egress lamina protein (NELP)	UL53	U37	UL31
Capsid			
Major capsid protein (pentons and hexons; MCP)	UL86	U57	UL19
Portal protein (PORT)	UL104	U76	UL6
Portal capping protein (PCP)	UL77	U50	UL25
Capsid triplex			
Monomer (TRI1)	UL46	U29	UL38
Dimer (TRI2)	UL85	U56	UL18
Small capsid protein (SCP) at hexon tips	UL48.5	U32	UL35
Tegument			
Encapsidation and egress protein (EEP)	UL103	U75	UL7
Myristoylated/palmitoylated cytoplasmic egress tegument protein (CETP)	UL99	U71	UL11
Virion protein kinase (VPK)	UL97	U69	UL13
Encapsidation chaperone protein (ECP)	UL95	U67	UL14
CETP-binding protein (CETPbp)	UL94	U65	UL16
Capsid transport tegument protein (CTTP)	UL93	U64	UL17
Cytoplasmic egress facilitator 2 (CEF2)	UL87	U58	UL21
Cell-to-cell fusion inhibitor	UL76	U49	UL24
Large tegument protein (LTP)	UL48	U31	UL36
LTP-binding protein (LTPbp)	UL47	U30	UL37

Cytoplasmic egress facilitator 1 (CEF1)	UL71	U44	UL51
Envelope			
Glycoprotein B (gB)	UL55	U39	UL27
Glycoprotein H (gH)	UL75	U48	UL22
Glycoprotein L (gL)	UL115	U82	UL1
Glycoprotein M (gM)	UL100	U72	UL10
Glycoprotein N (gN)	UL73	U46	UL49.5

Table from reference (20), adapted from information in references (74) and (68) and proteins named according to reference (65).

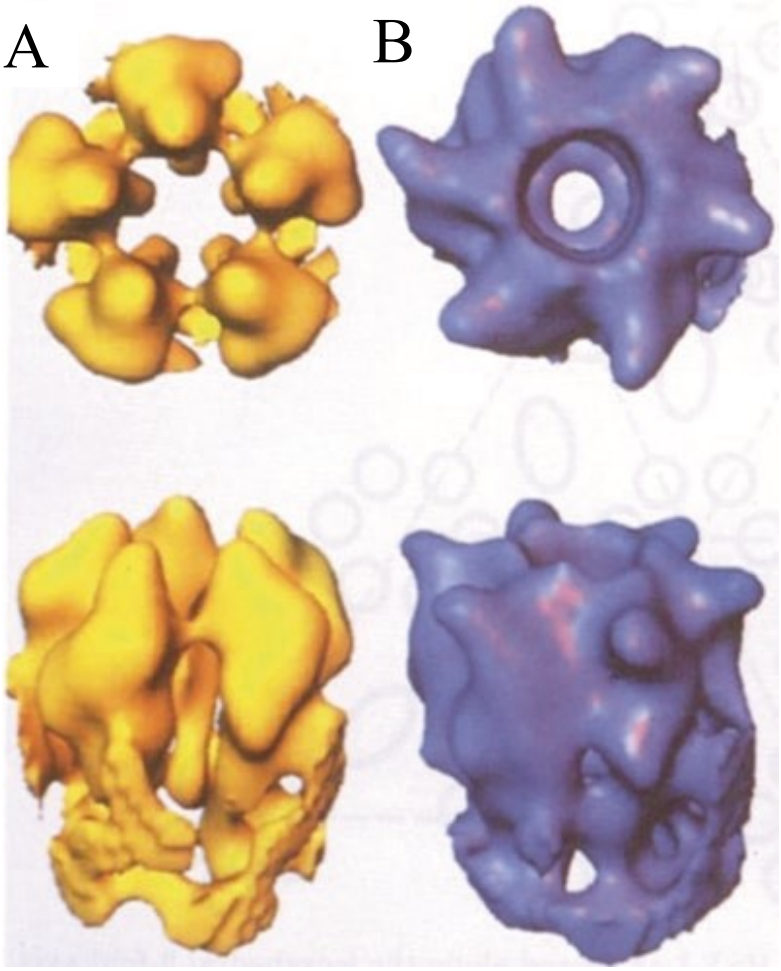

FIG 1-3 Three- dimensional reconstructions from cryoelectron microscopy images of herpes simplex virus type 1 pentons and hexons. (A) Top and side views of penton. (B) Top and side views of hexon. Images are representative of the analogous structures of HCMV. Image is from reference (114).

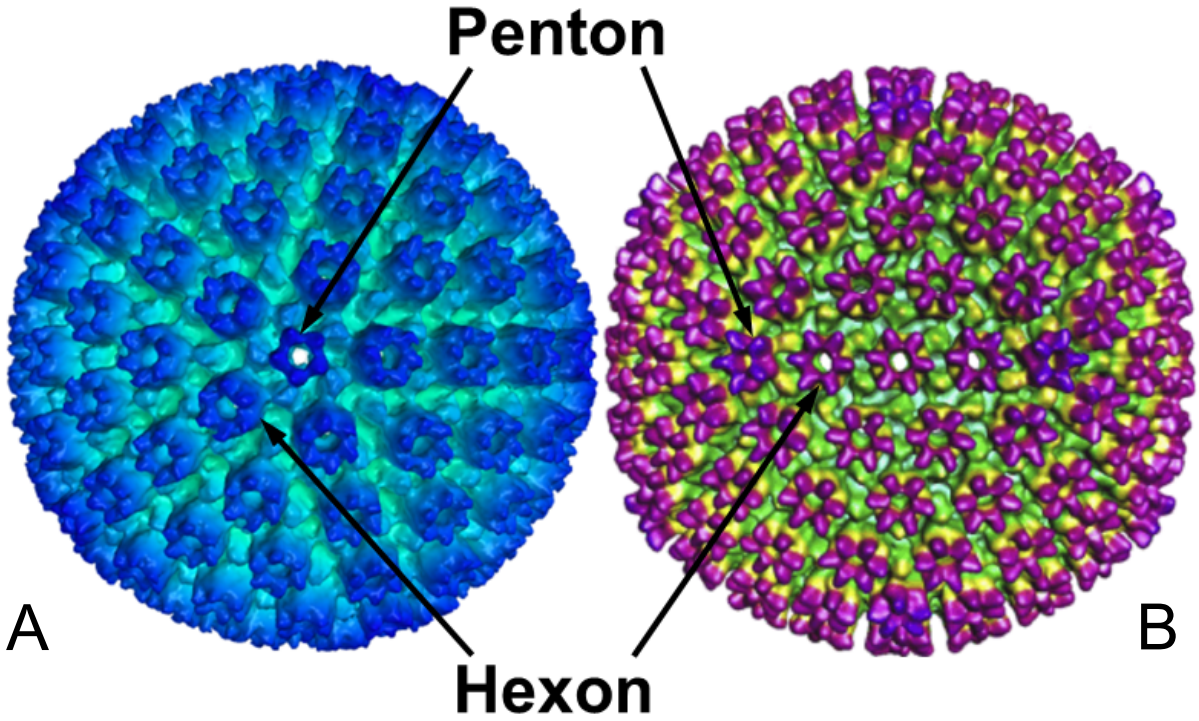

FIG 1-4 Three- dimensional reconstructions of cryoelectron microscopy images of HCMV nuclear capsids. (A) View along an icosahedral five-fold symmetry axis. (B) View along an icosahedral two-fold symmetry axis. Images are from references (24) and (112).

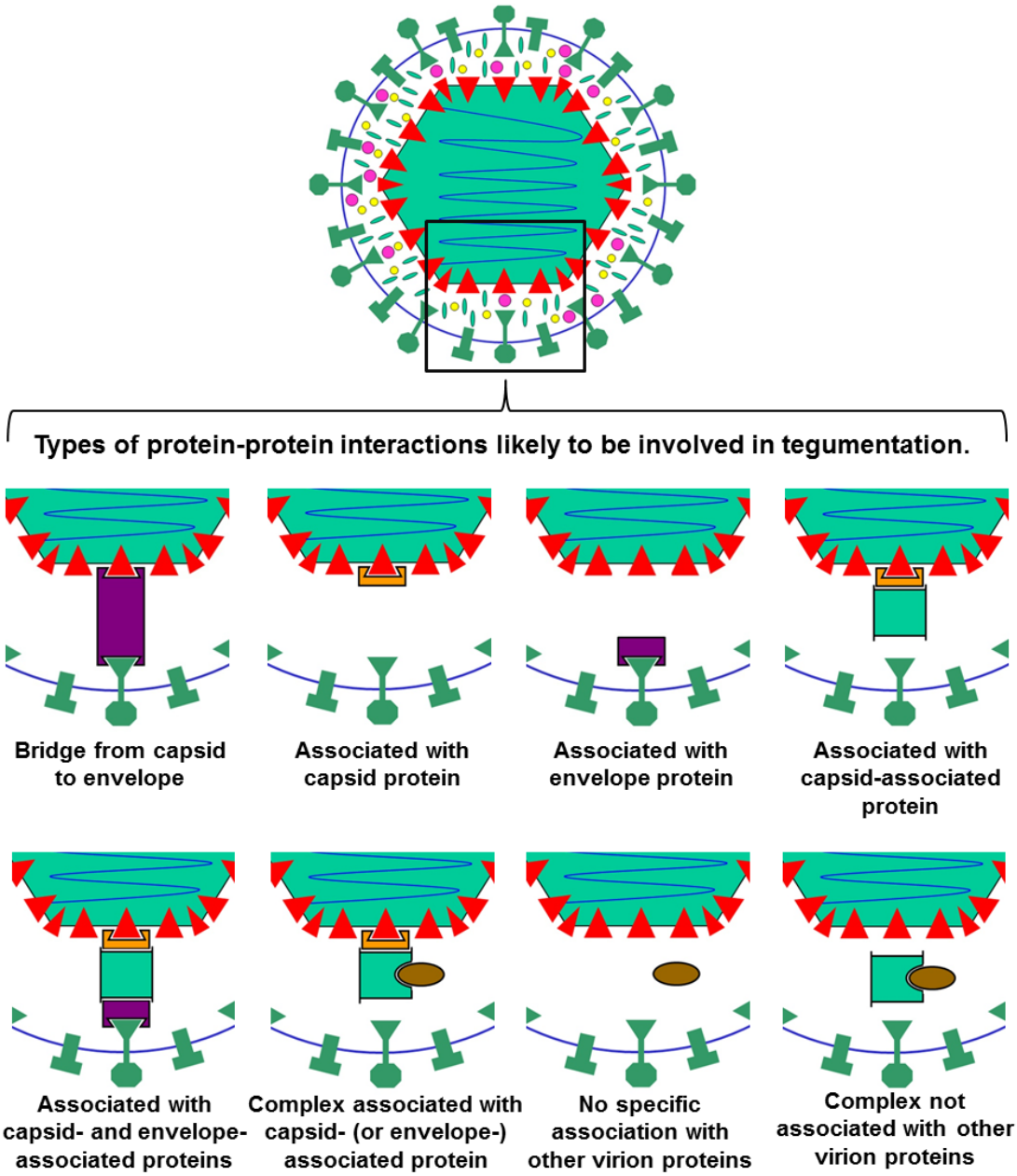

FIG 1-5 Protein-protein interactions within the tegument. The schematic represents a mature enveloped virion and the types of protein-protein interactions proven or predicted to occur during tegumentation. Image is from reference (20).

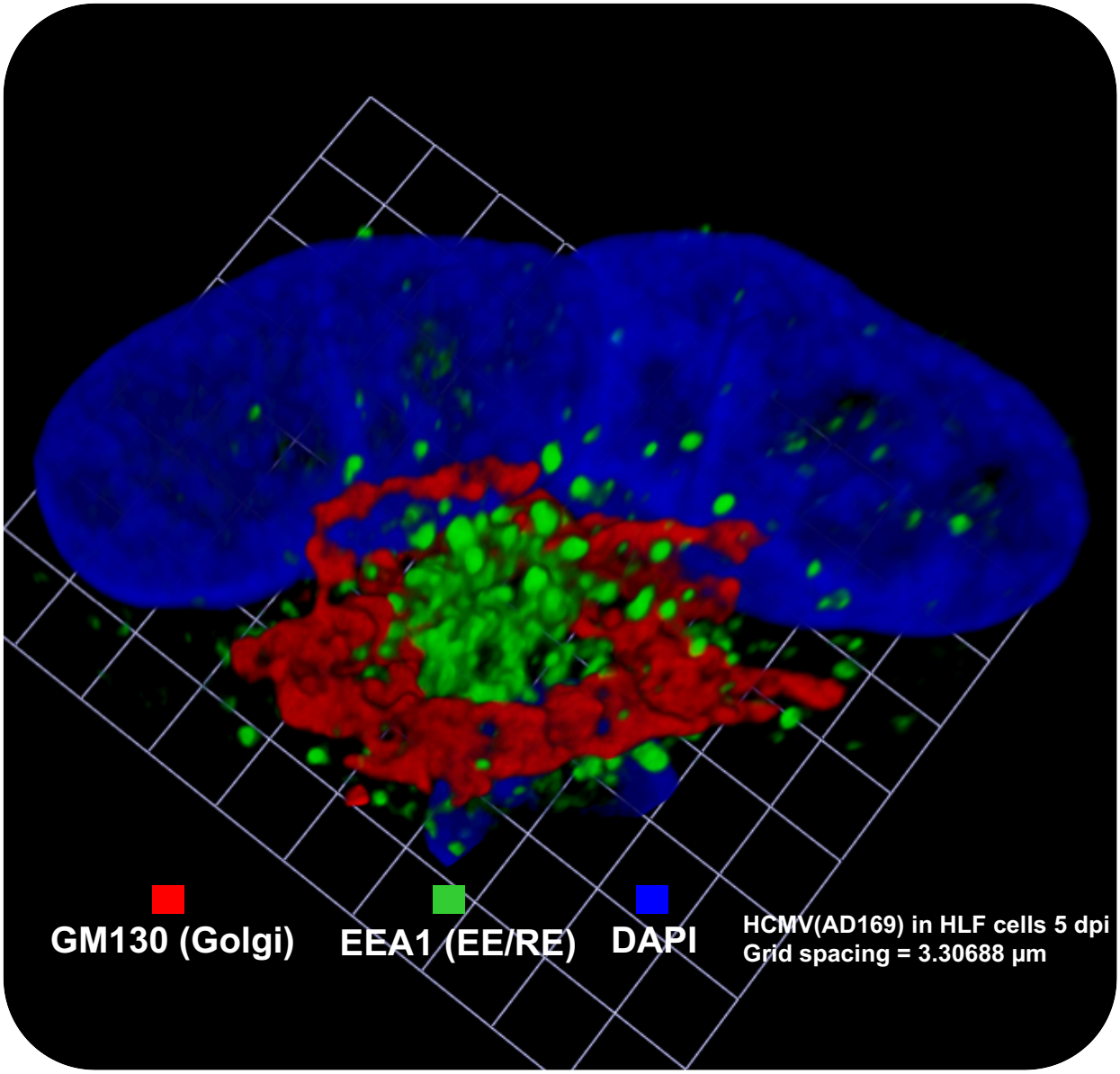

FIG 1-6 Immunofluorescence assay (IFA) image of the cVAC in an HCMV-infected cell. IFA of a confocal 3D reconstruction of a human lung fibroblast infected with HCMV showing a cVAC: a Golgi ring surrounding primarily early and recycling endosomes. Image is from reference (27).

Table 1-3 Characteristics of HCMV pUL103, HCMV pUL71 and their homologs HSV-1 pUL7, HSV-1 pUL51, EBV BBRF2, BSRF1, KSHV ORF42, and KSHV ORF55

Protein properties	HCMV pUL103	HCMV pUL71	HSV-1 pUL7	HSV-1 pUL51	EBV BBRF2	EBV BSRF1	KSHV ORF42	KSHV ORF55
In virions	+	+	+	+	+	+	+	+
Capsid-associated			+					
Localization:								
Nuclear	+		+		+ (in absence of BSRF1)			+
Cytoplasmic	+	+	+	+	+	+	+	+
with Golgi markers	+	+	+	+		+		+
Size (amino acids)	249	361	296	244	278	218	278	227
Expression kinetics	Late	Early	Late	Late	Late	Late	Late	
Interacts with:	pUL71	pUL103	pUL51	pUL7	BSRF1	BBRF2	ORF55	ORF42
Sub family	Beta	Beta	Alpha	Alpha	Gamma	Gamma	Gamma	Gamma
Virus activities	**HCMV pUL103**	**HCMV pUL71**	**HSV-1 pUL7**	**HSV-1 pUL51**	**EBV BBRF2**	**EBV BSRF1**	**KSHV ORF42**	**KSHV ORF55**
Virion egress	+	+	+	+				
Cell-to-cell spread	+	+	+	+	+	+	+	+
Secondary envelopment	+	+	+	+				
cVAC biogenesis	+	+		+*				
Maintenance of focal adhesions			+	+				
Stabilization of proteins			+	+	+			
Protein expression							+	
Intracellular virus production								+

+ indicates the protein has the given property. Blank fields indicate a lack of information on the given property for that protein. If Information in the table was obtained from the following references: HCMV pUL103: (1, 26, 35, 38, 64, 72, 74, 102); HCMV pUL71: (3, 38, 74, 87, 88, 102, 104, 110); HSV-1 pUL7: (3, 14, 71, 74, 79, 80); HSV-1 pUL51: (3, 14, 34, 70, 74, 79-81, 103) (Reference (103) shows pUL51 is involved in formation of single, viral protein dense areas similar to HCMV cVACs);EBV BBRF2: (47, 56, 60, 66, 108); EBV BSRF1: (47, 61, 66, 74, 108); KSHV ORF42: (13, 14, 60, 74); KSHV ORF55: (20, 52, 66, 74, 86)

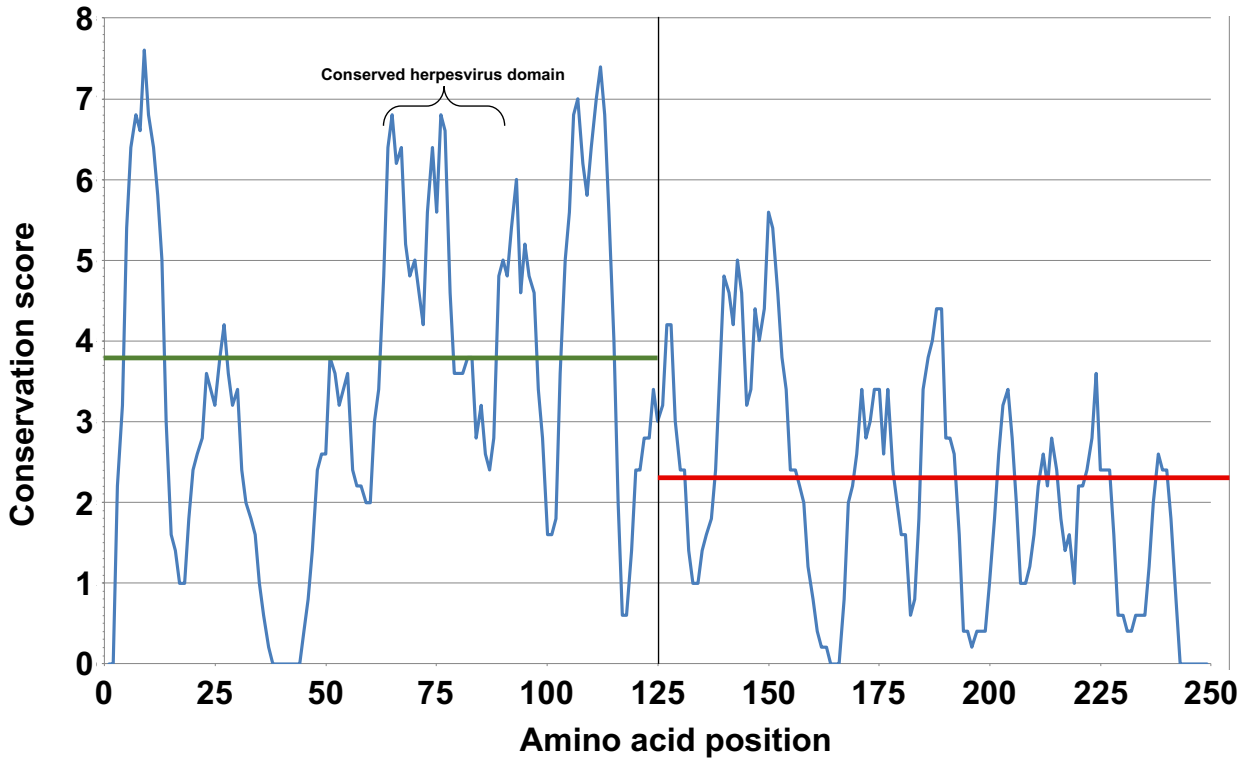

FIG 1-7 FIG 1-7 The N terminus (residues 1-125) of pUL103 is more conserved than the C terminus (residues 126-249). Similarity plot of pUL103 and 59 of its homologs from across the three subfamilies of the *Herpesviridae* with a sliding window of 5, generously provided by Dr. Richard Roller. *y*-axis scale: 0 (no similarity across residues) to 8 (identical residues). The green and red lines represent the average conservation scores of the N and C termini, respectively. The conserved herpesvirus domain consists of residues 65-91. Positions that have conservation scores of 0 represent gaps within the alignment

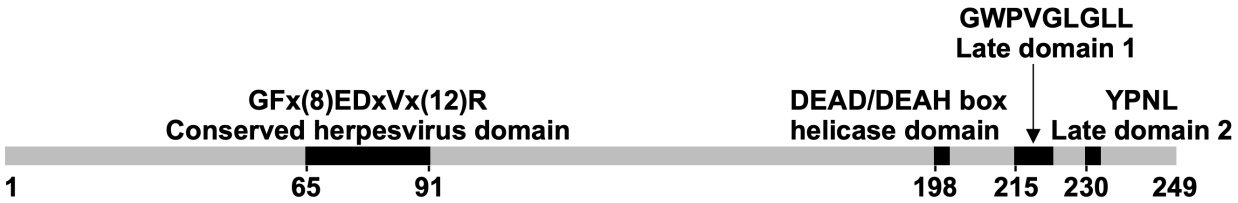

FIG 1-8 Diagram of the pUL103 amino acid sequence with confirmed and putative sequence domains. There are several literature sources documenting the conserved herpesvirus domain in pUL103 homologs. The putative DEAD/DEAH box and the late domains were recently identified in the pUL103 amino acid sequence by our laboratory. Further studies are needed to fully validate their functions. Image modified from that in reference (72).

CHAPTER 2: PROTEIN SEQUENCE ANALYSIS AND STRUCTURAL PREDICTION OF PUL103 HOMOLOGS

Introduction

HCMV has the largest genome of the nine human herpesviruses (33, 74) Its genome consists of 236 kb that encodes a multitude of gene products; these include at least 167 proteins, four large, noncoding RNAs, two oriLyt RNAs, and at least 23 microRNAs. HCMV is a master manipulator. This large genome and its corresponding products ensure that HCMV has everything necessary to prepare the intracellular environment for efficient production of new virions (20, 64). Preparing a cell to assemble and produce new virus particles requires a great deal of coordination by the virus's tegument.

The tegument is a layer of proteins and RNA that lies beneath the viral envelope and is present only in herpesviruses. Though it appears to lack a well-defined structure, there is some semblance of organization to this virion component (20). After the viral envelope fuses with the cellular membrane or an intracellular membrane after endocytosis, the tegument is immediately deposited into the cytoplasm where the components begin modulating the cellular environment to facilitate virion production (94). There are at least 38 different tegument proteins, present in various quantities, within HCMV virions. Tegument proteins are important for many processes throughout infection. Some of these proteins have roles in the virus life cycle such as trafficking of the capsid to and release of the viral DNA into the nucleus, transcriptional regulation, or virion assembly. Other tegument proteins have roles in modulating cellular processes and pathways. This includes but is not limited to the evasion of the cell's immune response, inactivation of host cell transcriptional repression mechanisms, alteration of the cell cycle, manipulation of intrinsic host defenses, cell death and signaling pathways (64). Many, if not most tegument proteins,

perform multiple, distinct functions. This is true for the tegument protein pUL103 that is the focus of my work.

pUL103 is a conserved HCMV tegument protein that has many roles throughout HCMV infection (26, 72, 74). In knockout experiments, pUL103 has been shown to be important for production of extracellular virus, with viral titers decreasing between 100 – 10,000 fold in its absence (35, 110). In experiments where pUL103 protein levels were able to be regulated, knockdown of pUL103 resulted in decreases in cVAC biogenesis, cell-to-cell spread, and secondary envelopment (26). The homologs of pUL103 in other herpesviruses have also been examined for their roles in virus infection. HSV-1 pUL7 plays roles in virion egress, cell-to-cell spread, secondary envelopment, stabilization of pUL51, and maintenance of focal adhesions (3, 98). EBV BBRF2 is needed for progeny virus infectivity and stabilization of BSRF1 (60). KSHV ORF 42 is necessary virion egress and late protein expression (13). Though several of the activities of these proteins are known, the exact mechanisms of how they perform their functions are unknown, along with their structures.

To address these unknowns, I analyzed the protein sequences of pUL103 and its homologs using the Protein BLAST through NCBI, a multiple sequence alignment, and the online server I-TASSER. The Protein BLAST is an alignment tool that can perform pairwise alignments between pairs of amino acid sequences and identify other proteins, potential homologs, with similar sequences to a query sequence (4). A multiple sequence alignment is an alignment of numerous sequences by similarity of their nucleotides or amino acids. Of the several programs that can be used to produce multiple sequence alignments, e.g., MUSCLE, MAFFT, CLUSTALW, T-Coffee, I used MUSCLE. From such analyses, conserved amino acids, motifs, regions, etc., can be identified (36, 57).

I-TASSER is an online server, created by the Zhang laboratory at the University of Michigan-Ann Arbor, that produces structure and function predictions from an amino acid sequence. The I-TASSER algorithm is quite complex but involves (1) retrieving protein structures from the protein data bank (PDB) and using them as templates for the predicted folds of the protein of interest, (2) complete models are built and undergo refinement, (3) the predicted structures undergo more refinement and structures with the least potential energies are identified, and (4) the final predicted structure is compared to crystallized structures of proteins to identify structural analogs and hints of possible functions (109). There are benefits and downsides to structural predictions. The benefits include having access to a structure, the ability to design and execute experiments *in silico*, and to obtain real results that can help guide and inform future experiments. The downsides include not knowing whether the predicted structure is the actual structure until the protein is crystallized and possibly wasting time generating erroneous data. I-TASSER addresses these downsides by participating in the Critical Assessment of Techniques for Protein Structure Prediction (CASP). CASP is a community-wide experiment that has been held every two years since 1994. This experiment tests the state-of-the-art of protein structure prediction software from investigators internationally. The experiment, in the form of a competition, is blinded so the structures of proteins to be tested are unknown to the participants. I-TASSER has participated in the CASP experiments since 2006. During that time, it has been the highest ranked server for automated protein structure prediction. It was only recently knocked into the number two position by another server produced by the Zhang lab, QUARK at CASP 13 in 2018. (51, 92, 109).

Methodology

Protein BLAST. The HCMV pUL103, HSV-1 pUL7, and EBV BBRF2 protein sequences were used as input into the Protein BLAST tool to perform pairwise alignments (19).

I-TASSER (Iterative Threading ASSEmbly Refinement). An on-line server that uses the I-TASSER algorithm to produce protein structure and function predictions from amino acid sequences (109). The amino acid sequences of pUL103 and its well-studied homologs pUL7 and BBRF2 were provided to the I-TASSER server and output, consisting of but not limited to predicted structures and structural analogs, was analyzed.

MegAlign Pro by DNAstar. A computer software program that performs multiple alignments of large DNA, RNA, and protein sequences and offers four alignment methods: MUSCLE, Mauve, MAFFT, and Clustal Omega. MUSCLE, using default parameters, was used to align the protein sequences of pUL103 and its homologs (36).

Nanoscale Molecular Dynamics and Visual Molecular Dynamics (NAMD/VMD). Computer software that simulates the dynamic molecular movements of large biomolecular systems. This allows for analysis of native, complexed, predicted, or known protein structures.

Phylogeny.fr. A simple to use web service dedicated to reconstructing and analyzing phylogenetic relationships between molecular sequences. The program runs and connects various bioinformatics programs to reconstruct a robust phylogenetic tree from a set of sequences (30, 31).

PRESTO (a Phylogenetic tReE viSualisaTiOn). An online phylogenetic tree visualization tool provided by the ATGC Montpellier Bioinformatics Platform. This platform provides several tools for conducting analyses in the following areas: phylogenetics, molecular evolution, genome dynamics, comparative and functional genomics, and transcriptomics.

PyMOL. A computer software program that allows for the visualization and manipulation of small molecules and biological macromolecules such as proteins (6). PyMOL was used to visualize and align the predicted structures of pUL103 and its homologs as well as crystalized structures of proteins.

Similarity plot. A similarity plot of pUL103 and homologs was created by obtaining sequences from the Protein BLAST search of the HCMV pUL103 protein sequence. The sequences were aligned using MUSCLE, with default parameters, from the European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI). Once aligned, the sequences were imported into Jalview, a sequence editor, where conservation scores at all the amino acid positions were observed. The conservation score is measured as a numerical index reflecting the conservation of physical and chemical properties of amino acids within the alignment. Identities score the highest followed by groups that contain substitutions to amino acids with similar properties. The conservation scores were imported into Microsoft Excel and graphed (4, 19, 36, 54, 57).

Results

pUL103 and its homologs share limited sequence identities. To determine how similar pUL103 homologs are, I performed pairwise alignments using the Protein BLAST tool through the National Center of Biotechnology Information (NCBI) on the protein sequences of HCMV pUL103, HSV-1 pUL7, and EBV BBRF2. pUL103 homologs contain a conserved domain known as the conserved herpesvirus domain that is also present in the topoisomerase III of fission yeast. It is located in between residues 65 and 91 in pUL103, between residues 101 and 127 in HSV-1 pUL7, and in between residues 74 and 100 in EBV BBRF2. In all sequences, this domain is located closer to the amino terminus than the carboxyl terminus. Using the Protein BLAST tool, I determined that pUL103 shares ~29% sequence identity with pUL7 and ~30% with BBRF2. pUL7 shares ~31% sequence identity with BBRF2 (Fig. 2-1 A-C). The majority of the shared identities between all three sequences are located within close proximity to the conserved herpesvirus domain. In addition, HCMV pUL103 and HCMV pUL7 share four amino acids within a stretch

of five residues near their carboxyl ends. The residues are rich in prolines with a putative motif of LPxPP. These proline-rich residues may be used to bind ESCRT machinery or SH3-domain containing proteins (2, 83). BBRF2 shares several residues with pUL103 at its carboxyl end (Fig. 2-1 A-B). From their identities, it appears that these homologs are not closely related. To better elucidate the homology of these sequences, I decided to perform a phylogenetic analysis of all the pUL103 homologs from the human herpesviruses as well as those from several animal cytomegaloviruses, a total of fourteen sequences (Fig. 2-2).

Phylogenetic analysis reveals relationships amongst pUL103 homologs. To better understand the ancestral relationships of the pUL103 homologs, I utilized bioinformatics software to produce a phylogenetic tree from their protein sequences (Fig. 2-3). The tree showed homologs from viruses within the same subfamilies clustered together. It also showed the branch lengths of homologs from closely related viruses human herpesviruses 6A and B and herpes simplex viruses 1 and 2 as being zero or very short, respectively. This information correlates with what is already known about the viruses. The tree showed that the homologs from the alphaherpesviruses are more closely related to the homologs from the gammaherpesviruses; whereas gammaherpesviruses themselves are more closely related to betaherpesviruses. This may be factual or the common ancestor for all three subfamilies goes back so far, the software had difficulty determining the proper branching points. Within the betaherpesvirus homologs, those belonging to the roseoloviruses cluster together, those from the rodent cytomegaloviruses cluster together, and those from the primate cytomegaloviruses cluster together. Within these clusters, the mouse pUL103 homolog is more closely related to the rat homolog than the one from GPCMV and pUL103 is more closely related to homolog pRh140 than the pUL103 homolog from CCMV. This information helped to highlight how the pUL103 homologs are related. After conducting a

phylogenetic analysis, I performed a sequence alignment of the homologs to look for conserved regions.

MUSCLE alignments show areas of conservation across pUL103 homologs. To identify conserved regions (motifs) and/or specific amino acids important for the function of pUL103, I aligned the protein sequences of the fourteen pUL103 homologs using MUSCLE (Multiple Sequence Comparison by Log-Expectation) (Fig. 2-4). The alignment showed (a) conserved side chain chemistry of the residues throughout the length of the pUL103 sequence, (b) numerous positions with perfectly conserved residues within and across the subfamilies, (c) the largest concentration of perfectly conserved residues lies within the conserved herpesvirus domain, (d) potential conserved motifs near the amino and carboxyl termini of the cytomegaloviruses (Fig. 2-5), (e) eight dileucine motifs present in the pUL103 sequence, and (f) the gamma herpesvirus subfamily has the most positions with perfectly conserved residues. One piece of information we hoped to gain from the alignment is the identification of residues suitable for mutagenesis. These would be residues that are conserved in two of the three herpesvirus subfamilies. The majority of these residues are present in the conserved herpesvirus domain (65-G, 66-Y, 75-E, 76-D, 91-R). However, there are some residues outside of this domain that are also conserved and would be interesting to study (7-R, 11-E, 53-Y, 63-F, 96-K, 103-F, 106-C, 113-E, 170-P, 191-F, 192-F, 214-D). This alignment shows how similar HCMV pUL103 and its homologs are. Prior to conducting my alignment, Dr. Richard Roller from the University of Iowa performed a MUSCLE alignment of pUL103 and 59 of its homologs and used their conservation scores to produce a similarity plot (Fig. 2-6). A similarity plot shows residue similarity at each position along the length of the alignment. This plot correlates with the information that is seen in my alignment; the peaks of the similarity plot align with the highly conserved residues I identified. Using the similarity plot from

Dr. Roller and a sliding window of 5, I showed that the N terminus of pUL103 homologs is more conserved than the C terminus. I determined the average conservation score of the N terminus (residues 1-125) to be 3.8, and greater than that of the C terminus (residues 126-249), which had an average conservation score of 2.3 (Fig. 2-7). After determining the extent of the similarity of these proteins through their sequences, I wanted to see how this information correlated with their structures.

I-TASSER structure and function predictions of pUL103 homologs provide insight into the functions of these proteins. When I began my analysis, there were no crystal structures of any of the pUL103 homologs. Thus, to gain insight into the structures of these proteins, I utilized I-TASSER (Iterative Threading ASSEmbly Refinement), an online server for protein structure and function prediction. Using I-TASSER, I produced predictions of the structures for HCMV pUL103, and its homologs HSV-1 pUL7 and EBV BBRF2 (Fig. 2-8 A-C). The predicted structure of pUL103 is modular with a primarily unstructured amino terminus with some beta strands and a primarily alpha-helical carboxyl terminus with an unstructured tail. The predicted structure of pUL7 resembles that of pUL103 and also contains an elongated, primarily alpha helical N terminal region. The predicted structure of BBRF2 contains many alpha helices all throughout the structure and does not share the same shape common to the structures predicted for pUL103 and pUL7. However, the predicted secondary structures of all three homologs, also produced by I-TASSER, are similar (data not shown). When the predicted structures of pUL103 and pUL7 are aligned using PyMOL, they align very well (Fig. 2-9 A). When the predicted structures of pUL103 and BBRF2 are aligned, there is hardly any overlap in the structures (Fig. 2-9 B).The differences between the predicted structures of pUL103, pUL7 and BBRF2 could be due to BBRF2 possibly having different or additional functions compared to those of pUL7 and pUL103, different tertiary

structures leading to similar functions, or the lack of appropriate proteins in the protein data bank for appropriate threading by I-TASSER. Having the predicted structure of pUL103 allowed me to conduct other analyses in addition to the structural alignments with its homologs.

In collaboration with the Kovari laboratory, I conducted molecular dynamics simulations using Nanoscale Molecular Dynamics and Visual Molecular Dynamics (NAMD/VMD) of the predicted structure of pUL103. This software simulates the dynamic molecular movements of large biomolecular systems, allowing for analysis of protein structures. The results of the analyses, root mean square deviation (RMSD), help to determine areas of rigidity or flexibility within the structure (76). These areas could be indicative of active sites, binding sites, catalytic sites, etc. Analysis of the predicted pUL103 structure showed that amino acids near the beginning of the conserved herpesvirus domain and those at the carboxyl terminal tail exhibit the most movement, as evidenced by their high RMSD values (red color in the heat map, high peaks on the graph). All throughout the protein, there are regions of alternating movement (Fig. 2-10). The most interesting sites to study based on these experiments are the very flexible regions at the beginning of the conserved herpesvirus domain and those at the C terminus. The conserved herpesvirus domain is worthy of study because of its conservation and possible connection to nucleic acids. We have previously shown that tags at the C terminus of pUL103 affect the ability of the virus to replicate (26). This suggests that the C terminus of pUL103 is important for its function and the biology of the virus. These could be sites for binding proteins, substrates, nucleic acids, etc.

In addition to providing secondary and tertiary structure predictions, I-TASSER also identified ten proteins, whose structures have been crystallized, that are similar to the pUL103 predicted structure. These suggested structural analogs are nucleotidyl transferases (Fig. 2-11). Nucleotidyl transferases are proteins that bind RNA and do not require a template to execute their

polyadenylation, uridylation, and other enzymatic activities (59). The top match for the predicted pUL103 structure (Fig. 2-12 A) was the catalytic domain of the human terminal uridylyl transferase 7 (TUT7) protein (Fig. 2-12 B). I aligned the predicted pUL103 structure with the TUT7 protein (Fig. 2-12 C) and the other nine structural analogs (data not shown) and they all aligned well with the pUL103 predicted structure. Given that the pUL103 predicted structure aligned well with the ten nucleotidyl transferases, I wanted to see if the pUL103 homologs shared similar protein chemistry with them. Thus, I conducted another MUSCLE alignment with the protein sequences of the nucleotidyl transferases and those of HCMV pUL103, HSV-1 pUL7, and EBV BBRF2 (Fig. 2-13). The alignment showed (a) numerous occurrences of shared protein chemistry between each of the herpesvirus homologs and at least half (five) of the nucleotidyl transferases, (b) the conserved herpesvirus domain has the greatest number of positions where all three pUL103 homologs share similar residue chemistry with the nucleotidyl transferases, followed by alpha helix 5, and (c) some active site residues of the human TUT7 protein are conserved across the pUL103 homologs as well. This suggests that the pUL103 homologs may be related to nucleic acid binding proteins.

However, upon closer examination, I observed that the pUL103 predicted structure does not contain the catalytic aspartate residues of nucleotidyl transferases. Thus, I proceeded to investigate whether proteins existed that resemble nucleotidyl transferases but lack catalytic activity. In my search I came across the LIM domain of the human TUT4 protein (Fig. 2-14 B). This domain is a nucleotidyl transferase domain that is catalytically inactive; it lacks the catalytic aspartate residues. It does not catalyze the polyadenylation, mono- or polyuridylation of RNA species. It instead binds RNA substrates and positions them for processing by the catalytic domain (107). When I aligned the TUT4 LIM domain with the predicted structure of pUL103 (Fig. 2-14

A), there was significant overlap between the structures (Fig. 2-14 C). Based on their predicted structures, pUL103 and its homologs may function to bind RNA species for catalysis. This data provides the foundation for studies into whether pUL103 is capable of binding nucleic acids.

Structure of HSV-1 pUL7 solved. The crystal structure of HSV-1 pUL7 was recently solved (Fig. 2-15 B). This crystallized structure contains residues 11-234 and 253-296. The N terminus consists of an equal mix of beta strands and alpha helices. The C terminus primarily consists of alpha helices. The investigators noted that the pUL7 structure has a novel fold not seen in other structures. The predicted structure of pUL7 (Fig. 2-15 A) lacks the N terminal beta strands, but contains a C terminus with many alpha helices. When the two structures are aligned, there is no obvious overlap, though they share similar secondary structures (Fig. 2-15 C). The crystallized structure of pUL7 was produced while in the presence of its binding partner pUL51. The predicted structure was produced in the absence of this binding partner; this could explain the differences in the two structures. The binding partner may have caused pUL7 to adopt a different conformation that the prediction software could not account for. It is also more than likely that the prediction is wrong. The pUL7 structure contains a novel fold; structures with this fold did not exist in the protein data bank for the most accurate structural prediction to be made by the software.

Discussion

To fill gaps relating to the structure and function of HCMV pUL103 and its homologs, I analyzed their protein sequences and produced structural predictions. One of the most interesting insights from doing this research is that more than one method was needed to explore the conservation of pUL103 and its homologs. Using only sequence identities to study these proteins would lead to the conclusion that they are not very similar. It is not until the sequence alignment is performed that relationships, outside of possessing the conserved herpesvirus domain, amongst

the proteins are visible. The conserved amino acids and possible motifs identified by this work provide the foundation for the development of novel experiments investigating their role in the function of pUL103 and HCMV biology as a whole. In addition, I also showed that the N terminus of pUL103 is more conserved than the C terminus. I hypothesized this to mean that the N terminal domain is more important for pUL103 function than the C terminal domain. However, given that each cytomegalovirus is species specific, it is possible that the C terminus contains species specific residues and explains the decrease in conservation in this area.

Given that a protein's structure can provide insight into its function, I sought to develop structural predictions of pUL103 and its homologs because theirs were not determined. The pUL103 homolog predicted structures, the nucleotidyl transferase structural analogs, and the sequence alignment of both suggest that pUL103 and its homologs may interact with nucleic acids. This correlates with the presence of the conserved herpesvirus domain within the homologs. Thus, these studies provide a logical pathway for the investigation of whether pUL013 and its homologs are capable of binding, modifying, etc. nucleic acids; however, these experiments were not tested in our lab. The crystallized HSV-1 pUL7 structure revealed there were differences between it and the predicted pUL7 structure. This could occur for several reasons: (a) the presence of the binding partner affects the conformation adopted by pUL7, (b) the reagents used to produce the crystals for X-ray crystallography may affect the conformations the protein adopts, (c) the protein structure produced in the crystals may not be what occurs naturally outside of the laboratory, or (d) the appropriate crystal structures did not exist for the most accurate structural prediction, and thus the structural prediction was wrong. In addition, the methods of structure and function prediction are constantly being improved to increase their accuracy. In the case of I-TASSER, the software is constantly being updated with new structures from the protein data bank, which help to improve

the threading and predictions. The Zhang lab that created I-TASSER has also developed auxiliary programs to work with I-TASSER to create improved structural predictions. With the pUL7 structure solved, some next steps would be to: (a) extract the secondary structure from the crystallized protein and compare with the a new prediction, (b) perform another prediction with the updated I-TASSER software and compare to the crystallized structure, and (c) use protein structure search engines to identify any proteins with solved structures similar to that of pUL7, using their annotations, if any, to design experiments to investigate pUL103 and its homologs for possible functions.

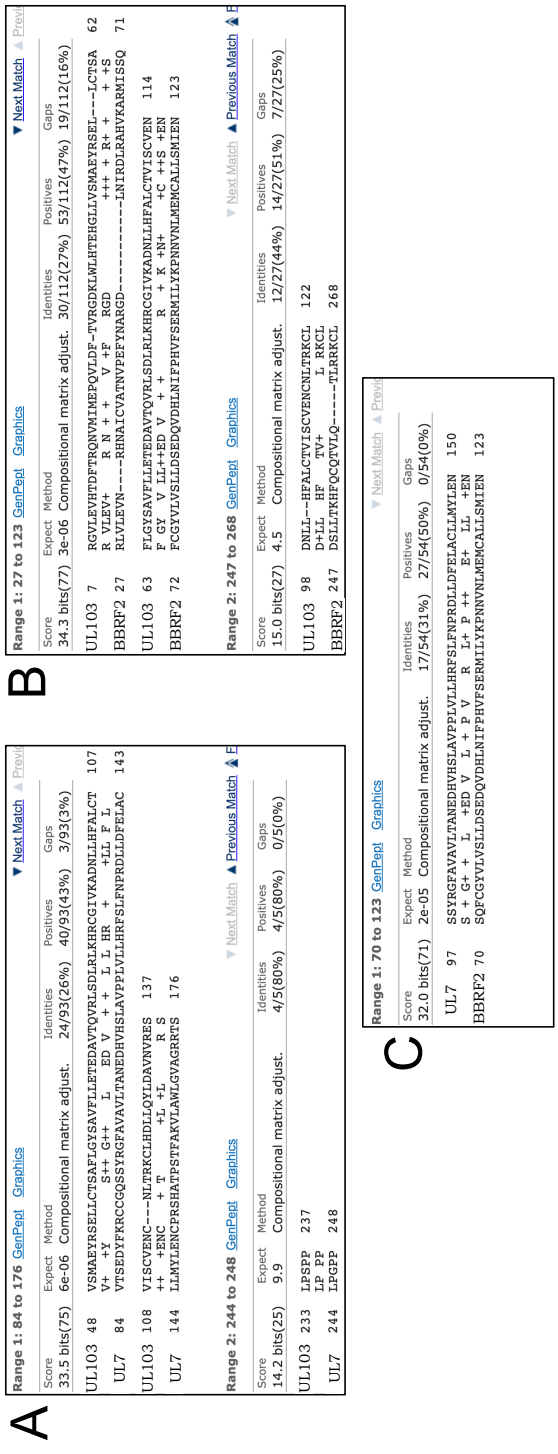

FIG 2-1 Human herpesvirus pUL103 homologs share limited sequence identities. NCBI BLAST analysis of HCMV pUL103, HSV-1 pUL7, EBV BBRF2, and VoHV1 ORF42. (A) Alignment information for HCMV pUL103 and HSV-1 pUL7. (B) Alignment information for HCMV pUL103 and EBV BBRF2. (C) Alignment information for HSV-1 pUL7 and EBV BBRF2.

Virus	Genus	Protein	Accession #
Betaherpesviruses			
Human cytomegalovirus (HCMV)	Cytomegalovirus	pUL103	YP_081549.1
Chimp cytomegalovirus (CCMV)		pUL103	NP_612734.1
Rhesus cytomegalovirus (RhCMV)		pRh140	YP_068231.1
Guinea pig cytomegalovirus (GPCMV)		pGP103	YP_007417869.1
Mouse cytomegalovirus (MCMV)		pM103	YP_214103.1
Rat cytomegalovirus (RCMV)		pE103	YP_007016499.1
Human herpesvirus 6A (HHV6A)	Roseolovirus	pU75	NP_042968.1
Human herpesvirus 6B (HHV6B)		pU75	NP_050254.1
Human herpesvirus 7 (HHV7)		pU75	YP_073815.1
Alphaherpesviruses			
Herpes simplex virus type 1 (HSV1)	Simplexvirus	pUL7	YP_009137081.1
Herpes simplex virus type 2 (HSV2)		pUL7	YP_009137158.1
Varicella zoster virus (VZV)	Varicellovirus	ORF53	NP_040175.1
Gammaherpesviruses			
Epstein-barr virus (EBV)	Lymphocryptovirus	BBRF2	YP_001129477.1
Kaposi's sarcoma-associated herpesvirus (KSHV)	Rhadinovirus	ORF42	YP_001129394.1

FIG 2-2 List of pUL103 homologs used for analysis. The 14 pUL103 homologs that were analyzed using several bioinformatics methods.

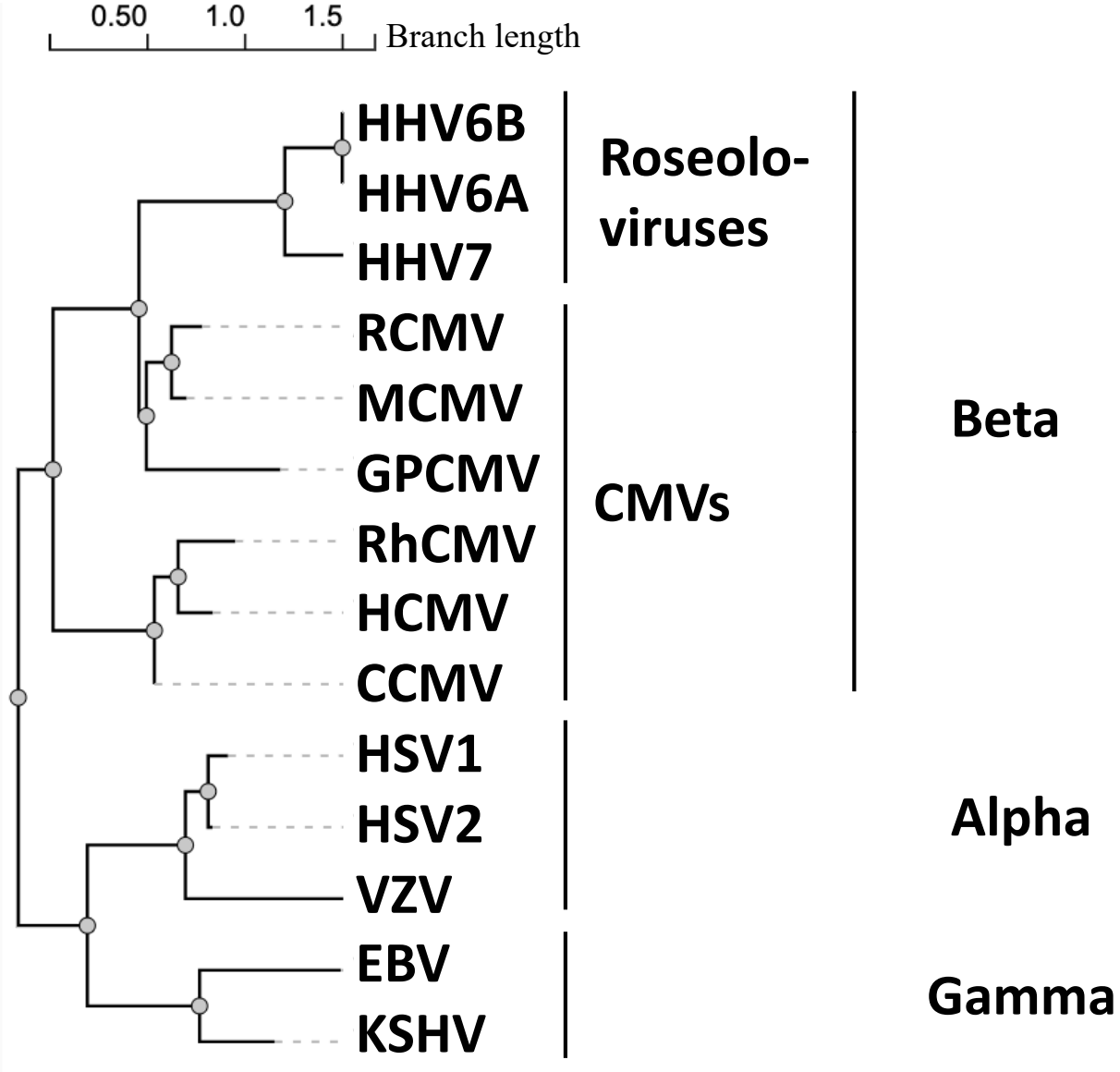

FIG 2-3 Phylogenetic analysis reveals relationships among pUL103 homologs. Phylogenetic tree of pUL103 and homologs produced using Phylogeny.fr (default settings using "One-Click" option). The tree was further modified using the PRESTO program from the ATGC bioinformatics platform.

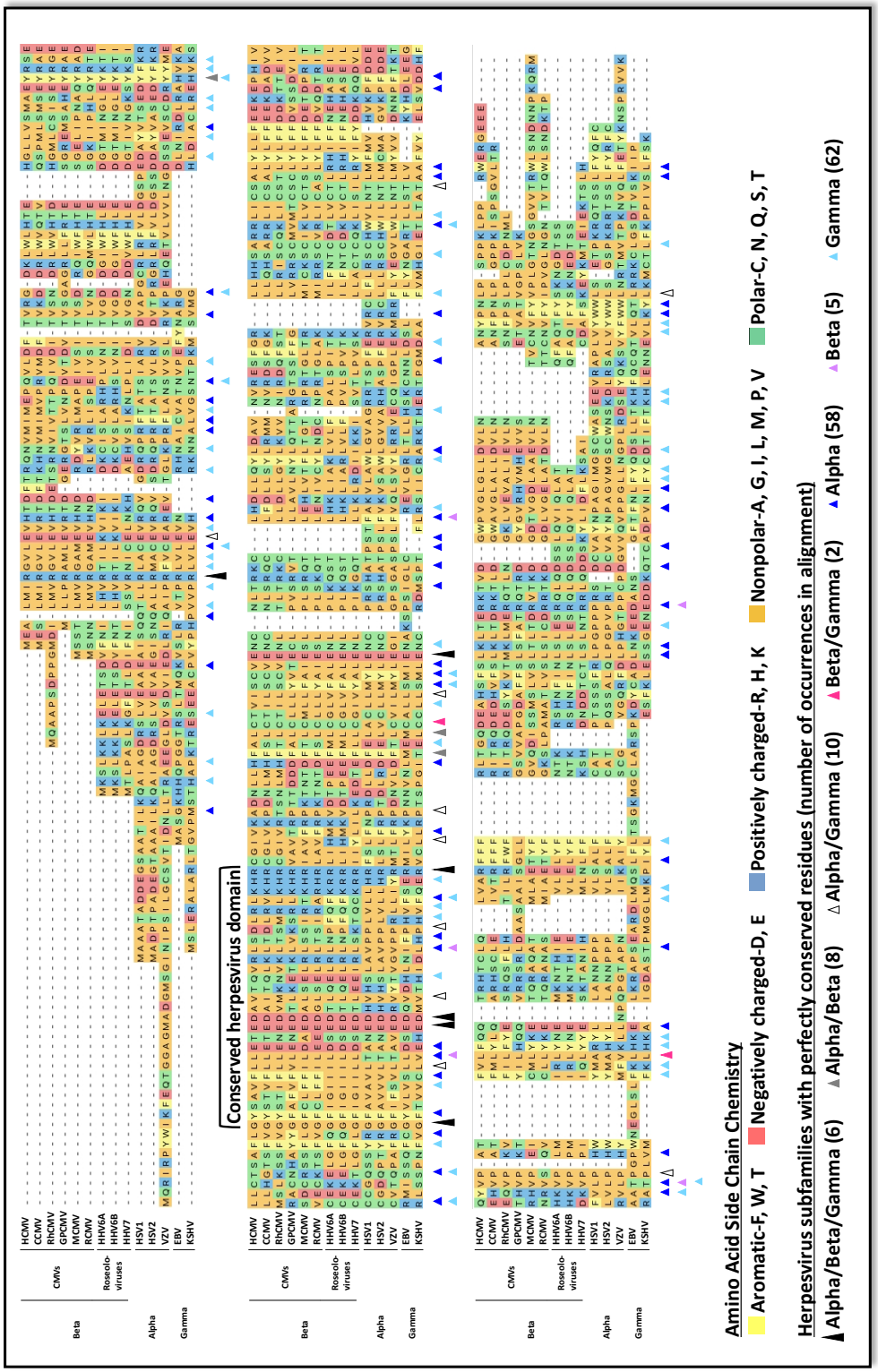

FIG 2-4 MUSCLE Alignment shows conserved protein chemistry across the 14 pUL103 homologs. DNASTAR MUSCLE alignment (using default settings) of the HCMV pUL103 protein sequence and 13 homologs from across the three subfamilies (alpha, beta, gamma) of herpesviruses.

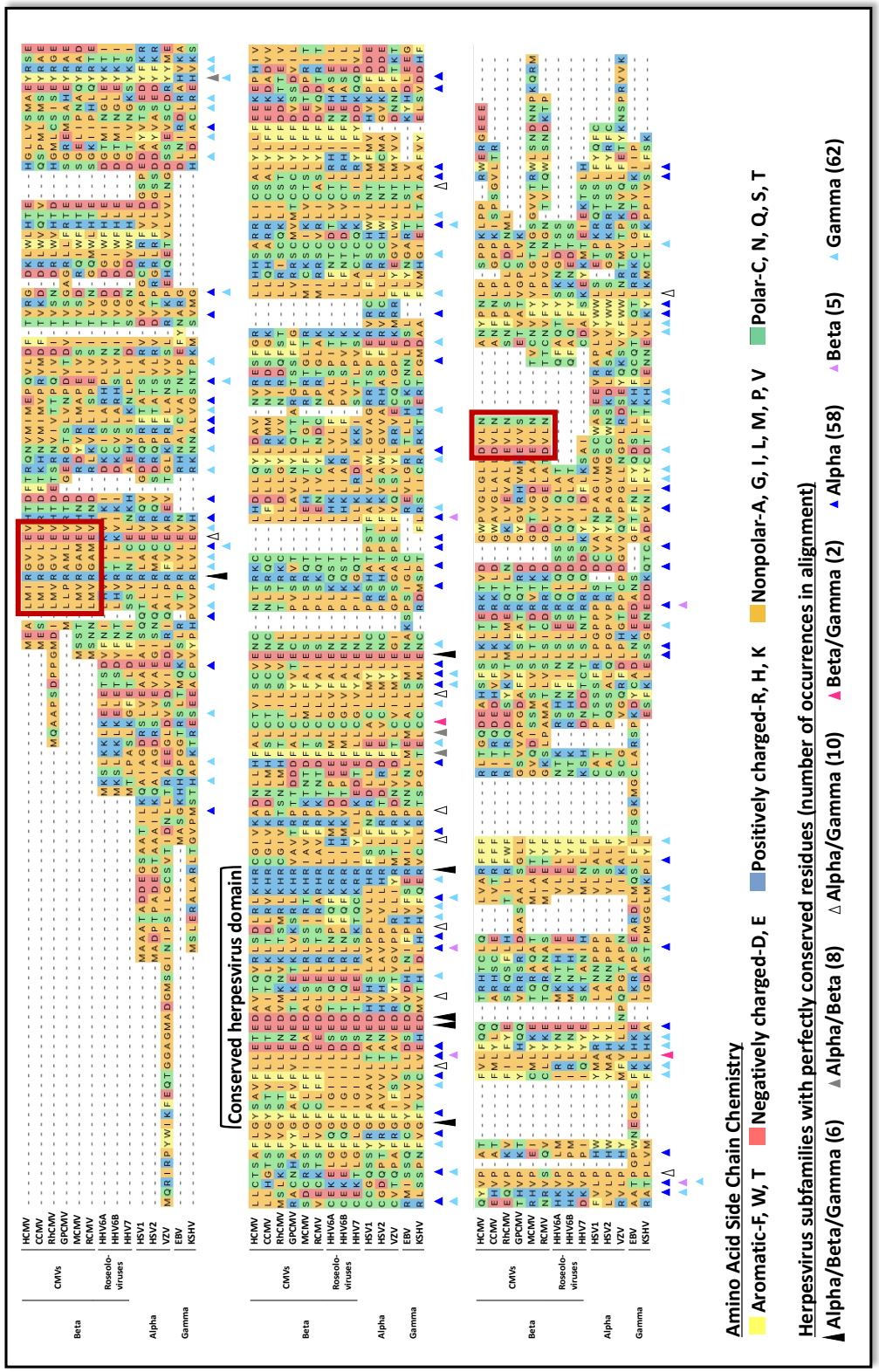

FIG 2-5 MUSCLE Alignment shows potentially conserved cytomegalovirus motifs. Red squares surround the putative amino and carboxyl terminal motifs of the cytomegaloviruses within the alignment.

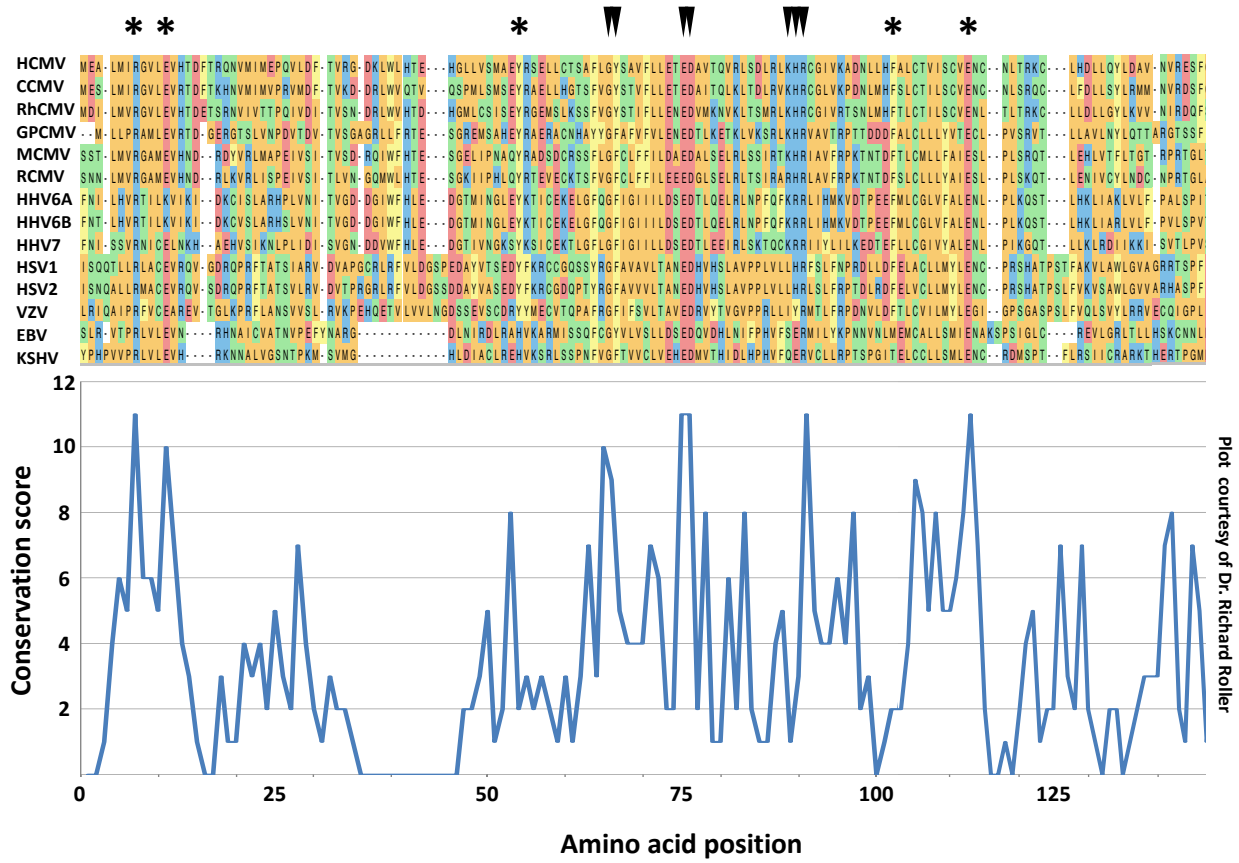

FIG 2-6 MUSCLE alignment produced in this study correlates with similarity plot from Dr. Richard Roller. Similarity plot of pUL103 and 59 of its homologs from across the three subfamilies of the *Herpesviridae*, generously provided by Dr. Richard Roller. *y*-axis scale: 0 (no similarity across residues) to 12 (identical residues). The conserved herpesvirus domain consists of residues 65-91; conserved residues within the conserved domain are marked with a black arrow ▼, residues outside of this domain are marked with an *. Positions that have conservation scores of 0 represent gaps within the alignment.

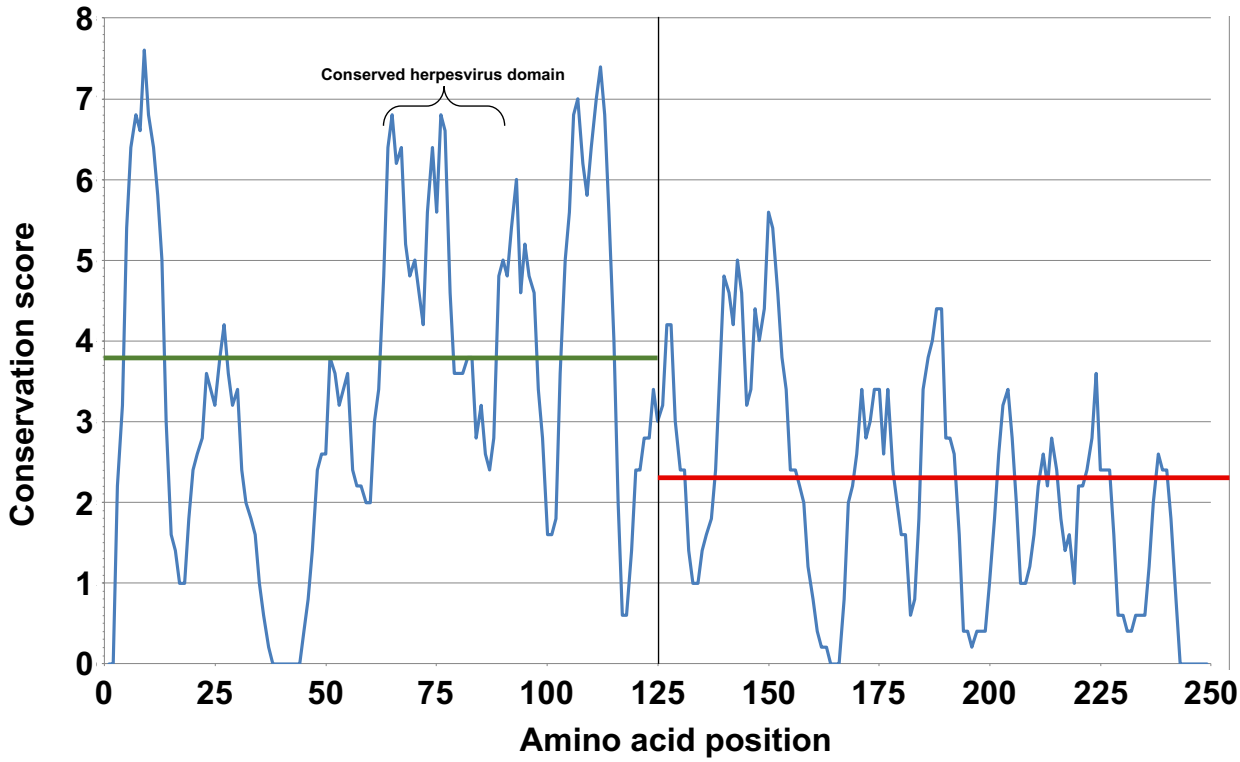

FIG 2-7 The N terminus (residues 1-125) of pUL103 is more conserved than the C terminus (residues 126-249). Similarity plot of pUL103 and 59 of its homologs from across the three subfamilies of the *Herpesviridae* with a sliding window of 5, generously provided by Dr. Richard Roller. *y*-axis scale: 0 (no similarity across residues) to 8 (identical residues). The green and red lines represent the average conservation scores of the N and C termini, respectively. The conserved herpesvirus domain consists of residues 65-91. Positions that have conservation scores of 0 represent gaps within the alignment

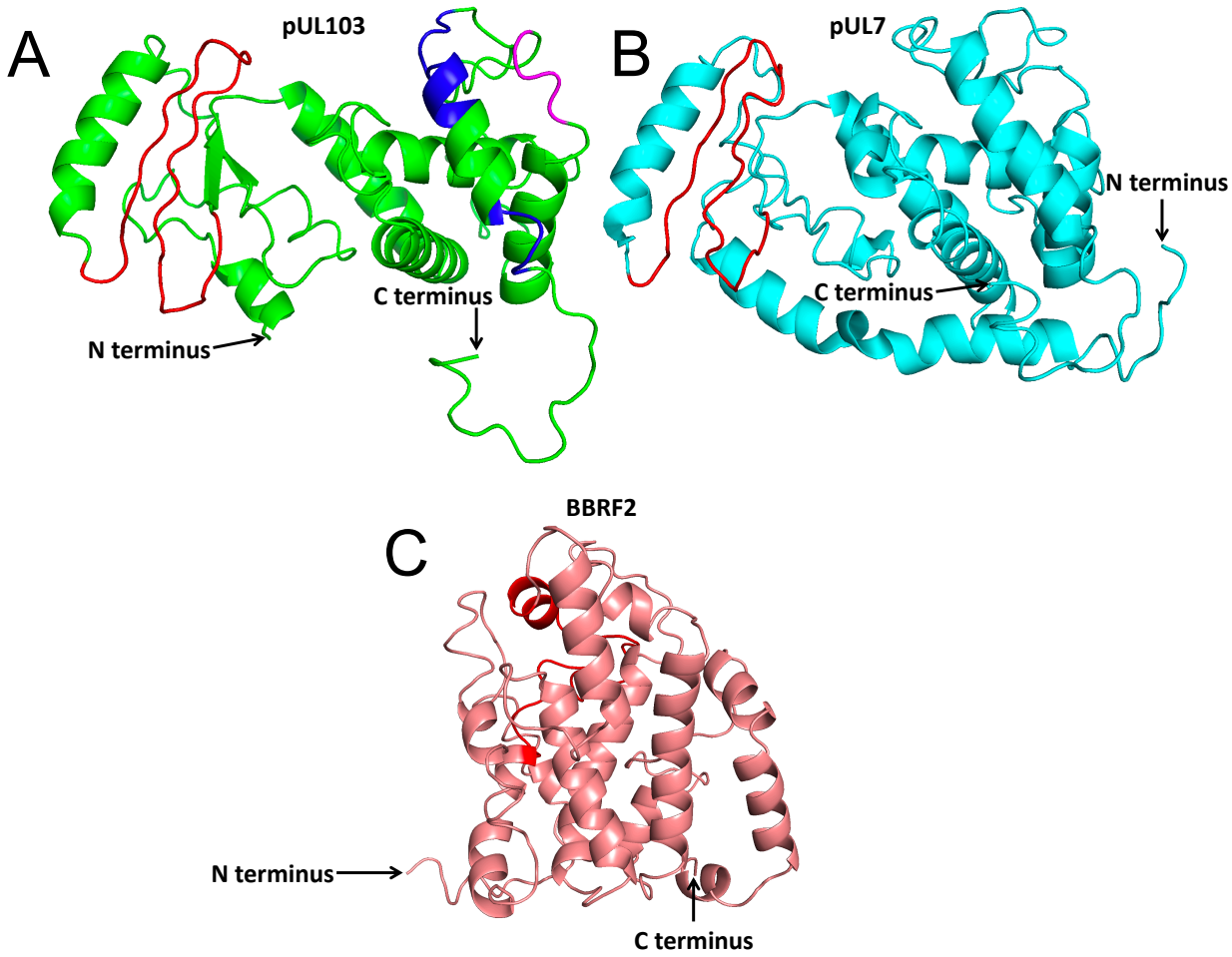

FIG 2-8 I-TASSER is a useful tool that provided predicted structures for the pUL103 homologs whose structures had not been crystallized. (A) Predicted structures of HCMV pUL103 in green, (B) HSV-1 pUL7 in blue, and (C) EBV BBRF2 in pink produced by I-TASSER. The conserved herpesvirus domain is indicated in each structure by the red colored residues. The light purple and blue colored residues within the predicted pUL103 structure represent the DEAD/DEAH box and late domains, respectively. The pUL103 and pUL7 predicted structures here were created in 2017. The BBRF2 predicted structure was created in 2018.

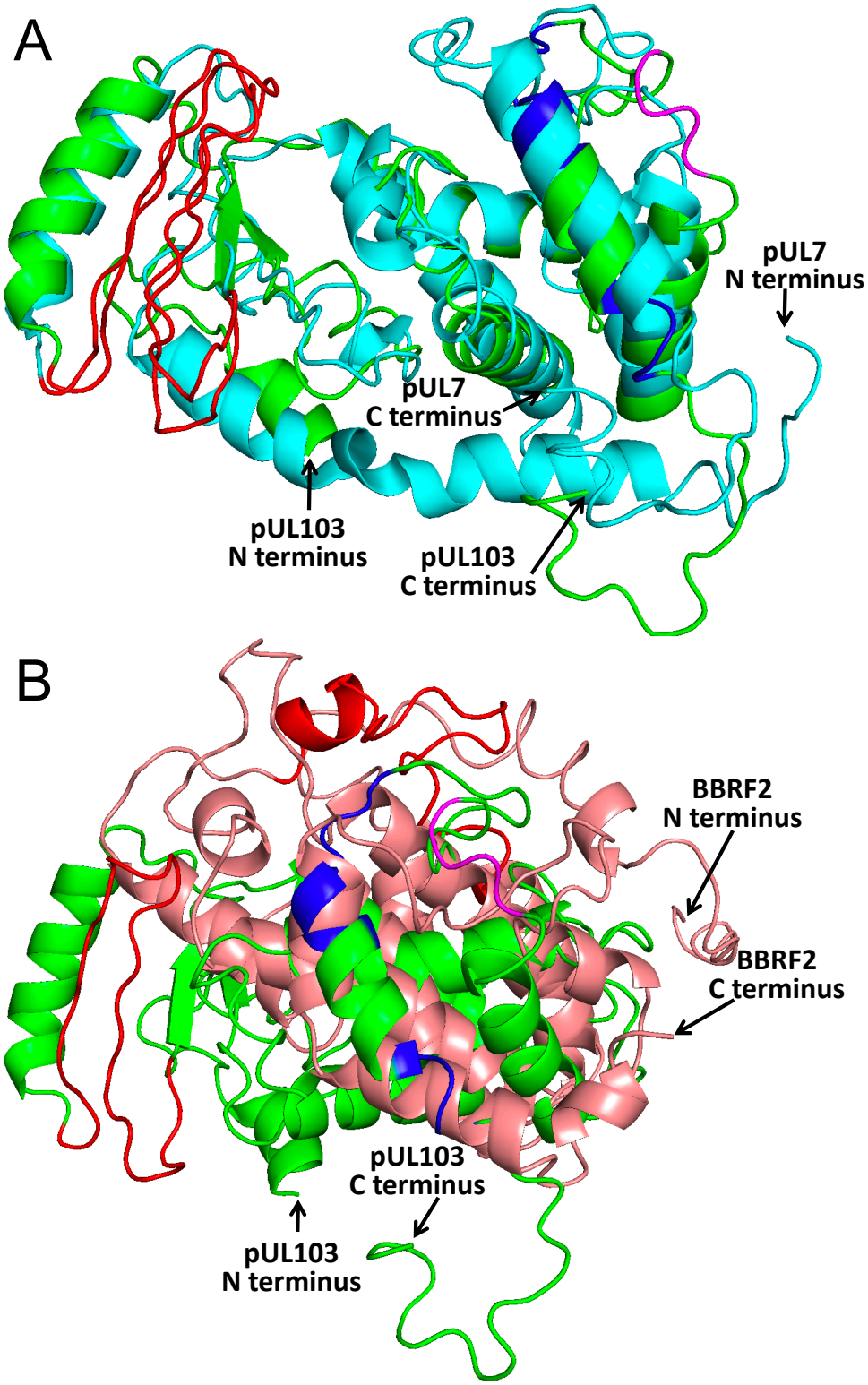

FIG 2-9 There are differences in the predicted structures of the pUL103 homologs. (A) Pymol super alignment of the predicted structures of pUL103 and pUL7 with an alignment score of 321. (B) Pymol super alignment of the predicted structures of pUL103 and BBRF2 with an alignment score of 203.

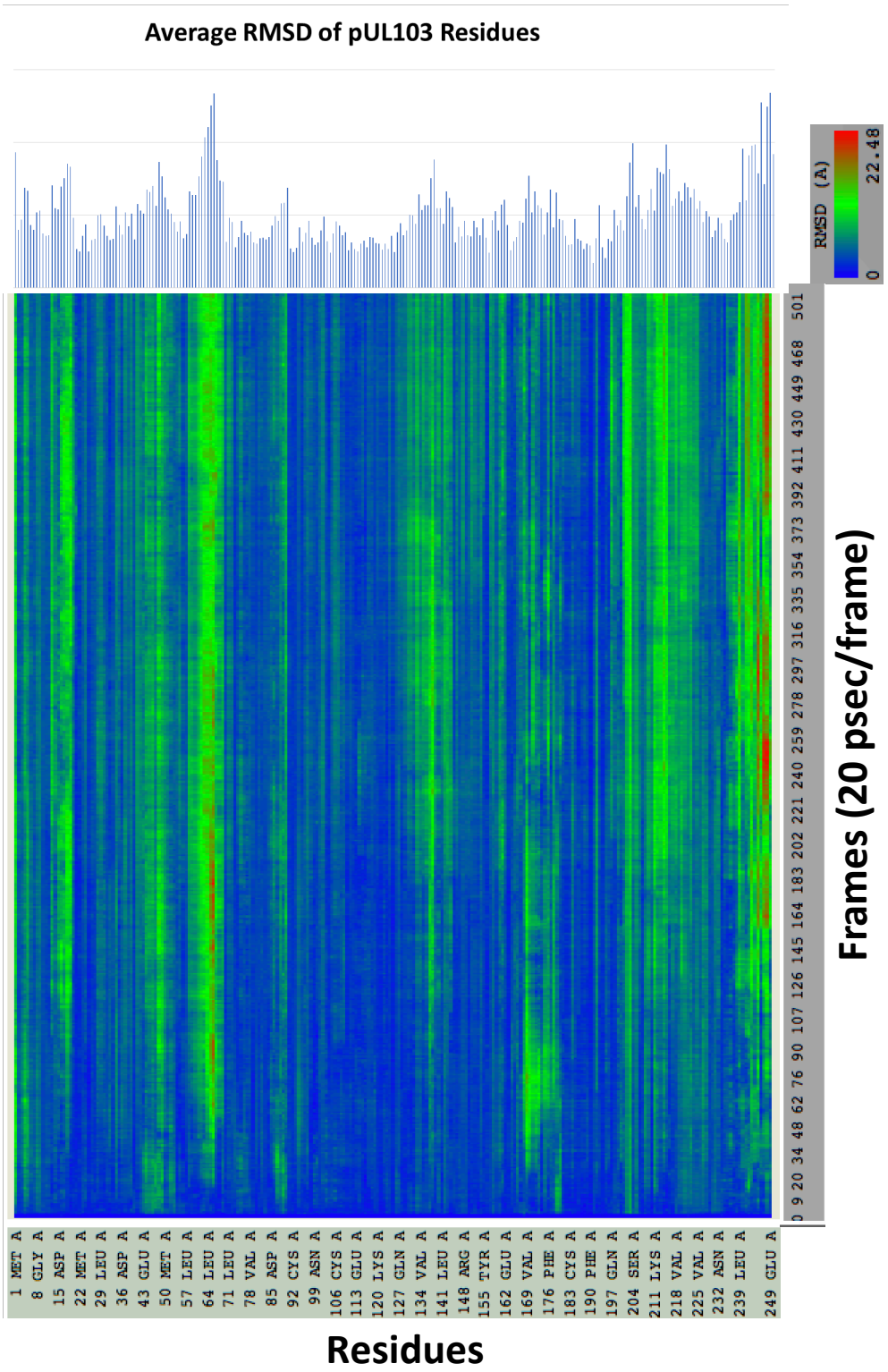

FIG 2-10 Heatmap showing alternating regions of movement within the predicted pUL103 structure. In the main panel, root mean square deviation (RMSD) for the individual residues of the predicted pUL103 structure for a total of 10 nanoseconds. In the side panel, the average RMSD for the residues of the predicted structure.

Protein	TM-Score	Species	Accession #
Terminal uridylyltransferase 7 catalytic module (TUT7)	0.869	*H. sapiens*	5W0B_A
Mitochondrial poly(A) RNA polymerase (mtPAP)	0.846	*G. gallus*	5A2V_A
Terminal uridylyltransferase 1 (TUT1)	0.825	*H. sapiens*	5WU1_A
Cid1 poly(U) polymerase (CID1)	0.800	*S. pombe*	4EP7_A
RNA editing TUTase 1 (RET1)	0.785	*T. brucei*	3HIY_A
RNA editing terminal uridylyl transferase 2 (RET2)	0.771	*T. brucei*	2B51_A
Terminal uridylyl transferase 4 (TUT4)	0.765	*T. brucei*	2Q0E_A
RNA Editing terminal uridylyl transferase 1 (RET1)	0.759	*T. brucei*	5HZD_A
Poly(A) polymerase complex GLD-2-GLD-3 (PAPGLD2GLD3)	0.759	*C. elegans*	4ZRL_A
Mitochondrial poly(A) polymerase (PAPD1)	0.757	*H. sapiens*	3PQ1_A

FIG 2-11 Nucleotidyl transferases make up the list of the top ten structural analogs of the predicted pUL103 structure. These proteins were identified by I-TASSER and come from a very diverse range of eukaryotes.

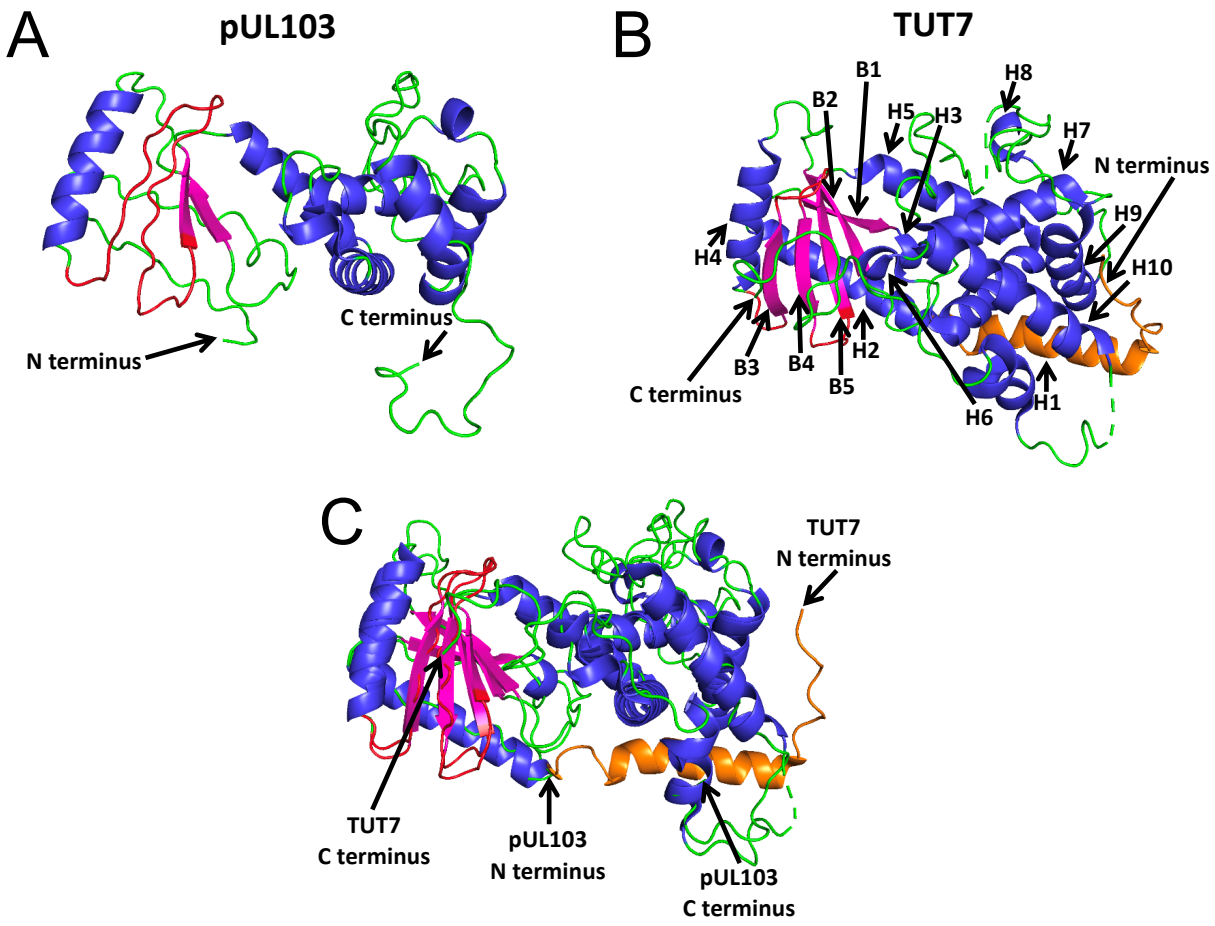

FIG 2-12 The predicted structure of pUL103 aligned well with the structure of the catalytic domain of the human terminal uridylyl transferase 7 (TUT7) protein. (A) Predicted structure of HCMV pUL103. (B) Structure of the catalytic domain of the human terminal uridylyl transferase 7 (TUT7) protein. (C) Pymol super alignment of the predicted structure of pUL103 and the structure of human TUT7 with an alignment score of 426. Alpha helices are colored deep purple, beta strands fuschia, loops green, and the conserved herpesvirus domain red. The predicted pUL103 structure pictured here was created in 2018, but does not vary greatly from the predicted structure created in 2017.

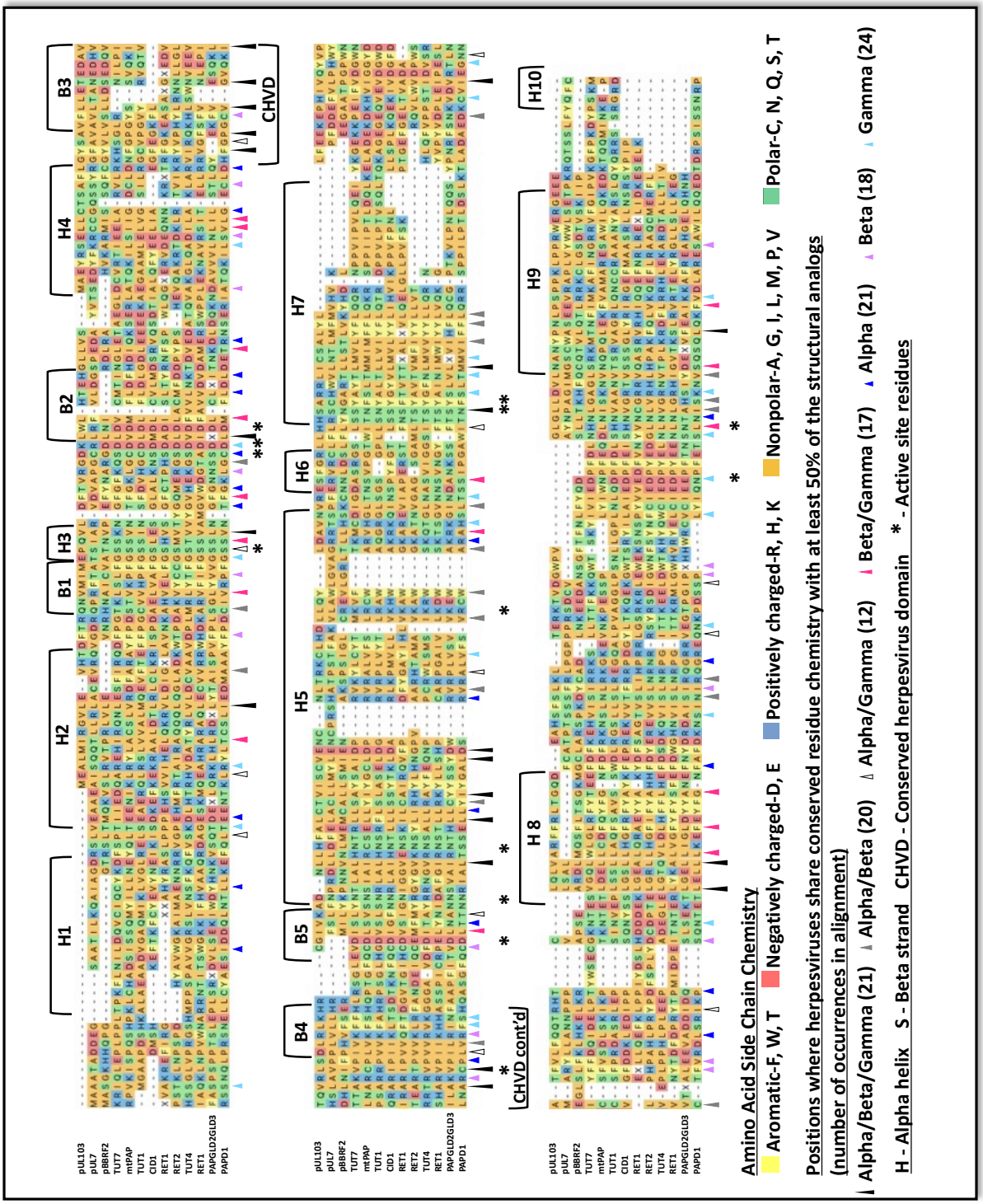

FIG 2-13 pUL103 homologs share protein chemistry with several nucleotidyl transferases. DNASTAR MUSCLE alignment (using default settings) of the HCMV pUL103, HSV-1 pUL7, and EBV BBRF2 protein sequences with 10 nucleotidyl transferases, structural analogs of predicted pUL103 structure, determined by I-TASSER. Regions that did not align with any of the herpesviruses were removed. The annotated helices and beta strands in the alignment are from the labels of the human TUT7 protein in Fig. 2-11 B.

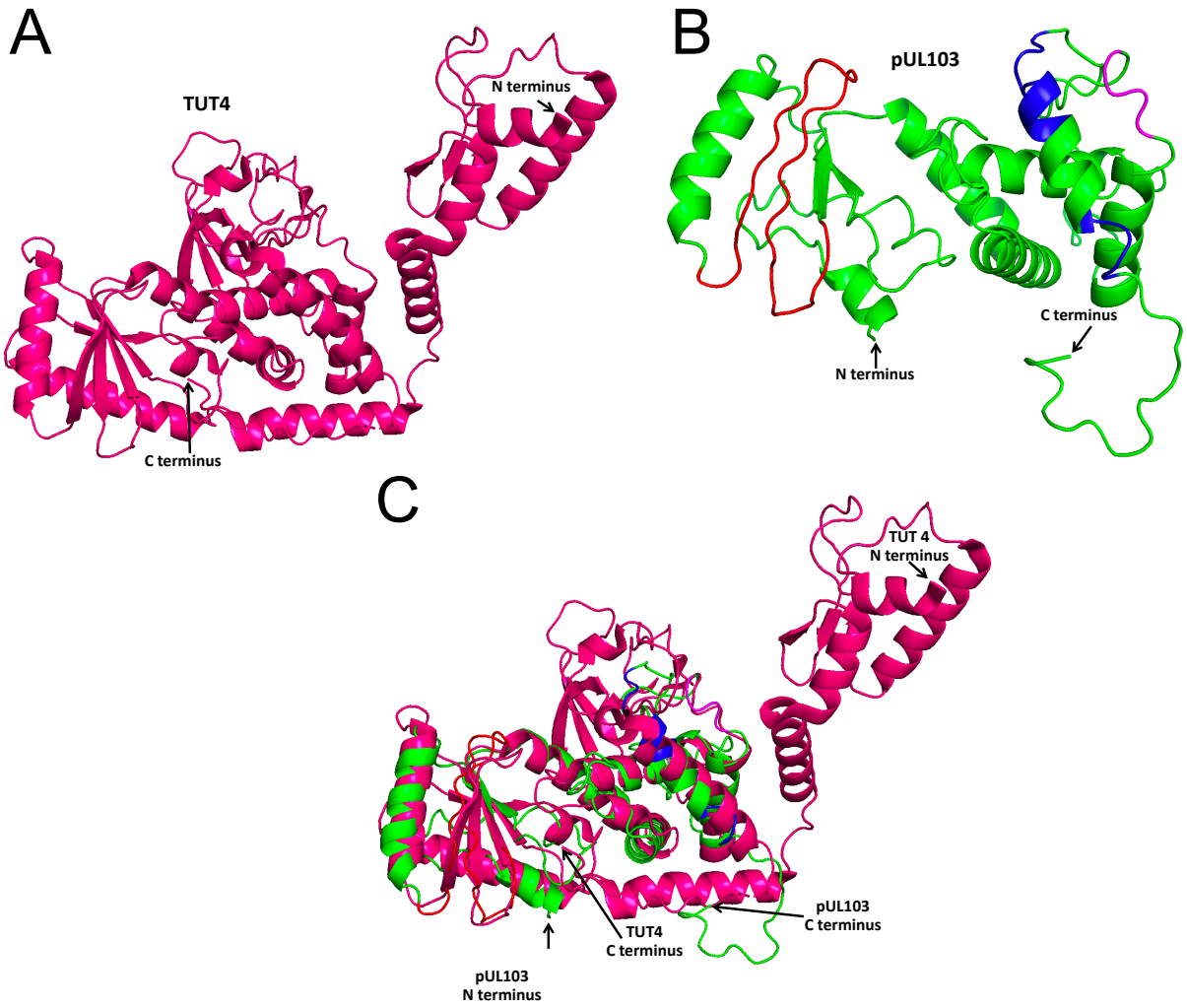

FIG 2-14 The predicted structure of pUL103 aligned well with the structure of the non-catalytic domain of the human terminal uridylyl transferase 4 (TUT4) protein. (A) Predicted structure of HCMV pUL103. (B) Structure of the non-catalytic domain of the human terminal uridylyl transferase 4 (TUT4) protein. (C) Pymol super alignment of the predicted structure of pUL103 and the structure of human TUT4 with an alignment score of 388. The light purple and blue colored residues within the predicted pUL103 structure represent the DEAD/DEAH box and late domains, respectively. The pUL103 predicted structure pictured here was created in 2017.

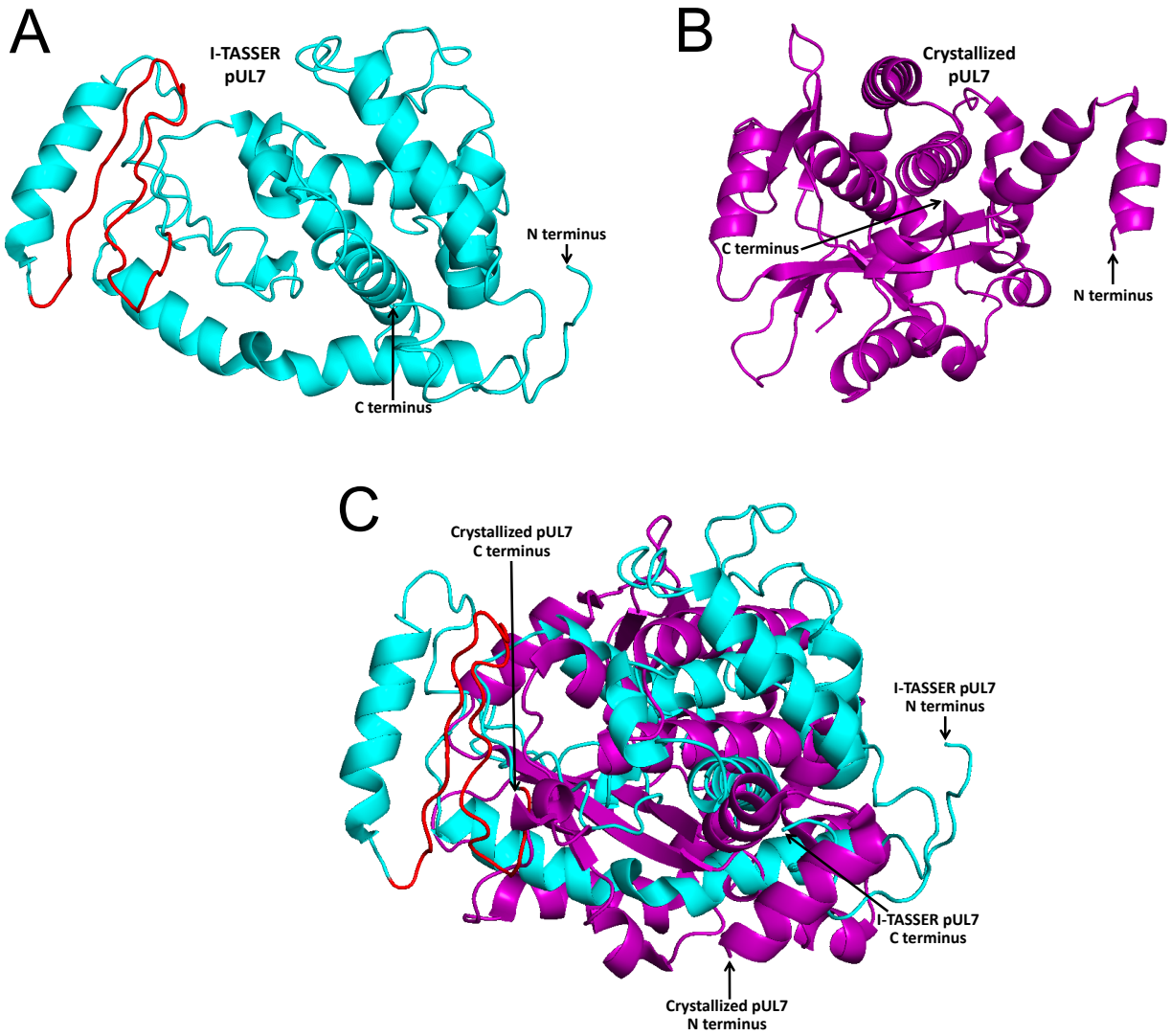

FIG 2-15 There are differences between the predicted pUL7 structure and the actual crystallized protein. (A) Predicted structure of HSV-1 pUL7. Conserved herpesvirus domain colored red. (B) Actual structure of the pUL7 protein from HSV-1. (C) Pymol super alignment of the predicted structure of pUL7 and the actual structure of pUL7 with an alignment score of 241.

CHAPTER 3: IDENTIFICATION OF THE TEMPORAL WINDOW WHEN PUL103 IS REQUIRED FOR MAXIMAL CYTOPLASMIC VIRION ASSEMBLY COMPARTMENT ABUNDANCE

Introduction

HCMV is a global, opportunistic pathogen that causes organ-specific damage and inflammation in immunocompromised (HIV/AIDS patients, transplant recipients) and immunodeficient (fetuses, infants) populations. Immunocompromised individuals experience organ-specific inflammatory conditions such as hepatitis, pancreatitis, pneumonitis, enteritis, retinitis meningitis and encephalitis. Immunodeficient individuals experience liver, lung, and spleen pathology, petechiae, hearing loss, and neurodevelopmental disabilities (64, 67). HCMV is the most common viral congenital infection, as well as the leading cause of birth defects and childhood disabilities in the United States (49, 67). More children suffer from CMV-related diseases than from more well-known congenital conditions such as Down syndrome, fetal alcohol syndrome, or spina bifida (16). The virus is also the most common, nonhereditary cause of hearing loss in children (43). Given its societal importance, the development of effective algorithms for early-life detection of congenital HCMV (77), and the limited array of therapeutic options for HCMV, more research is needed to uncover the mechanisms of HCMV replication and pathogenesis, to enable development of novel therapeutics.

HCMV has the largest genome of all human viruses (~ 236 kb) (33, 73, 74). This large and complex genome encodes at least 167 proteins, four large, noncoding RNAs, two oriLyt RNAs, and at least 23 microRNAs (64). Technology such as ribosome profiling has shown that the HCMV genome may have the capacity to encode at least 751 open reading frames; this includes short functional proteins of less than 100 amino acids (37, 96). HCMV virions are complex

particles consisting of four components: (a) an outer envelope made up of a lipid bilayer and embedded proteins, (b) a layer of proteins and RNA called the tegument, (c) a proteinaceous capsid, and (d) the viral genome that resides within the capsid (74). The HCMV virion consists of thousands of individual components that need to be intricately assembled to form an infectious particle. To enable this, HCMV reorganizes infected cells into virus-producing factories, in which the cellular endosecretory machinery, consisting mostly of Golgi membranes and early or recycling endosomes, is remodeled into the cytoplasmic virion assembly compartment (cVAC).

The cVAC is a hallmark of HCMV infection (84, 85). It is a juxtanuclear complex consisting of the Golgi reorganized into a ring-like structure with early and recycling endosomes primarily in the center of the ring (28). Other components of the secretory machinery, such as late endosomes and the endoplasmic reticulum (ER), are located at the periphery of the cVAC (12, 27, 113). Microtubules radiating from the microtubule organizing center (mTOC) located at the center of the cVAC, attach to the nuclear membrane via SUN and KASH domain proteins and induce nuclear bending that results in the characteristic reniform shape of the nucleus (5, 11). Within the cVAC, nascent capsids acquire their outer tegument, undergo secondary envelopment, and are then trafficked to the cellular membrane for release into the extracellular milieu (28, 89). Each infected cell has one cVAC, even cells with multiple nuclei. The cVAC is induced in various types of HCMV-infected cells in vitro (17, 23, 28, 53, 87, 89); unpublished data suggests this structure also exists in HCMV-infected tissues in vivo (C. Lim and P. Pellett).

cVAC biogenesis is a complex, multifactorial process. In prior studies we showed that cVAC formation leads to shifts in organelle identities, as defined by infection-induced changes in their composition, structure, localization, and function (27). cVAC biogenesis relies upon several viral components, including HCMV miRNAs and the late proteins pUL48, pUL94, pUL103, and

gpUL132 (26, 45, 106). cVACs still form but are less abundant when pUL103 is knocked down using either siRNAs or a system in which pUL103 stability is regulated (8, 26).

The stability regulation system employed a virus, UL103-FKBP-V5, in which the FKBP destabilization domain is fused to the C terminus of pUL103, to produce the fusion protein pUL103-FKBP-V5. In the absence of the stabilizing ligand Shield-1, the fusion protein is rapidly shuttled to the proteasome for degradation. When Shield-1 is present, the fusion protein is stabilized and able to carry out its functions. After knockdown of pUL103, cVACs still form, albeit in reduced numbers; this is likely due to redundancy among viral contributors to cVAC biogenesis and incomplete knockdown of pUL103 activity.

pUL103 is transcribed and translated from the late gene UL103, located within the unique long region of the HCMV genome (64). pUL103 is a small (249 amino acids, ~27.39 kDa), conserved tegument protein, with homologs in all members of the *Herpesviridae* (72, 74). As a tegument protein, pUL103 is present in virions and dense bodies. Within these particles, pUL103 makes up a very small fraction of the total amount of protein, around 0.1% (102). Thus, at the initiation of infection very little pUL103 is present. As a late gene product, pUL103 expression is low early in infection and ramps up late in infection. At 120 hours post-infection (hpi), the majority of pUL103 is in the cytoplasm within the cVAC, with a small amount in the nucleus (72).

Several studies have examined the roles of pUL103 during HCMV infection (Table 1). In pUL103 knockout experiments, viral titers were 10 - 10,000-fold less (1, 35, 110). In knockdown experiments, pUL103 was important for cell-to-cell spread and virion maturation, in addition to cVAC biogenesis (26). pUL103 may also have other roles throughout HCMV infection due to its presence in nuclei and interactions with cellular proteins involved in innate immune system signaling and intracellular trafficking (72).

pUL103 does not act alone during HCMV infection. Its most noted interaction partner is HCMV pUL71. These proteins form a complex that is thought to be important for proper cVAC formation, efficient secondary envelopment, and virion egress (14, 38, 72). pUL103 homologs in other herpesviruses have been studied as well. HSV-1 pUL7 is important for virion production, cell-to-cell spread, secondary envelopment, stabilization of pUL51, and maintenance of focal adhesions (3, 14, 98). EBV BBRF2 is important for cell-to-cell spread and stabilization of BSRF1 (60). KSHV ORF42 is important for virion production (13). These pUL103 homologs also interact with their respective pUL71 homologs (HSV-1 pUL51, EBV BSRF1, KSHV ORF55), an interaction that appears to be conserved across the *Herpesviridae* (3, 14, 60, 80, 108).

Given the many roles pUL103 has throughout HCMV infection, it has potential as an antiviral target. While we know pUL103 is involved in many activities during HCMV infection, its mechanisms of action are unknown. In-depth mechanistic studies of pUL103 are required to understand how this protein contributes to HCMV infection and to design inhibitors against it.

HCMV has a long reproductive cycle, during which it takes 48-72 hours for appreciable numbers of new virions to be produced (64), after which, virion production continues for several days. Because of the long and complex replication cycle of the virus, some mechanistic studies of pUL103 would be enhanced by knowing when it is required during HCMV infection. Thus, in this work we sought to identify the temporal window when pUL103 is required for maximal cVAC abundance and stability. We found that for maximal cVAC abundance, pUL103-FKBP-V5 needs to be stabilized no later than 72 to 84 hpi during a 120-hour infection. We also found that destabilization after 84 hpi results in production of fewer cVACs; it thus takes between 33 and 36 hours for the effects of pUL103-FKBP-V5 destabilization to become apparent. Our results suggest

that the period between 72 and 84 hpi would be highly informative for detailed study of the mechanistic roles of pUL103 in cVAC biogenesis.

Materials and Methods

Cells and viruses. Human foreskin fibroblasts (HFFs) passage $\leq$ 15 were used for experiments. HFFs were used for growing virus stocks and plaque assays. HFFs were grown in Dulbecco's modified Eagle medium (DMEM) containing high glucose, L-glutamine, and sodium pyruvate (HyClone, SH30243.FS, Pittsburgh, PA) supplemented with 10% fetal bovine serum (FBS, S11150, Atlanta Biologicals, Flowery Branch, GA), 1x GlutaMAX (Gibco, 35050061, Waltham, MA), and 1x nonessential amino acids (NEAA, HyClone, SH3023801, Pittsburgh, PA) known as complete DMEM (cDMEM). The bacterial artificial chromosome (BAC) containing HCMV pAD/Cre AD169 with 3' FKBP and V5 tags on UL103 (UL103-FKBP-V5) was constructed and used as previously described (26) (Table 1). A BAC containing HCMV pAD/Cre AD169 with 3' V5 and His tags on UL103 (UL103-V5-His) was also used in this study (72) (Table 1). Virus stocks were obtained by infecting 100% confluent HFFs at a multiplicity of infection (MOI) of 0.001. While growing the UL103-FKBP-V5 virus, Shield-1, at 1 μmol, was replenished every 72 hours. When 100% cytopathic effect (CPE) was reached, cells and supernatant were removed from flasks and centrifuged at 1280xg for 15 minutes at 4°C. All but ~1 ml of supernatant was removed. A milliliter of 10% 3x autoclaved milk was added to the supernatant and the pellet was resuspended. The cell suspension was sonicated three-times in a Branson analog 450 sonifier cup horn filled with ice water at an output control of 10 and 30% duty cycle for 10 seconds, followed by 10 seconds of rest. Sonicated material was centrifuged at 1280xg for 10 minutes at 4°C, and the supernatant was stored at -80°C. Virus titers were determined by performing plaque assays using confluent layers of HFFs.

Immunoblots. Unsynchronized HFFs were seeded at 80% confluence (4.22 x 10^5 cells) into 6-well plates. The following day, cells were left uninfected or infected with the virus UL103-FKBP-V5, in the absence or presence of Shield-1, at a multiplicity of infection (MOI) of 0.1 for 120 hpi. Shield-1 was replenished at 72 hpi. Every 24 hours, infections were stopped and harvested for protein using RIPA (radio-immunoprecipitation assay) buffer (10 mM HEPES [pH 7.4], 1x protease inhibitors (Roche, Indianapolis, IN or Thermo Fisher Scientific, A32961, Waltham, MA), 1% sodium deoxycholate, 150 mM NaCl, 1% Nonidet P-40, and 0.1% SDS). The RIPA-soluble protein fractions were collected and stored at -80°C. Protein was quantified by the Pierce bicinchoninic acid (BCA) assay (Thermo Fisher Scientific, 23227, Waltham, MA). For cell lysates, 30 µg of protein was solubilized in 4x SDS-Laemmli buffer followed by SDS-PAGE. Separated proteins were transferred to nitrocellulose membranes (Amersham Protran 0.1 µm pore size, 0600010, GE Healthcare Life Sciences, Chicago, IL), probed with primary antibodies (GAPDH, Thermo Fisher Scientific, MA5-15738; IE2, Millipore, MAB810; UL44, Virusys, G0729146; V5, Santa Cruz Biotechnology, sc-271944), and then reacted with horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG secondary antibody (Bio-Rad, 1706516, Hercules, CA). Reactions were detected with the SuperSignal West pico chemiluminescent substrate (Invitrogen, 34580, Waltham, MA) on autoradiography film (MidSci, BX810, St. Louis, MO).

Immunofluorescence assay (IFA). 8-well glass chamber slides (Nunc LabTek II-154534, Thermo Fisher Scientific, Waltham, MA) were incubated for 1 hour at 37°C with 300 µl of 0.2% gelatin in 1x PBS. Unsynchronized human foreskin fibroblasts (HFFs) were seeded at 80% confluence, equaling 3.52 x 10^4 cells, into the gelatin-coated slides. For checking for overlaps in the localization of the Golgi marker GM130 and the viral marker IE2, the following day the cells

were infected with UL103-V5-His at an MOI of 0.1 for 120 hpi. For the Shield-1 time course experiments, the following day the cells were infected with UL103-FKBP-V5 (in the absence of Shield-1, with Shield-1 present the entire infection, and with Shield-1 added at various times post infection) at an MOI of 0.1 for 120 or 144 hours. At the end of the infections, the cells were fixed with 300 µl of 4% paraformaldehyde in 1x PBS for 15 min., incubated with 300 µl of 50 mM ammonium chloride to quench autofluorescence for 15 min., permeabilized with 300 µl of blocking buffer (5%, glycine, 10% normal goat serum, 0.1% sodium azide in 1x PBS) containing 0.2% Triton X for 15 min., incubated with blocking buffer for 1 hour, removed media chamber and incubated with 75 µl of primary antibodies (GM130, BD Biosciences, 610823; IE2, Millipore, MAB810; V5, abcam, ab9116) diluted in blocking buffer for 1 hour, washed 2x for 7 min. in 1x PBS, incubated with 75 µl of fluorescence-tagged secondary antibodies (Alexa Fluor 488-conjugated goat anti-rabbit IgG (green), A-11034 and Alexa Fluor 568-conjugated goat anti-mouse IgG (red), A-11031, Invitrogen, Waltham, MA) diluted in blocking buffer for 1 hour, washed 2x for 7 min. in 1x PBS, mounted with VectaShield mounting medium with DAPI (Vector Laboratories, H1200, Burlingame, CA) and sealed with a coverslip (Fisherbrand, 12-548-5E, L-50 x W-22 mm, thickness 0.13 to 0.17mm, Waltham, MA) using clear top coat nail polish (wet n wild, Los Angeles, CA)(Loreal, Paris, France). Imaging was done on Nikon E800 and E600 microscopes.

Image analysis. The FIJI image processing program was used to visualize the immunofluorescent images and MIPAR image analysis software used to threshold and measure nuclei (90, 95). Infected cells late in infection and cVACs were counted and the percentage of cells with mature cVACs was calculated. During infection, HCMV causes cells to swell or become enlarged (cytomegaly) (69). Related to this, late in infection, infected cells have larger nuclear

cross-sections than uninfected cells. pUL103 is a product of the UL103 late gene and is indicative of cells late in infection as well. In our experiments, cells late in infection were defined as having (a) a nuclear area greater than 200 μm^2 (based on comparisons of histograms of nuclear areas between infected and uninfected cells) and (b) visible IE2 and pUL103-FKBP-V5 staining. cVACs were defined by the presence of a juxtanuclear Golgi ring. At least 30 late infection cells per condition were counted and statistical analysis performed.

Statistical analysis. Chi-square analyses were performed using GraphPad Prism version 8.4.3 for MacOS, GraphPad Software, San Diego, California USA, www.graphpad.com.

Results

Experimental overview. To identify the period when pUL103 is needed for maximal cVAC abundance, as detailed in Materials and Methods, we determined the relative abundance of cVACs produced as a function of pUL103 stabilization with Shield-1, using the UL103-FKBP-V5 virus, which allows regulation of pUL103-FKBP-V5 stability, and thereby cVAC abundance (26).

Time course of HCMV protein expression in cells infected with the UL103-FKBP-V5 virus in the presence and absence of Shield-1. To extend from our prior studies, we used immunoblots to examine the time course of expression of HCMV proteins IE1 and IE2 (immediate early), UL44 (early), and pUL103-FKBP-V5 (late) in infections with the UL103-FKBP-V5 virus, in the presence and absence of Shield-1, every 24 hours during a 120-hour infection (Fig. 3-1). As expected, in the presence of Shield-1, pUL103-FKBP-V5 (a late protein) abundance increased over the course of the infection; we attribute the residual pUL103-FKBP-V5 in the absence of Shield-1 to its incomplete the protein's incomplete degradation by the proteasome. This could be due to the proteasome being either overloaded or otherwise affected by HCMV or the FKBP fusion protein is not delivered to the proteasome. The presence or absence of Shield-1 did not affect IE1

levels until 120 hpi, when they decreased in the absence of Shield-1. IE2 and UL44 levels were reduced throughout the infection in the absence of Shield-1. This is likely due to less cell-to-cell spread during low MOI infections when pUL103-FKBP-V5 is not stabilized (26). This could also be due to pUL103-FKBP-V5 having potential effects on protein expression due to its conserved herpesvirus domain (Fig. 1-8); this domain is also present in the topoisomerase III of yeast and may allow binding to nucleic acids. This further illustrates the importance of pUL103 in HCMV biology. Collectively, this data demonstrates that in the presence of Shield-1, the UL103-FKBP-V5 virus gene expression program proceeds as expected.

The immunofluorescence signals of cytoplasmic Golgi marker GM130 and nuclear viral marker IE2 do not overlap in HCMV-infected cells. The Golgi ring is a defining characteristic of the cVAC (5, 26, 28, 87, 93). As part of developing a more robust method for quantification of cVACs, we validated simultaneous use of antibodies against the Golgi marker GM130, and a nuclear marker of infection IE2. To rule out the possibility that IE2 and GM130 staining overlap, we stained separate wells with one or the other mouse primary antibody, and then used the same fluorescently- tagged secondary antibody, to look for colocalization with DAPI fluorescence (DNA stain). IE2 staining overlapped and colocalized with nuclear DAPI staining, with no IE2 signal in the cytoplasm (Fig. 3-2). GM130 staining is only present in the cytoplasm and does not colocalize with nuclear DAPI. Thus, as expected, the two proteins unambiguously localized to different locations within infected cells, validating use of the same secondary antibody to detect them simultaneously.

A tool for identifying mature cVACs in HCMV infections. To further reduce subjectivity while enumerating Golgi rings, we collected images of numerous Golgi structures present in IE2-positive cells and arranged them into a gradient from less to more certain

assignments as cVACs (Fig. 3-3). Using this tool as a reference while counting cVACs helped to more reliably identify cVACs in a quick, relatively objective manner.

Changes in the stability of pUL103-FKBP-V5 affect cVAC abundance. To verify that the UL103-FKBP-V5 virus allows manipulation of cVAC abundance in response to variation in pUL103-FKBP-V5 stability, we measured relative cVAC abundance in the presence and absence of Shield-1. Previously, using the UL103-FKBP-V5 virus, we showed pUL103-FKBP-V5 stability is important for cVAC biogenesis (26). Thus, we can use cVAC biogenesis as a biological endpoint for pUL103-FKBP-V5 activity. As expected, in the presence of Shield-1, Golgi rings form and there is pUL103-FKBP-V5 staining in the center of those rings, indicative of cVACs. In the absence of Shield-1, pUL103-driven cVAC biogenesis is disrupted (Fig. 3-4 A). The decrease in relative cVAC abundance in the absence of Shield-1 is quantifiable (Fig. 3-4 B). In general, there are about twice as many cVACs produced in the presence of Shield-1 compared to when it is absent.

Shield-1 time course experiments. To identify the period when pUL103 is required for maximal cVAC abundance, we varied the period of Shield-1 presence during infection. In each of the following experiments, positive controls included Shield-1 throughout infection and negative controls to which Shield-1 was never added; an example is shown in Fig. 3-4. After 120 or 144 hours, the infections were stopped and analyzed using IFA (Figs. 3-5 B-E).

Identification of the beginning of the period when pUL103-FKBP-V5 is required for maximal cVAC abundance. To determine the latest time of pUL103-FKBP-V5 stabilization for maximal cVAC abundance during 144-hour infections, we added Shield-1 at 12-hour intervals from 12 to 60 hpi (Fig. 3-5 A, Experiment 1). There was a statistically significant difference

between the positive and negative controls, but no difference between the positive control and any of the experimental time points (Fig. 3-5 B).

Since Experiment 1 showed no changes in cVAC abundance when Shield-1 was added as late as 60 hpi, we tested addition of Shield-1 at time points between 72 and 108 hpi during 120-hour infections (Fig. 3-5 A, Experiment 2). There were progressively fewer cVACs present when Shield-1 was added at time points later than 84 hpi, but in comparison to the positive control, the differences were not statistically significant (Fig. 3-5 C).

To more precisely determine the latest time for pUL103-FKBP-V5 stabilization to obtain maximal cVAC abundance during 120-hour infections, Shield-1 was added every three hours between 84 and 96 hpi (Fig. 3-5 A, Experiment 3). Compared to the positive control, fewer cVACs were produced at each experimental time point (Fig. 3-5 D). There was a statistically significant difference between the positive control and the addition of Shield-1 at 96 hpi.

Our results indicate that the latest time for initiation of pUL103-FKBP-V5 stabilization for maximal cVAC abundance during 120-hour infections is between 72 and 84 hpi.

How long does pUL103-FKBP-V5 need to be stabilized for maximal cVAC abundance? For infections initiated under conditions that stabilize pUL103-FKBP-V5, we determined how long pUL103-FKBP-V5 needed to be stabilized for maximal cVAC abundance during 120-hour infections. Thus, in Experiment 4, Shield-1 was present from time zero and then removed (pUL103-FKBP-V5 destabilized) at time points after maximal cVAC formation (72-84 hpi) and the relative cVAC abundance was determined at 120 hpi (Fig. 3-5 A, Experiment 4). cVAC abundance was significantly reduced when Shield-1 was removed at 84 hpi, to a level similar to that of the negative control, in which Shield-1 was absent for the entire 120-hour infection (Fig. 3-5 E). When Shield-1 was absent from 87 to 93 hpi, cVAC abundance was slightly,

but not significantly reduced compared to the positive control. Our results indicate that for infections initiated under conditions that stabilize pUL103-FKBP-V5, it takes between 33 and 36 hours for the effects of its destabilization on cVAC abundance to become apparent.

Discussion

The work described here addresses two main areas. The first relates to the biochemical and biological functions of pUL103, a protein conserved across the herpesvirus family that may be involved in many processes from entry to egress, including, cell-to-cell spread, cVAC biogenesis, and secondary envelopment. The second relates to the processes and mechanisms involved in HCMV cVAC biogenesis.

pUL103 and the important processes it is involved in represent potential targets for antiviral development. For that to be realized, detailed studies are needed to determine the mechanisms of actions of pUL103. As a prelude to such work, it is important to identify the periods or states when pUL103 is serving a particular purpose during infection. Our objective here was to identify the temporal window when pUL103 is required for maximal cVAC abundance.

In Experiments 1, 2, and 3, we found that for maximal cVAC abundance, initiation of pUL103-FKBP-V5 stabilization must begin no later than 72 to 84 hpi during a 120-hour infection. In the absence of stabilization of pUL103 during this period, it is possible that key protein interactions required to create the cVAC are lost. In addition, pUL103 may exert regulatory effects by binding nucleic acids via its conserved putative topoisomerase domain (41).

When Shield-1 was present from the beginning of infection (Experiment 4), we found that it takes between 33 and 36 hours for the effects of pUL103-FKBP-V5 destabilization to become apparent. This suggests that during a 120-hour infection, pUL103 must be present after 84 hpi to maintain maximal cVAC abundance.

The combined results from experiments 1-4 suggest that the period between 72 and 84 hpi would be an informative period for study of the mechanistic roles of pUL103 in cVAC biogenesis and activity. Such studies might include analysis of virion assembly, intra- and extracellular viral titers, characterization of identity shifting of endosecretory organelles and machinery prior during cVAC biogenesis (27), live cell imaging of viral and cellular proteins, and deep temporal and spatial proteomics analysis of viral and cellular proteins during this process. We note that, while stabilization of pUL103 before 72 to 84 hpi is not required to achieve the maximum number of cVACs, it is possible that it may function early in infection in ways that are unrelated to cVAC biogenesis in cell culture, such as interactions with antiviral proteins of the innate immune system and cell trafficking (72).

Due to lack of antibodies, we have not been able to detect native pUL103 during infections. However, with the presence of C terminal V5 and His tags, we are able to visualize pUL103 with these modifications (pUL103-V5-His). Over a 120-hour infection, pUL103-V5-His is most abundant at between 96 and 120 hpi (unpublished data). Thus, during the temporal window of 72 - 84 hpi when we determined pUL103 is required for maximal cVAC abundance, the protein itself may not be at its maximal abundance.

Table 3-1 Names and properties of viruses used to study HCMV pUL103

Virus	Schematic representation of pUL103 construct	Growth	Reference/Source
pAD/Cre HCMV AD169 BAC: wild-type		10^5 - 10^6 pfu/ml	Das et al., 2014
UL103 deletion: Towne strain with UL103 gene deleted from start to stop codon		100 - 1000-fold less	Dunn et al., 2003
UL103-DH-7: wt virus with 5' transposon replacement of 102 nt after codon 44		> 10,000-fold less	Yu et al., 2003
UL103-DH-68: wt virus with 5' transposon replacement of 43 nt after codon 101		> 10,000-fold less	Yu et al., 2003
UL103-TS-288: wt virus with 3' transposon (3.66 kb) insertion after codon 247		No data available	Yu et al., 2003
Δ53-759**UL103**: Towne strain with 5' Kanamycin cassette replacement of 698 nt after codon 17		10 - 1000-fold less	Ahlqvist et al., 2011
UL103-Stop-F/S: Towne strain with replacement of codon 52 with stop codon and insertion of 7 nt to create frameshift mutation		10 - 1000-fold less	Ahlqvist et al., 2011
TB-103-STOP: TB40/E strain with point mutations at codon 24 introducing a stop codon and a subsequent frame shift at codon 25		100 - 1000-fold less	Jens von Einem, PhD
TB-103-FLAG: TB40/E strain with addition of FLAG tag after codon 249		No data available	Jens von Einem, PhD
UL103-FKBP: wt virus with addition of FKBP domain after codon 249		10 - 100-fold less	Das et al., 2014
UL103-FKBP-V5: wt virus with addition of FKBP domain and V5 tag after codon 249		10 - 1000-fold less	Das et al., 2014
UL103-V5-His: wt virus with addition of V5 and His tags after codon 249		10 - 1000-fold less	Ortiz et al., 2016
UL103-GFP: wt virus with addition of GFP tag after codon 249		No data available	Nat Moorman, PhD

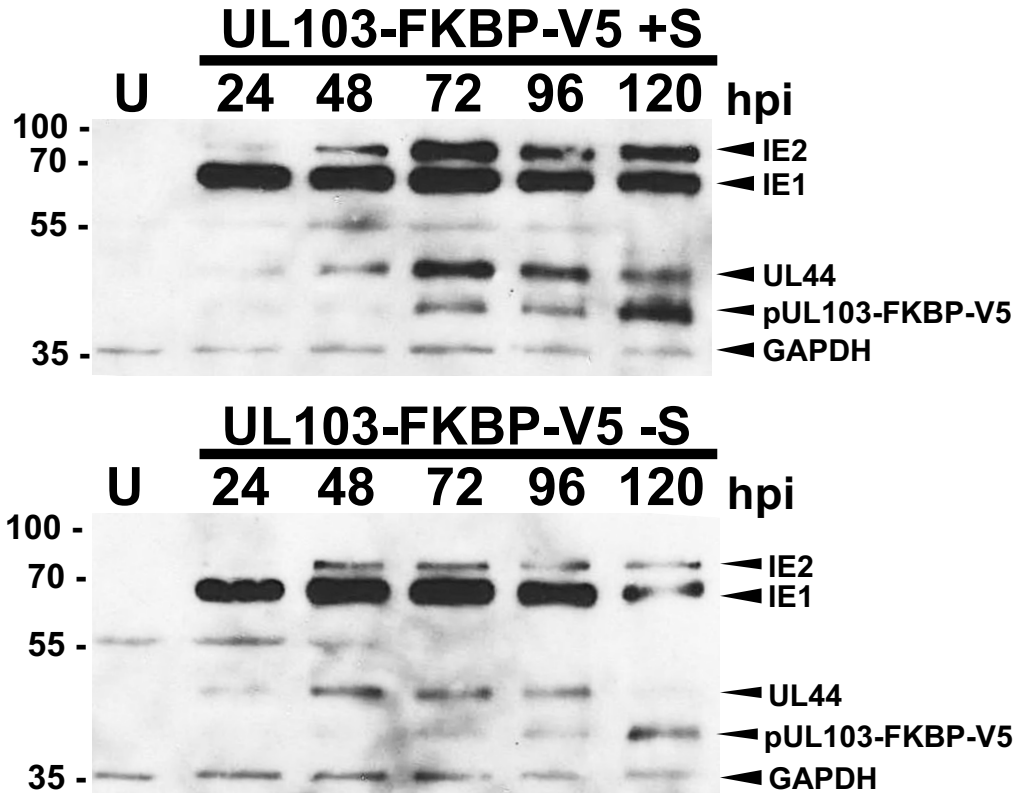

FIG 3-1 Time course of HCMV protein expression in cells infected with the UL103-FKBP-V5 virus in the presence and absence of Shield-1. HFFs were uninfected (U) or infected with HCMV pAD/Cre UL103-FKBP-V5 in the absence (-S) and presence (+S) of Shield-1 at an MOI of 0.1 for 120 hpi. Shield-1 was replenished at 72 hpi. Infections were stopped, protein harvested with RIPA buffer, and separated into RIPA-soluble and RIPA-insoluble fractions every 24 hours throughout the infection. 30 µg of the RIPA-soluble samples were analyzed using immunoblotting. GAPDH (glyceraldehyde-3-phosphate dehydrogenase) was used as the loading control. Samples were probed for immediate-early (IE1, IE2), early (UL44), and late (pUL103-FKBP-V5) viral proteins using IE2, UL44, and V5 primary antibodies followed by goat anti-mouse IgG conjugated to horseradish peroxidase.

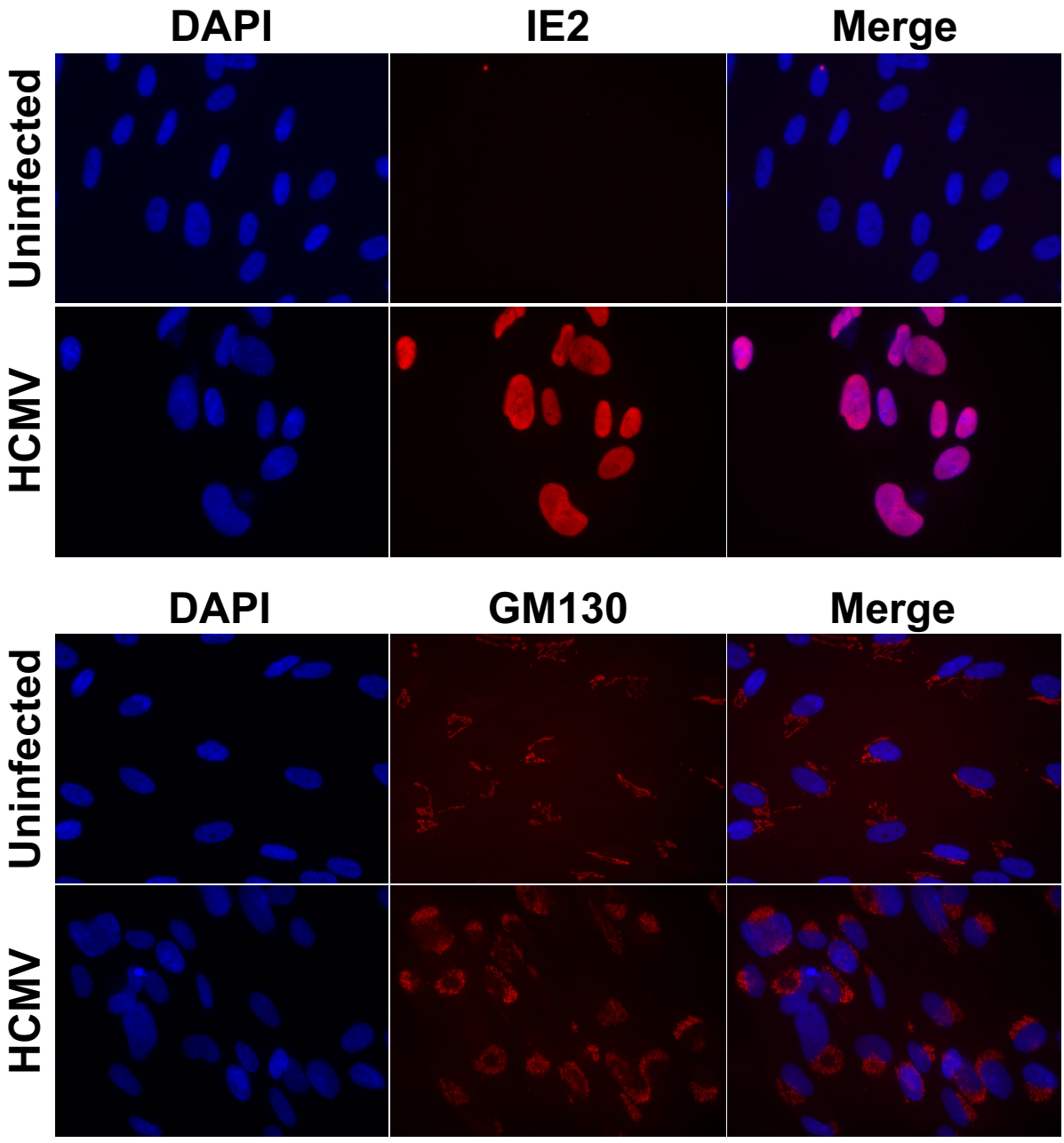

FIG 3-2 Immunofluorescence signals of viral nuclear marker IE2 and cellular cytoplasmic Golgi marker GM130 do not overlap in HCMV-infected cells. HFFs were uninfected or infected with UL103-V5-His at an MOI of 0.1 for 120 hpi. IFA was used to assess the localization patterns of both proteins. DAPI stains DNA, primarily in the nucleus, and is blue. IE2 is present in the nucleus and detected using an IE2 primary antibody followed by Alexa Fluor Goat anti-mouse 568 secondary antibody that is red. GM130 is present in the cytoplasm and detected using a GM130 primary antibody and Alexa Fluor Goat anti-mouse 568 secondary antibody that is red. IFA performed by Taylor Vensko.

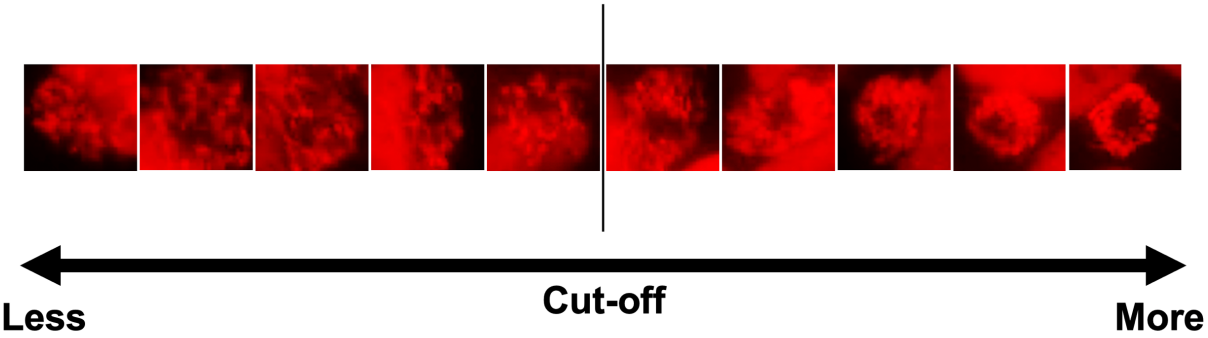

FIG 3-3 A Golgi ring gradient: a tool created to assist with discernment of mature cVACs. Images of Golgi rings and other Golgi structures from one IFA image of an HCMV infection used as a tool for assessing the presence or absence of Golgi rings representative of cVACs. HFFs were infected with UL103-FKBP-V5 in the presence of Shield-1 at an MOI of 0.1 for 120 hpi. Shield-1 was replenished at 72 hpi. Golgi is in the cytoplasm and was detected using a GM130 primary antibody followed by Alexa Fluor Goat anti-mouse 568 secondary antibody that is red.

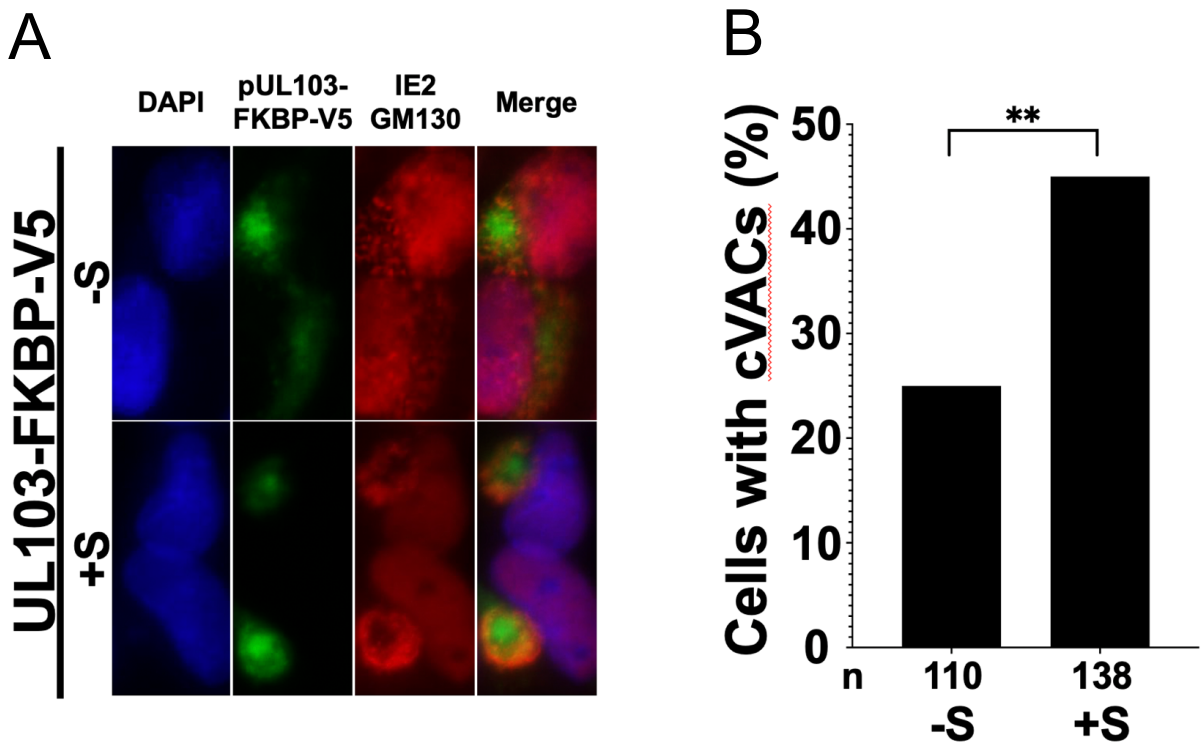

FIG 3-4 Changes in the stability of pUL103-FKBP-V5 affect cVAC abundance. (A) IFA images of UL103-FKBP-V5 virus infections at an MOI of 0.1 for 120 hpi in the absence (-S) and presence (+S) of Shield-1. Shield-1 was replenished at 72 hpi. DAPI stains DNA, primarily in the nucleus, and is blue. pUL103-FKBP-V5 is present in the cytoplasm and detected using a V5 primary antibody followed by Alexa Fluor Goat anti-rabbit 488 secondary antibody that is green. IE2 is present in the nucleus and detected using an IE2 primary antibody followed by Alexa Fluor Goat anti-mouse 568 secondary antibody that is red. GM130 is present in the cytoplasm and detected using a GM130 primary antibody and Alexa Fluor Goat anti-mouse 568 secondary antibody that is red. (B) Quantification of cVACs from infections in (A). Statistical significance was measured using the chi-square test (**: $p < 0.01$,). n, number of cells counted. hpi, hours post infection.

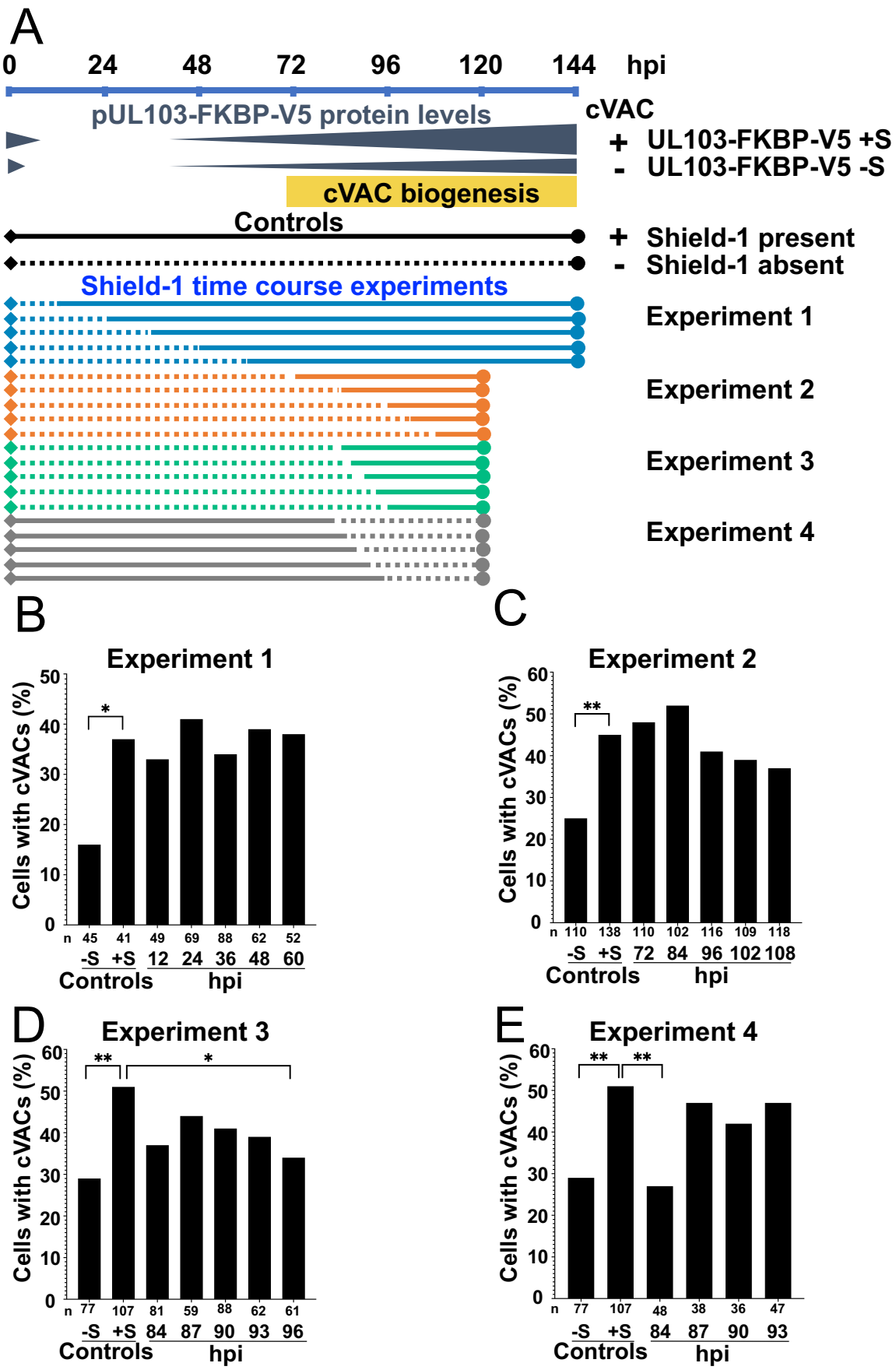
A
0
24
48
72
96
120
144
hpi
pUL103-FKBP-V5 protein levels
cVAC
+ UL103-FKBP-V5 +S
- UL103-FKBP-V5 -S
cVAC biogenesis
Controls
+ Shield-1 present
- Shield-1 absent
Shield-1 time course experiments
Experiment 1
Experiment 2
Experiment 3
Experiment 4
B
Experiment 1
Cells with cVACs (%)
n 45 41 49 69 88 62 52
-S +S 12 24 36 48 60
Controls hpi
C
Experiment 2
Cells with cVACs (%)
n 110 138 110 102 116 109 118
-S +S 72 84 96 102 108
Controls hpi
D
Experiment 3
Cells with cVACs (%)
n 77 107 81 59 88 62 61
-S +S 84 87 90 93 96
Controls hpi
E
Experiment 4
Cells with cVACs (%)
n 77 107 48 38 36 47
-S +S 84 87 90 93
Controls hpi

FIG 3-5 pUL103-FKBP-V5 is needed for maximal cVAC abundance between 72 and 84 hpi. (A) Schematic of events that occur over the infection period including changes in pUL103-FKBP-V5 protein levels (blue-gray triangles), cVAC biogenesis (yellow bar), and Shield-1 presence (solid lines) and absence (dotted lines). Shield-1 was replenished at 72 hpi. Diamonds at the beginning of lines indicate the beginning of infection. The circles at the end of lines indicate when the infection was stopped. Positive and negative controls for experiments were Shield-1 present and absent from time zero, respectively. In experiments 1, 2, and 3, Shield-1 was added between 12 and 60 hpi, 72 and 108 hpi, and 84 and 96 hpi, respectively. In experiment 4, Shield-1 was removed between 84 and 93 hpi. (B) Quantification of cVACs from experiment 1, illustrated in (A). (C) Quantification of cVACs from experiment 2, illustrated in (A). (D) Quantification of cVACs from experiment 3, illustrated in (A). (E) Quantification of cVACs from experiment 4, illustrated in (A). Statistical significance was measured using the chi-square test (*: $p < 0.05$, **: $p < 0.01$,). n, number of cells counted. hpi, hours post infection. Experiment 1 was performed by Ma Christina Lim.

CHAPTER 4: FINAL CONCLUSIONS AND DISCUSSION

History of HCMV. Human cytomegalovirus has been recognized as a human pathogen since as far back as 1881, when several German scientists visualized images of HCMV-infected cells with nuclear (owl's eye) and cytoplasmic inclusions (Fig. 4-1 A-C). However, these scientists mistakenly thought the phenotype was caused by protozoans. By 1932, scientists continued to observe cells with these inclusions, but had not yet identified their origin. During this time a lethal, congenital condition was being characterized. The patients presented with petechiae, swelling of the liver and spleen (hepatosplenomegaly), and intracerebral calcification. In the 25 patients with this condition, they all had cells with nuclear and cytoplasmic inclusions. Because of this, a group of scientists from St. Louis, MO began referring to the condition as generalized cytomegalic inclusion disease (CID). In 1956 and 1957 three independent labs cultured the agent responsible for causing nuclear and cytoplasmic inclusions in various cell types and the agent was termed human cytomegalovirus (44).

Diseases caused by HCMV. From the time HCMV was identified, much has been learned about the types of diseases it causes. For those with intact immune systems, HCMV usually does not cause severe disease. Healthy individuals may experience mononucleosis-like symptoms or none at all. Typically, HCMV requires a deficiency in the host's immunity for severe disease to develop. Thus, the populations most affected by HCMV infections are those with immunodeficiencies (fetuses, infants) and those who are immunocompromised (transplant recipients, HIV/AIDS patients). Immunodeficient populations may suffer from liver, lung, and spleen problems, petechiae, deafness, intracerebral calcification, microcephaly, and/or mental retardation. Immunocompromised populations may suffer from widespread inflammation of the

eyes (choroid and retina), lungs, esophagus, colon, small intestine, brain and meninges, spinal cord, pancreas, and/or liver (64, 67).

HCMV statistics. HCMV is the most common viral congenital infection, as well as the leading cause of birth defects and childhood disabilities in the United States (49, 67). More children suffer from CMV-related diseases than from more well-known congenital conditions such as Down syndrome, fetal alcohol syndrome, or spina bifida (16). The virus is also the most common, non-hereditary cause of hearing loss in children (43).

Current treatments for HCMV infections. Since the discovery of HCMV, there have been five main drugs developed for treating these infections: ganciclovir (GCV), valganciclovir (VGCV), cidofovir (CDV), foscarnet (FOS), and letermovir (LTV). GCV, VGCV, CDV, and FOS all target viral DNA replication. This is problematic because antiviral resistance to one of these medications can lead to resistance to most, if not all. In addition, they can be toxic, resulting in unfavorable side effects for the patients (9, 10, 44, 64, 67). This is partially due to drug interactions between these treatments and other medications taken by the patients (46). The toxicity of these drugs may also be due to off-target effects on host cell DNA replication as well. LTV is the newest and most promising prophylactic treatment for HCMV infections in stem cell transplant recipients. This drug has been shown to be effective against several clinical and GCV-resistant strains and is well tolerated. LTV affects viral DNA packaging and specifically targets the viral terminase complex. There is no known human analog of the terminase complex, which may explain why the drug is so well-tolerated (10). This is the direction the development of new HCMV treatments should be heading. We need to be working towards more novel, specific, diverse, and less toxic therapeutics.

Future HCMV remedies. To effectively prevent and treat HCMV infections we need treatments targeting both lytic and latent infections. The current HCMV treatments target lytic replication and have no effect on the virus in its latent state. For treatments of lytic infections, most, if not all, steps of the replication cycle should be targeted. This strategy has several benefits. The first is the development of antiviral cocktails consisting of multiple drugs for more potent and effective treatments. The second is plenty of drug treatment options available if patients develop antiviral resistance or experience side effects. Preferably, the antiviral targets should have no human analogs, decreasing the chances of off-target effects. pUL103, a tegument protein, has the potential to be an antiviral target for HCMV infections.

pUL103: potential antiviral target. pUL103 has the potential to be an antiviral target for treating HCMV lytic infections. As a tegument protein, pUL103 is multifunctional and has been shown to have multiple roles throughout the life cycle of the virus (1, 26, 72). Thus, multiple steps of the HCMV replication cycle would be affected with targeting pUL103. Several studies showed in the absence of pUL103, HCMV is attenuated (1, 35, 110). Inhibition of pUL103 would weaken the virus and allow the host's immune system time to control the infection. There is no known human analog of pUL103, so chances of patient side effects would be low. In addition, pUL103 is conserved across herpesviruses (74). Treatments targeting it may treat other herpesvirus infections. However, more research of pUL103 is needed before studies investigating its usefulness as an antiviral target can begin. There is still a great deal unknown about pUL103. This research helps to further the knowledge of this protein.

Summary of key research results from chapters 2 and 3. Currently the function, mechanisms of action, and structure of pUL103 are unknown. The goal of our studies is to provide additional information about pUL103 to help fill these gaps in knowledge. In chapter 2, I identified

several residues and two potential motifs that are conserved across the three herpesvirus subfamilies and within the cytomegalovirus genus, respectively. The conservation of these amino acids indicates they have purpose. It suggests they are necessary for the function of pUL103; otherwise, they would not be maintained. These single amino acids are suitable for mutagenesis because each one can be systematically examined for its contribution to the function of pUL103 and HCMV biology. In chapter 2, I also created a structural prediction of pUL103 that aligned well with catalytic and catalytically inactive nucleotidyl transferase domains. This suggested, along with the presence of its conserved herpesvirus domain, that pUL103 may bind nucleic acids, though that hypothesis was not tested in these studies. These experiments are necessary to understanding whether or not pUL103 interacts with DNA or RNA as one of its many functions. In chapter 3, I identified 72-84 hpi as the temporal window when pUL103 is required for maximal cVAC abundance and stability during a 120-hour infection. This is when pUL103 is actively engaged in activity beneficial to the virus. Given the long replication cycle of HCMV, it is difficult to know when mechanistic studies of pUL103 should be performed; this data shows when they should happen. The information in these chapters is necessary for future studies of pUL103, which are critical if the goals of determining its function, mechanisms of action, and structure are to be achieved.

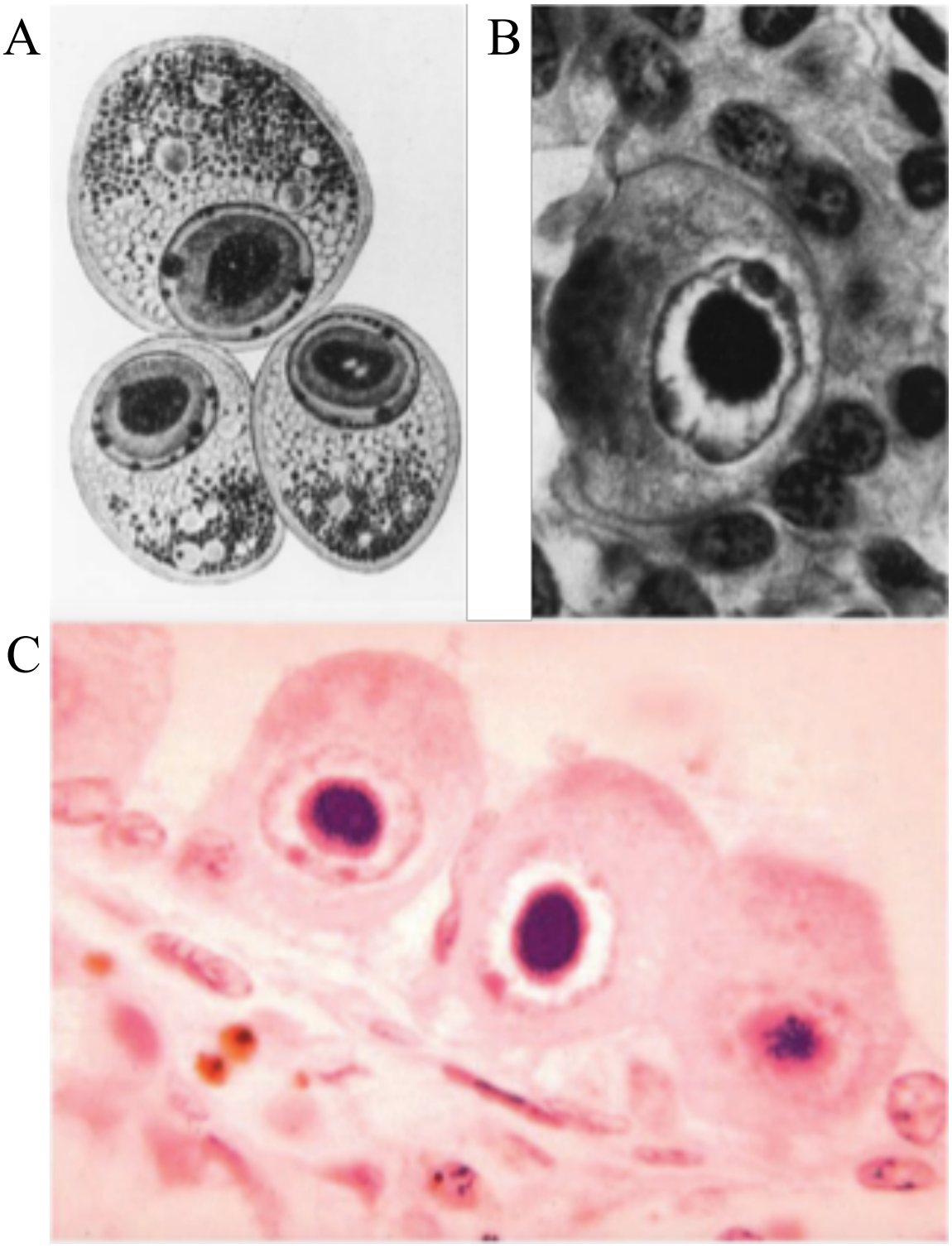

FIG 4-1 Various images of HCMV-infected cells with characteristic nuclear inclusions with halos (owl's eye nuclei). (A) Cells that German scientists mistook for parasite-infected cells. (B) Cell with cytoplasmic inclusion on left side, in addition to nuclear inclusion. Images (A) and (B) from reference Ho, 2008. (C) Centered: cell with basophilic nuclear inclusion body. Image from reference (67).

REFEERENCES

1. Ahlqvist J, Mocarski E. 2011. Cytomegalovirus UL103 controls virion and dense body egress. J Virol 85:5125-35.
2. Aitio O, Hellman M, Kazlauskas A, Vingadassalom DF, Leong JM, Saksela K, Permi P. 2010. Recognition of tandem PxxP motifs as a unique Src homology 3-binding mode triggers pathogen-driven actin assembly. Proc Natl Acad Sci U S A 107:21743-8.
3. Albecka A, Owen DJ, Ivanova L, Brun J, Liman R, Davies L, Ahmed MF, Colaco S, Hollinshead M, Graham SC, Crump CM. 2017. Dual function of the pUL7-pUL51 tegument protein complex in herpes simplex virus 1 infection. J Virol 91.
4. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. J Mol Biol 215:403-10.
5. Alwine JC. 2012. The human cytomegalovirus assembly compartment: a masterpiece of viral manipulation of cellular processes that facilitates assembly and egress. PLoS Pathog 8:e1002878.
6. Anonymous. The PyMOL Molecular Graphics System, Version 2.0.2 Schrödinger, LLC,
7. Ashford P, Hernandez A, Greco TM, Buch A, Sodeik B, Cristea IM, Grunewald K, Shepherd A, Topf M. 2016. HVint: A Strategy for Identifying Novel Protein-Protein Interactions in Herpes Simplex Virus Type 1. Mol Cell Proteomics 15:2939-53.
8. Banaszynski LA, Chen LC, Maynard-Smith LA, Ooi AG, Wandless TJ. 2006. A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. Cell 126:995-1004.
9. Biron KK. 2006. Antiviral drugs for cytomegalovirus diseases. Antiviral Res 71:154-63.

10. Britt WJ, Prichard MN. 2018. New therapies for human cytomegalovirus infections. Antiviral Res 159:153-174.

11. Buchkovich NJ, Maguire TG, Alwine JC. 2010. Role of the endoplasmic reticulum chaperone BiP, SUN domain proteins, and dynein in altering nuclear morphology during human cytomegalovirus infection. J Virol 84:7005-17.

12. Buchkovich NJ, Maguire TG, Paton AW, Paton JC, Alwine JC. 2009. The endoplasmic reticulum chaperone BiP/GRP78 is important in the structure and function of the human cytomegalovirus assembly compartment. J Virol 83:11421-8.

13. Butnaru M, Gaglia MM. 2019. The Kaposi's sarcoma-associated herpesvirus protein ORF42 is required for efficient virion production and expression of viral proteins. Viruses 11:1-24.

14. Butt BG, Owen DJ, Jeffries CM, Ivanova L, Hill CH, Houghton JW, Ahmed MF, Antrobus R, Svergun DI, Welch JJ, Crump CM, Graham SC. 2020. Insights into herpesvirus assembly from the structure of the pUL7:pUL51 complex. Elife 9.

15. Campadelli-Fiume G. 2007. The egress of alphaherpesviruses from the cell. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press, Cambridge, United Kingdom.

16. Cannon MJ, Davis KF. 2005. Washing our hands of the congenital cytomegalovirus disease epidemic. BMC Public Health 5:70.

17. Cavaletto N, Luganini A, Gribaudo G. 2015. Inactivation of the Human Cytomegalovirus US20 Gene Hampers Productive Viral Replication in Endothelial Cells. J Virol 89:11092-106.

18. Chandran B, Hutt-Fletcher L. 2007. Gammaherpesviruses entry and early events during infection. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press, Cambridge, United Kingdom.

19. Clamp M, Cuff J, Searle SM, Barton GJ. 2004. The Jalview Java alignment editor. Bioinformatics 20:426-7.

20. Close WL, Anderson AN, Pellett PE. 2018. Betaherpesvirus virion assembly and egress. Adv Exp Med Biol 1045:167-207.

21. Committee on Taxonomy of Viruses I. 2012. Herpesvirales, p 99-107. Elsevier/Academic Press.

22. Compton T, Feire A. 2007. Early events in human cytomegalovirus infection. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press, Cambridge, United Kingdom.

23. Cruz L, Streck NT, Ferguson K, Desai T, Desai DH, Amin SG, Buchkovich NJ. 2017. Potent Inhibition of Human Cytomegalovirus by Modulation of Cellular SNARE Syntaxin 5. J Virol 91.

24. Dai X, Yu X, Gong H, Jiang X, Abenes G, Liu H, Shivakoti S, Britt WJ, Zhu H, Liu F, Zhou ZH. 2013. The smallest capsid protein mediates binding of the essential tegument protein

pp150 to stabilize DNA-containing capsids in human cytomegalovirus. PLoS Pathog 9:e1003525.

25. Damania BA, Cesarman E. 2013. Kaposi's A. Sarcoma–Associated Herpesvirus. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), Fields Virology, 6th ed, vol 2. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, Pennsylvania.

26. Das S, Ortiz DA, Gurczynski SJ, Khan F, Pellett PE. 2014. Identification of human cytomegalovirus genes important for biogenesis of the cytoplasmic virion assembly complex. J Virol 88:9086-99.

27. Das S, Pellett PE. 2011. Spatial relationships between markers for secretory and endosomal machinery in human cytomegalovirus-infected cells versus those in uninfected cells. J Virol 85:5864-79.

28. Das S, Vasanji A, Pellett PE. 2007. Three-dimensional structure of the human cytomegalovirus cytoplasmic virion assembly complex includes a reoriented secretory apparatus. J Virol 81:11861-9.

29. Davis ZH, Verschueren E, Jang GM, Kleffman K, Johnson JR, Park J, Von Dollen J, Maher MC, Johnson T, Newton W, Jager S, Shales M, Horner J, Hernandez RD, Krogan NJ, Glaunsinger BA. 2015. Global mapping of herpesvirus-host protein complexes reveals a transcription strategy for late genes. Mol Cell 57:349-60.

30. Dereeper A, Audic S, Claverie JM, Blanc G. 2010. BLAST-EXPLORER helps you building datasets for phylogenetic analysis. BMC Evol Biol 10:8.

31. Dereeper A, Guignon V, Blanc G, Audic S, Buffet S, Chevenet F, Dufayard JF, Guindon S, Lefort V, Lescot M, Claverie JM, Gascuel O. 2008. Phylogeny.fr: robust phylogenetic analysis for the non-specialist. Nucleic Acids Res 36:W465-9.

32. Dietz AN, Villinger C, Becker S, Frick M, von Einem J. 2018. A tyrosine-based trafficking motif of the tegument protein pUL71 Is crucial for human cytomegalovirus secondary envelopment. J Virol 92.

33. Dolan A, Cunningham C, Hector RD, Hassan-Walker AF, Lee L, Addison C, Dargan DJ, McGeoch DJ, Gatherer D, Emery VC, Griffiths PD, Sinzger C, McSharry BP, Wilkinson GWG, Davison AJ. 2004. Genetic content of wild-type human cytomegalovirus. J Gen Virol 85:1301-1312.

34. Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch DJ. 1998. The genome sequence of herpes simplex virus type 2. J Virol 72:2010-21.

35. Dunn W, Chou C, Li H, Hai R, Patterson D, Stolc V, Zhu H, Liu F. 2003. Functional profiling of a human cytomegalovirus genome. Proc Natl Acad Sci U S A 100:14223-8.

36. Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792-7.

37. Finkel Y, Stern-Ginossar N, Schwartz M. 2018. Viral Short ORFs and Their Possible Functions. Proteomics 18:e1700255.

38. Fischer D. 2012. Dissecting functional motifs of the human cytomegalovirus tegument protein pUL71. Doctoral thesis. University Medical Center Ulm Institute of Virology, Ulm, Germany.

39. Flint J, Racaniello V, Rall G, Skalka A, Enquist L. 2015. Principles of Virology 4th Edition, vol 1. ASM Press, Washington, DC.

40. Frank D. 2016. The inhibition of the cGAS-STING signaling pathway by KSHV: investigating the relationship between ORF55 and STING. Bachelor of Science. Florida State University.

41. Fuchs W, Granzow H, Klopfleisch R, Klupp BG, Rosenkranz D, Mettenleiter TC. 2005. The UL7 gene of pseudorabies virus encodes a nonessential structural protein which is involved in virion formation and egress. J Virol 79:11291-9.

42. Gilman B, Tijerina P, Russell R. 2017. Distinct RNA-unwinding mechanisms of DEAD-box and DEAH-box RNA helicase proteins in remodeling structured RNAs and RNPs. Biochem Soc Trans 45:1313-1321.

43. Grosse SD, Ross DS, Dollard SC. 2008. Congenital cytomegalovirus (CMV) infection as a cause of permanent bilateral hearing loss: a quantitative assessment. J Clin Virol 41:57-62.

44. Ho M. 2008. The history of cytomegalovirus and its diseases. Med Microbiol Immunol 197:65-73.

45. Hook LM, Grey F, Grabski R, Tirabassi R, Doyle T, Hancock M, Landais I, Jeng S, McWeeney S, Britt W, Nelson JA. 2014. Cytomegalovirus miRNAs target secretory pathway genes to facilitate formation of the virion assembly compartment and reduce cytokine secretion. Cell Host Microbe 15:363-73.

46. Jacobsen T, Sifontis N. 2010. Drug interactions and toxicities associated with the antiviral management of cytomegalovirus infection. Am J Health Syst Pharm 67:1417-25.

47. Johannsen E, Luftig M, Chase MR, Weicksel S, Cahir-McFarland E, Illanes D, Sarracino D, Kieff E. 2004. Proteins of purified Epstein-Barr virus. Proc Natl Acad Sci U S A 101:16286-91.

48. Kaleta EF. 1990. Herpesviruses of birds--a review. Avian Pathol 19:193-211.

49. Kenneson A, Cannon MJ. 2007. Review and meta-analysis of the epidemiology of congenital cytomegalovirus (CMV) infection. Rev Med Virol 17:253-76.

50. Krug LT, Pellett PE. 2014. Roseolovirus molecular biology: recent advances. Curr Opin Virol 9:170-7.

51. Kryshtafovych A, Moult J, Bales P, Bazan JF, Biasini M, Burgin A, Chen C, Cochran FV, Craig TK, Das R, Fass D, Garcia-Doval C, Herzberg O, Lorimer D, Luecke H, Ma X, Nelson DC, van Raaij MJ, Rohwer F, Segall A, Seguritan V, Zeth K, Schwede T. 2014. Challenging the state of the art in protein structure prediction: Highlights of experimental target structures for the 10th Critical Assessment of Techniques for Protein Structure Prediction Experiment CASP10. Proteins 82 Suppl 2:26-42.

52. Liang Q, Wei D, Chung B, Brulois KF, Guo C, Dong S, Gao SJ, Feng P, Liang C, Jung JU. 2018. Novel Role of vBcl2 in the Virion Assembly of Kaposi's Sarcoma-Associated Herpesvirus. J Virol 92.

53. Lim MCR. 2017. Human cytomegalovirus cytoplasmic virion assembly complex: structure in vivo and role of pUL03 in its biogenesis. Master of Science. Wayne State University School of Medicine, Detroit, MI.

54. Livingstone CD, Barton GJ. 1993. Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. Comput Appl Biosci 9:745-56.

55. Longnecker R, Neipel F. 2007. Introduction to the human gamma-herpesviruses. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), Human herpesviruses: biology, therapy, and immunoprophylaxis. Cambridge University Press, Cambridge, United Kingdom.

56. Longnecker RM, Kieff E, Cohen JI. 2019. Epstein-Barr Virus. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), Fields Virology, vol 2. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, Pennsylvania.

57. Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R. 2019. The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636-w641.

58. Marschang RE. 2011. Viruses infecting reptiles. Viruses 3:2087-126.

59. Martin G, Doublie S, Keller W. 2008. Determinants of substrate specificity in RNA-dependent nucleotidyl transferases. Biochim Biophys Acta 1779:206-16.

60. Masud HMAA, Yanagi Y, Watanabe T, Sato Y, Kimura H, Murata T. 2019. Epstein-Barr Virus BBRF2 Is required for maximum infectivity. Microorganisms 7:1-14.

61. McKenzie J, Lopez-Giraldez F, Delecluse HJ, Walsh A, El-Guindy A. 2016. The Epstein-Barr Virus immunoevasins BCRF1 and BPLF1 are expressed by a mechanism independent of the canonical late pre-initiation complex. PLoS Pathog 12:e1006008.

62. Meissner CS, Suffner S, Schauflinger M, von Einem J, Bogner E. 2012. A leucine zipper motif of a tegument protein triggers final envelopment of human cytomegalovirus. J Virol 86:3370-82.

63. Meng B, Lever AM. 2013. Wrapping up the bad news: HIV assembly and release. Retrovirology 10:5.

64. Mocarski ES, Shenk, T., Griffiths, P.D., Pass, R.F. 2013. Cytomegaloviruses. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), Fields virology, 6th ed, vol 2. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, Pennsylvania.

65. Mocarski Jr E. 2007. Betaherpes viral genes and their functions. *In* Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K (ed), Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis. Cambridge University Press, Cambridge, United Kingdom.

66. Murata T. 2018. Encyclopedia of EBV-Encoded Lytic Genes: An Update. Adv Exp Med Biol 1045:395-412.

67. Murray PR, Rosenthal KS, Pfaller MA. 2016. Medical microbiology, 8th ed. Elsevier, Philadelphia, PA.

68. Nicholas J. 1996. Determination and analysis of the complete nucleotide sequence of human herpesvirus. J Virol 70:5975-89.

69. Nokta M, Fons MP, Eaton DC, Albrecht T. 1988. Cytomegalovirus: sodium entry and development of cytomegaly in human fibroblasts. Virology 164:411-9.

70. Nozawa N, Daikoku T, Koshizuka T, Yamauchi Y, Yoshikawa T, Nishiyama Y. 2003. Subcellular localization of herpes simplex virus type 1 UL51 protein and role of palmitoylation in Golgi apparatus targeting. J Virol 77:3204-3216.

71. Nozawa N, Daikoku T, Yamauchi Y, Takakuwa H, Goshima F, Yoshikawa T, Nishiyama Y. 2002. Identification and characterization of the UL7 gene product of herpes simplex virus type 2. Virus Genes 24:257-66.

72. Ortiz DA, Glassbrook JE, Pellett PE. 2016. Protein-protein interactions suggest novel activities of human cytomegalovirus tegument protein pUL103. J Virol 90:7798-810.

73. Paredes AM, Yu D. 2012. Human cytomegalovirus: bacterial artificial chromosome (BAC) cloning and genetic manipulation. Curr Protoc Microbiol Chapter 14:Unit14E.4.

74. Pellett PE, Roizman B. 2013. Herpesviridae. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), Fields virology. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, Pennsylvania.

75. Peng L, Ryazantsev S, Sun R, Zhou ZH. 2010. Three-dimensional visualization of gammaherpesvirus life cycle in host cells by electron tomography. Structure 18:47-58.

76. Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kalé L, Schulten K. 2005. Scalable molecular dynamics with NAMD. J Comput Chem 26:1781-802.

77. Razonable RR, Inoue N, Pinninti SG, Boppana SB, Lazzarotto T, Gabrielli L, Simonazzi G, Pellett PE, Schmid DS. 2020. Clinical Diagnostic Testing for Human Cytomegalovirus Infections. J Infect Dis 221:S74-S85.

78. Ren X, Hurley JH. 2011. Proline-rich regions and motifs in trafficking: from ESCRT interaction to viral exploitation. Traffic 12:1282-90.

79. Roizman B, Knipe DM, Whitley RJ. 2013. Herpes simplex viruses. *In* Knipe DM, Howley PM, Cohen JI, Griffin DE, Lamb RA, Martin MA, Racaniello VR, Roizman B (ed), Fields

virology, 6th ed, vol 2. Wolters Kluwer Health/Lippincott Williams & Wilkins, Philadelphia, Pennsylvania.

80. Roller RJ, Fetters R. 2015. The herpes simplex virus 1 UL51 protein interacts with the UL7 protein and plays a role in its recruitment into the virion. J Virol 89:3112-22.

81. Roller RJ, Haugo AC, Yang K, Baines JD. 2014. The herpes simplex virus 1 UL51 gene product has cell type-specific functions in cell-to-cell spread. J Virol 88:4058-68.

82. Rozenblatt-Rosen O, Deo RC, Padi M, Adelmant G, Calderwood MA, Rolland T, Grace M, Dricot A, Askenazi M, Tavares M, Pevzner SJ, Abderazzaq F, Byrdsong D, Carvunis AR, Chen AA, Cheng J, Correll M, Duarte M, Fan C, Feltkamp MC, Ficarro SB, Franchi R, Garg BK, Gulbahce N, Hao T, Holthaus AM, James R, Korkhin A, Litovchick L, Mar JC, Pak TR, Rabello S, Rubio R, Shen Y, Singh S, Spangle JM, Tasan M, Wanamaker S, Webber JT, Roecklein-Canfield J, Johannsen E, Barabasi AL, Beroukhim R, Kieff E, Cusick ME, Hill DE, Munger K, Marto JA, Quackenbush J, Roth FP, et al. 2012. Interpreting cancer genomes using systematic host network perturbations by tumour virus proteins. Nature 487:491-5.

83. Samson RY, Obita T, Freund SM, Williams RL, Bell SD. 2008. A role for the ESCRT system in cell division in archaea. Science 322:1710-3.

84. Sanchez V, Greis KD, Sztul E, Britt WJ. 2000. Accumulation of virion tegument and envelope proteins in a stable cytoplasmic compartment during human cytomegalovirus replication: characterization of a potential site of virus assembly. J Virol 74:975-86.

85. Sanchez V, Sztul E, Britt WJ. 2000. Human cytomegalovirus pp28 (UL99) localizes to a cytoplasmic compartment which overlaps the endoplasmic reticulum-golgi-intermediate compartment. J Virol 74:3842-51.

86. Sander G, Konrad A, Thurau M, Wies E, Leubert R, Kremmer E, Dinkel H, Schulz T, Neipel F, Sturzl M. 2008. Intracellular localization map of human herpesvirus 8 proteins. J Virol 82:1908-22.

87. Schauflinger M. 2012. Characterization of the human cytomegalovirus protein pUL71 and its impact on viral morphogenesis in fibroblasts, endothelial cells and macrophages. Doctoral thesis. Ulm University Hospital Institute of Virology, Ulm, Germany.

88. Schauflinger M, Fischer D, Schreiber A, Chevillotte M, Walther P, Mertens T, von Einem J. 2011. The tegument protein UL71 of human cytomegalovirus is involved in late envelopment and affects multivesicular bodies. J Virol 85:3821-32.

89. Schauflinger M, Villinger C, Mertens T, Walther P, von Einem J. 2013. Analysis of human cytomegalovirus secondary envelopment by advanced electron microscopy. Cell Microbiol 15:305-14.

90. Schindelin J, Arganda-Carreras I, Frise E, Kaynig V, Longair M, Pietzsch T, Preibisch S, Rueden C, Saalfeld S, Schmid B, Tinevez JY, White DJ, Hartenstein V, Eliceiri K, Tomancak P, Cardona A. 2012. Fiji: an open-source platform for biological-image analysis. Nat Methods 9:676-82.

91. Sehrawat S, Kumar D, Rouse BT. 2018. Herpesviruses: harmonious pathogens but relevant cofactors in other diseases? Front Cell Infect Microbiol 8:177.

92. Senior AW, Evans R, Jumper J, Kirkpatrick J, Sifre L, Green T, Qin C, Zidek A, Nelson AWR, Bridgland A, Penedones H, Petersen S, Simonyan K, Crossan S, Kohli P, Jones DT, Silver D, Kavukcuoglu K, Hassabis D. 2019. Protein structure prediction using multiple deep neural networks in the 13th Critical Assessment of Protein Structure Prediction (CASP13). Proteins 87:1141-1148.

93. Seo JY, Britt WJ. 2007. Cytoplasmic envelopment of human cytomegalovirus requires the postlocalization function of tegument protein pp28 within the assembly compartment. J Virol 81:6536-47.

94. Smith RM, Kosuri S, Kerry JA. 2014. Role of human cytomegalovirus tegument proteins in virion assembly. Viruses 6:582-605.

95. Sosa JM, Huber DE, Welk B, Fraser HL. 2014. Development and application of MIPAR™: a novel software package for two- and three-dimensional microstructural characterization. Integrating Materials and Manufacturing Innovation 3:123-140.

96. Stern-Ginossar N, Weisburd B, Michalski A, Le VT, Hein MY, Huang SX, Ma M, Shen B, Qian SB, Hengel H, Mann M, Ingolia NT, Weissman JS. 2012. Decoding human cytomegalovirus. Science 338:1088-93.

97. Streblow D, Varnum S, Smith R, Nelson J. 2006. A Proteomics Analysis of Human Cytomegalovirus Particles. *In* Reddehase M (ed), Cytomegaloviruses: Molecular Biology and Immunology. Caister Academic Press, Mainz, Germany.

98. Tanaka M, Sata T, Kawaguchi Y. 2008. The product of the Herpes simplex virus 1 UL7 gene interacts with a mitochondrial protein, adenine nucleotide translocator 2. Virol J 5:125.

99. Tang H, Mori Y. 2018. Glycoproteins of HHV-6A and HHV-6B. Adv Exp Med Biol 1045:145-165.

100. Tesini BL, Epstein LG, Caserta MT. 2014. Clinical impact of primary infection with roseoloviruses. Curr Opin Virol 9:91-6.

101. Ueda K. 2018. KSHV genome replication and maintenance in latency. Adv Exp Med Biol 1045:299-320.

102. Varnum SM, Streblow DN, Monroe ME, Smith P, Auberry KJ, Pasa-Tolic L, Wang D, Camp DG, 2nd, Rodland K, Wiley S, Britt W, Shenk T, Smith RD, Nelson JA. 2004. Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. J Virol 78:10960-6.

103. White S, Kawano H, Harata NC, Roller RJ. 2020. HSV Forms an HCMV-like Viral Assembly Center in Neuronal Cells. bioRxiv doi:10.1101/2020.04.22.055145:2020.04.22.055145.

104. Womack A, Shenk T. 2010. Human cytomegalovirus tegument protein pUL71 is required for efficient virion egress. MBio 1.

105. Wozniakowski G, Samorek-Salamonowicz E. 2015. Animal herpesviruses and their zoonotic potential for cross-species infection. Ann Agric Environ Med 22:191-4.

106. Wu H, Ballestas M, Sanchez V, Mach M, Britt W. 2019. HCMV envelope protein gpUL132 controls viral production through formation of the viral assembly compartment, Knoxville, TN.

107. Yamashita S, Nagaike T, Tomita K. 2019. Crystal structure of the Lin28-interacting module of human terminal uridylyltransferase that regulates let-7 expression. Nat Commun 10:1960.

108. Yanagi Y, Masud H, Watanabe T, Sato Y, Goshima F, Kimura H, Murata T. 2019. Initial characterization of the Epstein(-)Barr Virus BSRF1 gene product. Viruses 11:1-17.

109. Yang J, Zhang Y. 2015. I-TASSER server: new development for protein structure and function predictions. Nucleic Acids Res 43:W174-81.

110. Yu D, Silva MC, Shenk T. 2003. Functional map of human cytomegalovirus AD169 defined by global mutational analysis. Proc Natl Acad Sci U S A 100:12396-401.

111. Yu X, Shah S, Atanasov I, Lo P, Liu F, Britt WJ, Zhou ZH. 2005. Three-dimensional localization of the smallest capsid protein in the human cytomegalovirus capsid. J Virol 79:1327-32.

112. Yu X, Shah S, Lee M, Dai W, Lo P, Britt W, Zhu H, Liu F, Zhou ZH. 2011. Biochemical and structural characterization of the capsid-bound tegument proteins of human cytomegalovirus. J Struct Biol 174:451-60.

113. Zeltzer S, Zeltzer CA, Igarashi S, Wilson J, Donaldson JG, Goodrum F. 2018. Virus Control of Trafficking from Sorting Endosomes. mBio 9.

114. Zhou ZH, Prasad BV, Jakana J, Rixon FJ, Chiu W. 1994. Protein subunit structures in the herpes simplex virus A-capsid determined from 400 kV spot-scan electron cryomicroscopy. J Mol Biol 242:456-69.

ABSTRACT

BIOLOGICAL AND COMPUTATIONAL STUDIES OF THE STRUCTURE AND FUNCTION OF PUL103, A HUMAN CYTOMEGALOVIRUS TEGUMENT PROTEIN

by

ASHLEY ANDERSON

August 2020

Advisor: Philip E. Pellett, PhD

Major: Immunology and Microbiology

Degree: Doctor of Philosophy

Human cytomegalovirus (HCMV) is an enveloped, single segment, double-stranded DNA virus. HCMV infection causes disease in immunocompromised (HIV patients, transplant recipients) and immunodeficient (fetuses, neonates) populations. Current treatments are effective but are either limited in use or can lead to organ damage and/or antiviral resistance, and no vaccines are available. Additional antiviral targets are needed. HCMV pUL103 is a potential antiviral target. pUL103 is a conserved herpesvirus protein present in the tegument, layer of proteins and RNA between the envelope and capsid of HCMV virions. pUL103 helps reorganize cellular secretory machinery (Golgi, endosomes) to form the cytoplasmic virion assembly compartment (cVAC) that is hypothesized to facilitate efficient virion assembly and egress. pUL103 is also important for cell-to-cell spread and virion maturation. The structure, function, and mechanisms of action of pUL103 are unknown.

To identify amino acids important for pUL103 functions, I aligned the sequences of 14 pUL103 homologs using MUSCLE (Multiple Sequence Comparison by Log-Expectation). My data showed 12 conserved amino acids outside of the conserved herpesvirus domain. Within the

N terminus there are 13 conserved residues, which shows the importance of this region and these residues in pUL103 function.

To guide biological studies of pUL103, I used I-TASSER (Iterative Threading ASSEmbly Refinement) to predict structures of pUL103 and its homologs HSV-1 pUL7 (pUL7) and EBV BBRF2 (BBRF2) using their amino acid sequences. My data showed (a) their predicted secondary structures are similar, (b) the predicted structure of BBRF2 differs from pUL103 and pUL7, which are similar, and (c) the predicted pUL103 structure aligned with the non-catalytic LIM domain of human terminal uridylyl transferase 4 (TUT4), a nucleotidyl transferase domain lacking catalytic residues. This data suggests pUL103 may bind nucleic acids; experiments can be designed to test this hypothesis.

To identify when pUL103 mechanistic studies should be performed, I determined when pUL103 is necessary for maximal cVAC abundance. During 120-hour infections, I showed pUL103 must be present between 72 and 84 hpi. Such studies might include analysis of virion assembly, intra- and extracellular viral titers, and deep temporal and spatial proteomics analysis of viral and cellular proteins during this process.

AUTOBIOGRAPHICAL STATEMENT

My name is Ashley Nicole Anderson. I was born Tuesday, September 26, 1989 at Women and Infants Hospital in Providence, RI. I lived on the east coast in Massachusetts with my mother, brother, and father until I was 6 years old; then we moved to Detroit, Michigan.

I excelled academically from elementary all throughout high school, despite adversity. In 2005, I was diagnosed with systemic lupus erythematosus (SLE, lupus). SLE is an autoimmune disease that disproportionately affects women, specifically women of color. Lupus gave me strength and changed my career trajectory for the better.

As an undergraduate at the University of Michigan-Ann Arbor (U of M), I continued to excel in my academics and had the opportunity to conduct lupus research in the laboratory of Dr. James Shayman under the tutelage of Dr. Robert Kelly. In April 2011, I graduated with my Bachelor of Science degree, with honors, in Neuroscience. After receiving my undergraduate degree, I began my predoctoral studies at Wayne State University.

As a doctoral candidate in the laboratory of Dr. Philip Pellett, my dissertation work focused on the human cytomegalovirus tegument protein pUL103. I utilized biological and computational studies to fill gaps in knowledge relating to the structure, function, and mechanisms of action of this protein. It has been a long journey, but I have finally done all the work necessary to achieve my PhD in immunology and microbiology. Though this phase of my life has come to an end, I am excited about what lies ahead of me.

I accepted a post-doctoral position at the University of Virginia in the laboratory of Dr. Shu Man Fu. There, I will study lupus, specifically the origin of lupus-related auto-antibodies (auto-Ab). My project will focus on identifying the origin of auto-Ab to Ro60. Anti-Ro60 are the most common auto-Ab in healthy normal individuals.