Characterization of Structural and Functional Flexibility in Herpesvirus Outer Tegument Proteins

A thesis submitted by

Claire Marie Metrick

in partial fulfillment of the requirements for the degree of

PhD

in

Biochemistry

Tufts University

Sackler School of Graduate Biomedical Sciences

February 2018

Adviser: Ekaterina Heldwein, PhD

Abstract

Herpesviruses are large, enveloped, DNA viruses that infect nearly all vertebrates from fish to humans and cause diseases ranging from mild to fatal. They are morphologically complex and sandwiched between their genome-filled capsid and glycoprotein-studded envelope, herpesviruses feature a dynamic layer of viral proteins called the tegument. Tegument proteins play important structural roles in virion morphogenesis, but a detailed understanding of how these proteins operate is lacking. It is also becoming clear that many tegument proteins play regulatory roles in addition to their canonical functions in viral structure and assembly.

Herpes simplex virus type 1 (HSV-1) UL11 is the smallest tegument protein. It is conserved among all herpesviruses and is found associated with cytoplasmic membranes through acyl modifications at its N-terminus. UL11 binds UL16, a tegument protein in HSV-1 that is also completely conserved. As UL16 binds capsids in the cytoplasm, the interaction between these two proteins is thought to form a physical link between the capsid and the envelope and play a structural role in secondary envelopment, the step of viral maturation and egress in which nucleocapsids gain their final envelope. In HSV-1, this interaction requires tegument protein UL21. UL21 is conserved among alphaherpesviruses and is found in the cytoplasm and nucleus in transfected and infected cells. With capsid-associated UL21, UL16 and UL11 bind glycoprotein E, forming an even more stable link between the capsid and the envelope. HSV-1 UL51 is another conserved tegument protein that is associated with membranes by lipid modification. It too is thought to play structural roles in virion morphogenesis through interactions with conserved tegument proteins UL7 and UL14 and glycoprotein E to affect secondary envelopment.

In addition to these structural roles that are not completely understood, there is evidence that UL11, UL16, UL21, and UL51 also function independently of their complexes. To provide context for exploration of the multiple roles of these proteins, we have characterized them structurally and biochemically. In addition to determining the crystal structures of HSV-1 UL21 domains, which revealed novel protein folds, we have characterized each of these proteins as being conformationally flexible, which likely contributes to their multifunctionality. Current work focuses on carefully characterizing binding interactions, including two novel interactions with nucleic acids, and assessing their structural implications to better unravel the mechanisms of the many functions these proteins play during viral replication.

Acknowledgments

Firstly, I need to thank my thesis mentor, Dr. Ekaterina "Katya" Heldwein, PhD. Joining Katya's lab felt intuitive to me due to the similarities between her research and my previous lab experience as a technician—in fact, so intuitive that (in a very risky move) I committed to Tufts before meeting her or inquiring about her availability to take on new students. Things worked out in my favor and I joined her lab after a successful first rotation. Six years, two crystal structures, three papers, a pre-doctoral NIH fellowship, and too much SAXS data later, I think she'd report that taking me on worked out for her too. Katya taught me so much, but she also allowed me freedom and independence, which gave me invaluable practice in teaching myself and others.

Despite the propensity to provide distractions (both good and bad), my labmates helped me look forward to coming in each day. All of my thanks to the Heldwinners, in order of appearance with regards to my tenure: Elvira Vitu, Jessica Silverman, Sam Stampfer, Jared Pitts, Henry Rogalin, Heidi Burke (my lab twin), Janna Bigalke, Rebecca Cooper, Andrea Koenigsberg (my "yay! bay" bae), Xuanzong Guo, Ellen White, Lizzy Draganova, Mike Thorsen, Anshu Mittal, Adam Hilterbrand, and Alex Vorperian. Thank you for your advice, questions, and most of all, camaraderie. I'm proud to say that I met some of my best friends in Katya's lab.

Thanks to my thesis advisory committee Drs. Andrew Bohm, Michael Forgac, and Claire Moore for providing guidance in experiments and training over my graduate career and to Dr. Thomas Schwartz for reviewing my thesis as my outside examiner. Neither the defense process nor other crowning achievements during my graduate career would have been possible without the rest of the faculty, staff, and students of the Biochemistry and ISP programs, the CMDB and Microbiology departments, and

everyone in the Dean's office. I found the Sackler School to be a truly supportive and special community and feel fortunate to have been a part of it.

I must thank my rotation students and trainees Lamyaa Shaban, Aaron Bernstein, Dan Golden, Kathy Nevola, and Jose Martin Ramirez not only for generating useful data for my project, but also for giving me the opportunity to act as a mentor—a vital experience for both my career and personal growth.

As physical endeavors were imperative to maintaining my (relative) sanity during graduate school, I must thank my favorite athletic trainers Ali Isdaner, Melanie O'Brien, Jaime Petricca, Caitlyn Graham Visconte, and Yovanny Henningsen for keeping me going by helping me to release frustration and generate so many second (and third, fourth, fifth…) winds.

Finally, I must thank my parents Lee and Linda Metrick, my sister, Lisa Metrick, my especially curious aunt and uncle, Terry and Tim Collins, my close-as-family friend, Kelsey Brodersen, and my previous advisor Dr. Kurt Swanson for always believing in me, being sources of unending support, and most importantly, providing the context, perspective, and motivation that kept me going through my entire graduate career.

Table of Contents

List of Tables

List of Figures

List of Copyrighted Materials

Figures from the following publications were reprinted, some with listed adaptations, in the introduction:

Coen, D.M. and P.A. Schaffer, Antiherpesvirus drugs: a promising spectrum of new drugs and drug targets. Nat Rev Drug Discov, 2003. 2(4): p. 278-88.

Mettenleiter, T.C., B.G. Klupp, and H. Granzow, Herpesvirus assembly: an update. Virus Res, 2009. 143(2): p. 222-34.

Cooper, R.S. and E.E. Heldwein, Herpesvirus gB: A Finely Tuned Fusion Machine. Viruses, 2015. 7(12): p. 6552-69.

Johnson, D.C. and J.D. Baines, Herpesviruses remodel host membranes for virus egress. Nat Rev Microbiol, 2011. 9(5): p. 382-94.

Owen, D.J., C.M. Crump, and S.C. Graham, Tegument Assembly and Secondary Envelopment of Alphaherpesviruses. Viruses, 2015. 7(9): p. 5084-114.

Leege, T., et al., Effects of simultaneous deletion of pUL11 and glycoprotein M on virion maturation of herpes simplex virus type 1. J Virol, 2009. 83(2): p. 896-907.

Starkey, J.L., et al., Elucidation of the block to herpes simplex virus egress in the absence of tegument protein UL16 reveals a novel interaction with VP22. J Virol, 2014. 88(1): p. 110-9.

Sarfo, A., et al., The UL21 tegument protein of herpes simplex virus type-1 is differentially required for the syncytial phenotype. J Virol, 2017.

Han, J., et al., Function of glycoprotein E of herpes simplex virus requires coordinated assembly of three tegument proteins on its cytoplasmic tail. Proc Natl Acad Sci U S A, 2012. 109(48): p. 19798-803.

Albecka, A., et al., Dual Function of the pUL7-pUL51 Tegument Protein Complex in Herpes Simplex Virus 1 Infection. J Virol, 2017. 91(2).

Oda, S., et al., The Interaction between Herpes Simplex Virus 1 Tegument Proteins UL51 and UL14 and Its Role in Virion Morphogenesis. J Virol, 2016. 90(19): p. 8754-67.

Portions of the following manuscripts previously published by the author were reprinted in Chapter 2.2, Chapter 4, and Chapter 7.3:

Metrick, C.M., P. Chadha, and E.E. Heldwein, The unusual fold of herpes simplex virus 1 UL21, a multifunctional tegument protein. J Virol, 2015. 89(5): p. 2979-84.

Metrick, C.M. and E.E. Heldwein, Novel Structure and Unexpected RNA-Binding Ability of the C-Terminal Domain of Herpes Simplex Virus 1 Tegument Protein UL21. J Virol, 2016. 90(12): p. 5759-69.

List of Abbreviations

10H, His10	10X histidine tag
6H, His6	6X histidine tag
A260	absorption 260 nm
A280	absorption 280 nm
BHV-1	bovine herpesvirus 1
BHV-5	bovine herpesvirus 5
BS3	Bis(sulphosuccinimidyl)suberate
CD	circular dichroism
CHOL	cholesterol
CMV	cytomegalovirus
CTD	C-terminal domain
CV	column volume
CVSC	capsid specific vertex complex
DHB	dihydroxybenzoic acid
Dmax	maximum dimension
dsRNA	double stranded RNA
DTT	dithiothreitol
E	elution
EBV	Epstein Barr virus
EDTA	ethylenediaminetetraacetic acid
EHV-1	equine herpesvirus 1
ELU	elution
FT	flow through
gB, H, etc.	glycoprotein B, glycoprotein H, etc.
GFP	green fluorescent protein
GSH	glutathione
GST	glutathione-s-transferase

HEPES	4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid
HHV-1	human herpesvirus 1
HHV-2	human herpesvirus 2
HHV-3	human herpesvirus 3
HHV-4	human herpesvirus 4
HHV-5	human herpesvirus 5
HHV-6	human herpesvirus 6
HHV-7	human herpesvirus 7
HHV-8	human herpesvirus 8
HRP	horseradish peroxidase
HRV3C, PP	human rhinovirus 3C protease, PreScission protease
HSV-1	herpes simplex virus type 1
HSV-2	herpes simplex virus type 2
HSW	high salt wash
ICP	infected cell protein
IDR	intrinsically disordered region
IgG	immunoglobulin G
INSOL	insoluble
IPTG	isopropyl-β-d-thiogalactopyranoside
kD	kilodalton
kDa	kilodalton
KSHV	Kaposi's sarcoma associated herpesvirus
LB	luria broth
LSW	low salt wash
LYS	lysate
MALDI-TOF	matrix assisted laser desorption/ionization time of flight
MLV	multilamellar vesicle
MOPS	3-(N-morpholino)propanesulfonic acid

MPD	2-methyl-2,4-pentanediol
MRE	mean residue ellipticity
MRW	mean residue weight
MW	molecular weight
MWCO	molecular weight cut off
NEC	nuclear egress complex
NS1	non-structural 1
NTD	N-terminal domain
OD	optical density
PA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
PAGE	polyacrylamide gel electrophoresis
PC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
PCR	polymerase chain reaction
POPA	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate
POPC	1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine
POPS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
PRV	pseudorabies virus
PS	1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine
PSM	proteolytic site mutant
PVDF	polyvinylidene fluoride
Rg	radius of gyration
SAD	single-wavelength anomalous dispersion
SAXS	small-angle X-ray scattering
SDS	sodium dodecyl sulfate
SEC	size exclusion chromatography
SeMet	selenomethionine
SOE	single overlap extension
SOL	soluble

StII	StrepII tag
TB	terrific broth
TBST	tris-buffered saline with tween
TCEP	tris(2-carboxyethyl)phosphine
TLCK	tosyl-L-lysyl-chloromethyl ketone
TPCK	tosyl phenylalanyl chloromethyl ketone
UL	unique long
UL16C	UL16 C-terminal domain
UL16CS	UL16 cysteine to serine mutant
UL16N	UL16 N-terminal domain
UL21C	UL21 N-terminal domain
UL21N	UL21 N-terminal domain
US	unique short
vhs	virion host shutoff
VP	viral protein
VZV	varicella zoster virus
W1, W2…	wash fraction 1, 2…
Xa	factor Xa protease
α-TIF	alpha trans-inducing factor

Chapter 1: Introduction

1.1 General Virology

1.1.1 Overview of viruses

Viruses are tiny packages of biomolecules that, despite their size and apparent simplicity, can cause devastating disease to a host as side effects of their only goal: to replicate. For example, the well-known influenza virus is a small, enveloped virus that consists of an RNA genome that encodes fourteen different proteins and usually causes acute respiratory illness [1], but in the 1918 pandemic, 50-100 million people succumbed to their influenza infections [2]. They cause many types of disease ranging from the typically asymptomatic infections of adenoviruses, to massively fatal hemorrhagic fever caused by filoviruses including ebolavirus, and come in many shapes, sizes, and types. For example poliovirus comprises an RNA genome wrapped only in a tiny proteinaceous capsid with total diameter about 20 nm [1], whereas herpesviruses are made of a double-stranded DNA genome packed into a proteinaceous capsid, surrounded by a complex layer of structural and regulatory proteins, and encompassed in an envelope with a total particle diameter larger than 200 nm [3].

In order to replicate, viruses must enter the host cell, make copies of their genomic material, and repackage these new genomic copies into particles that function as new infectious agents (Figure 1-1), all while evading the host's defensive strategies.

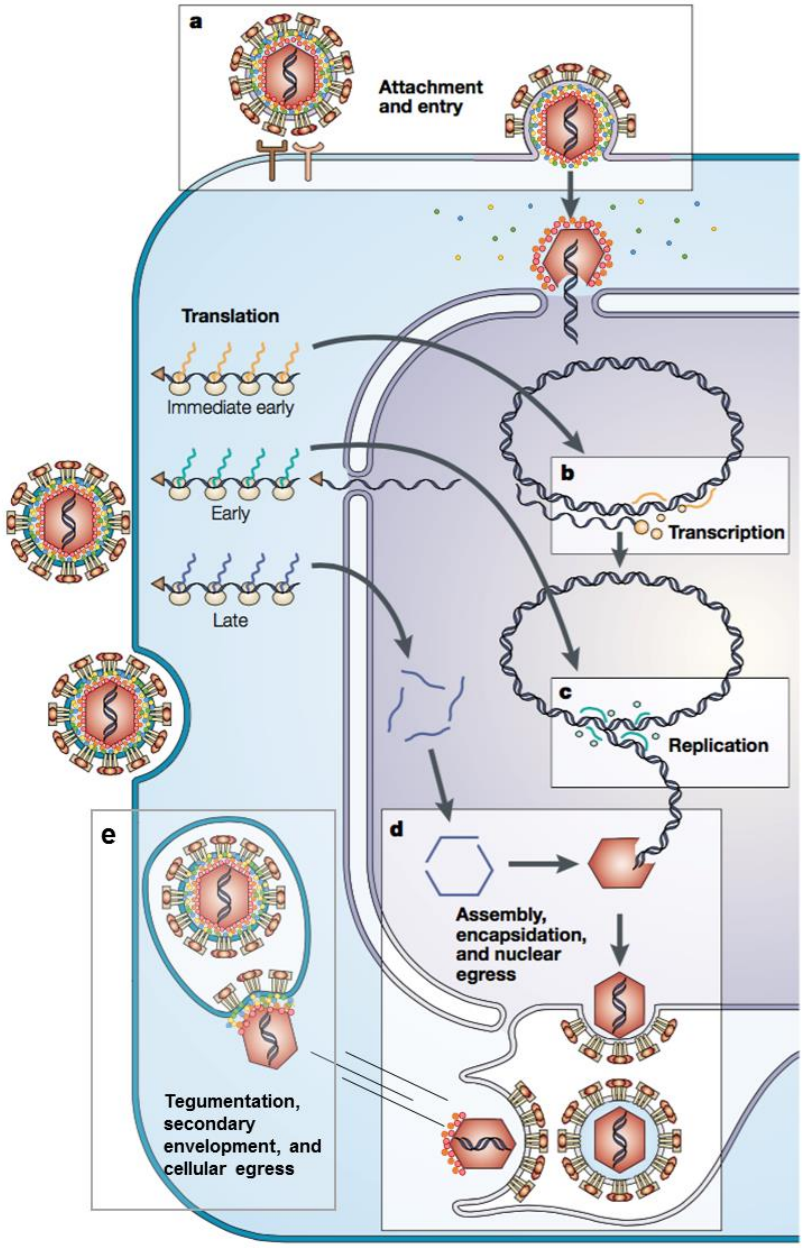


Figure 1-1. Overview of the herpesvirus replication cycle

The virus (a) attaches and fuses its membrane with that of the host cell to allow the genetic material access to the nucleus. Note that in specific cases, fusion can occur within endosomes (not shown). In the nucleus, viral genes are (b) transcribed for translation, viral DNA is (c) replicated, and (d) assembly of progeny virions begins. The egress of the new virions starts in the nucleus. After they exit the nucleus, they continue to mature by (e) traveling through the cytoplasm and gaining an envelope on the way out of the cell. Mature particles can also be delivered directly to neighboring cells (not shown).Reprinted with permission from [4]. Adapted to add representation of tegument proteins (colored circles) and cytoskeletal networks (black lines).

All viruses are reliant on the host cell and use a combination of virally-encoded proteins and hijacked host proteins to perform every function necessary for replication These required tasks can be split into two general categories: regulatory functions that help the virus to replicate successfully in a host cell and structural functions that control assembling these newly replicated materials and turning them into infectious progeny. Since viruses are so small, their genome size and therefore coding capacity is physically limited. In addition to coopting host proteins to perform many of the tasks required for replication, many viral proteins are multi-taskers. For example, VP35 from filoviruses is an essential structural and regulatory cofactor to the viral RNA-dependent RNA polymerase L [5-7] during viral replication, but also inhibits antiviral response by binding and sequestering double-stranded RNA (dsRNA) [8] that would typically trigger an innate immune response if detected by the cell. Similarly, Influenza virus non-structural protein 1 (NS1) binds dsRNA to prevent an immune response, but also may play a role in anti-apoptotic mechanisms through activation of phosphoinositide 3-kinase (PI3K), and has been shown to interact with a number of other cellular proteins with unknown implications [9]. Multifunctional proteins are a common feature of many different viruses and the research presented in this thesis focuses on characterizing the roles of multifunctional proteins from a large, complex virus: Herpes simplex virus type 1 (HSV-1).

1.1.2 Overview of herpesviruses

Herpesviruses are everywhere. There are hundreds of types of herpesviruses that infect many vertebrates ranging from human to catfish, and within the past fifteen years, even an oyster herpesvirus was characterized [10]. These viruses are causative agents of

many diseases with wildly divergent symptoms depending on the virus and host and have an extreme impact on human life, even when they are not directly infecting humans. For example, herpesviruses can cause upper respiratory infections in many domesticated animals including cats and horses, and lessen a favorite pet's quality of life. They can also have an impact on our food. Pseudorabies virus (PRV), a veterinary herpesvirus that infects pigs, and oyster herpesvirus have extremely high mortality rates that lead to massive loss of farmed livestock.

Most of the hundreds of herpesviruses that have yet been identified are grouped into three subfamilies: alpha-, beta-, and gammaherpesviruses. Among these many viruses, there are eight herpesviruses that infect humans: Herpes simplex virus types 1 and 2 (HSV-1/HHV-1, HSV-2/HHV-2), Varicella-zoster virus (VZV/HHV-3), Epstein-Barr Virus (EBV/HHV-4), Human herpesvirus 6 (HHV-6), Human herpesvirus 7 (HHV-7), and Kaposi's sarcoma-associated herpesvirus (KSHV/HHV-8). These pathogens generally cause rather mild disease symptoms that can be exacerbated when the patient is immunocompromised. For example, HSV-1 often causes orofacial lesions commonly known as cold sores, but in severe cases, this can become keratitis, from which approximately 2% of patients go blind [11], or encephalitis that can be fatal. Similarly, VZV causes patients to break out with the well-known dermal lesions of chicken pox or shingles, but infection can also lead to neuralgia and fatal pneumonia [12, 13]. Most commonly though, these viral infections are asymptomatic, which allows for their spread to go unnoticed and unimpaired.

One hallmark of herpesvirus infection is the ability to maintain two distinct stages of infection—lytic and latent, so a herpesvirus infection lasts forever. After primary

infection in epithelial cells, alphaherpesviruses (which include human herpesviruses HSV-1, HSV-2, VZV, veterinary pathogen PRV, and others) can infect neurons to establish a latent, lifelong infection. In latency, the genome is maintained in the nucleus and no viral progeny is produced until some stressor initiates reactivation of a lytic infection. Herpesviruses in other subfamilies also establish latency but in different cell types. The specifics of latency maintenance and reactivation are not fully understood. In addition to the fact that treatment options are limited to a single vaccine (for VZV) and a handful of antivirals to combat human herpesviruses, this lifelong infection also contributes to the persistence of herpesviruses within an individual and amongst the population.

1.1.3 Herpesviral morphology

Herpesviruses are large, complex particles. Members of the order herpesvirales, which includes all identified herpesviruses, are classified by being relatively large enveloped viruses with a linear, double-stranded DNA genome packaged in a nucleocapsid that is surrounded by a protein matrix called the tegument [14] (Figure 1-2). Human herpesviral genomes are also relatively large as they encode as few as approximately 70 proteins (HSV-1, -2, and VZV) (Table 1-1) to over 150 proteins in the case of CMV [15]. The physical organization of these protein-coding genes in the genome of HSV-1 allows for their classification and naming based on whether they exist in the unique long (UL) or unique short (US) portion of the genome [1]. The genome is densely packed into an icosahedral capsid that is formed in HSV-1 mainly by viral protein UL19/VP5 organized in hexamers (hexons) on the faces of the capsid and pentamers (pentons) at the vertices. UL35/VP26 associates with hexons, UL38/VP19C

and UL18/VP23 form triplexes that sit between hexons and pentons, and UL36 (or part of it) may be in complex with either pentons or the triplex-associated capsid-specific vertex component (CVSC)[16] (Figure1-2). The CVSC is made of UL25 and UL17 and is thought to mediate DNA expulsion from and retention in the nucleocapsid on viral entry and egress, respectively.

Table 1-1. Selected herpesviral proteins, virion locations, and functions

protein	location	previously described potential function(s)
glycoprotein B	envelope	fusion machinery: master fusogen
glycoprotein D	envelope	fusion machinery: receptor binding, signals to gB through gH/gL
glycoprotein E/ Us8	envelope	fusion machinery: cell spread; bridge capsid to membrane with UL11/UL16/UL21, UL51, other tegument proteins
glycoprotein H/ UL22	envelope	fusion machinery: gB trigger with gL
glycoprotein K/ UL53	envelope	fusion machinery: accessory; binds UL37 to regulate secondary envelopment?
glycoprotein L/ UL1	envelope	fusion machinery: gB trigger with gH
ICP0	tegument	initiates viral gene transcription
UL6	capsid	portal vertex member

UL7	tegument	regulate secondary envelopment through binding UL51, unknown mitochondrial function, regulate infected cell morphology with UL51
UL11	tegument (membrane-associated)	regulate secondary envelopment through binding UL16, regulate cell spread through binding gE with UL16/UL21, unknown nuclear function (potentially primary envelopment)
UL13	tegument	viral kinase
UL14	tegument	regulate secondary envelopment through binding UL51, apoptosis inhibition?
UL15	not included	terminase subunit
UL16	tegument	regulate secondary envelopment through binding UL11 and UL21, regulate cell spread through binding gE with UL11/UL21, chaperone-like activities?, unknown nuclear and mitochondrial functions?
UL17	capsid	capsid-specific vertex component (CVSC) member
UL18/VP23	capsid	triplex component
UL19/VP5/ ICP5	capsid	major capsid protein, many interactions with tegument
UL20	envelope	regulates secondary envelopment with UL37, fusion modulator
UL21	tegument	regulates secondary envelopment with UL16/UL11, regulates cell spread with UL16/UL11 on gE , fusion

		modulator, unknown nuclear function (potentially primary envelopment)
UL25	capsid	capsid-specific vertex component (CVSC) member, facilitates viral DNA expulsion from/retention in capsid
UL26	capsid	capsid scaffold, protease
UL28	not included	terminase subunit
UL31	not included	nuclear egress complex subunit
UL33	not included	terminase subunit
UL34	not included	nuclear egress complex subunit
UL35/VP26	capsid	outer capsid component, many interactions with tegument
UL36/VP1/2	tegument	major inner tegument protein ("hub"), interacts with capsid and many tegument proteins, required for secondary envelopment, deubiquitinase activity used in anti-host defense mechanisms, implicated in capsid trafficking
UL37	tegument	inner tegument protein, required for secondary envelopment, and implicated in capsid trafficking
UL38/VP19C	capsid	triplex component
UL41/vhs	tegument	"viral host shutoff" degrades host RNA early in infection to favor viral transcripts, potential role in assembly

UL46/VP11/12	tegument	potential role in assembly, potential regulator of UL48
UL47/VP13/14	tegument	potential role in assembly, potential regulator of UL48, nucleocytoplasmic shuttle, binds ssRNA, potentially packages RNA into virions
UL48/VP16/ α-TIF	tegument	"trans-inducing factor" nucleates a transcription complex to initiate viral gene expression, outer tegument "hub" imperative for tegument assembly, potential role in secondary envelopment
UL49/VP22	tegument	potential roles in assembly and secondary envelopment, binds RNA
UL51	tegument (membrane-associated)	regulates cell spread, secondary envelopment, and with UL7, infected cell morphology
US3	tegument	viral kinase

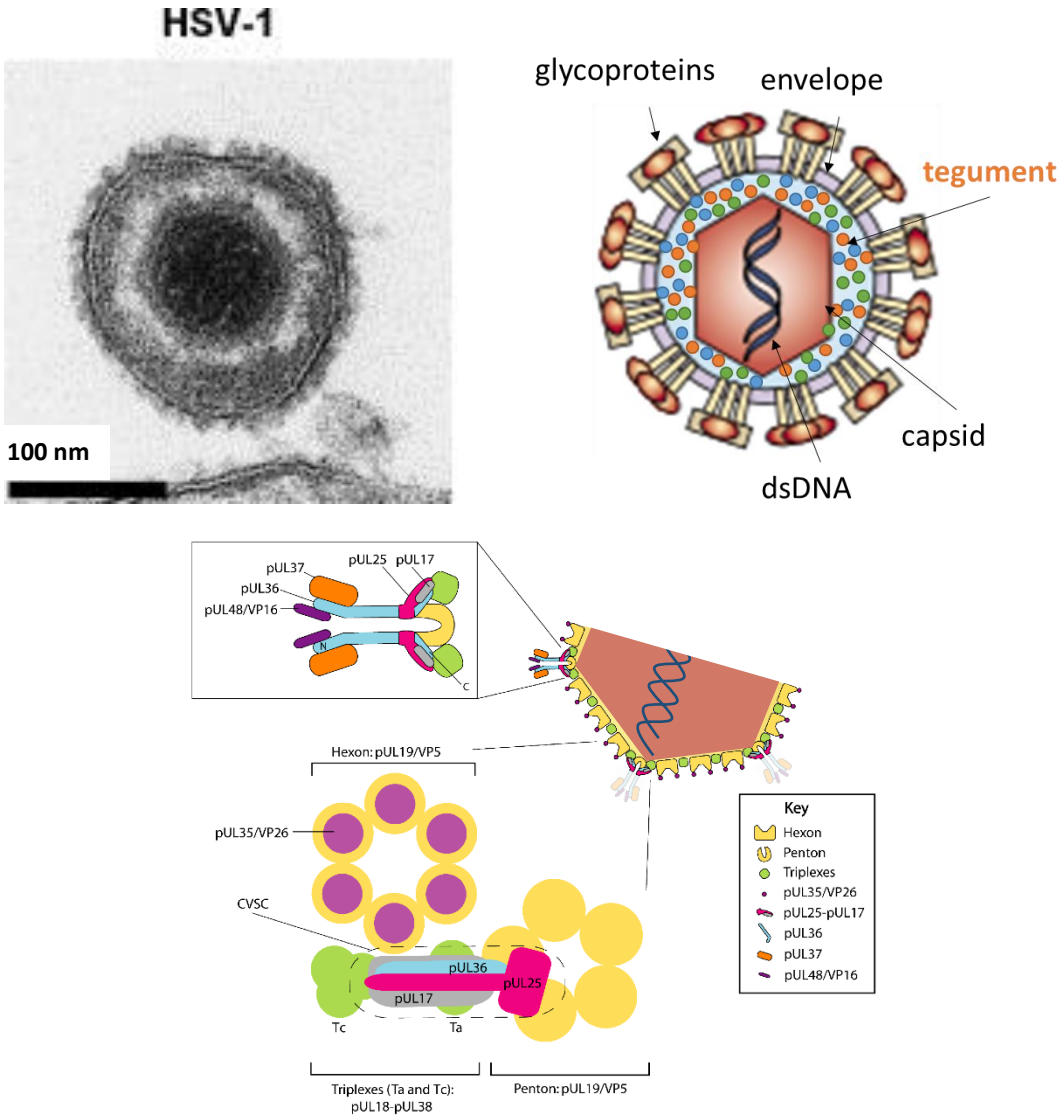


Figure 1-2. Herpesvirus morphology

A microscopic image of an extracellular HSV-1 virion (L) compared to a cutaway model (R). (Bottom) shows a model of how viral proteins are organized to form the capsid. (L) was reprinted with permission from [14] and adapted to remove images of other herpesviruses and add dimension to the scale bar. (R) was reprinted with permission from [4] and adapted to add colored circles representing tegument proteins and labels for each compartment. (Bottom) was reprinted with permission from [16] and adapted to recolor the capsid to match the scheme of the rest of the figure.

The entire particle is surrounded by a membrane derived from the host cell (Figure 1-2). This membrane is studded with viral glycoproteins including glycoproteins B, H, L, D, and E (gB, gH, gL, gD) in HSV-1 that encompass the machinery required for

initial fusion with the host cell [17] and other types of fusion that allow for specialized delivery of progeny viruses (discussed below).

Between the capsid and the envelope is a dynamic layer called the tegument (Figure 1-2) that includes both host and viral proteins [18] as well as host and viral coding and non-coding RNAs [19]. Many viral proteins reside here (discussed in depth below) to play a structural role in linking the capsid to the envelope. This layer also allows the virus to bring along a toolbox of host and viral proteins to perform any function necessary to mount a productive infection that the target cell can't perform. It is likely that many structural and regulatory functions are performed by the same tegument proteins.

1.2 Herpesvirus replication

1.2.1 Entry

Like other enveloped viruses, in order to replicate, the herpesvirus must first enter the cell (Figure 1-1a). The first step in entry after the virus has found an appropriate host cell is for the viral envelope to fuse with the host membrane. This is where the relative complexity of herpesvirus replication begins. While most enveloped viruses use only one protein to affect receptor binding and fusion [20], all herpesviruses use three conserved glycoproteins gB, gH, and gL for membrane fusion. Different herpesviruses also require divergent accessory proteins to modulate fusion events including tropism. In the case of alphaherpesviruses, one required accessory protein is gD, which, upon binding a cellular receptor, transmits a signal through a heterodimeric complex of gH/gL that triggers gB to begin a process of conformational change that brings together and fuses the host cell and viral membranes, forming a pore through which the tegumented capsid containing the

dsDNA genome can enter the host cell [17, 21] (Figure 1-3). This fusion can occur in a pH dependent or independent fashion and at the plasma membrane or in endocytic vesicles, depending on cell type [22].

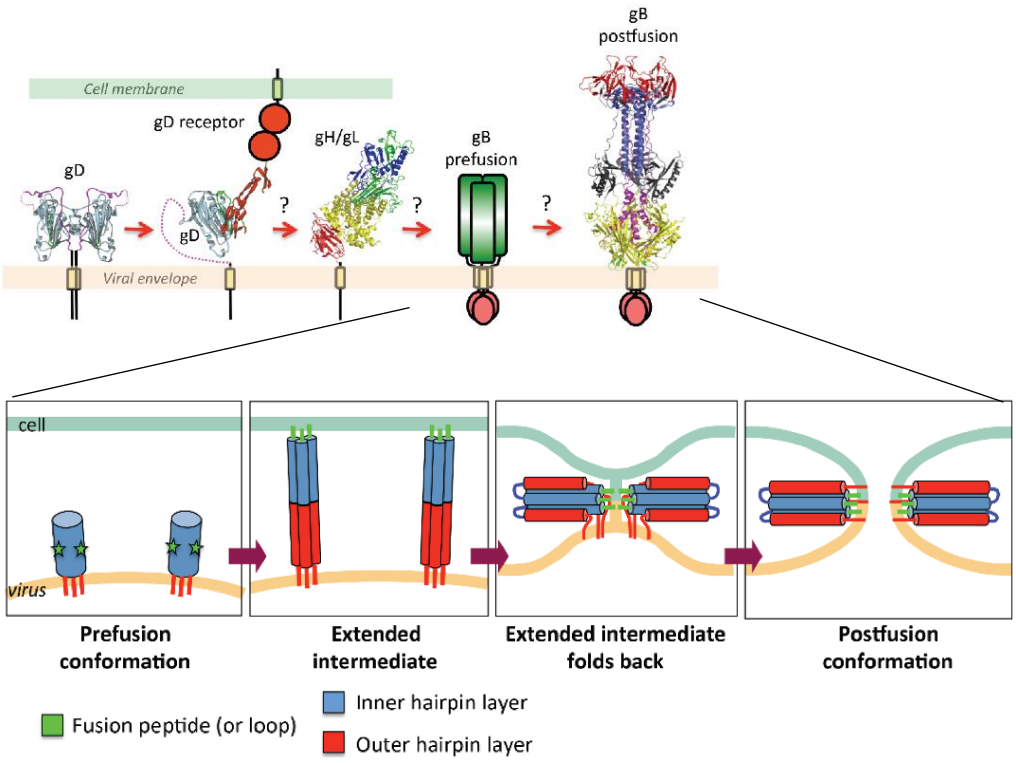


Figure 1-3. Alphaherpesviral entry

Alphaherpesviral entry. Above shows the crystal structures of alphaherpesviral fusion machinery where gB is in the postfusion state. Below is a model of how gB (and other viral fusogens) is thought to refold from prefusion to postfusion to fuse viral envelope with host cell membrane. Adapted with permission from [17]. Changes include merging two figures into one.

Once the capsid enters the cytoplasm, it sheds its coat of incumbent tegument proteins (Figure 1-1a). For HSV-1 this processes is catalyzed by the phosphorylation of outer tegument proteins by viral kinases UL13, which is conserved in all herpesviruses and US3, which is conserved in alphaherpesviruses in addition to cellular kinases [22]. Some of these shed tegument proteins stay in the cytoplasm, like alphaherpesvirus-specific UL41 (viral host shutoff, vhs)[23], while others travel to the nucleus, like

alphaherpesvirus-specific UL48 (VP16, α-trans inducing factor, α-TIF). The specific roles of these dispersed tegument proteins are discussed below, but in general they prepare the cell for a productive infection. After the removal of the outer tegument from the capsid, only the inner tegument remains, as it has been shown that completely conserved tegument proteins UL36 and UL37 remain associated after entry of PRV [24, 25] and HSV-1[26]. The mostly-uncoated capsid then has to travel to the nucleus. Trafficking on cytoskeletal proteins is thought to occur by redundant interactions between the capsid and inner tegument proteins and microtubule-associated motor proteins dynein/dynactin [27, 28] and in support of this, HSV-1 capsid protein UL35 [29] and PRV UL36 [30] have been shown to interact directly with these motors. Once the capsid is directed to the nucleus, a process likely driven by UL36 (and in HSV-1 potentially UL14)[31, 32], it docks at the nuclear pore complex by virtue of UL6 (a member of the portal vertex) and UL25 interacting with nuclear pore protein Nup214. This aligns the portal for DNA exit from the nucleocapsid with an entry to the nucleus. The viral DNA is then released from the nucleocapsid into the nucleus, a process in which UL36 and UL25 likely play roles (reviewed in [22] and [16]).

1.2.2 Replication

After the genome has entered the nucleus, viral genes are transcribed by host RNA polymerases in three phases: immediate-early, early, and late (Figure 1-1b). Genes are described by the kinetic class in which they are transcribed. In general, each phase produces proteins that are required for the next phase of gene expression. That is, immediate-early genes are expressed shortly after infection and promote the expression of early genes. Early genes encode products that are needed for viral DNA replication,

and late genes are expressed after and depend on viral DNA synthesis. This process is initiated by UL48 that is shed from the incoming virion upon entry and travels to the nucleus to form a transcriptional activation complex and promote transcription of viral intermediate-early genes including ICP0. This protein moves transcription into the early phase by removing repression of viral genes by host-encoded REST/CoREST complex and allowing their transcription (reviewed in [22]). The products of early genes, in cooperation with host proteins, support viral DNA replication and allow for the production of concatameric herpesviral genomes, from which late genes are transcribed [1].

The late phase of gene expression results in the production of viral proteins that play structural roles in capsid formation and assembly. In a simplified model, capsid protein UL19 assembles around a viral scaffold, UL26, to make a procapsid. A multifunctional viral protein complex called the terminase (likely UL15, UL28, and UL33) both cleaves concatameric DNA into unit length and forces genomes into procapsids using energy from ATP hydrolysis. Concurrently, the scaffold protein UL26, which also functions as a protease, self-digests as its space within the procapsid is filled with viral DNA. The previously-spherical procapsid, now filled with densely packed DNA, undergoes conformational changes to become the rigid, stereotypical, icosahedral nucleocapsid (reviewed in [33]).

1.3 Herpesviral assembly and spread

1.3.1 Primary envelopment and de-envelopment

Once the nucleocapsid has been assembled inside the nucleus, it must go through a number of unique membrane budding events to become a mature, infectious herpesviral

particle [34] (Figure 1-1 d,e)(Figure 1-4). These events include primary envelopment at the inner nuclear membrane, de-envelopment at the outer nuclear membrane, and secondary envelopment at cytoplasmic membranes derived from the golgi/endosomal compartment.

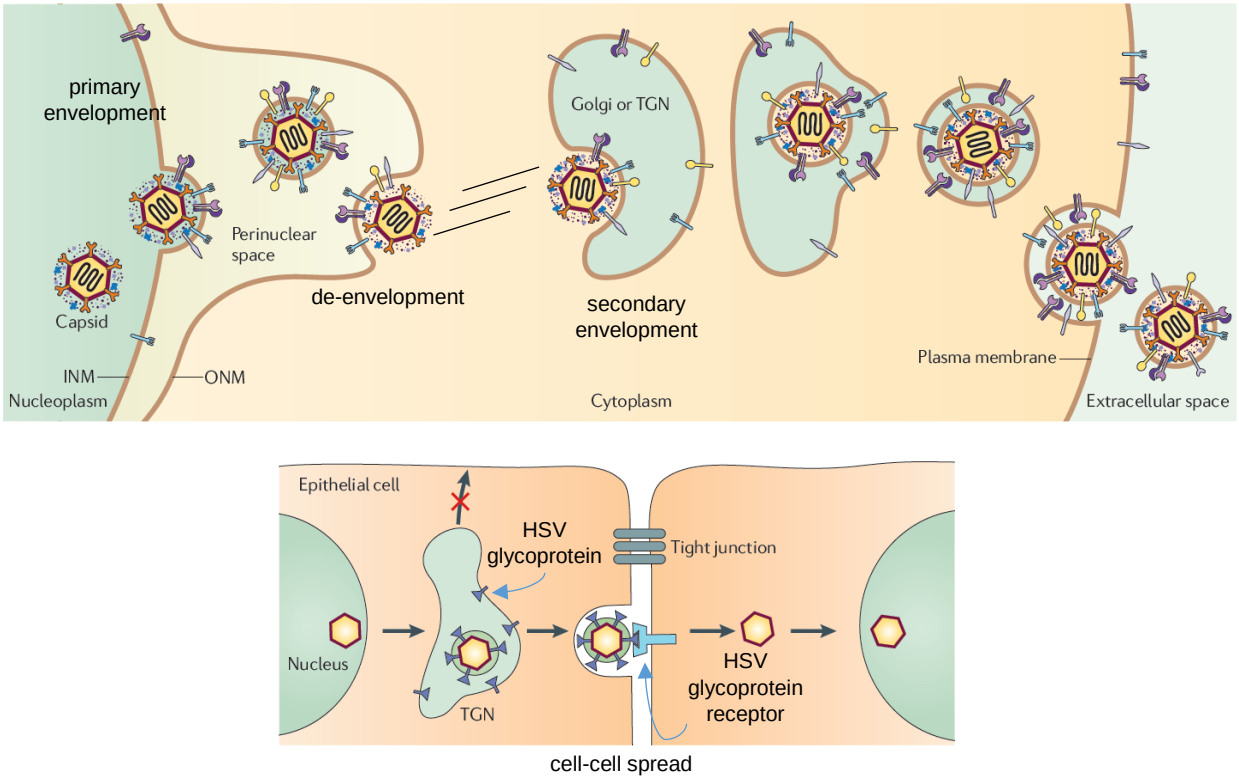


Figure 1-4. Herpesvirus egress and spread

Herpesvirus egress and spread. Above shows the process of membrane deformation steps from the nucleus, through the cytoplasm, to the extracellular space that result in a mature virion. Below shows directed delivery of mature particles to neighboring cells in cell-cell spread. Adapted with permission from [34]. Changes include adding labels for each step of envelopment, adding microtubules in (above), and labeling glycoprotein and receptor in (below).

The concurrent maturation and egress of the progeny virus begins with escaping the nucleus. The viral capsid is too large to traverse the nuclear pore complex, yet herpesviruses do not rupture the nuclear membrane for capsids to exit. Instead, viral proteins orchestrate a process of the nucleocapsid budding into and being enveloped by

the inner nuclear membrane, traversing the perinuclear space as an enveloped particle, and fusing envelopes with the outer nuclear membrane to enter the cytoplasm as an unenveloped nucleocapsid [35]. This is thought to occur with the help of tegument protein kinases US3 (in alphaherpesviruses) and UL13, and cellular kinases, which together phosphorylate nuclear lamin proteins to dissolve the structural lamina coating the inner nuclear membrane. This nuclear egress process also requires viral proteins UL31 and UL34 which together make the nuclear egress complex (NEC). The NEC is necessary for nucleocapsid nuclear egress and sufficient to deform and perform scission of membranes *in vivo* [36, 37] and *in vitro* [38]. The newly enveloped nucleocapsid soon loses its envelope after a process of dissolving of the NEC scaffold potentially mediated by viral kinase US3, and a membrane fusion event that is not well understood but may involve gB and other glycoproteins necessary for the initial fusion with the host cell (reviewed in [35]).

1.3.2 Tegumentation and secondary envelopment

A mature virion must acquire a layer of tegument before gaining its external envelope (Figure 1-1e)(Figure 1-4). In the most simplistic view, soluble inner tegument proteins are added in the cytoplasm while outer tegument proteins wait at the membranes at which secondary envelopment will occur; however, it is possible that some outer tegument proteins are added before reaching the site of secondary envelopment and as an example, HSV-1 UL16 has been found on cytoplasmic capsids [39, 40]. It is also important to note that many other viral proteins are present in the nucleus and may associate with the capsid in the nucleus, such as UL36 and UL37, and on primary enveloped virions, including UL11, UL48, and 49, although the functions of these

interactions aren't fully understood and other reports have contradicted some of these findings (reviewed in [16] [22]).

Newly de-enveloped virions must travel to the site of secondary envelopment (Figure 1-1e)(Figure 1-4). They do this by hitching a ride on microtubules (that have been massively remodeled due to infection) [41] and likely through not yet identified interactions between cytoskeletal proteins and the inner tegument protein UL36, an early addition to the nucleocapsid in terms of tegumentation. Secondary envelopment occurs at cytoplasmic membranes derived from the endosomal/golgi network (that has also been massively remodeled due to infection) [34]. This is where the remainder of the tegument proteins comprising the outer tegument is believed to be added, as many of these proteins have been found at golgi-derived membranes in transfected and infected cells. These connections are potentially mediated by outer tegument "hub" UL48 that interacts with UL36, as it has been shown to be that other tegument proteins are decreased in packaging when UL48 is absent and UL48 has many binding partners within the tegument [16]. Finally, the tegumented nucleocapsid is wrapped by an unknown mechanism with this cytoplasmic membrane that also contains integrated viral glycoproteins, the newly enveloped, vesicle-like particle is separated from the cytoplasmic membrane, potentially with help from cellular ESCRT machinery as interruption of these processes block replication of some herpesviruses [42, 43], and the mature virion egresses the cell.

1.3.3 Routes to infect new cells

Once a mature, infectious herpesviral particle has made it through secondary envelopment, there are three mechanisms by which it can infect a host cell and begin a new round of the replication cycle [44] (Figure 1-4). The most conceptually

straightforward mechanism is for the particle simply to be released from the cell via exocytosis where it can then attach to receptors on a new cell and start the whole process all over again [34]. Herpesviruses also take advantage of at least two more specialized methods of particle delivery that allows them to infect neighboring cells more efficiently while evading detection by neutralizing antibodies. During normal infection, herpesviruses can promote the formation of syncytia, or multinucleated cells. In general, this occurs when an infected cell fuses with a neighboring cell by virtue of the viral glycoprotein fusion machinery residing at its plasma membrane. In cell-to-cell spread, instead of exiting the cell at the plasma membrane, newly matured viral particles are delivered to points of contact between cells and transmission occurs directly to the neighboring cell. It has been shown that glycoprotein E (gE), and in particular its cytoplasmic tail, is critical for this method of spread in alphaherpesviruses [44], and its function is likely mediated by tegument protein complexes bound to its cytoplasmic tail [45, 46], and requires proper localization of the protein. Similarly, the repurposed activity of the typical fusogens to form syncytia is thought to be regulated by their cytoplasmic tails as mutations leading to hypersyncytial phenotypes have been mapped mostly to the intraviral portions of gB and gD [47].

1.4 Multiple roles of the tegument

1.4.1 Interactions important for virion structure

The tegument layer of a herpesvirus is full of virally-encoded proteins that have been shown to make numerous, redundant interactions [14, 16, 22]. These interactions serve the purpose of bringing the capsid in contact with the envelope through many direct and indirect interactions between capsid-bound proteins and proteins that are associated with the membrane either by lipid-modification or, in the case of glycoproteins whose

cytoplasmic tails extend into the tegument, integration (Figure 1-5). The list of these "bridging" interactions that are thought to bring the tegument and membrane together is ever-growing, but there are a number of tegument protein complexes that have been relatively well-studied and were found to be conserved among the different families of herpesviruses that suggest their relative importance in assembly of the particle.

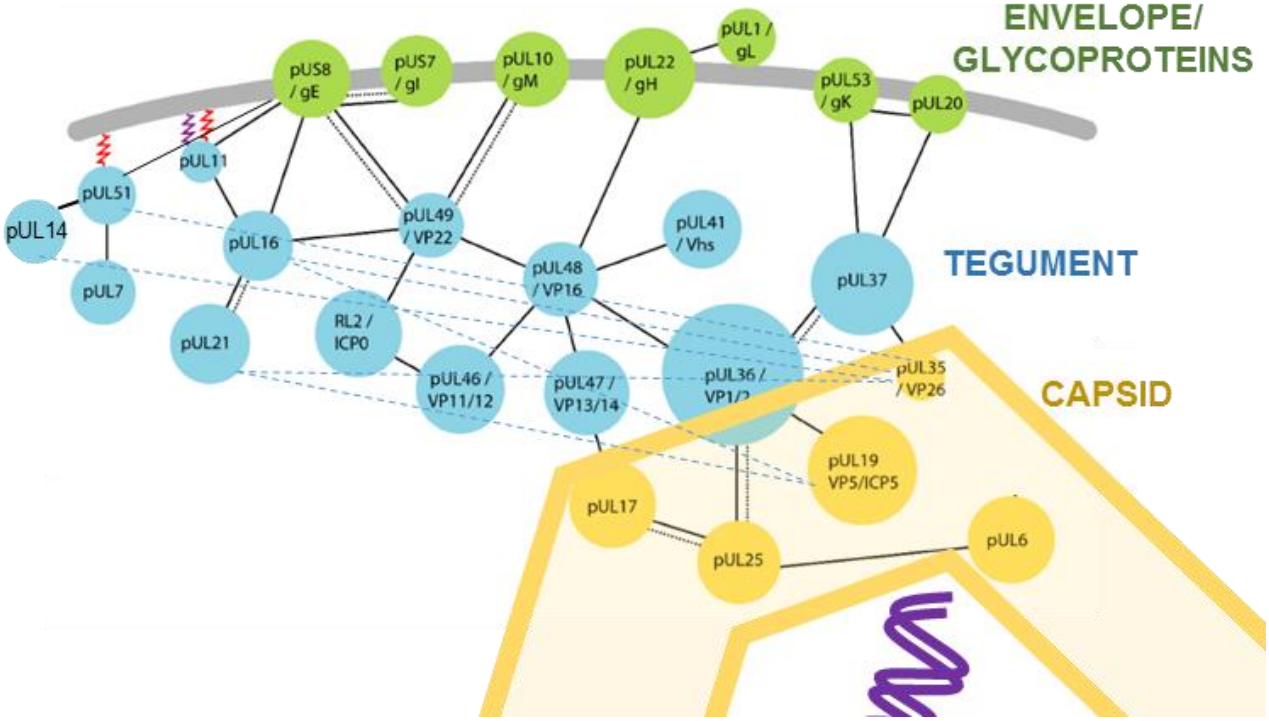


Figure 1-5. Tegument organization

A partial diagram of tegument organization, based on known binding interactions from HSV-1. A purple squiggle denotes myristylation and a red squiggle denotes palmitylation. The dotted blue lines indicate that the interaction hasn't been validated. The dotted black lines indicate that the interaction has been seen with PRV homologs. Adapted with permission from [16]. Changes include labeling and delineating substructure boundaries, adding the recently described UL51-UL14 interaction, adding the potential interactions UL21-UL19, UL21-UL35, UL16-UL19, UL16-UL35, UL14-UL19, and UL51-UL19 from [48] and the potentially indirect interaction UL51-gE from [46].

The inner tegument comprises UL36, bound to the capsid by interactions with UL19/VP5 and UL25, and UL37, through its conserved interaction with UL36 and potentially UL35/VP26 [16]. Without these proteins secondary envelopment is hampered, resulting in an accumulation of capsids in the cytoplasm. Furthermore, no other tegument proteins are added. Therefore, UL36 can be thought of as the "hub" of tegumentation in the inner tegument [14]. Similarly, UL46, UL47, UL49/VP22, and UL48/VP16 in particular can be thought of as the "hubs" of the inner tegument as they are the most abundant tegument proteins, make many connections with tegument proteins and glycoprotein tails, and have implications for the packing of other tegument proteins. UL48/VP16 is crucial as it makes contacts with the inner tegument through UL36 and relatedly its absence in HSV-1 and PRV results in extremely low viral yield (reviewed in [16]).

Conserved tegument proteins UL16 and UL11 have been shown to interact in HSV-1 [49] and their homologs from other herpesviruses have been shown to interact as well, which suggests that this interaction may be conserved amongst all herpesviruses. As HSV-1 UL16 is capsid associated [39, 40], potentially through an interaction with capsid proteins UL19/VP5 and UL35/VP26 [48] and HSV-1 UL11 is membrane associated by virtue of myristylation and palmitoylation at its N-terminus[50], this protein-protein interaction is thought to play an important and conserved bridging function for the virus. This bridge may be strengthened by other interactions with other tegument and membrane-associated proteins as UL16 has been shown to bind UL49/VP22 [51] and both UL11 and UL16 have been shown to bind gE [45, 52]. Furthermore, alphaherpesviruses encode tegument protein UL21 [53, 54], that is capsid associated[55],

potentially through interactions with capsid proteins UL19/VP5 and UL35/VP26 like UL16 [48], and has been shown to bind UL16[56]. These proteins have all been shown to be required for efficient virion production as viral titers are lowered 5-100 times in HSV-1, and consistent with these proteins forming a bridging interaction, deletion of UL11, UL16, UL21, or most homologs results in defects in secondary envelopment (Shown for HSV-1 in Figure 6). In HSV-1, the ultrastructural phenotypes of these deletions are slightly different (discussed below). Interestingly, deletion of pp28/UL99 (the CMV UL11 homolog) [57] or UL94 (the CMV UL16 homolog) have reportedly identical phenotypes in halting secondary envelopment, and these deletions or disruption of their interaction [58] renders the virus defective for replication. KSHV shows an intermediate dependence on the UL11-UL16 interaction as deletions of the UL16 and UL11 homologs (ORF33 and ORF38 respectively) result in identical inhibitory effects to secondary envelopment, but only a 10-fold decrease in progeny [59].

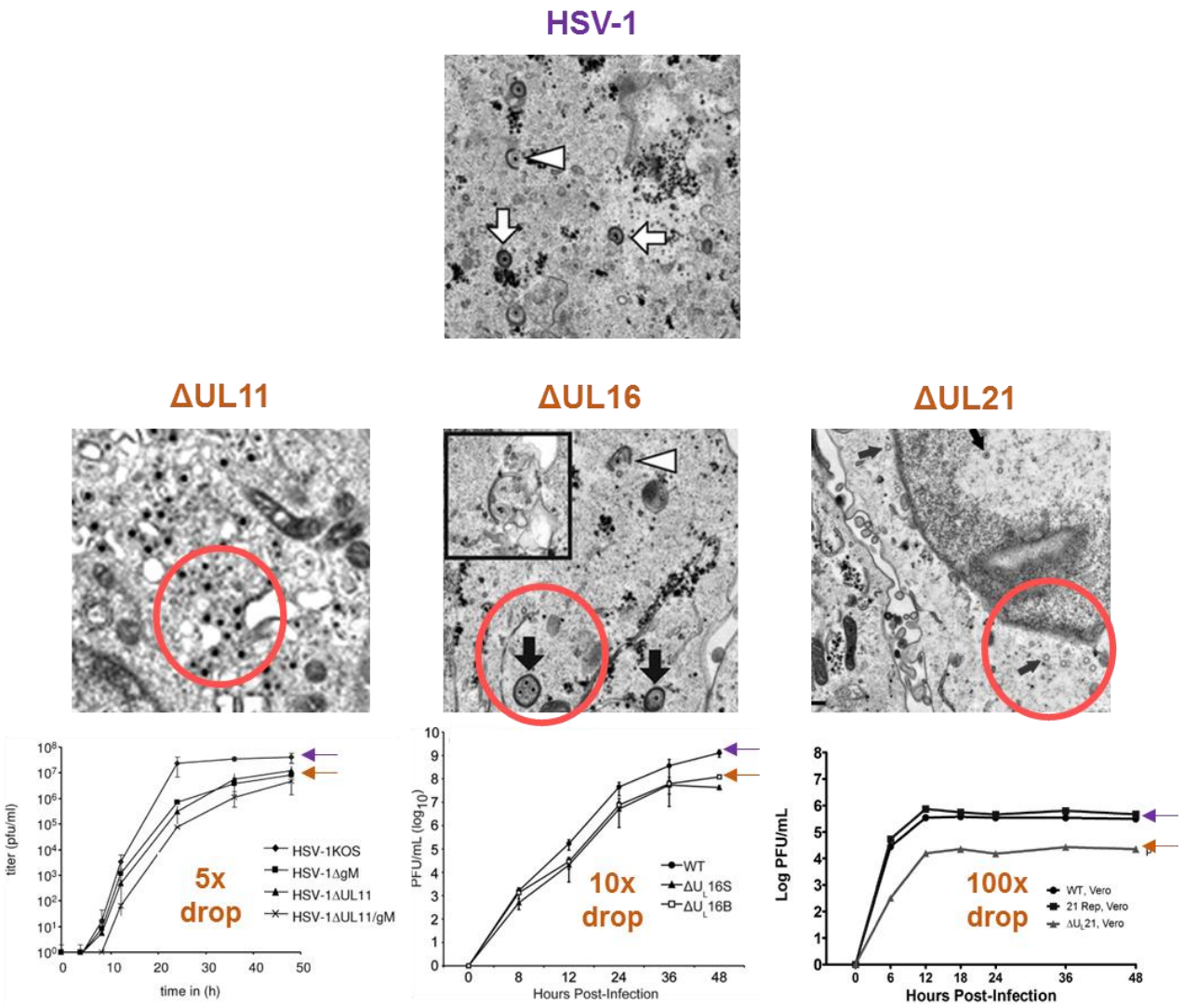


Figure 1-6. Effects of deletion of UL11/UL16/UL21 on secondary envelopment

Ultrastructural analysis of Vero cells infected with deletions of UL11, UL16 or UL21 in the HSV-1 KOS background all show defects in secondary envelopment, but with distinct characteristics (pink circles) and differing effects on viral titer. Not to scale. Adapted with permission from (11) [60], (16) [51], (21) [61]. Changes include removing and rearranging not discussed panels, merging the images into one figure, circling characteristic structures, adding arrows to point out wild type (purple) versus deletion (gold), and text description on effect on titer.

The HSV-1 UL16-UL11 interaction is not very efficient *in vitro* or *in vivo* unless the C-terminal half of UL16 is missing, cysteine to serine mutations are made in the C-terminus of UL16, or UL21 is present [45, 56]. Therefore, the interaction between UL21

and UL16 in HSV-1 is thought to remove repressive activity of UL16C on UL16N and activate the molecule for binding to UL11 (Figure 1-7). The interaction between UL16 and gE is not very efficient either, but including UL11 facilitates the UL16-gE interaction. UL11 has also been shown to bind gE [52]. The entire HSV-1 "tripartite complex" of UL11/UL16/UL21 has been shown to bind the cytoplasmic tail of gE and guide its proper localization [45] (Figure 1-7). Indeed, these four proteins colocalize with extreme efficiency to a juxtanuclear compartment in transfected cells, pulling each protein except UL11 (and some gE) away from its original location [45] (Figure 1-7). This localization is similar to what is seen in infected cells as UL11's distribution remains the same, whereas UL16 and UL21 are found in both the nucleus and the cytoplasm.

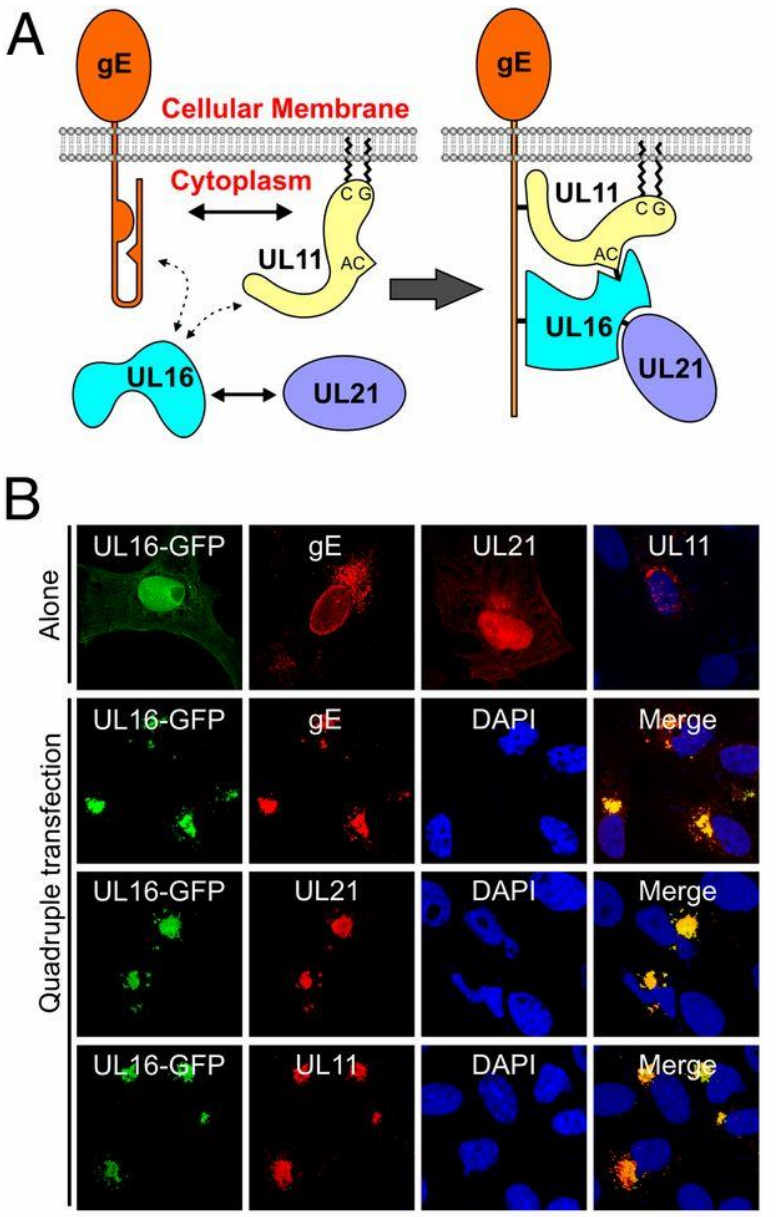


Figure 1-7. UL11, UL16, and UL21 form a complex on the tail of gE

A) model of binding of the tripartite complex. B) Vero cells transfected with UL16-GFP, UL21, UL11 or gE alone and together. Adapted with permission from [45]. Changes include removing not discussed panels.

UL51 is conserved among all herpesviruses and is important for assembly as deletion of this protein from PRV, CMV, and HSV-1 has resulted in impaired secondary envelopment [46, 62, 63]. In particular, when this protein is deleted from HSV-1, titer is

reduced 10-100x in different cell types, plaque size is reduced 75%, and an accumulation of naked capsids in the cytoplasm is seen [64]. This protein is palmitylated at a cysteine near its N-terminus and uses this modification to localize to the golgi in transfected cells and to a juxtanuclear region partially overlapping with golgi markers in HSV-1 infected cells [62]. This allows the protein to associate with membranes, but not in lipid rafts [65]. It has recently been shown that UL51 interacts directly with another conserved tegument protein UL7 [64, 66]. Deletion of UL7 from HSV-1 also results in a 10-100x reduction in titer and 75% reduction in plaque size [64] and defects in secondary envelopment have been seen with UL7-null PRV [67]. When both of these proteins were removed from HSV-1, the effects on growth and plaque size were nearly identical to those of the single mutants and microscopy showed an accumulation of nonenveloped capsids in the cytoplasm near deformed membrane structures (Figure 51)[64].

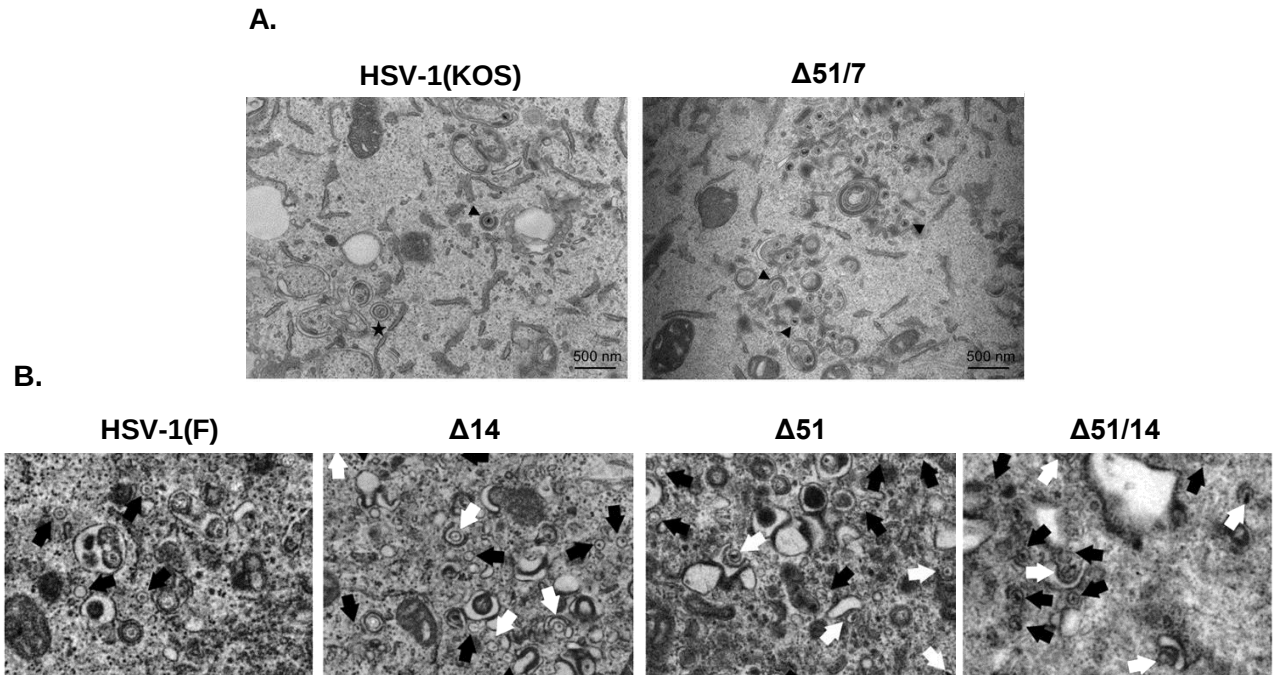


Figure 1-8. Effects of deleting UL7/UL14/UL51 on secondary envelopment

Ultrastructural analysis of infected cells. A and B are approximately to scale. A) HFF-Tert cells infected with HSV-1 (KOS) wild type or lacking UL51 and UL7. Arrowheads denote nonenveloped or partially wrapped cytoplasmic nucleocapsids, star denotes a properly enveloped nucleocapsid. Reprinted with permission from [64]. B) Vero cells infected with HSV-1 (F) wild type or lacking UL14, UL51, or UL14 and UL51. White arrows denote partially enveloped capsids and black arrows denote unenveloped capsids. Reprinted with permission from [68]. Adapted to reposition images in B with respect to each other, remove undiscussed panels, and scale panels in B to match A.

UL51 was found to interact directly with UL14, another conserved tegument protein [68]. In HSV-1 lacking either of these proteins the total proportion of partially enveloped and unenveloped capsids in the cytoplasm increased from 5% in wild type infection to approximately 40% in either deletion. Like in the case of UL51 and UL7, when both UL51 and UL14 were removed from the virus or their interaction was abolished, the proportion of capsids in the cytoplasm was similar to that of either single deletion, suggesting that these proteins primarily work in a complex to regulate secondary envelopment [68]. Since UL51 is membrane-associated itself and has been

shown to interact with membrane-associated gE [46] and both UL51 and UL14 may interact with UL35/VP26 [48], these complexes could provide a mechanism connecting the capsid to the envelope and a role in secondary envelopment is not surprising. Furthermore, in the absence of either UL51 complex, discrepancies were seen at the membrane wrapping step of secondary envelopment; that is, capsids were observed in close proximity to deformed cytoplasmic membranes rather than solely free in the cytoplasm (Figure 1-8), suggesting that these complexes or UL51 are central to regulating the "completion" of secondary envelopment rather than travel to the site of envelopment or stabilizing the membrane interaction.

1.4.2 Non-structural roles of tegument proteins

In addition to these structural roles, most of the tegument proteins described above have been found to play regulatory roles during herpesviral infection. A major example is the inner tegument protein UL36, which has come up a number of times during this chapter. In a structural function, this protein is intimately related with the capsid and provides a scaffold for many bridging interactions with tegument proteins (Figure 1-5). Additionally, this protein may bind cellular motor proteins and assist with capsid trafficking during entry and egress, facilitate delivery of the viral genome to the nucleus, and stabilize packaging of newly replicated genomes. Furthermore, UL36 has been shown to have deubiquitinase activity in its N-terminus [69] that interferes with cellular antiviral defense systems including the translesion synthesis-based DNA damage response [70] and the interferon β immune pathway [71].

UL37 binds membrane associated gK and UL20 [72] and capsid-associated UL36 and it has been proposed that this linking is important for secondary envelopment [72]

(Figure 1-5), alluding to its structural role in replication. In another regulatory role, the C-terminal domain of UL37 has been shown to harbor deamidase activity that interferes with RIG-I sensing of dsRNA and immune response [73]. Previous work in our lab has shown that the UL37 N-terminal domain from PrV [74] and HSV-1 [75] are structurally similar to the CATCHR subunit of a multi subunit tethering complex, a protein that tethers cargo to the correct recipient membrane, and might therefore play a role in directing the capsid to the correct site of secondary envelopment [74]. Furthermore UL48/VP16, the inner tegument hub, kicks off viral gene production by forming a transcription complex just after viral entry, and UL47 is also a nucleocytoplasmic shuttle that preferentially binds single-stranded polyadenylated RNAs, including some viral mRNAs, and may help incorporate them into nascent virions.

In addition to binding gE in a potentially structural role, the tripartite complex of UL11/UL16/UL21 is thought to mediate the function of gE in regulating cell-cell spread [45]. Delays in viral replication have been seen for UL11-null HSV-1 in examining viral growth curves [76-78], UL16-null HSV-2 in examining the expression of viral genes [79], and UL21-null viruses while examining the transcription and expression of viral genes and growth curves for both HSV-1 [80] and HSV-2 [54]. This observation suggests that these proteins may play roles, and possibly separate ones, in regulating viral gene expression. These roles could be direct, potentially by interacting with transcription or translation machinery, or indirect, such as regulating the delivery of necessary factors. Furthermore, UL16, UL21 [45], and potentially UL11 [81] are found in the nucleus, yet no nuclear functions of these proteins have yet been characterized.

These three proteins are also likely to have independent functions and these functions may be virus specific. Firstly, lack of UL11 reduces HSV-1 viral titers to 20% [82], lack of UL16 reduces viral titers to 10% [51], and lack of UL21 reduces viral titers to 1% [61] that recover over time. Secondly HSV-2 UL16 and UL21 have been shown to be necessary for nuclear egress, while homologs from HSV-1 with high sequence identity show no defects in nuclear egress. The most obvious (because it's visual) evidence for independent roles of these proteins comes from ultrastructural observations of UL11-, UL16-, and UL21-null infected cells (Figure 1-6).

Like other tegument proteins, it is probable that UL51, UL14, and UL7 also have non-assembly roles and functions independent of their complexes. HSV-2 UL14 has been reported to have heat-shock protein-like characteristics [83] that may function to inhibit apoptosis [84] and relocalizes capsid protein UL35/VP26 and trans-inducing factor UL48/VP16 to the nucleus in transfected cells [31, 32]. UL7 partially localizes to mitochondria in HSV-1 infected cells and has been found to bind adenine nucleotide translocator 2, though the function of this interaction is unknown [85], and UL51 and UL7 were not seen to colocalize at mitochondria in early infection [66]. Despite the importance of the UL51/UL7 complex for secondary envelopment, these proteins do not completely colocalize [66]. Taken together this suggests that UL7 and UL51 may have independent functions in addition to their functions as a complex. In addition to their putative role as a complex in assembly, UL7 and UL51 codependently localize to focal adhesions in infected cells in a gE-independent manner and may serve to regulate morphology of infected cells as cells infected with the double mutant virus appeared more rounded and made less contacts with neighbors [64]. The UL51/UL7 complex may

affect cell-cell spread by stabilizing cell-cell contacts necessary for this process to occur [64]; however, UL51 interacts with gE [46], an important regulator of cell-cell spread, so it is possible that this protein functions in multiple ways to promote cell-cell spread. Consistent with this idea, infection with HSV-1 lacking most of the UL51 protein resulted in plaques that were one log smaller than the already shrunken plaques arising from infection with gE-null HSV-1 [46].

1.5 Research Goals

The above observations have inspired the research presented in this thesis. We believe that many more tegument proteins, including but not limited to UL11, UL16, UL21, and UL51 have as-of-yet undiscovered regulatory functions in addition to their documented potential roles in secondary envelopment. Furthermore, their structural contributions are not yet fully understood either. Since the tegument is a dense network full of redundant interactions, complete characterization of each function of each protein is difficult in the context of infection. We aim to characterize these proteins "from the ground up", and instead characterize them *in vitro* alone and in relevant complexes to provide structural and biochemical context for interpretation of *in vivo* observations.

Specifically, we aimed to describe the independent and dependent functions of two tegument protein complexes: UL11, UL16, and UL21 and UL51, UL7 and/or UL14. Research presented in this thesis describes structural and biochemical characterization of UL11, UL21, UL16, and UL51. This includes the crystal structures of either globular domain of HSV-1 UL21 and structural characteristics of HSV-1 UL11, UL21, and UL51 in solution. We also show that UL11 and UL21 have the capacity to bind RNA. The work on UL16 suggests that it may need stabilization by a binding partner for *in vitro*

characterization. Finally, we show that UL11, UL21, and UL51 use flexibility to accomplish their multiple functions. These observations help to unravel the complexity of herpesvirus replication and assembly.

Chapter 2: Methods and Materials

2.1 Methods to study UL11

2.1.1 Sequences and analysis

Sequences for UL11 homologs from 8 human herpesviruses: HSV-1 strain 17 UL11 (YP_009137085.1), HSV-2 strain HG52 UL11 (YP_009137162.1), VZV strain Dumas ORF49 (NP_040171.1), EBV strain B95-8 BBLF1 (YP_401686.1), CMV strain AD169 UL99 (CAA35335.1), HHV-6 strain Z29 U71 (NP_050250.1), HHV-7 strain JI U71 (YP_073811.1), and KSHV strain GK18 ORF38 (YP_001129391.1) were aligned in Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and adjusted manually. Expasy ProtParam (http://web.expasy.org/protparam/) was used to calculate protein size and identify numbers of each residue in each sequence. CSSpalm (http://csspalm.biocuckoo.org/online.php) was used to predict palmitylation sites. Myristylation was marked on glycine 2 based on previous studies (maclean 1989, 1992 bowzard 2000) and conservation. PsiPred and DISOpred (bioinf.cs.ucl.ac.uk/psipred) were used to predict secondary structure and disordered regions, respectively.

2.1.2 Cloning and expression constructs

A plasmid encoding full-length UL11 from HSV-1 strain 17 preceded by a His_6 tag and a thrombin cleavage site (6H-UL11) in a pET28 vector was a gift from John Wills (PennState U.). This plasmid contains an A8T mutation in UL11, but as this region of protein is poorly conserved, the mutation was not considered problematic. Furthermore, UL11 with this mutation successfully rescues a UL11-null virus [76]. UL11 was amplified from the pET28 vector using primers CM121 and CM122 (Table 1), digested with NcoI and XhoI, and subcloned into the NcoI/XhoI-digested pGEX-6P-1 backbone to produce plasmid pCM054 which encodes UL11 preceded by a Glutathione

S-transferase tag (GST), a human rhinovirus (HRV) 3C (PreScission) protease site, and a Leu-Gly-Ser linker, GST-HRV3C-UL11. To add a C-terminal Strep-tag II (StII), single overlap extension (SOE) PCR was performed on pCM054 using primers pGEX3, pGEX5, CM131, and CM132 (Table 2-1). The insert was digested with NcoI and XhoI and subcloned into the NcoI/XhoI-digested pCM054 vector to produce plasmid pCM056 encoding GST-HRV3C-UL11-StII. Finally, the same insert was subcloned into the original pET-28 vector, again using NcoI and XhoI, to produce plasmid pCM055 encoding UL11-StII. To produce a construct encoding the predicted core of UL11 followed by a C-terminal StII tag, SOE was performed on pCM055 using primers AK34, AK37, CM136, and CM137 (Table 2-1). The insert was digested with NcoI and XhoI and subcloned into pCM055 to create plasmid pCM084 encoding UL11(1-65)-StII.

Table 2-1. UL11 cloning primers

Primer	Sequence	Description
AK34	5’- AAAGGAAGGGAAGAAAGCGAAA-3’	pET downstream primer, reverse
AK37	5’- ATACCGCGAAAGGTTTTGCG-3’	pET upstream primer, forward
CM121	5’-TAATCGATTACCATGGGCCTCTCGTTCTCC-3’	UL11 N-terminus, forward
CM122	5’-GGGATCTAGTCTCGAGTTATTATTCGCTATCGGACATGGGG-3’	UL11 C-terminus, reverse
CM131	5’-AAATTGTGGGTGGCTCCACTCGCTATCGGACATGGG-3’	UL11-StII, reverse
CM132	5’-CACCCACAATTTGAGAAGTAATAACTCGAGCGGCC-3’	UL11-StII, forward
CM136	5’-TGGGTGGCTCCAACGCAGGCGCTG-3’	UL11(1-65)-StII, reverse

CM137	5’-CAGCGCCTGCGTTGGAGCCACCCA-3’	UL11(1-65)-StII, forward
pGEX3	5’-GTCAGAGGTTTTCACCGTCAT-3’	pGEX downstream primer, reverse
pGEX5	5’-ATAGCATGGCCTTTGCAGGGCT-3’	pGEX upstream primer, forward

2.1.3 Recombinant protein expression and purification

6H-UL11 was expressed in *E. coli* strain T7 cells in LB for 4-8 hours at 37°C using induction with 1 mM isopropyl-β-d-thiogalactopyranoside (IPTG) at an optical density at 600 nm (OD600) of 0.6 to 1.0. GST-HRV3C-UL11 was expressed in *E. coli* strain Rosetta cells using the autoinduction protocol [86]. Briefly, a small culture of cells was grown overnight at 37°C in terrific broth (TB) with 1% glucose and 2 mM $MgSO_4$ before inoculation at a ratio of 1:100 into TB with 0.2% lactose and 2 mM $MgSO_4$ at 37°C for 4 hours then 25°C for 18-22 hours. StII-tagged constructs were expressed in *E. coli* strain Rosetta or LoBStr [87] cells at 37°C using IPTG induction as above.

Cells were harvested by centrifugation at 5,000×g and 4°C for 30 minutes, resuspended in buffer A [50 mM HEPES, 100 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)] with added 0.1 mM phenylmethanesulfonyl fluoride (PMSF), 1× cOmplete protease inhibitor cocktail (Roche), and optional 1 mM dithiothreitol (DTT), and lysed using a Microfluidizer. After removal of insoluble material from lysates by centrifugation at 20,000×g and 4°C for 30 minutes, each construct was purified using affinity chromatography in buffer A as follows. 6H-UL11 was captured on Ni Sepharose 6 Fast Flow (GE Healthcare Life Sciences), washed sequentially with 30 mM and 50 mM imidazole, and eluted with 300 mM imidazole in buffer A.

GST-HRV3C-UL11 was captured on Glutathione Sepharose 4B (GE Healthcare Life Sciences) and eluted with 10 mM reduced glutathione in buffer A. The GST tag was removed from GST-HRV3C-UL11 to make GPLGS-UL11 (where GPLGS is a linker) by incubating the protein with a recombinant, GST-tagged PreScission protease at a 1:30 protease:protein molar ratio without stirring or shaking at 4°C overnight. Uncleaved GST-HRV3C-UL11, excess GST-tag, and PreScission protease were removed from GPLGS-UL11 by passing the mixture over a standalone glutathione resin column and collecting the flow through and/or including Glutathione Sepharose Fast Flow HiTrap columns (GSTrap FF, GE Healthcare Life Sciences) placed in line with size-exclusion chromatography (below).

Prior to purification, UL11-StII-containing lysates were supplemented with small amounts of egg white avidin (Sigma), and protein was captured on Strep-Tactin Sepharose resin (GE Healthcare Life Sciences) and eluted with 5 mM *d*-desthiobiotin (Sigma) in buffer A. UL11-StII was subsequently separated from copurifying nucleic acids using a Heparin Sepharose HiTrap column (GE Healthcare Life Sciences) and eluted with a gradient of 0.1 to 1.0 M NaCl in buffer A.

All constructs were further purified using size-exclusion chromatography using Superdex 75 (GE Healthcare Life Sciences), in buffer A. The column was calibrated using blue dextran, aldolase, conalbumin, ovalbumin, and ribonuclease A (Gel Filtration Calibration Kits, GE Healthcare Life Sciences). Calibration curves were generated to calculate the apparent molecular weight and Stoke's radius of UL11 using the following formulas, according to manufacturer's instructions:

$K_{av} = \frac{V_e - V_o}{V_c - V_o}$ proportional to log($molecular\ weight$), where K_{av} is the partition coefficient, V_e is the elution volume, V_c is the geometric column volume, and V_o is the void volume.

Throughout purification, UL11 samples were concentrated in Amicon Ultra centrifugal concentrators (Millipore) with a 10,000 Da molecular weight cut off and stored with 0.1 mM PMSF and 1× Halt protease inhibitor cocktail (Pierce). Final samples were evaluated for sample purity and concentration using SDS-PAGE and spectrophotometrically using the calculated extinction coefficient 1490 M^{-1} cm^{-1} and molecular weight 12679.96 Da for 6H-UL11, 1490 M^{-1} cm^{-1} and 10928.10 Da for cleaved GST-HRV3C-UL11, and 6990 M^{-1} cm^{-1} and 11556.79 Da for UL11-StII. It should be noted that both 6H-UL11 and the cleaved product of GST-HRV3C-UL11 lack tryptophans and thus have weak absorbance at 280 nm, which results in less reliable estimates of concentration based on A280. By contrast, the addition of the Strep-II affinity tag to UL11-StII provided a tryptophan residue for more reliable concentration determination based on spectroscopic measurements.

Western blots probing for the His_6 tag were performed by transferring proteins from polyacrylamide gels to nitrocellulose membrane in a semi-dry transfer apparatus for 30 minutes at 25 volts in transfer buffer [50 mM Tris, 37.5 mM glycine, 0.0375% sodium dodecyl sulfate (SDS), 30% methanol]. Blots were blocked for 30 minutes at room temperature in 5% non-fat dry milk prepared in TBST [Tris-buffered saline (150 mM NaCl, 20 mM Tris pH 7.5) with 0.05% Tween-20] before adding Penta-His HRP conjugate (Qiagen) at a ratio of 1:5000 and incubation overnight at 4°C. After multiple washes in TBST, blots were developed using an ECL western blotting substrate (Pierce).

2.1.4 Membrane-binding co-sedimentation assay

Multilamellar vesicles (MLVs) were prepared and the co-sedimentation assay was performed as described previously [38]. Briefly, POPA [1-palmitoyl-2-oleoyl-sn-glycero-3-phosphate (sodium salt)], POPC (1-palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine), POPS [1-palmitoyl-2-oleoyl-sn-glycero-3-phospho-L-serine (sodium salt)] (Avanti), or cholesterol (Sigma), were combined in various ratios from 10 mg/ml stocks in chloroform and dried into films. The films were rehydrated in buffer A with shaking at 37°C to form MLVs. Pre-cleared protein samples in 6 µg aliquots were incubated with 30 µg MLVs for 30 minutes at room temperature before separating the soluble portion from the MLV-associated insoluble pellet with centrifugation at 16,000×g and 4°C for 20 minutes. Binding was assayed with SDS-PAGE.

2.1.5 Crystal Screening

UL11 constructs were screened at various concentrations in 9 screens [in-house Harrison Lab Grid Screen, Classics Suite (Qiagen), Protein Complex Suite (Qiagen), Index (Hampton Research), Peg/Ion (Hampton Research), SaltRx (Hampton Research), Top 96 (Anatrace), Wizard 1-4 (Rigaku)] using 96-well sitting drop vapor diffusion format with drops containing 0.2 µL crystallization solution and 0.2 µL protein dispensed by Crystal Phoenix liquid handling robot (Art Robbins). Plates were stored at room temperature and were evaluated using a stereo microscope daily for a week and months later.

2.1.6 Limited proteolysis, mass spectrometry, and N-terminal sequencing

Limited proteolysis was performed as described previously [88]. Briefly, 5.5 µg purified UL11-StII was incubated for 1 hour at room temperature with 1.5, 3, 6, 12, 24, 48, 96, or 192 ng TPCK-treated trypsin (Sigma) in 0.5X buffer A. After an hour, excess

phenylmethanesulfonyl fluoride (PMSF, 10 mM) was added to stop the reaction, and reactions were analyzed with SDS-PAGE. For mass spectrometry, 55 µg purified UL11-StII was digested with 2.2 ng trypsin as above and analyzed by MALDI-TOF on a Voyager DE-PRO instrument using a dihydroxybenzoic acid (DHB) matrix at the Tufts University Core Facility. For N-terminal sequencing, 33 µg purified UL11-StII was digested with 18 ng trypsin as above, separated on a 4-15% SDS-PAGE, and transferred to polyvinylidene fluoride (PVDF) membrane in 50 mM CAPS supplemented with 10% methanol using a semi-dry transfer apparatus. The membrane was stained in 40% methanol with 0.05% Coomassie blue R-250, followed by destaining in 50% methanol and then water. Pieces cut from the dried membrane were subjected to N-terminal sequencing by Edman degradation at the Tufts University Core Facility.

2.1.7 Circular Dichroism

Secondary structure content of UL11 was estimated using circular dichroism. Purified, nucleic-acid-free UL11-StII in buffer A was exchanged into CD buffer (20 mM sodium phosphate, 100 mM sodium fluoride, pH 8.0) using a PD SpinTrap G-25 column (GE Healthcare Life Sciences) and diluted to various concentrations. Far-UV spectral scans from 185-300 nm were taken in a 1 mm cuvette at 20°C on a JASCO J-815 spectropolarimeter, continuously scanning at 50 nm/min with 1 nm band width and 1 second data integration time. Five spectra were collected, averaged, and buffer-subtracted for each sample. Machine data was converted to mean residue ellipticity using the equation below

$MRE = \theta_{mrw} = \frac{MRW \cdot \theta}{10 \cdot c \cdot l}$ where MRW = mean residue weight = molecular mass /[(number of amino acids)-1], theta is ellipticity (mdeg) c is concentration in g/L and l is pathlength (cm)

and analyzed using CDSSTR [89] with reference set 6 [90] or SP175 [91] in Dichroweb [92] (http://dichroweb.cryst.bbk.ac.uk/html/home.shtml).

2.1.8 Size-exclusion Chromatography Small Angle X-ray Scattering (SEC-SAXS)

SEC-SAXS experiments were carried out at beamline G1 at the Cornell High Energy Synchrotron Source (Ithaca, NY). UL11-StII at approximately 18 mg/ml was pre-cleared of aggregates by centrifugation at 16,000×g and 4°C for 5 minutes, and a 100 ul sample was injected onto a Superdex 200 Increase 5/150 (GE Healthcare Life Sciences) in buffer A at 0.3 ml/minute and 4°C and fed directly into a quartz flow cell. Samples were irradiated with a 9.8833 keV (1.254489 Å) beam with 5.7 x 1011 photons/s flux and diameter of 250 µm x 250 µm, and 2-second images were recorded through the duration of the run on a Pilatus 100K-S detector (Dectris) at a sample-to-detector distance giving 0.006 $Å^{-1} < q < 0.8$ $Å^{-1}$ where scattering vector $q = 4\pi\sin(\theta)/\lambda$.

Data was processed using the RAW software package (https://sourceforge.net/projects/bioxtasraw/) [93]by averaging appropriate frames with a consistent Rg value and subtracting averaged buffer frames. Subtracted curves were further analyzed using programs in the ATSAS software package (https://www.embl-hamburg.de/biosaxs/software.html) [94] including DATGNOM[95], DAMMIN [96], DAMAVER[97], and EOM 2.0 [98]. DATGNOM was used to generate and automatically evaluate the pair distance distribution function [p(r)] from the subtracted experimental data before manual adjustment of the automatically assigned D_{max} value to

improve p(r) shape. In DAMMIN, the p(r) function from GNOM was used to compile a bead model in slow, fine bead mode with no known symmetry or anisometry. The DAMAVER suite was used to superimpose and average 10 *ab initio* bead models from DAMMIN. One model was automatically removed by the program and two models were removed after manual assessment of inability to superimpose. In EOM, a pool of 10,000 random native-like structures representing most of the structural space available to the sequence of UL11-StII were generated. Theoretical SAXS curves were generated for random combinations of these models and the ensemble was optimized with a genetic algorithm to best fit the experimental data. No symmetry information or structured domains were provided. The initial pool and final ensemble are given flexibility metrics relating directly to the amount of information entropy in the system. Furthermore, the Rg and Dmax values calculated from the species in each group are represented in histograms. Lysozyme data was supplied with the RAW software as an example and processed similarly.

2.1.9 Nuclease digestion

Nuclease digestion assays were performed as described previously [99]. Briefly, after StrepTactin resin purification, copurifying endogenous *E. coli* nucleic acids (NAs) in complex with UL11-StII were extracted by phenol-chloroform precipitation. Either slightly acidic phenol-chloroform to preferentially isolate RNA (phenol/chloroform/isoamyl alcohol 25:24:1, pH 6.7, Fisher Scientific) or slightly basic phenol-chloroform (phenol/chloroform/isoamyl alcohol 25:24:1, pH 7.9, Ambion) were mixed with aliquots of UL11-StII-NA complex in a 1:1 volume ratio and centrifuged at 16,000×g and ambient temperature for 5 minutes. The upper aqueous layer was removed,

added to 10 μg glycogen (Life Technologies), 1/10 volume 3 M sodium acetate, and 10 volumes isopropanol, and incubated for 10 minutes at -80 °C. Pellets containing nucleic acids were collected by centrifugation at 16,000 × g and 4°C for 20 min and washed with 75% ethanol prior to resuspension in water for analysis. Nucleic acid samples were digested in 1x Turbo DNase buffer with either Turbo DNase (Ambion, 2U per 2 μg NA) or RNase A (Invitrogen, 0.4 μg per 2 μg NA) with 2 mM calcium chloride for 30 min at 37 °C and analyzed on a denaturing RNA gel. The loading buffer comprised 5× RNA sample buffer {4 mM EDTA, 2.7% formaldehyde, 30.8% formamide, 20% glycerol, 40% 10× MOPS buffer [200 mM 3-(N-morpholino)propanesulfonic acid (MOPS), 30 mM sodium acetate trihydrate, 10 mM EDTA; pH adjusted to 7.0 with sodium hydroxide]} to 1x, 25% FORMAzol (Molecular Research Center, Inc.), and 0.05 mg/ml ethidium bromide. After heating at 85 °C for 1 to 3 min and cooling to 4 °C, samples were immediately loaded onto a 1.2% agarose gel prepared in 1× MOPS buffer with 5% formaldehyde, and run at 75 V in 1× MOPS buffer.

2.1.10 Crosslinking

Chemical crosslinking was performed as described previously [38]. Briefly, Bis(sulphosuccinimidyl)suberate (BS^3, Fisher Scientific) was added to 100 pmol purified UL11-StII in water at 0, 50, 100, 200, and 400 times molar ratio crosslinker to protein and incubated in amber tubes in the dark at ambient temperature for 20 minutes, at which point 1M Tris pH 8.0 was added to 50 mM and incubated for 15 minutes to stop the reaction. Purified HSV-1 nuclear egress complex (NEC220) was used as a positive control. Half of the reaction was analyzed by SDS-PAGE and Coomassie staining.

2.2 Methods to study UL21

2.2.1 Sequences and prediction

Sequences of 16 alphaherpesvirus UL21 homologs [P10205, P89444, Q5Y0T2, E2IUE9, Q91IG6, J9QYM9, D1FXW1, P28972, Q5PPA1, A4GRJ2, P09289, Q6X232, Q77CC3, Q782S5, Q9IBV4, and Q9DPR5 (UniProtKB accession numbers)] were aligned in Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Expasy ProtParam (http://web.expasy.org/protparam/) was used to calculate protein size and identify numbers of each residue in each sequence. PsiPred (bioinf.cs.ucl.ac.uk/psipred) was used to predict secondary structure.

2.2.2 Cloning and expression constructs

Plasmids encoding full-length wild type HSV-1 strain 17 UL21 or UL21 with proteolytic site mutations (PSM, R216G/R217S/R220S/K234G/R235S) in a pET24a vector preceeded by an N-terminal Strep-II (StII) tag were gifts from John W. Wills.

Insert corresponding to UL21N(1-216) was amplified using primers CM001 and CM002 (Table 2-2) and subcloned into the original pET21a vector with Nde/HindIII. Inserts corresponding to other various UL21 domain fragments were constructed by overlap extension PCR [100] from the full-length PSM gene. Primer sequences are listed in Table 2-2. SOE PCR products were cleaved by NdeI and HindIII and subcloned into the NdeI/HindIII cleaved backbone of the pET24a containing UL21. This generated constructs to produce various fragments of UL21N followed by a C-terminal StrepII tag or of UL21C preceded by an N-terminal StrepII tag and a Gly-Ser linker. To generate StII-HRV3C-UL21C (275-535) construct, an HRV3C (PreScission) protease site was inserted between the affinity tag and UL21C.

StII-HRV3C-UL21(PSM) was cloned by using SOE PCR with CM020, CM021, CM023, and CM029 (Table 2-2) with StII-UL21(PSM) as a template. StII-UL21(PSM) was amplified with CM090 and CM093 (Table 2-2) and subcloned into pET24 with NdeI/HindIII to construct PSM-StII. PSM(1-532)-HRV3C-StII was constructed by subcloning a purchased geneblock (IDT) encoding UL21(516-532)-HRV3C-StII flanked by KpnI and XhoI sites into pET24a with a UL21 PSM insert after digestion in NEBuffer 1 with 0.1% bovine serum albumin.

Table 2-2. UL21 cloning primers

Primer	Sequence	Description
CM001	5’- CGAAATCATATGTGGAGCCACCCGCAG -3'	StII primer, forward
CM002	5’-GCATAAGCTTATTAGCGGCCCGTGATCACCAC -3'	StII-UL21N (1-216) primer, reverse
CM003	5’- GCGGCCAAGCTTATTACACAGACTGTCC -3'	Downstream primer, reverse
CM004	5’- GAAAAAGGTAGCGCCACCGTCAGC -3'	StII-UL21C (237-535), forward
CM005	5’- GCTGACGGTGGCGCTACCTTTTTC -3'	StII-UL21C (237-535), reverse
CM006	5’- GAAAAAGGTAGCGGCCCCACGCTA -3'	StII-UL21C (281-535), forward
CM007	5’- TAGCGTGGGGCCGCTACCTTTTTC -3'	StII-UL21C (281-535), reverse
CM008	5’- GAAAAAGGTAGCCAGGATTCCGCC-3'	StII-UL21C (275-535), forward
CM009	5’- GGCGGAATCCTGGCTACCTTTTTC -3'	StII-UL21C (275-535), reverse
CM016	5’-GAAAAACTAGAGGTTCTGTTCCAAGGCCCCCAGGAT -3'	StII-HRV3C-UL21C (275-535), forward

CM017	5’- ATCCTGGGGGCCTTGGAACAGAACCTCTAGTTTTTC -3'	StII-HRV3C-UL21C (275-535), reverse
CM020	5’- GTGTCCGGTATCTCGACGCTCTCCCTTATG -3'	pET forward, upstream
CM021	5’- GCGGTGTCAGCAGCCAACTCAGCTTCC -3'	pET reverse, downstream
CM023	5’- TCCCTGGAACAGCACCTCTAGTTTCTCGAACTGCGG -3'	StII-HRV3C-UL21(PSM) internal primer, reverse
CM029	5’- GTGCTGTTTCAGGGACCGATGGAGCTTAGTTACGCC -3’	StII-HRV3C-UL21(PSM) internal primer, forward
CM090	5’-GAAGGAGACTCATATGGAGCTTAGCTACGC -3'	UL21-StII, forward
CM091	5’-GAAGGAGACTGGATCCATGGAGCTTAGCTACGCCACCACCATGCACTAC -3’	GST-UL21, forward
CM092	5’-CTTCTTCTTTGCGGCCGCTCGAGTTATTACACAGACTGTCCGTGTTGGGAGCG -3’	GST-UL21 (C) (275-535), reverse
CM093	5’-GCTCGTAAGCTTATTATTTCTCGAACTGCGGGTGGCTCCAACTACCGACAGACTGTCCGTGTTGGGAG -3'	UL21-StII, reverse
CM096	5’- GTGCCAGGGGCCGGTAGTTGGAGC -3'	UL21N(1-274)-StII, internal forward
CM097	5’- GCTCCAACTACCGGCCCCTGGCAC -3'	UL21N(1-274)-StII, internal reverse
CM098	5’- CGCGTGGGCCCCGGTAGTTGGAGC -3'	UL21N(1-253)-StII, internal forward
CM099	5’- GCTCCAACTACCGGGGCCCACGCG -3'	UL21N(1-253)-StII, internal reverse

CM147	5’- CCCTCCACTCCATGGGACAGGATTCCGCCC -3’	GST-UL21C(275-535), forward

2.2.3 Recombinant protein expression, purification, and interaction studies

All UL21 constructs were expressed in Rosetta E. coli (Novagen) using overnight induction at OD600 0.6-1.0 with 1 mM isopropyl-B-D-thiogalactopyranoside (IPTG) at 16 °C. Cell pellets were resuspended in buffer A [50 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 100 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)] supplemented with 1 mM ethylenediaminetetraacetic acid (EDTA), 0.1 mM phenylmethanesulfonylfluoride (PMSF), 1X cOmplete protease inhibitor cocktail (Roche), and egg white avidin (Sigma) and lysed using either sonication, French press, or a microfluidizer. StII-tagged UL21 was captured from the clarified lysate using StrepTactin Sepharose resin (GE Healthcare) and eluted with 5 mM d-desthiobiotin (Sigma) in buffer A. GST-tagged UL21C was captured from the clarified lysate using Glutathione Sepharose 4B (GE Healthcare Life Sciences) and eluted with 10 mM reduced glutathione in buffer A.

Prior to size exclusion chromatography, UL21 or UL21C were separated from co-purifying E. coli nucleic acids using a Heparin sepharose HiTrap column (GE Healthcare) in buffer A. UL21C was eluted from heparin in buffer A using a NaCl gradient from 0.1 to 1.0 M. Constructs containing the PreScission protease site were cleaved with recombinant GST-tagged HRV3C (PreScission) protease at a 10:1 protein:protease mass ratio at 4 °C overnight without continued stirring or rocking. After cleavage, the mixture was passed over StrepTactin sepharose and glutathione sepharose (GE Healthcare) to remove the uncleaved protein, cleaved StrepII tag, and PreScission

protease, and the unbound fraction containing untagged UL21C was collected. All UL21 constructs were further purified using size-exclusion chromatography (Superdex 200 or Superdex 75, in buffer A), and stored with 1X Halt Protease Inhibitor Cocktail (Pierce).

Affinity purification and size exclusion were also used to investigate the interactions between the domains of UL21. 100-150 µg of various UL21N and UL21C constructs were incubated in a total of 250 µl buffer A alone or together at room temperature for 10-15 minutes before incubation with appropriate affinity resin or injection onto Superdex 75 size exclusion chromatography.

2.2.4 UL21N crystallization and structure solution

Crystals were grown by vapor diffusion in hanging drops (2 µL protein at ~5 mg/ml, 2 µL crystallization solution [(0.8-1.2 M ammonium sulfate, 100 mM HEPES pH 7.5, 7-10% MPD (2-methyl-2,4-pentanediol)]) at room temperature and flash frozen in crystallization solution with 15% MPD for data collection. Hg derivative crystals were prepared from native with soaking in crystallization solution with 1.7 mM thimerosal. X-ray diffraction data were collected at 100 K on beamline X25 at the National Synchrotron Light Source and processed in HKL2000 [101] (Table UL21Nstats). SAD was used to generate experimental electron density in autoSHARP [102]. The model was built manually in Coot [103] and refined against native data including gradient minimization refinement of XYZ coordinates, individual thermal parameters, and TLS parameters, with optimization of X-ray/stereochemistry and X-ray/ADP weights, all in *phenix.refine* [104] (Table UL21Nstats), with 5% of reflections set aside as a reference. The final Rwork is 16.6% and Rfree is 22.1%.

2.2.5 UL21C crystallization and structure solution

Native and SeMet UL21C crystals were grown overnight at room temperature by vapor diffusion in hanging drops (2 μL protein at ~3 mg/ml, 2 μL crystallization solution [17-23% polyethylene glycol 3350, 100 mM sodium citrate pH 3.5-4.0, 0.3-0.4 M non-detergent sulfobetaine-256 (Hampton Research), and 25 mM sodium fluoride]). Crystals were soaked in a cryoprotective solution consisting of crystallization solution supplemented with 18% sorbitol and flash frozen in liquid nitrogen. Diffraction data were collected at 100 K on beamline 24-ID-C at the Advanced Photon Source at Argonne National Labs and processed using RAPD software (https://rapd.nec.aps.anl.gov/rapd). To locate selenium sites, two highly redundant isomorphous Se SAD data sets (400° rotation each) were merged using RAPD. The SAD pipeline within RAPD was used to locate selenium sites (SHELX Sheldrick [105]), calculate phases, and generate a preliminary electron density map (AutoSol [104]). An initial model was built manually into the experimental electron density using Coot [103] and improved through multiple rounds of density modification and phase combination in RESOLVE [104]. The resulting model was refined against native data in *phenix.refine* [104] to 2.7 Å. Like for UL21N, model refinement included gradient minimization refinement of XYZ coordinates, individual thermal parameters, and TLS parameters, with optimization of X-ray/stereochemistry and X-ray/ADP weights, all as implemented in *phenix.refine*. The final R_{work} is 21.77% and R_{free} is 23.43%. All statistics are listed in Table UL21Cstats.

2.2.6 Limited proteolysis and N-terminal sequencing

Limited proteolysis and N-terminal sequencing were performed as for UL11.

2.2.7 Size Exclusion Chromatography Small Angle X-ray Scattering (SEC-SAXS)

SEC-SAXS experiments with UL21[(PSM)1-532-HRV3C-STII] were carried out at beamline 18ID at the Advanced Photon Source (Ithaca, NY). UL21[(PSM)1-532-HRV3C-STII] at approximately 8 mg/ml was pre-cleared of aggregates and approximately 300 µL was injected through a Superdex 75 10/300 in buffer A at ambient temperature and fed directly into a flow cell. Samples were irradiated with a 12 keV (1.254489 Å) beam and 2 second images were taken through the duration of the run with a Pilatus 3 1M detector (Dectris) at a distance giving $0.006\ Å^{-1} < q < 0.35\ Å^{-1}$ with scattering vector $q = 4\pi\sin(\theta)/\lambda$. Data was processed in PRIMUS [106] within the ATSAS software package (https://www.embl-hamburg.de/biosaxs/software.html) [94] by averaging appropriate frames with a consistent Rg value and subtracting averaged buffer frames.

SEC-SAXS experiments with StII-UL21(PSM) and UL21(PSM)-StII were carried out at beamline G1 at the Cornell High Energy Synchrotron Source (Ithaca, NY). Samples at approximately 5 mg/ml were pre-cleared of aggregates by centrifugation and 100 ul aliquots were injected through a Superdex 200 Increase 5/150 (GE Healthcare Life Sciences) in buffer A at 0.3 ml/minute and 4°C or 22°C and fed directly into a quartz flow cell. Samples were irradiated with a 9.8833 keV (1.254489 Å) beam with 5.7 x 1011 photons/s flux and diameter of 250 µm x 250 µm, and 2 second images were taken through the duration of the run with a Pilatus 100K-S detector (Dectris) at a distance giving $0.006\ Å^{-1} < q < 0.8\ Å^{-1}$ with scattering vector $q = 4\pi\sin(\theta)/\lambda$. Data was processed in the RAW software package (https://sourceforge.net/projects/bioxtasraw/) [93] by

averaging appropriate frames with a consistent Rg value and subtracting averaged buffer frames.

Subtracted curves were further analyzed using programs in the ATSAS software package including GNOM [107], DAMMIF[108], DAMMIN [96], DAMAVER [97], CORAL [109], and EOM 2.0 [98]. GNOM was used to generate the pair distance distribution function [p(r)] from the subtracted experimental data. In DAMMIN, the p(r) function from GNOM was used to compile a bead model in slow, fine bead mode with no known symmetry or anisometry. The DAMAVER suite was used to superimpose and average *ab initio* bead models from DAMMIN, DAMMIF, or CORAL. In CORAL, the crystal structures of UL21N and UL21C were used to create rigid body models from experimental SAXS data. In EOM, a pool of 10,000 random native-like structures representing most of the structural space available to the sequence of UL21 were generated. Theoretical SAXS curves were generated for random combinations of these models and the ensemble was optimized with a genetic algorithm to best fit the experimental data. No symmetry information or structured domains were provided.

2.2.8 Nucleic acid precipitation, nuclease digestion, and formaldehyde gels.

StII-UL21 bound to endogenous *E. coli* nucleic acids was expressed and purified from 1 L Rosetta *E. coli*. The complex was split into three aliquots from which DNA, RNA, or total nucleic acids were isolated. Total nucleic acids were isolated by phenol:chloroform precipitation. Briefly, an equal volume of phenol:chloroform (125:24, Life Technologies) was added to each sample. The upper aqueous layer was mixed with 10 µg glycogen (Life Technologies), 1/10 volume 3M sodium acetate, and 3 volumes ethanol, and incubated at -20 °C for 30 minutes. Nucleic acids were pelleted at 16,000 x g

at 4 °C for 20 minutes and washed with 75% ethanol before resuspension in water for analysis. To isolate RNA only, the precipitation was performed using acidic phenol:chloroform, pH 4.5. To isolate DNA only, the complex was treated with RNase A at 1:1000 RNase A (Invitrogen):UL21C volume ratio for 20 minutes at 37 °C prior to nucleic acid extraction with a neutral pH phenol:chloroform mixture. Two µg of precipitated nucleic acid fractions resuspended in water were treated with 2 units of TURBO DNase (Ambion) or 0.4 µg RNase A (Invitrogen) in 1X TURBO DNase buffer with 2 mM calcium chloride for 30 minutes at 37 °C. Samples were analyzed on a denaturing RNA gel. Samples were prepared with 5X RNA sample buffer (4mM EDTA, 2.7% formaldehyde, 30.8% formamide, 20% glycerol, 40% 10X MOPS buffer [200 mM 3-(N-Morpholino)propanesulfonic acid (MOPS), 30 mM sodium acetate trihydrate, 10 mM EDTA, pH adjusted to 7.0 with sodium hydroxide]) diluted to 1X, 25% FORMAzol (Molecular Research Centers, Inc.), and 0.05 mg/ml ethidium bromide. Samples were heated at 85°C for 1-3 minutes, cooled on ice, and loaded immediately onto a 1.2% agarose gel prepared in 1X MOPS buffer with 5% formaldehyde and run at 75V in 1X MOPS buffer.

2.3 Methods to study UL16

2.3.1 Sequences and prediction

Sequences of UL16 homologs from human herpesviruses [YP_009137090.1, CAB06776.1, CAA27927.1, YP_401691.1, CAA35368.1, NP_050245.1, AAC54727.1, YP_001129386.1 (RefSeq or GenBank accession numbers)] were aligned in Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/). Expasy ProtParam (http://web.expasy.org/protparam/) was used to calculate protein size. PsiPred (bioinf.cs.ucl.ac.uk/psipred) was used to predict secondary structure. Alignment,

conservation, and secondary structure figure was made with ESPRIPT (http://espript.ibcp.fr/ESPript/ESPript/).

2.3.2 Cloning and expression constructs

Plasmids encoding full-length, residues 1-155, or residues 156-373 of wild type HSV-1 strain 17 UL16 in a pET11a vector preceeded by an N-terminal Strep-II (StII) tag and factor Xa cleavage site were gifts from John W. Wills. Codon optimized DNA encoding full length wild type HSV-1 strain 17 UL16 or UL16 with 12 non-conserved cysteines mutated to serines (C69, 78, 93, 125, 142, 193, 210, 283, 294, 315, 328, 343S, Figure 16-1A) were purchased from GeneArt, amplified with CM024 and CM028, cut with BamHI and XhoI, and subcloned into a pET24b vector encoding an N-terminal His_6-SUMO tag and human rhinovirus 3C protease (HRV3C, PreScission) cleavage site to produce His_6-SUMO-HRV3C-UL16CS, a pET24b vector encoding an N-terminal StII-SUMO-HRV3C tag to produce StII-SUMO-HRV3C-UL16, or into pGEX-6P-1 vector to produce GST-HRV3C-UL16 or GST-HRV3C-UL16CS. StII-UL16 was constructed by amplifying the codon optimized UL16 gene with CM086 and CM028, digesting the product with NdeI and XhoI, and subcloning it into pET24a backbone. To create UL16CS-His_6 or UL16CS(41-373)-His_6, UL16CS was amplified with primers CM123 and CM110 or CM113, cut with NcoI and XhoI, and subcloned into a pET28a backbone. To construct UL16CS(156-373)-StII, UL16CS-StII was first made using SOE PCR with primers AK37, CM138, CM139, and AK34 using a template featuring full length UL16CS in a pET24b vector. The amplified insert was subcloned into a pET-24a backbone with NcoI and XhoI. The insert representing UL16CS(156-373)-StII was amplified from UL16CS-StII with primers CM140 and CM141, digested with NdeI and

BamHI, and subcloned into pET11 vector. To construct UL16CS(1-155)-GFP-His_6, SOE PCR was performed using primers AK37 and CM142 to amplify UL16CS(1-155) from UL16CS-His_6 and primers CM143 and CM144 to amplify GFP from vector pEGFP-N2 and add a His_6 tag and XhoI restriction site. The two amplicons were further amplified with primers AK37 and CM144, cut with NcoI and XhoI, and subcloned into pET28 vector. All primers are listed in Table 2-3.

Table 2-3. UL16 cloning primers

Primer	Sequence	Description
AK34	5’- AAAGGAAGGGAAGAAAGCGAAA -3’	Downstream pET primer, reverse
AK37	5’- ATACCGCGAAAGGTTTTGCG -3’	Upstream pET primer, forward
CM024	5’- ATATATCTGGGATCCATGGCACAGCTG -3'	UL16 N-terminus, forward
CM028	5’- CATGCTCTCGAGTTATTATTCAGGATCGC -3'	UL16 C-terminus, reverse
CM086	5’-GAGATATACATATGTGGAGTCACCCTCAGTTCGAAAAAGGTAGCATGGCCCAGCTAGGTCCG -3'	StII-UL16 C-terminus, forward
CM110	5’- AAAAACGCTCGTCCATGGCACAGCTGGGT -3’	UL16 N-terminus, forward
CM113	5’-CAAAAACGCTCGTCCATGGGCGATGTTGATAGCATTGCACGT -3’	UL16 residue 41, forward
CM123	5’-GACTGTAGCTCTCGAGTTATTAGTGGTGATGATGGTGATGGCTGCCTTCCGGATCGCTGC -3’	UL16 C-terminus His_6, reverse
CM138	5’- AAATTGTGGATGTGACCATTCCGGATCGCTGCT -3’	UL16CS-StII internal primer, reverse
CM139	5’-CATCCACAATTTGAGAAGTAATAACTCGAGCACCACCAC -3’	UL16CS-StII internal primer, forward

CM140	5’- GCGATACAACATATGGAAGAAACCCCTGATCCG -3’	UL16CS(156-373), forward
CM141	5’- CGCGCGCGGATCCTTATTATTTCTCGAATTGTGGGTG -3’	StII C-terminus, reverse
CM142	5’- GCCTTTGCTCACCATCGGAGGGGTTGCTGC -3’	UL16(1-155)-GFP internal primer, reverse
CM143	5’- GCAGCAACCCCTCCGATGGTGAGCAAAGGC -3’	UL16(1-155)-GFP internal primer, forward
CM144	5’- GAGTGGAGGGCTCGAGTTATTAGTGGTGATGATGGTGAT GCTTGTACAGTTCGTC -3’	GFP-His_6, reverse

2.3.3 Recombinant protein expression

Unless otherwise noted, all UL16 constructs were expressed in Rosetta E. coli (Novagen) using overnight induction at OD600 0.6-1.0 with 1 mM isopropyl-B-D-thiogalactopyranoside (IPTG) at 16 °C or 4 hour induction at 37 °C. Cell pellets were resuspended in buffer A [50 mM 4-(2-hydroxyethyl)piperazine-1-ethanesulfonic acid (HEPES), 100 mM NaCl, 0.5 mM tris(2-carboxyethyl)phosphine (TCEP)] supplemented with 0.1 mM phenylmethanesulfonylfluoride (PMSF) and 1X cOmplete protease inhibitor cocktail (Roche), and lysed using a Microfluidizer for purification scale harvest, or lysozyme, or multiple cycles of freezing and thawing for solubility assays. Some constructs were expressed by autoinduction as in reference [86]. Briefly, a small culture of cells was grown overnight at 37°C in terrific broth (TB) with 1% glucose and 2 mM MgSO4 before inoculation at a ratio of 1:100 into TB with 0.2% lactose and 2 mM MgSO4 at 37°C for 4 hours then 25°C for 18-22 hours.

2.3.4 Expression and solubility assays

For assessing expression and solubility in small scale, cell density was measured for each culture and a sample equivalent to 1 mL of OD600 = 1.0 was collected. The cells were pelleted for 30 seconds at 16,000 x g at room temperature and resuspended in 120 µL buffer of choice. For assessing expression, 30 µL 5X reducing sample buffer was added to the sample and boiled for 2-10 minutes at 90-95°C. For assessing solubility, cells were either lysed using multiple cycles of freezing and thawing or with lysozyme. If frozen and thawed, samples were submerged in liquid N_2 until frozen and heated at 37°C until thawed. This process was repeated for a total of 5 cycles of freezing and thawing. If lysed by lysozyme, the original sample size was…The insoluble and soluble portions were separated by centrifugation at 16,000 x g at 4°C for 20 minutes and the soluble portion was placed in a second tube. The insoluble pellet was resuspended in 150 µL 1X reducing sample buffer, 30 µL 5X reducing sample buffer was added to the 120 µL soluble portion, all samples were boiled for 2-10 minutes at 90-95°C, and 10 µL each sample was run on SDS-PAGE for staining or western blotting.

2.3.5 Protein Purification

After lysis in the Microfluidizer, insoluble material was spun out 20,000 x g and 4°C for 30 minutes. StII-tagged constructs were captured from the clarified lysate using StrepTactin Sepharose resin (GE Healthcare) and eluted with 5 mM d-desthiobiotin (Sigma) in buffer A. GST-tagged constructs were captured from the clarified lysate using Glutathione Sepharose 4B (GE Healthcare Life Sciences) and eluted with 10 mM reduced glutathione in buffer A.

For His_6-tagged constructs, all buffers were degassed before use. If purifying in a reducing environment, buffers had a total of 2 mM TCEP. His_6-tagged constructs were captured from the clarified lysate using Ni Sepharose 6 Fast Flow (GE Healthcare Life Sciences) and eluted with 300 mM imidazole in buffer A.

When the construct was subjected to heparin purification, the protein was passed over a heparin sepharose HiTrap column (GE Healthcare) that had previously been charged with buffer B and equilibrated in buffer A. The protein-laden column was washed with buffer A, and eluted using a NaCl gradient from 0.1 to 1.0 M (100% buffer A to 100% buffer B). If ion exchange was performed, the same procedure was done using Q sepharose (HiTrap Q HP, GE Healthcare Life Sciences). If proteins were separated from the affinity tag by virtue of the PreScission protease site, recombinant GST-tagged HRV3C (PreScission) protease was added at an approximately 50:1 protein:protease mass ratio at 4 °C overnight without continued stirring or rocking. After cleavage, the mixture was passed over glutathione sepharose (GE Healthcare) to remove the uncleaved protein, cleaved GST tag, and PreScission protease, and the unbound fraction was collected. Any size-exclusion chromatography (Superdex 200 or Superdex 75) was performed in buffer A.

2.3.6 Western blotting

Western blots were performed by transferring proteins from polyacrylamide gels to nitrocellulose membrane in a semi-dry transfer apparatus for 30 minutes at 25 volts in transfer buffer [50 mM Tris, 37.5 mM glycine, 0.0375% sodium dodecyl sulfate (SDS), 30% methanol]. To probe for the StII tag, blots were blocked for 30 minute at room temperature in 3% bovine serum albumin (New England Biolabs) in TBST [Tris-buffered

saline (150 mM NaCl, 20 mM Tris pH 7.5) with 0.05% Tween-20] before adding StrepTactin-HRP conjugate (Bio-Rad) at a ratio of 1:3000 and incubation overnight at 4°C. To probe for the GST tag, GFP tag, or UL16, blots were blocked for 30 minutes in 5% non-fat dry milk before addition of the HRP-conjugated GST antibody (Abcam, 1:x), rabbit polyclonal GFP antibody (Santa Cruz, 1:3000), or rabbit polyclonal UL16 or UL16C antibody (Gifts from John Wills, 1:3000) overnight at 4°C. For unconjugated primary antibodies, after several washes in TBST, an HRP-conjugated goat antibody recognizing rabbit IgG (Bio-Rad, 1:10,000) was added for an hour at room temperature. After multiple washes in TBST, blots were developed using an ECL western blotting substrate.

2.4 Methods to study UL51

2.4.1 Sequences and prediction

Sequences for HSV-1 (F) UL51 (ADD60010.1), HSV-2 (HG52) UL51 (YP_009137204.1), VZV (Dumas) ORF7 (NP_040130.1), EBV (B95-8) BSRF1 (YP_401663.1), CMV (AD169) UL71 (DAA00167.1), HHV-6 (Z29) U44 (NP_050225.1), HHV-7 (RK) U44 (YP_073784.1), and KSHV (GK18) ORF55 (YP_001129408.1) were aligned in Clustal Omega (http://www.ebi.ac.uk/Tools/msa/clustalo/) and adjusted manually to show conservation of palmitoylation site, cysteine 9 in HSV-1. PsiPred and DISOpred (bioinf.cs.ucl.ac.uk/psipred) were used to predict secondary structure and disordered regions, respectively.

2.4.2 Cloning and expression constructs

Plasmids encoding full length HSV-1 strain F UL51 preceeded by an N-terminal His_{10}-SUMO tag and a human rhinovirus (HRV) 3C (PreScission) protease site in pET-

24b (His_{10}-SUMO-HRV3C-UL51 or His_{10}-SUMO-UL51) or UL51 (12-244) preceeded by a Glutathione S-transferase tag (GST) and PreScission site in pGEX-6P-1 [GST-HRV3C-UL51(12-244) or GST-UL51(12-244)] were previously cloned in the lab by Xuanzong Guo (Figure 51.1). To construct His_{10}-SUMO-UL51 with a C-terminal Strep II tag (StII) (His_{10}-SUMO-HRV3C-UL51-StII or His_{10}-SUMO-UL51-StII) (Figure 51.1), single overlap extension (SOE) PCR was performed on His_{10}-SUMO-UL51 using primers CM20, CM21, CM182, and CM183, as listed in Table 2-4.

Table 2-4. UL51 cloning primers

Primer	Sequence	Description
CM20	5'-GTGTCCGGTATCTCGACGCTCTCCCTTATG-3'	pET upstream primer, forward
CM21	5'-GCGGTGTCAGCAGCCAACTCAGCTTCC-3'	pET downstream primer, reverse
CM182	5'-TCACATCCACAATTTGAGAAGTAATAAAAGCTTGCG-3'	HS-UL51-StII internal primer, forward
CM183	5'-TTGTGGATGTGACCAGGAACCTTGACCCAAAAC-3'	HS-UL51-StII internal primer, reverse

2.4.3 Recombinant protein expression and purification

UL51 constructs were expressed in LoBStr *E. coli* cells in LB after induction at OD 0.6-0.8 with 0.2 mM IPTG. Cells were harvested by centrifugation for 30 minutes at 5,000 xg in a table top centrifuge and lysed in a fluidizer in buffer C (50 mM HEPES pH 7.0, 500 mM NaCl, 10% glycerol, 1 mM TCEP) with 1X complete, 0.1 mM PMSF, and optional scant DNase. For C-terminally strep-tagged constructs, the pH of buffer C was increased to 7.5. Lysates were cleared of insoluble aggregates by centrifugation and the clarified lysates were retained. His_{10}-tagged constructs were passed over Ni Sepharose 6 Fast Flow (GE Healthcare Life Sciences) resin equilibrated in buffer C, washed with 20, 30, then 50 mL buffer C with 20-30 mM imidazole, and eluted with 300 mM imidazole in buffer C. If the UL51 construct encoded an additional StII tag, the elution from nickel resin was passed directly over StrepTactin resin equilibrated in buffer C, washed with 20 then 50 mL buffer C, and eluted with buffer C with 5 mM d-desthiobiotin. Clarified lysates containing GST-UL51 (12-244) were passed over glutathione resin equilibrated in buffer C, washed with 2 x 50 mL buffer C, and eluted with buffer C with 10 mM reduced glutathione.

Elution fractions were incubated with 0.5mM EDTA, 1mM MgCl2, 10 µM ATP, and PreScission protease at 4°C overnight without movement. For elution fractions that were impure, approximately 800 µg PreScission was added per liter of expression. For elution fractions that contained mostly pure UL51, PreScission was added at a 1:30 PreScission:protein molar ratio. The UL51 cleavage products will herein have prefix of "GPGS-." The UL51 cleavage reactions were concentrated in a 50 kDa MWCO concentrator before size exclusion purification on superdex S75 in buffer D (20 mM

HEPES pH 7.0, 100 mM NaCl, 1.0 mM TCEP) with GSTrap columns inline to capture PreScission and free GST-tag if present. For StII-tagged constructs, the pH was adjusted to 7.5. Fractions just after the void containing UL51 were collected.

For His_{10}-SUMO-UL51, the pool from S75 was subjected to ion exchange purification on Q sepharose (HiTrap Q HP, GE Healthcare Life Sciences). The Q columns were equilibrated with 5 column volumes (CV) buffer D, charged with 5CV buffer E (20 mM HEPES pH 7.0, 1.0 M NaCl, 1.0 mM TCEP), and washed with 5CV buffer D before applying protein. The UL51-bound columns were then washed with 5CV buffer C, a 10CV gradient from 100% buffer D to 100% buffer E, 5CV buffer E, and 5 CV buffer D. Fractions containing pure UL51 (mostly from the flow through) were pooled and concentrated for another round of size exclusion purification on superdex S200 in buffer D.

To calibrate size exclusion columns, blue dextran, aldolase, conalbumin, ovalbumin, and ribonuclease A (Gel Filtration Calibration Kits, GE Healthcare Life Sciences) were injected over the column and calibration curves were generated to calculate the apparent molecular weight and of UL51 using the following formulas, according to manufacturer's instructions.

$K_{av} = \frac{V_e - V_o}{V_c - V_o}$ proportional to $\log(molecular\ weight)$, where K_{av} is the partition coefficient, V_e is the elution volume, V_c is the geometric column volume, and V_o is the void volume.

2.4.4 Western blotting

Western blots probing for the StII tag were performed by transferring proteins from polyacrylamide gels to nitrocellulose membrane in a semi-dry transfer apparatus for 30 minutes at 25 volts in transfer buffer [50 mM Tris, 37.5 mM glycine, 0.0375% sodium dodecyl sulfate (SDS), 30% methanol]. Blots were blocked for 30 minute at room temperature in 3% bovine serum albumin (New England Biolabs) in TBST [Tris-buffered saline (150 mM NaCl, 20 mM Tris pH 7.5) with 0.05% Tween-20] before adding StrepTactin-HRP conjugate (Bio-Rad) at a ratio of 1:3000 and incubation overnight at 4°C. After multiple washes in TBST, blots were developed using an ECL western blotting substrate.

2.4.5 Thermofluor assay

To perform the Thermofluor assay [110], fluorescent Sypro Orange dye (Invitrogen) was diluted 1000x in GPGS-UL51 or GPLGS-UL51(12-244) at 0.2 mg/ml in CD buffer. The dye and protein was mixed at a 1:1 ratio with a various buffers at a total of 20 µL in a qPCR plate. After sealing and centrifugation for 1 minute at 500 x g, the plate was subjected to a temperature gradient in a Roche LightCycler from 25-95°C with continuous measurements at an acquisition rate of 3 per °C with an excitation wavelength of 425 nm and an emission wavelength of 610 nm. The resulting data was analyzed with the ThermoQ software, which presented melting curves, derivative curves, and melting points for each condition. StII-UL21C(275-535) and PreScission (HRV3C) protease were used as a controls for well-folded protein.

2.4.6 Circular dichroism

Circular dichroism was performed on GPGS-UL51 at various concentrations as for UL11.

2.4.7 Limited proteolysis, N-terminal sequencing, and mass spectrometry

Limited proteolysis was performed as for UL11 but 3 µg GPGS-UL51 was incubated with 0, 0.7, 1.9, 3.75, 7.5, 15, 30, 60, or 120 ng TLCK-treated chymotrypsin (Sigma) in 1X buffer D. For N-terminal sequencing and mass spectrometry, 450 µg GPGS-UL51 was reacted with 3 µg chymotrypsin at room temperature for an hour before stopping with 1.0 mM PMSF. A portion of the sample containing 75 µg GPGS-UL51 was separated with SDS-PAGE on a 12% before transferring to PVDF in preparation for N-terminal sequencing, as was done for UL21. For mass spectrometry, a portion of the same sample containing 350 µg cleaved GPGS-UL51 was transferred into water, as was done for UL11.

2.4.8 Size Exclusion SAXS

SEC-SAXS was performed as for UL11 except for 100 µL GPGS-UL51 at ~16 mg/mL was injected over a Superdex 200 Increase 10/300 at 0.6 mL/min in buffer D. Data was processed in the same way.

2.4.9 Co-sedimentation assay

The co-sedimentation assay was performed as for UL11 and UL21.

2.4.10 Crosslinking assay

Chemical crosslinking was performed as for UL11.

Chapter 3: UL11

3.1 Introduction

UL11 is the smallest tegument protein in HSV-1 [16] and it is conserved throughout all subfamilies of herpesviruses. It is important for efficient replication of HSV-1 as deletion of UL11 results in a 10-fold reduction in viral titer [76, 111, 112], with similar effects on replication seen after deleting the VZV homolog ORF49 in some cell types [113] or KSHV homolog ORF38 [59]. UL99/pp28, another UL11 homolog, is essential for replication of CMV as deletion of UL99 renders the virus incapable of releasing particles into the media with a concurrent accumulation of unenveloped capsids in the cytoplasm [57]. HSV-1 UL11 is modified by a myristyl group on conserved glycine residue 2 [50, 114] at its N terminus, and myristylation has been confirmed for UL11 homologs from HSV-2 [115], EBV [116], and CMV[117]. This modification is likely conserved among all other homologs [50, 118, 119] and highlighted by the observation that this is one of few conserved residues in UL11, myristylation is likely central to a conserved role of UL11. This modification allows UL11 homologs to associate with cytoplasmic membranes and, in one report, potentially nuclear membranes of HSV-1 [81]. Additionally, myristylation of HSV-1 UL11, EBV homolog BBLF1, and likely HSV-2 UL11 precedes palmitylation of these proteins [65, 116, 120, 121], which occurs on at least one cysteine between residues 11-13 in UL11 and on cysteine 8 in BBLF1. This second modification is thought to stabilize the localization of UL11 homolog with membranes at the site of secondary envelopment and is required for association of HSV-1 or -2 UL11 with lipid rafts.

Membrane-associated UL11 interacts with capsid-associated tegument protein UL16 in a conserved interaction. Through this interaction (that in alphaherpesviruses,

also occurs with capsid-associated UL21 and glycoprotein gE), UL11 is thought to bridge the capsid and envelope during secondary envelopment. Additionally, UL11 (with UL16 and UL21) has also been shown to modulate the function of gE in cell-cell and potentially virus-host cell fusion [45]. Furthermore, there are some as of yet undescribed functions of UL11 that don't rely on its association with membranes as G2A mutations resulting in a non-acylated, soluble version of the protein can partially rescue replication of a UL11-null HSV-1 virus [76].

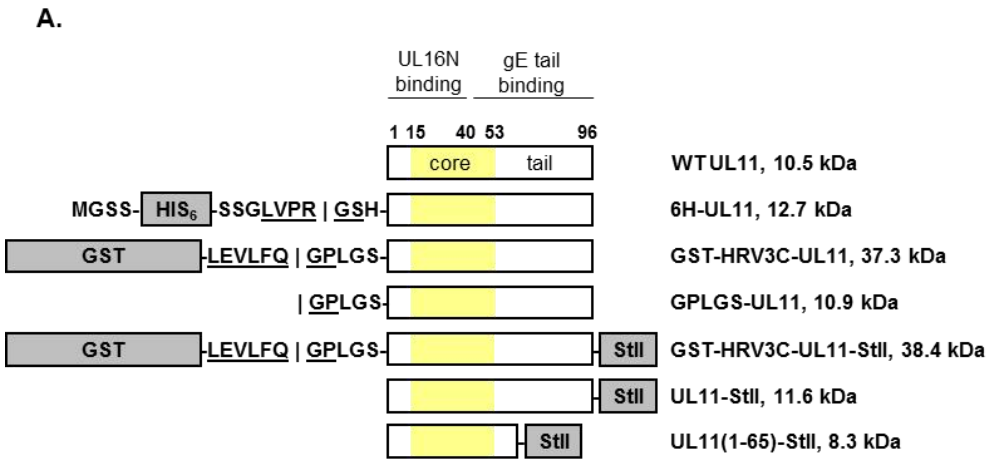
To begin to characterize the multiple functions and binding interactions of HSV-1 UL11 and provide guidelines for future mutagenesis studies, we sought to characterize it structurally and biochemically. We found that UL11 has a small N-terminal core and a long unstructured C terminus that has characteristics of an intrinsically disordered protein. UL11 is a flexible protein that adopts multiple conformations in solution. Furthermore, we found that UL11 binds ribosomal RNA.

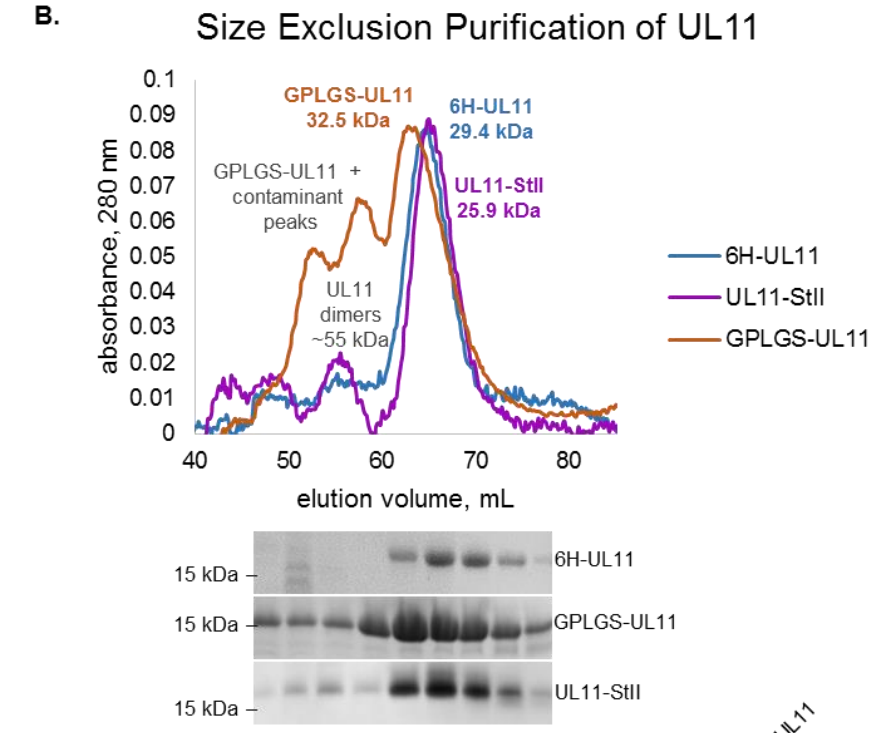
3.2 Results

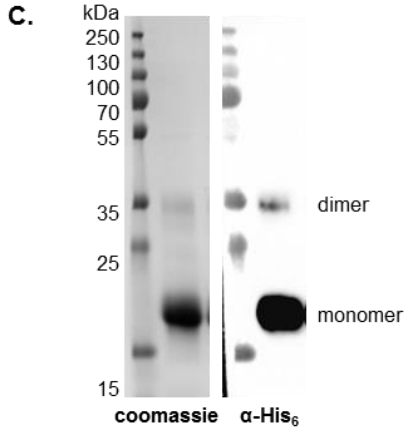
3.2.1 Expression and purification of UL11 constructs

HSV-1 UL11 is a 96-amino acid protein with calculated molecular weight of 10486.62 Da (Figure 3-1A) before any post-translational modifications. To begin characterizing UL11 from HSV-1, 6H-UL11 was expressed in *E. coli* in unlipidated form [122] and purified. This construct was soluble and eluted from size-exclusion chromatography predominantly at a volume consistent with a globular molecule of approximately 30 kD, a weight between the calculated masses of a 6H-UL11 dimer and trimer (Figure 3-1B). On SDS-PAGE, this construct migrated slightly larger than the 15 kD marker. Larger multimeric species around 35 kD were occasionally observed in size-exclusion

chromatograms and by SDS-PAGE (Figure 3-1B-C). GPLGS-UL11 and UL11-StII constructs were also purified using size-exclusion chromatography. Elution volumes were similar to 6H-UL11 and the calculated molecular weights also fell near those of a dimer or a trimer, suggesting that 6H-UL11 is dimeric or trimeric in solution. Since non-globular proteins can elute aberrantly from size exclusion leading to errors in molecular weight and multimer assignment, UL11 could alternately be taking the form of an elongated monomer.


A.
UL16N binding
gE tail binding
1 15 40 53 96
core
tail
WT UL11, 10.5 kDa
MGSS- HIS6 -SSGLVPR | GSH-
6H-UL11, 12.7 kDa
GST -LEVLFQ | GPLGS-
GST-HRV3C-UL11, 37.3 kDa
| GPLGS-
GPLGS-UL11, 10.9 kDa
GST -LEVLFQ | GPLGS-
StII
GST-HRV3C-UL11-StII, 38.4 kDa
UL11-StII, 11.6 kDa
UL11(1-65)-StII, 8.3 kDa


B.
Size Exclusion Purification of UL11
absorbance, 280 nm
elution volume, mL
GPLGS-UL11
32.5 kDa
6H-UL11
29.4 kDa
UL11-StII
25.9 kDa
GPLGS-UL11 + contaminant peaks
UL11 dimers ~55 kDa
6H-UL11
UL11-StII
GPLGS-UL11
15 kDa
6H-UL11
GPLGS-UL11
UL11-StII


C.
kDa
250
130
100
70
55
35
25
15
dimer
monomer
coomassie
α-His6

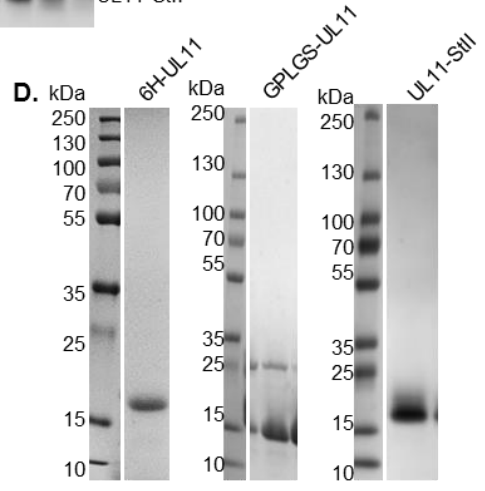

D.
kDa
6H-UL11
GPLGS-UL11
UL11-StII
250
130
100
70
55
35
25
15
10

Figure 3-1. Expression and purification of UL11 constructs
A) Linear maps of UL11 constructs including predicted structural features "core" and "tail". Amino acid boundaries are indicated. WT, wild type. Protease recognition sites are underlined with a vertical line identifying the cleavage site. B) Size exclusion chromatography A280 traces, gels, and calculated molecular weights for three UL11 constructs. C) (L) Coomassie gel shows that 6H-UL11, shown as a representative, has a predominant monomer band at 15 kD and a second band at 35 kD. (R) An anti-His_6 western Blot shows that this band is likely a UL11 dimer as it also contains the affinity tag. D) Coomassie gels of three main UL11 constructs indicate that, with the exception of GPLGS-UL11 and retained GST, UL11 preparations were extremely pure.

3.2.2 UL11 is a flexible molecule and undergoes phase separation

During affinity purification of 6H-UL11, the elution fraction appeared cloudy; however, centrifugation of the sample did not result in precipitate formation. Furthermore, after size-exclusion chromatography, the most concentrated fractions of the UL11-containing peak also appeared cloudy macroscopically. Under the microscope, such cloudy samples had obvious liquid-liquid phase separation (Figure 3-2A). This phenomenon occurred in a wide range of buffer conditions and was observed in 34% of 752 sitting drop crystal screening conditions. Furthermore, phase separation was also observed with GP-UL11 and UL11-StII samples and was thus independent of the affinity tag and reflected and intrinsic property of UL11. Proteins seen to undergo liquid-liquid phase separation often contain long intrinsically disordered regions (IDRs)[123]. Consistent with UL11 being disordered, proteins with IDRs often run at elevated molecular weights on SDS-PAGE [124].

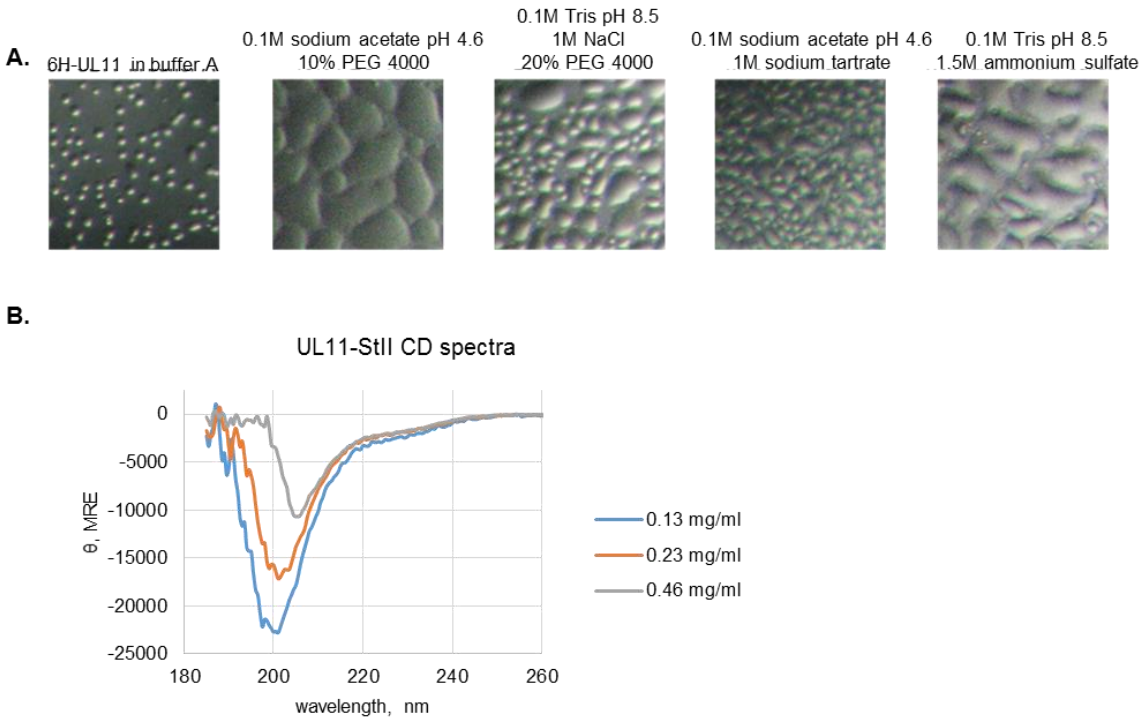


Figure 3-2. UL11 is a flexible molecule

A) Microscopic images of 6H-UL11 in solution show that it undergoes phase separation under widely varying conditions. (L) 6H-UL11 in purification buffer. (all R) 6H-UL11 in purification buffer set up 1:1 in various crystallization conditions. B) CD spectra of UL11-StII at varying concentrations show random coil characteristics. C) Sequences of UL11 homologs from human herpesviruses aligned to HSV-1 UL11 with residue number marked. The sequences are shown out of numerical order due to the comparatively long length of CMV UL99. Conserved residues are marked with an asterisk. Ģ, shown or predicted myristylation site. Ç, shown or predicted palmitylation site. Ĝ, conserved glycine likely to play a structural role. Italics denote disorder predicted by DISOPRED3. Straight underlined residues denote beta strands predicted by PSIPRED. Dashed underlined residues denote alpha helices predicted by PSIPRED. The shown or predicted acidic cluster is bolded. Disorder promoting residues are colored: A/G/S/P/Q are shown in purple, E in red, and R/K in blue.

3.2.3 UL11 contains a some secondary structure characteristics

To further characterize the structure of UL11, we used circular dichroism (CD). While visual inspection of the CD spectra at multiple protein concentrations appeared consistent with a random coil polypeptide (Figure 3-2B), CD analysis program CDSSTR using three different reference sets estimated that UL11 only comprises 40-62% polypeptide without regular structure (Table 3-1). This range in estimates is dependent on concentration, not reference set, indicating that concentrating UL11 may promote the formation of regular secondary structural elements. It is important to note that the magnitude of the CD signal decreases with increasing concentration, which is a phenomenon often associated with protein aggregation, though neither gross precipitation nor phase separation was obvious after removing samples from the instrument. These results suggest that there are some regular secondary structural elements within UL11 and their formation may be induced with concentration. Together with output from secondary structure, turn, and disorder prediction algorithms, we predict that the N-terminal half of HSV-1 UL11 is a folded domain whereas the C-terminal half is unstructured (Figure 3-2C). Unexpectedly, the CD analysis suggested the presence of alpha helices, which was not predicted for HSV-1 UL11 by secondary structure prediction software, even though it was predicted for several UL11 homologs (Figure 3-2C).

Table 3-1. Secondary structure in UL11-StII

Concentration (mg/mL)	Reference Set	Analysis Program	NRMSD	Helix1	Helix2	Strand1	Strand2	Turns	Unordered
0.46	SP175	CDSSTR	0.043	0.08	0.12	0.12	0.11	0.17	0.4
	Set 6	CDSSTR	0.056	0.03	0.04	0.21	0.09	0.17	0.46
	Set 7	CDSSTR	0.041	0.06	0.1	0.19	0.09	0.21	0.35
				0.06	**0.09**	**0.17**	**0.10**	**0.18**	**0.40**
				alpha	**0.14**	**beta**	**0.27**		
0.23	SP175	CDSSTR	0.034	0.01	0.07	0.17	0.13	0.16	0.46

	Set 6	CDSSTR	0.023	0.04	0.03	0.13	0.08	0.13	0.59
	Set 7	CDSSTR	0.024	0.02	0.04	0.12	0.07	0.14	0.58
				0.02	**0.05**	**0.14**	**0.09**	**0.14**	**0.54**
				alpha	**0.07**	**beta**	**0.23**		
0.13	SP175	CDSSTR	0.019	0.02	0.06	0.15	0.12	0.17	0.49
	Set 6	CDSSTR	0.022	0.02	0.04	0.11	0.07	0.12	0.64
	Set 7	CDSSTR	0.023	0.02	0.04	0.07	0.05	0.1	0.72
				0.02	**0.05**	**0.11**	**0.08**	**0.13**	**0.62**
				alpha	**0.07**	**beta**	**0.19**		

The sequence alignment of HSV-1 UL11 and its homologs from other human herpesviruses (Figure 3-2C) shows that despite limited sequence similarity between homologs, residues near the N terminus are more conserved than those at the C terminus, as previously suggested [50, 114]. Furthermore, adding secondary structure predictions to this analysis suggest that the N terminus of these molecules forms a structured core with predicted regular secondary structural elements and conserved features including lipidation sites, acidic clusters, and one conserved glycine that is likely to be necessary for structural flexibility. In contrast, the C termini vary widely in sequence identity and length. Nevertheless, even the C termini may share flexibility as they all show predicted disorder and an increase in disorder promoting residues [125], notably positively charged ones, in the C-terminal region in comparison to the N-terminal region (Figure 3-2C). This observation suggests that instead of the C terminus being non-conserved and dispensable, the C-terminus of UL11 may be conserved in its flexibility.

To test the hypothesis that UL11 consists of a folded core, we subjected UL11-StII to limited proteolysis. In the presence of increasing amounts of trypsin (Figure 3-3) and chymotrypsin (not shown), a single proteolytically-resistant core with apparent molecular weight between 10 and 15 kilodaltons was detected. Based on predicted sites

of proteolysis and previously described predicted secondary structure, we expected this putative core to map to residues 1-65. Larger scale proteolysis was performed and analyzed with N-terminal sequencing and mass spectrometry to determine the boundaries of this putative core. The predominant peaks in the digested UL11 mass spectra represent species encompassing residues 15-53, and 78-96-StII of UL11-StII. The apparent ~15-kD full length band, the prominent ~12 kDa core band, and <10 kDa smear were subjected to N-terminal sequencing and identified as beginning at residues 2, 15, and 73, respectively. Additionally, each of these bands had a secondary read assignment associated with a species beginning at residue 78. In combination, these results suggest that residues 15-53 form a proteolytically-resistant, structured core, in agreement with predictions, while the residues 78-96 also resist tryptic digest.

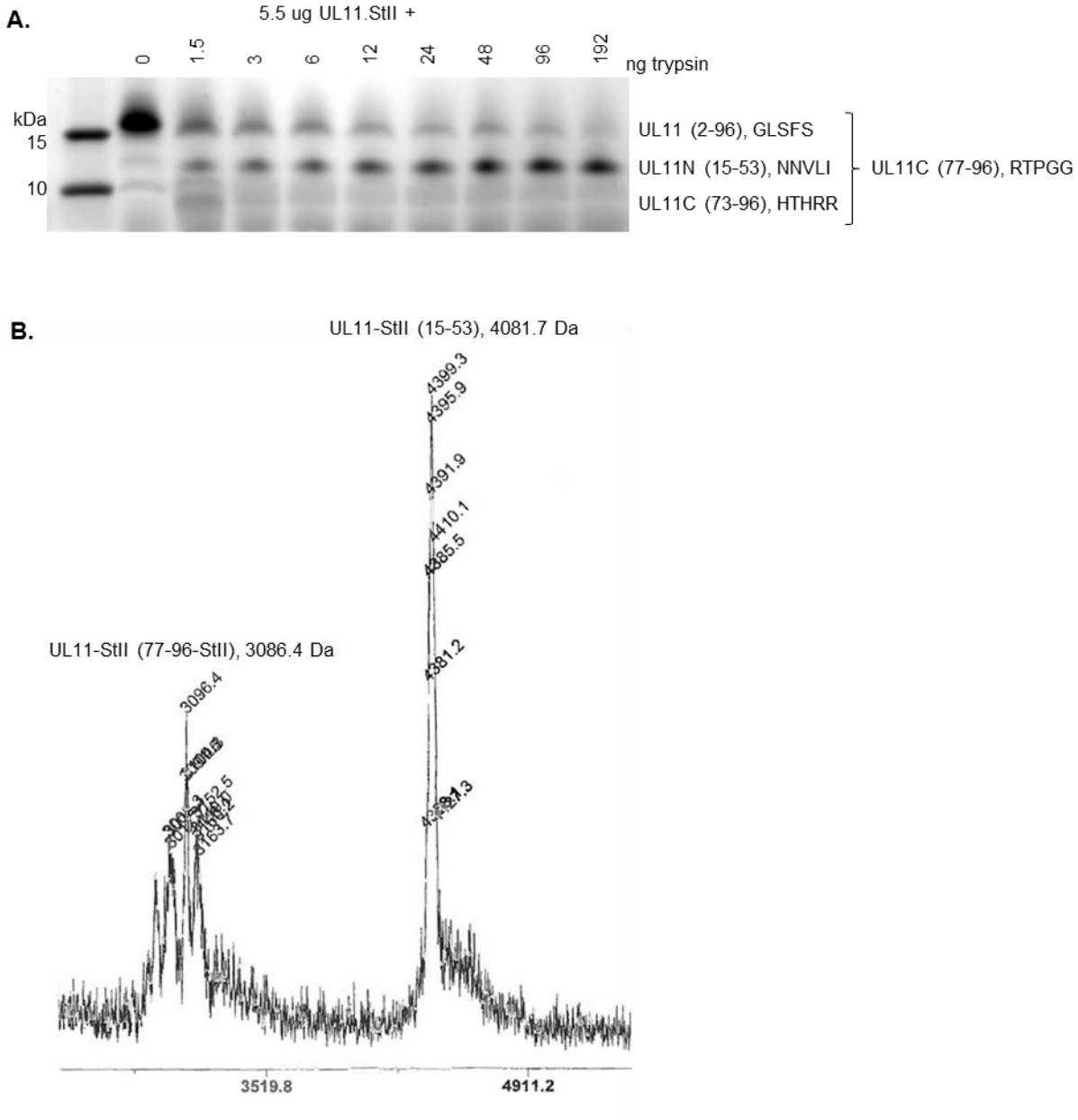


Figure 3-3. UL11 has a protected core

A) Tryptic digest of UL11-StII reveals the presence of a major protease-resistant core at an apparent molecular weight of 12 kDa that begins with residues NNLVI via N-terminal sequencing. N-terminal sequencing also identified species beginning with HTHRR and RTPGG. B) Mass spectrum of digested UL11-StII identifies major protected species of UL11 as residues 15-53 and 77-96.

3.2.4 UL11 is an extended, flexible molecule

To further characterize the overall structure and flexibility of UL11 in solution, we performed small angle x-ray scattering coupled with size exclusion chromatography (SEC-SAXS) on UL11-StII. Lysozyme – a folded, compact, globular molecule – was used as a standard. Whereas the 14.3-kDa lysozyme had radius of gyration (R_g) of 14 Å and a maximum dimension (D_{max}) of 41 Å, the 11.6-kDa UL11-StII had Rg of 25 Å and Dmax of 82 Å (Figure 3-4A-B), which suggested that UL11 adopted an extended rather than globular conformation. The p(r) function derived from a SAXS curve describes the distribution of all pairwise distances between atoms of the molecule in solution. In agreement with the dimensional characteristics, the p(r) function from UL11-StII SAXS data trailed off to large distances, a characteristic of an elongated molecule as opposed to the symmetric parabola expected of a globular molecule (Figure 3-4B).

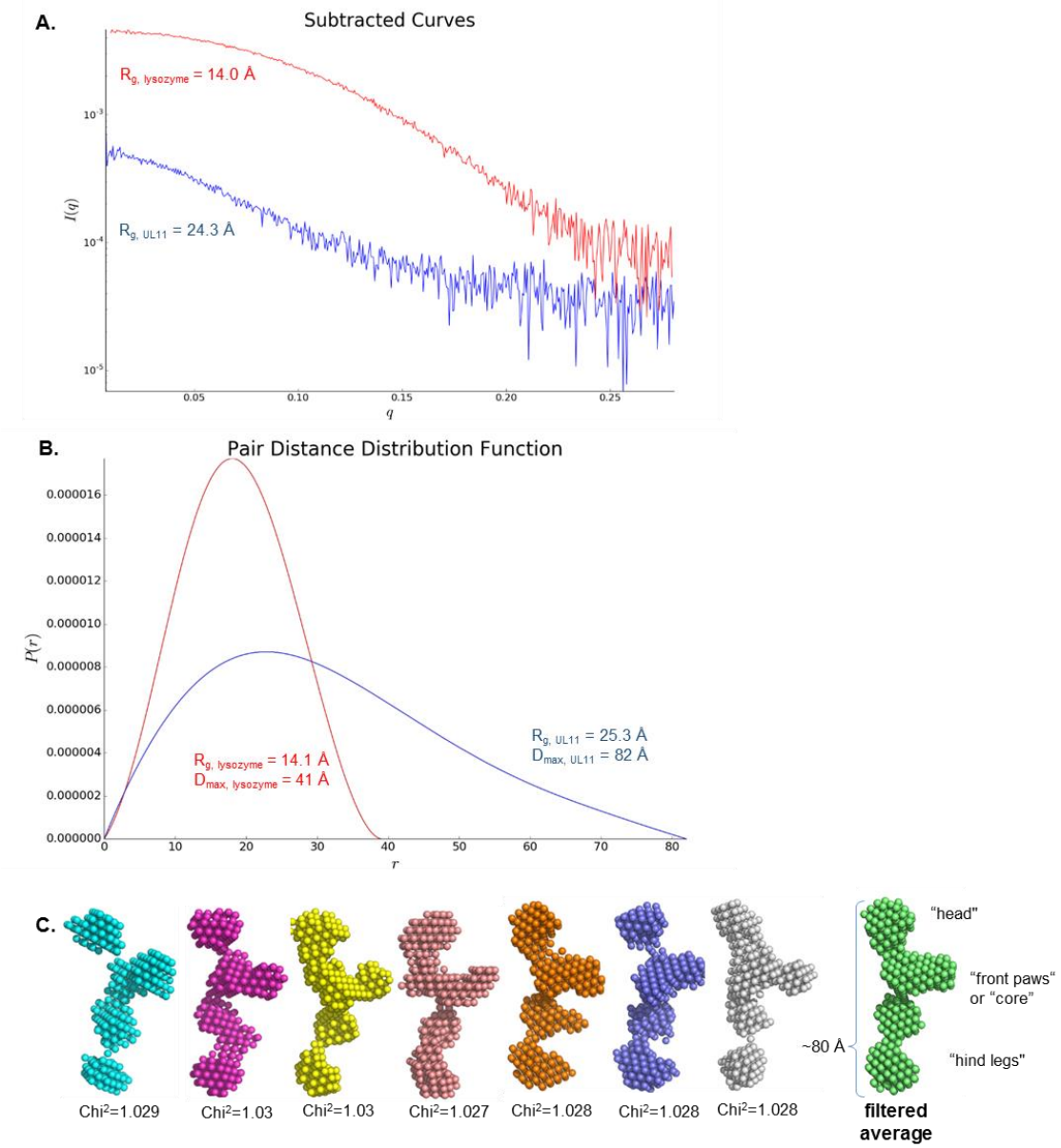


Figure 3-4. UL11 can be an extended molecule
A) Subtracted SAXS curves for UL11-StII and example data for globular lysozyme. Comparing the Guinier approximations of Rg between these molecules of similar molecular weight suggests that UL11 adopts a more extended conformation. B) Pair distance distribution functions [p(r)] for UL11-StII and lysozyme with corresponding Rg and D_{max}. The non-symmetric, trailing curve and corresponding dimension values for UL11-StII compared to the parabolic curve of lysozyme also indicate an elongated conformation of UL11. C) Seven rigid body models of UL11-StII constructed from the UL11-StII p(r) in DAMMIN with respective chi^2 values and averaged model show low-resolution structural characteristics of UL11, including and overall extended conformation comprising a small "head", a structured "front paws" or "core", and elongated "tail".

Ab initio bead modeling of the molecule represented by the p(r) function yielded an averaged bead model of UL11-StII resembling a dancing circus poodle. Although the predicted secondary structure elements of UL11 cannot be assigned to the specific features of the averaged bead model with certainty, it is tempting to speculate that the N-terminal extension of UL11 forms the flexible "head", the structured core forms the "front paws", and the unstructured C-terminal tail in its most extended conformation forms the "hind legs" (Figure 3-4C).

The Kratky plot derived from the UL11-StII SAXS curve (Figure 3-5A) shows a plateau at large values of q rather than the bell shape associated with a globular body. This shape of the Kratky plot is characteristic of an unfolded or disordered protein, so in addition to being elongated, this characteristic of the SAXS data suggests that full-length UL11 is also flexible, which is consistent with electrophoretic mobility, disorder prediction, and observed liquid phase separation described above. To further describe the flexibility of UL11, we analyzed the SAXS data using Ensemble Optimization Modeling (EOM). This is a method that models what species make up a heterologous SAXS sample mixture by determining the combination (or "ensemble") of molecules needed to make a theoretical SAXS curve that matches the experimental data and then describing the distribution of their radii (Rg) and maximal dimensions (Dmax) as compared to a random pool of molecules. Comparing the entropy levels of the random pool with the optimized ensemble also describes the flexibility of the sample. In the case of UL11-StII, there was no appreciable decrease in the entropy levels between the random pool (87.76) and optimized ensemble (86.28) (Table 3-2), suggesting that the optimized ensemble and starting pool are equally random and that UL11 is flexible.

Table 3-2. Dimensions of UL11 in solution

	R_g	D_{max}	Curves (out of 14)
compact	20.86	67.83	1
	21.16	59.76	6
average	**21.1**	**60.8**	**50%**
intermediate	26.23	80.5	2
	29.33	81.32	3
average	**28.1**	**81.0**	**36%**
extended	34.37	118.57	2 (14%)

Dimensional characteristics of the final optimized pool show that UL11-StII in solution has three conformations: one narrow peak representing 50% of the ensemble at an average of 21.1 Å, a wider peak representing 36% of the ensemble at an average of 28.1 Å, and small peak associated with 14% of the ensemble at 34.4 Å (Figure 3-5B, Table 3-2). The ensemble similarly shows three Dmax bins—50% at an average of 60.8 Å, 36% at 81.0 Å, and 14% at 118.6 Å (Figure 5C, Table 3-2). We anticipate that the smallest dimensioned population is the most compact version of UL11, potentially with the tail in contact with the core. Indeed this value is closer to the predicted hydrodynamic radius of UL11-StII (18.3 Å) if it were folded and globular [126]. We interpret the intermediate bin to correlate with a more extended conformation of UL11 as the Rg of 28.1 is approaching the predicted Rg for UL11 if it were an IDP (approximately 28 Å) [127]. Finally, the most extended peak could represent residual dimer as UL11 has shown some proclivity to multimerize (above), but it could alternately represent a mostly unfolded molecule lacking any secondary structure as the Rg value of 34.4 approaches the lower boundary for predicted Rg of a chemically denatured protein of the same size (38.9 Å) [127]. The results above have shown that UL11 is a multimodal protein with a

structured core and a disordered tail that may associate with the core or transiently become structured. While the rigid bead modeling of UL11 described in Figure 3-4 seems to contradict this conclusion that UL11 is a flexible molecule that takes on multiple conformations, it has been suggested that bead modeling of a flexible system selects a real conformation from the mixture that is elongated without contacts between globular domains [128].

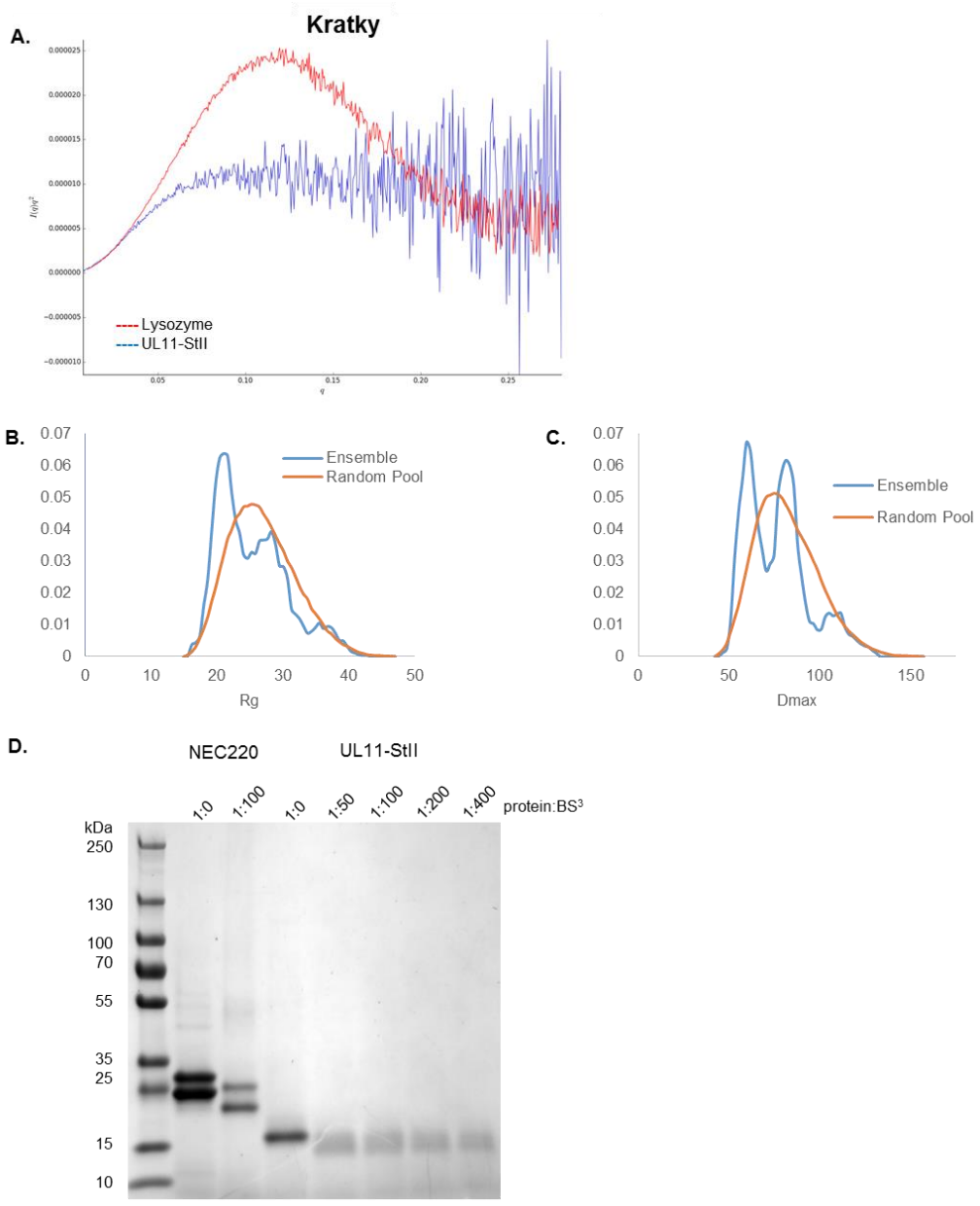


Figure 3-5. UL11 is flexible with compact and extended conformations
A) Compared to the parabola in the Kratky plot of globular lysozyme, the Kratky plot of UL11-StII plateaus at large values of q, indicating flexibility of this molecule. B) Histogram comparing the distribution of Rg values between the random pool and optimized ensemble of structures from analyzing the UL11-StII SAXS curve using EOM indicates the presence of three conformations. C) Histogram comparing the distribution of Dmax values between the random pool and optimized ensemble of structures from analyzing the UL11-StII SAXS curve using EOM also shows three conformations. D) Incubating UL11-StII with a crosslinking reagent results in an increased electrophoretic mobility, suggesting an intramolecular interaction. NEC220, comprising UL31 and UL34, was used as a positive control.

3.2.5 UL11 tail crosslinks to the structured core *in vitro*

To test the hypothesis that the putative disordered tail of UL11 interacts with the structured core of the molecule, UL11-StII was incubated with BS^3, a homobifunctional crosslinking reagent that reacts with primary amines, including the N-terminus and lysine side chains of a protein molecule. A crosslinked polypeptide would be expected to migrate faster than a fully denatured polypeptide due to its inability to unfold completely in the presence of SDS. In the presence of increasing amounts of crosslinker, UL11-StII appears smaller on SDS-PAGE (Figure 3-5D). If the UL11 tail interacts with the structured UL11 core, we would expect intramolecular crosslinking to occur, so this result further supports a model in which the UL11 tail interacts with the core of the molecule, particularly because the only reactive amines in this molecule are the N-terminus and a single lysine residue in the C-terminal StII tag (Figure 3-2C).

3.2.6 Full length UL11 does not interact with membranes *in vitro* without acyl tags

In vivo, UL11 associates with cytosolic membranes [50, 120, 121]. Deletion or mutation of the lipidation sites abrogates membrane binding suggesting that membrane binding is mediated solely by lipid anchors. However, these mutated variants also become destabilized and are not detectable in infected cells [52, 58, 76, 116] raising the possibility that lack of membrane interaction in the absence of lipid anchors may be due to misfolding caused by mutagenesis. To test membrane interactions of unlipidated WT UL11, we incubated purified UL11-StII and GPLGS-UL11 with synthetic multilamellar vesicles prepared using various lipid compositions and assessed binding by a cosedimentation assay. Purified soluble nuclear egress complex known to bind acidic liposomes [38] was used as a positive control. As shown in Figure 3-6, UL11 does not

bind synthetic liposomes *in vitro* (Figure 3-6), and we conclude that lipid anchors are required for membrane binding, in accordance with *in vivo* studies.

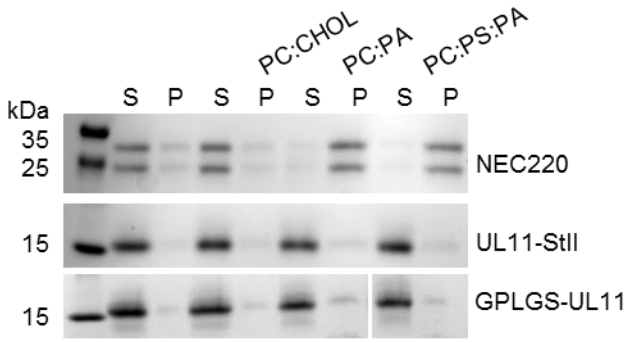


Figure 3-6. UL11 does not bind lipids without post-translational modification
Cosedimentation analysis of UL11-StII and GPLGS-UL11 with MLVs of varying composition suggest that recombinant UL11 stays in the supernatant (S) and does not coprecipitate (P) with and therefore does not interact with lipids. In contrast, positive control NEC220 shifts dramatically to the precipitate in the presence of acidic lipids.

3.2.7 UL11 copurifies with ribosomal RNA from *E. coli*

Upon initial purification of UL11-StII, protein eluting from StrepTactin affinity resin appeared pure by SDS-PAGE gel; however, the 260/280 absorbance ratio was 1.9 indicative of the sample containing approximately 60% nucleic acids (NAs) [129]. To separate UL11 from co-purifying NAs, we added a heparin affinity purification step, which reduced the 260/280 absorbance ratio to 0.8, which corresponds to 2% NA content (Figure 3-7A). UL11 has not yet been reported to bind NAs, but given the high NA content of UL11 samples even after heparin affinity step, we asked whether UL11 interacted with NAs specifically. To determine whether UL11 binds DNA or RNA, we used nuclease digestion. After StrepTactin affinity purification, bound NAs were isolated from UL11/NA complex samples using phenol:chloroform extraction and incubated with DNase, or RNase, or left untreated. UL11-bound NAs were sensitive to RNase but not

DNase treatment (Figure 3-7C). Furthermore, the size and banding pattern of the UL11-bound NAs is consistent with ribosomal RNA. The RNA-binding ability of UL11 requires its C-terminal tail, residues 66-96, because UL11(1-65)-StII does not bind heparin resin (Figure 3-7B). We conclude that UL11 preferentially binds ribosomal RNA, likely through its flexible C-terminal tail.

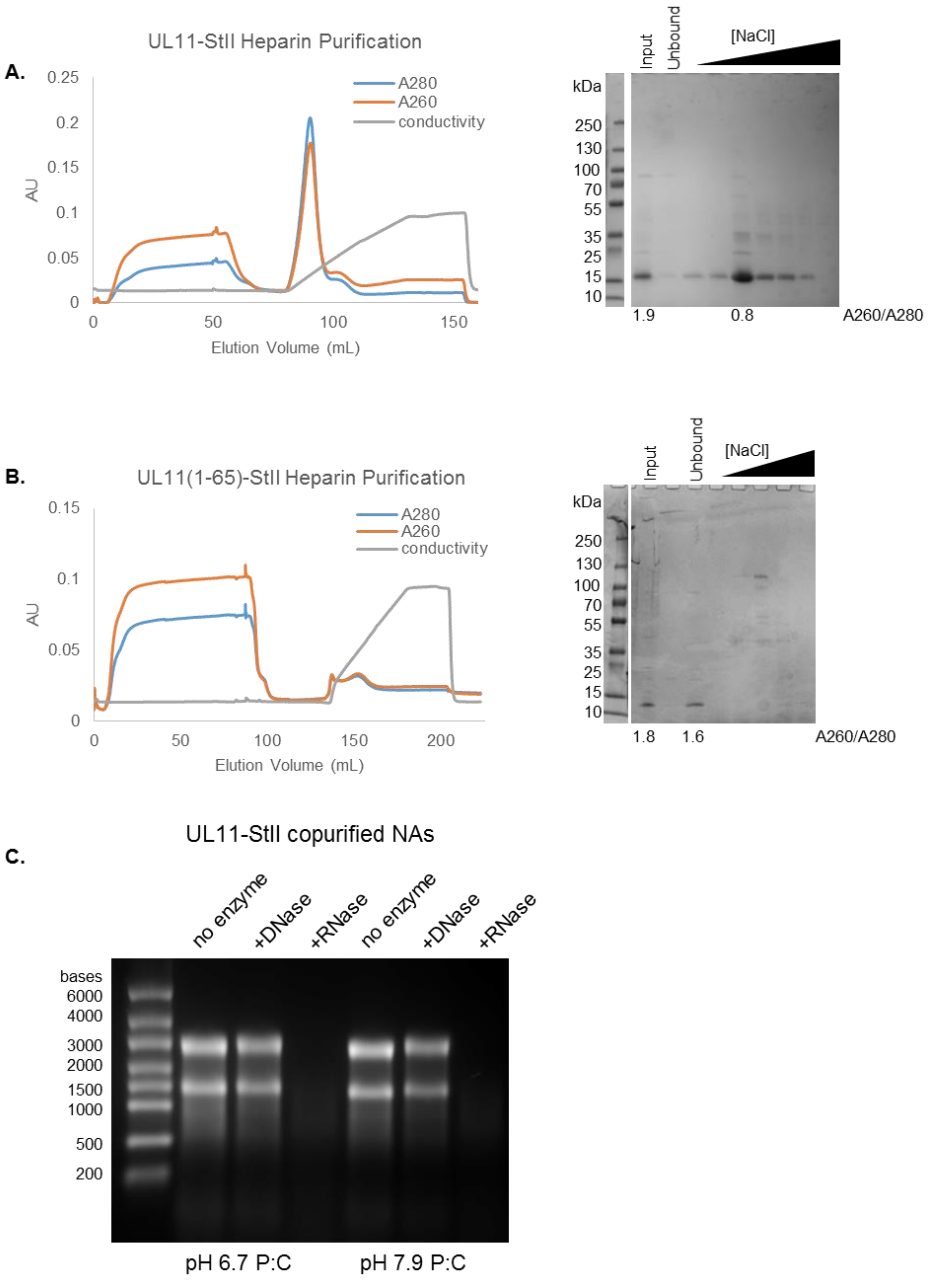


Figure 3-7. UL11 binds RNA through its C-terminal tail
A) UL11-StII is separated from copurifying nucleic acids on heparin resin. (L) Chromatogram of heparin purification shows protein is bound to heparin and eluted with a salt gradient. (R) Coomassie gel shows that bound protein includes UL11-StII and that the eluted protein is cleared of nucleic acid, as indicated by A260/A280 ratio. B) UL11(1-65)-StII does not bind heparin resin as there is no protein eluted from the column (L) and UL11 is found in the unbound fraction (R). C) Nucleic acids copurified with UL11-StII are susceptible to digestion by RNase, not DNase and show banding pattern of ribosomal RNA. Samples were resolved on a formaldehyde agarose gel and stained with ethidium bromide.

3.3 Discussion

UL11 is a conserved across all herpesviruses and has major implications on viral replication as without it, many herpesviruses cannot undergo secondary envelopment to release progeny. Furthermore, it likely has a role in cell spread and potentially unknown roles in the nucleus. We set out to characterize its basic biochemical and structural properties in order to better understand its many roles in the viral replication cycle and found that HSV-1 UL11 is a dynamic, flexible molecule comprising an N-terminal structured core and an intrinsically disordered C-terminal region that binds ribosomal RNA. Future work will focus on careful characterization of the multiple structures UL11 makes alone and in complex with binding partners UL16, lipids, and RNA and how these multiple conformations impact tegument organization and regulatory processes during replication.

Chapter 4: UL21

4.1 Introduction

UL21 is a tegument protein that is conserved among the members of the alphaherpesvirus subfamily [53, 54], with sequence identity ranging between 27 and 84% and sequence similarity ranging between 57 and 94%. Beta- and gammaherpesviruses encode positional homologs of UL21 within their genomes (e.g., HCMV UL87 and KSHV ORF24), but in the absence of obvious sequence similarity, whether these proteins share functions is unclear. UL21 is required for efficient viral replication although the extent of this dependency is debated. In HSV-1 and HSV-2, a lack of UL21 results in the delay in production of viral mRNA, viral proteins, and virions [54, 80, 130]. In the absence of UL21, viral titers are decreased 5-10 fold in HSV-1 [61, 80] and PRV [131, 132], and it has recently been shown that a lack of UL21 results in an increase in empty capsids in the cytoplasm [61], which may be related to early reports from PRV that UL21 plays a role in DNA packaging [133, 134]. In HSV-2, UL21 appears essential for replication as nuclear egress is halted in its absence and no detectable virus is made unless the multiplicity of infection is greatly increased [54]. HSV-1 virions lacking UL21 have reduced amounts of tegument protein UL16 [39] while PRV virions lacking UL21 incorporate lower amounts of tegument proteins UL46 and UL49, and viral kinase US3 [135]. UL21 can bind capsids [55, 133] and, through its interaction with UL16 [118, 132, 136], membrane-associated UL11 and gE [45]. Thus, UL21 may promote secondary envelopment by bridging the capsid to the enveloping membrane [34, 45]. In line with this hypothesis, when cells are infected with PRV lacking both UL16 and UL21, secondary envelopment is hindered and clusters of unenveloped particles are found in the cytoplasm [132]. Together, these observations suggest that in addition to being a

component of the tegument, UL21 has a role in viral morphogenesis. Also in complex with UL16 and UL11, UL21 is thought to regulate cell-cell spread by moderating the function of HSV-1 gE[45], and it has recently been shown that HSV-1 UL21 is required for syncytia formation mediated by gB [61], so UL21 is important for dissemination of infection. Additionally, UL21 has been shown to interact with and promote the polymerization of microtubules [55]. Finally, replacing the mutated amino acids in UL21 found in the PRV vaccine strain Bartha with wild type residues rescues the defect in retrograde transit [137] and restores virulence [138], which may suggest a trafficking role for UL21.

Although these roles imply cytoplasmic localization, in both infected and transfected cells, the majority of UL21 localizes to the nucleus [45, 55, 133, 136, 139], likely to the nuclear rim [54]. Except for the observation that nuclear egress is inhibited in HSV-2 lacking UL21 [54], no potential nuclear roles of UL21 have been investigated. These characteristics suggest that UL21 may have even more distinct functions in the herpesvirus replication cycle than is currently appreciated.

A systematic exploration of the various activities of this multifunctional protein would benefit from a detailed roadmap in the form of its three-dimensional structure. Here, we report the crystal structures of the N- and C-terminal domains of UL21 and map surface regions of potential functional importance. We also describe an unanticipated ability of UL21 to bind RNA, which may hint at a yet unexplored function. These novel structures of UL21N and UL21C enable structure-guided mutagenesis and functional exploration of the multiple roles of UL21 in the replication and pathogenesis of alphaherpesviruses. In line with the observation that these domains do not have obvious

affinity for each other, we also describe the characterization of full length UL21 structure in solution as a flexible molecule.

4.2 Results

4.2.1 UL21 is composed of two stable domains.

Full-length HSV-1 strain 17 UL21 expressed in *E. coli* with an N-terminal StII tag (Figure 4-1) underwent spontaneous proteolysis by a trypsin-like protease, generating N- and C-terminal fragments (Figure 4-2B). The stability of these fragments in combination with predicted secondary structure and sequence alignment suggested that UL21 was composed of two ordered domains—a conserved N-terminal domain (sequence identity 9.7%, similarity 25%) [53] and a more variable C-terminal domain (sequence identity 1.2%, similarity 8.4% [99])—connected by a flexible, protease-sensitive linker containing a stretch of conserved sequence predicted to form a β-strand (Figure 4-1) (Figure 4-3,4). N-terminal sequencing revealed that these fragments correspond to UL21N (1-216) [53], UL21C (237-535), and UL21C (281-535) (Figure 4-1) [99].

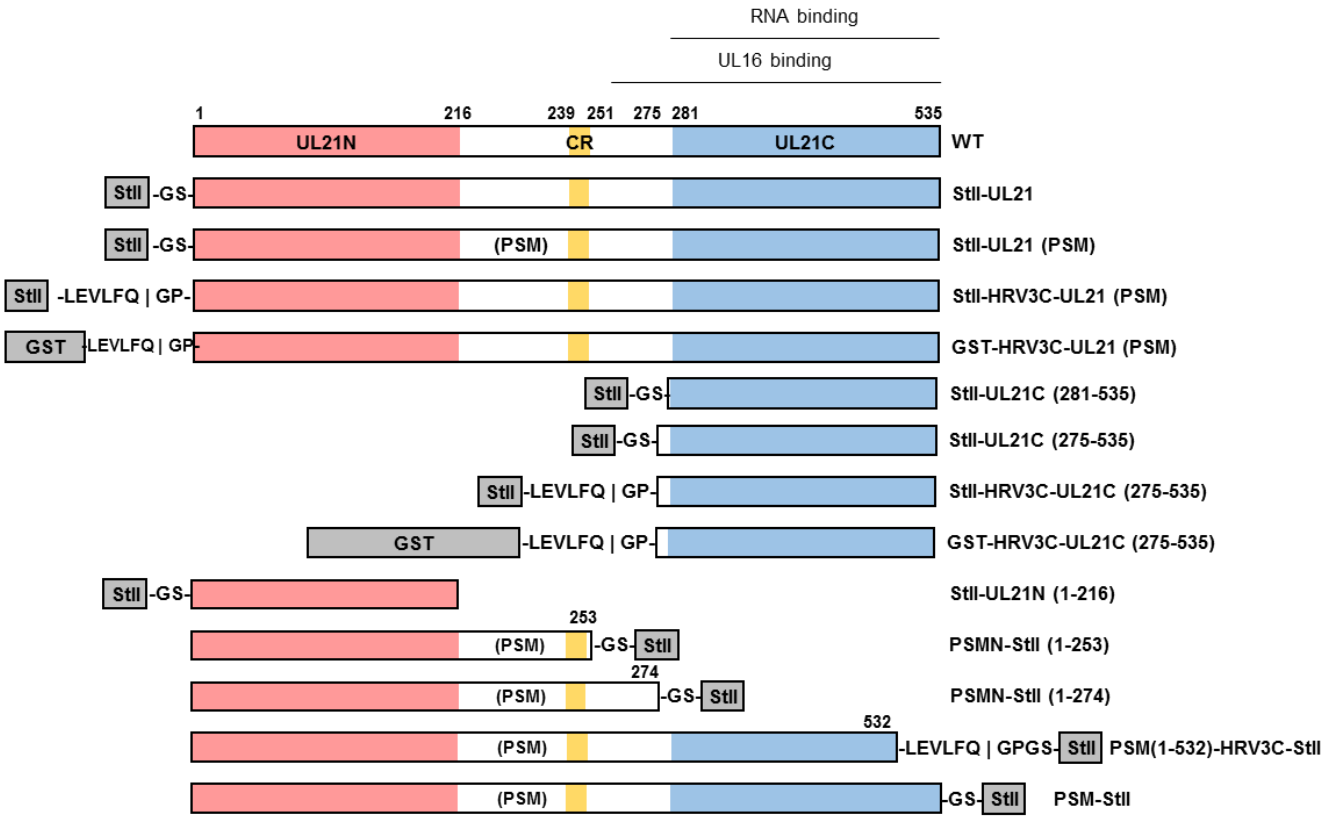


Figure 4-1. UL21 constructs
Linear diagrams of the predicted secondary structure and domains of UL21. Amino acid boundaries are indicated. WT, wild type; CR, the conserved region (residues 239 to 251) between UL21N (residues 1 to 216) and UL21C(281-535) that is predicted to form a β-strand.

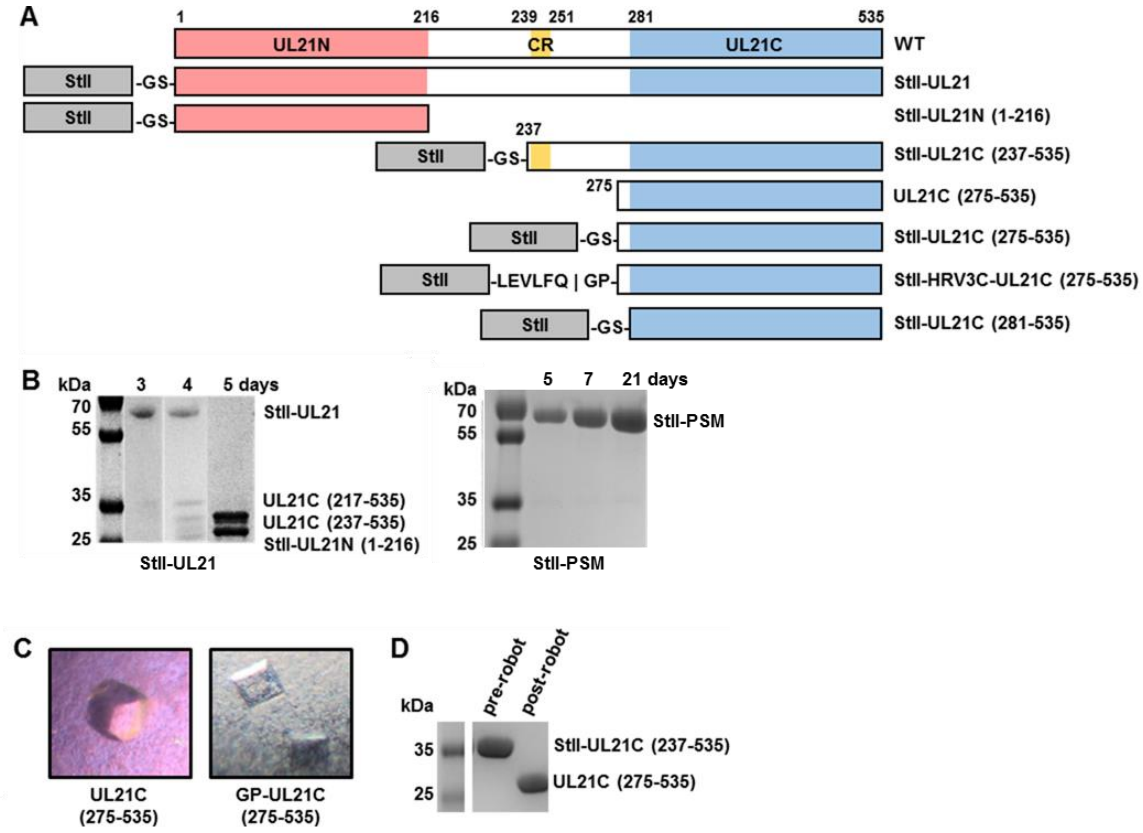


Figure 4-2. Full length UL21 is susceptible to proteolysis
(A) A portion of Figure 4-1. (B) UL21 is proteolytically cleaved during purification by contaminating proteases. (C) UL21C crystals formed by spontaneous cleavage of UL21C during initial screening of crystallization conditions (left) or by cleavage due to an inserted protease cleavage site (right). (D) UL21C was cleaved during initial screening of crystallization conditions. Adapted with permission from [99]. Changes include addition of PSM stability in B.

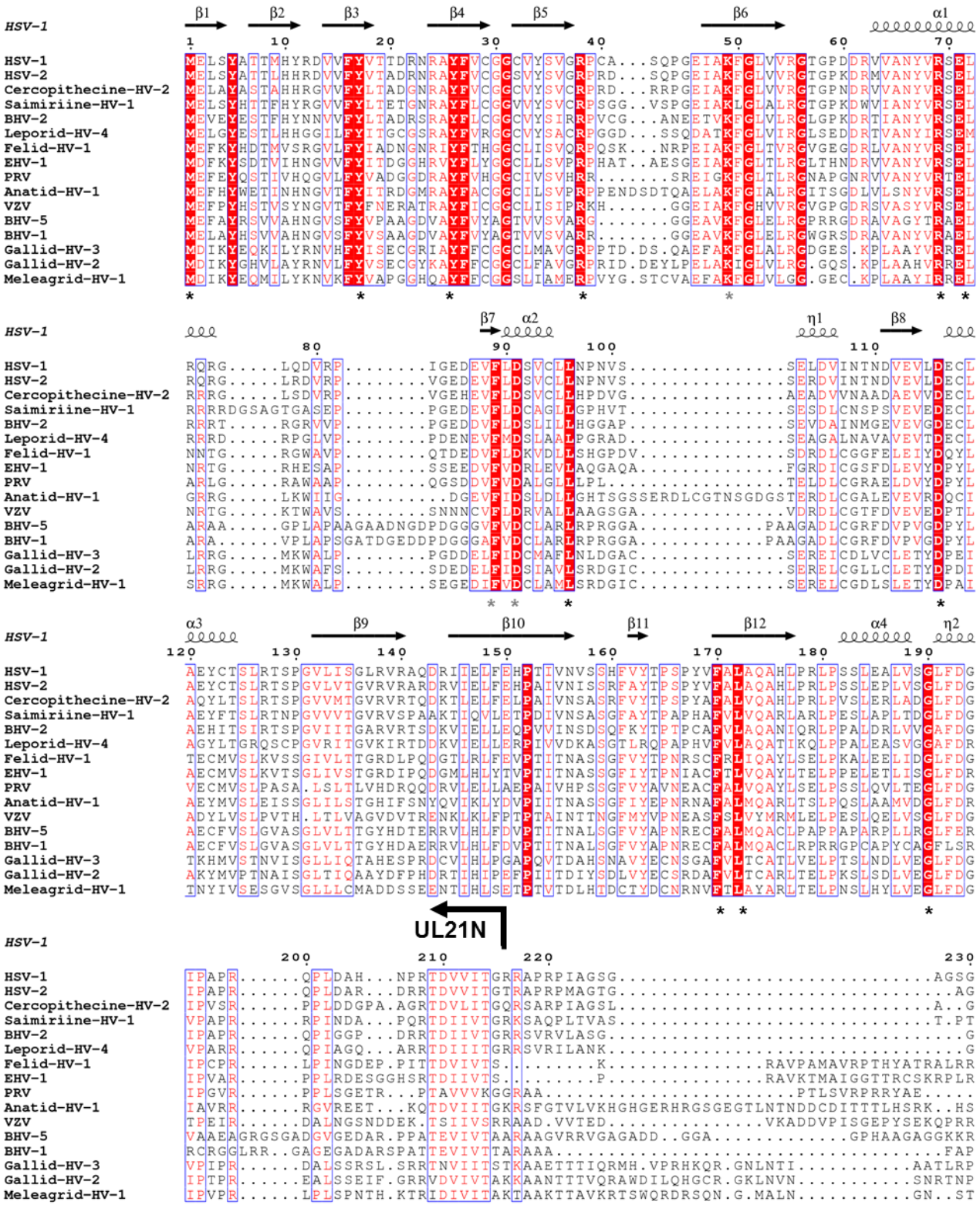

HSV-1
β1 β2 β3 β4 β5 β6 α1
HSV-1
HSV-2
Cercopithecine-HV-2
Saimiriine-HV-1
BHV-2
Leporid-HV-4
Felid-HV-1
EHV-1
PRV
Anatid-HV-1
VZV
BHV-5
BHV-1
Gallid-HV-3
Gallid-HV-2
Meleagrid-HV-1
β7 α2 η1 β8
β9 β10 β11 β12 α4 η2
α3
UL21N

Figure 4-3. Sequence conservation in UL21N
Multiple-sequence alignment of UL21 homologs from 16 alphaherpesviruses. Only the alignment of residues corresponding to residues 1 to 216 of HSV UL21N is shown. The secondary structure of HSV-1 UL21 is shown above the aligned sequences. Similar residues are shown in red text. Identical residues are boxed in red with white text, and those exposed on the surface of UL21N are marked by asterisks. Gray asterisks identify conserved residues that are surface exposed in the model but are likely obscured in the protein by the unresolved loop containing residues 76 to 87. Reprinted with permission from [53].

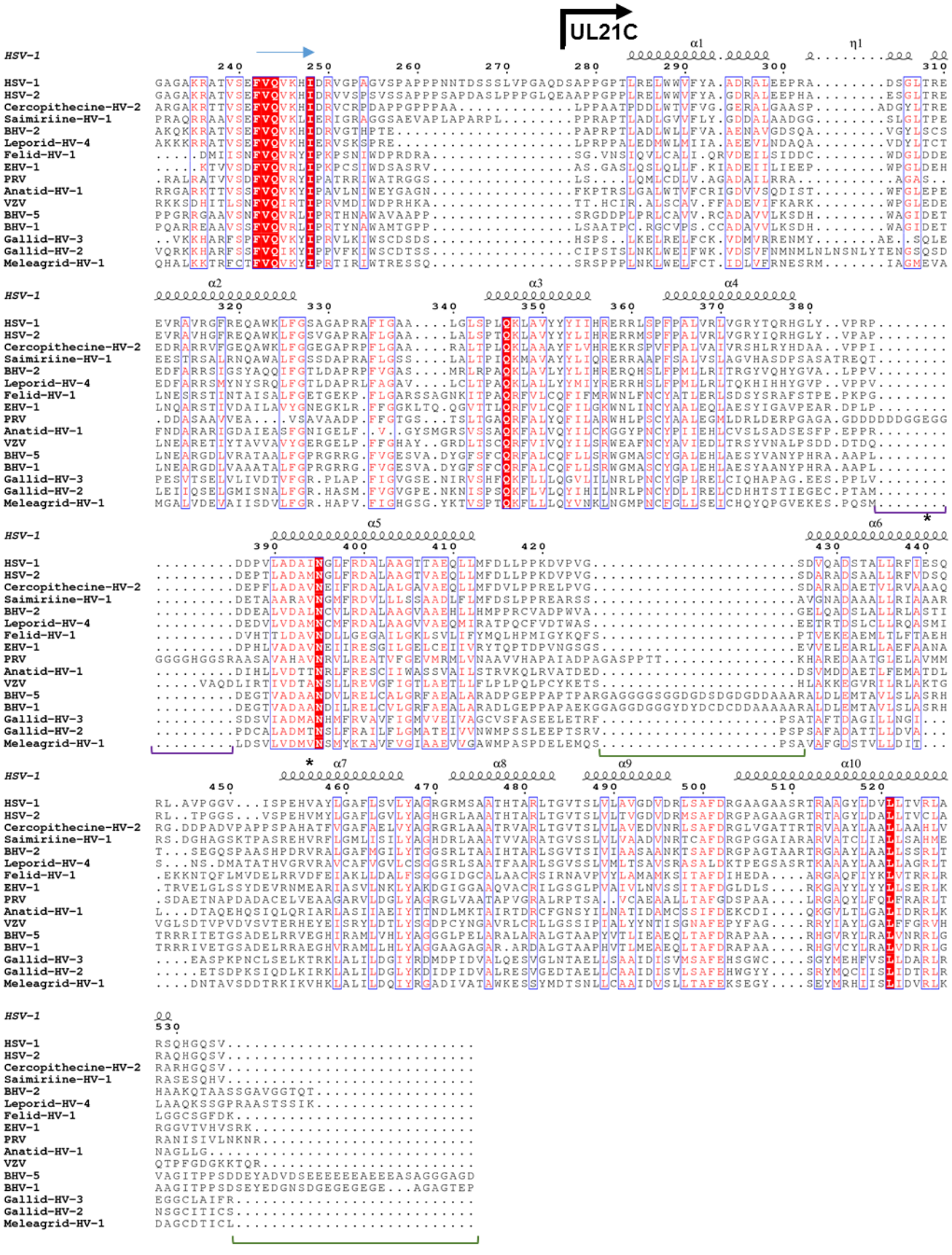

UL21C
HSV-1
HSV-2
Cercopithecine-HV-2
Saimiriine-HV-1
BHV-2
Leporid-HV-4
Felid-HV-1
EHV-1
PRV
Anatid-HV-1
VZV
BHV-5
BHV-1
Gallid-HV-3
Gallid-HV-2
Meleagrid-HV-1

Figure 4-4. Sequence conservation in UL21 linker and UL21C
Shown is a multiple-sequence alignment of UL21 homologs from 16 alphaherpesviruses. Only the alignment of residues corresponding to residues 231 to 535 of HSV-1 UL21 is shown. The secondary structures of HSV-1 UL21 and the start of UL21C are marked above the alignment. Identical residues are represented by white letters on a red background. Similar residues are represented by red letters boxed in blue. UL21C sequences share 1.2% identity and 8.4% similarity. Insertions in BHV-1, BHV-5, and PRV homologs are indicated by green or purple brackets, respectively, below the alignment. Residues E355 and V375, mutated in PRV Bartha UL21, are indicated by asterisks below the alignment. Reprinted with permission from [99].

4.2.2 Characterization of UL21 domains

4.2.2.1 UL21 domain construct design and UL21N crystallization.

UL21N (1-216), UL21C (237-535), and UL21C (281-535) (Figure 4-1) were subcloned with N-terminal StII tags, expressed and purified. The expression, purification, and concentration of StII-UL21N (1-216) was straightforward and this construct could be concentrated upwards of 20 mg/ml and formed football-shaped crystals around 5 mg/ml in ~1M ammonium sulfate and 0.1M HEPES pH 7.5. These crystals were optimized with the inclusion of ~10% MPD (2-methyl-2,4-pentanediol) and cryopreserved by flash freezing in 15% MPD. Crystals took space group $P6_322$ and diffracted to ~2 Å. A mercury derivative of these crystals was generated by soaking native crystals in thimerosal and was used to solve the structure of UL21N by single-wavelength anomalous dispersion (SAD). There is one UL21N molecule in the asymmetric unit. The final structure was refined against a 2.05-Å native data set, contains residues 1-198 of UL21, the N-terminal StrepII affinity tag, and the Gly-Ser linker. Residues 38-46, 76-87, and 199-216 were not resolved despite being present in the crystals. This model has R_{work} = 16.55% and R_{free} = 22.14% (Table 4-1).

Table 4-1. Data collection and refinement statistics (UL21N)

	Native	**Thimerosal**
Data collection[a]		
Space group	$P6_322$	$P6_322$
Cell dimensions		
a, *b*, *c* (Å)	108.35, 108.35, 65.24	109.48, 109.48, 65.29
α, β, γ (°)	90, 90, 120	90, 90, 120
Resolution (Å)	41.68-2.05 (2.09-2.05)	47.41-2.80 (2.90-2.80)
R_{sym} or R_{merge}	0.061 (0.475)	0.111 (0.651)
$I/\sigma I$	23.26 (2.52)	23.96 (5.78)
Completeness (%)	99.78 (99.58)	100.00 (100.00)
Redundancy	5.3 (5.5)	18.4 (19.1)
Refinement		
Resolution (Å)	30.81-2.05	
No. reflections (free)	14225 (713)	
R_{work} / R_{free}	16.55/22.14	
No. atoms		
Protein	1643	
Water	66	
B-factors		
Protein	51.80	
Water	49.40	
R.m.s. deviations		
Bond lengths (Å)	0.007	
Bond angles (°)	1.08	
Ramachandran plot[b]		
Favored (%)	98.49	
Allowed (%)	1.51	
Outliers (%)	0.0	

[a]Values in parentheses are for highest-resolution shell.

[b]As determined using Molprobity [140].

4.2.2.2 UL21C crystallization.

Although StII-UL21C (281-535) was insoluble above 1 mg/ml, StII-UL21C (237-535) could be concentrated to 4 mg/ml and formed barrel-shaped crystals in 100 mM sodium acetate pH 4.2, 1.2 M ammonium phosphate (Figure 4-2C). However, on a Coomassie-stained gel, both the crystallized protein and the protein that had been exposed to the dispensing nozzle of the crystallization robot migrated as a shorter species than the originally purified protein stock (Figure 4-2C), indicating that StII-UL21C (237-535) was cleaved upon exposure to the protease-contaminated dispensing nozzle. N-terminal sequencing of the protein stock revealed the crystallized species to be UL21C (275-535). UL21C (275-535) was subcloned with N-terminal StII tag, but the resulting StII-UL21C (275-535), like StII-UL21C (281-535), was insoluble above 1 mg/ml. To mitigate poor protein solubility, presumably caused by suboptimal tag placement, an HRV3C (PreScission) protease cleavage site was inserted between the StII tag and residue 275 to generate construct StII-HRV3C-UL21C (275-535). This construct was expressed, purified, separated from the StII tag, concentrated to ~3 mg/ml, and found to form square prisms in 100 mM sodium citrate pH 3.5, 19% w/v PEG 3350 (Figure 4-2C). Crystals took space group $P4_32_12$ and diffracted to 2.7 Å. A selenomethionine derivative of StII-HRV3C-UL21C (275-535) was prepared, crystallized under native conditions, and used to determine the structure of UL21C by single-wavelength anomalous dispersion (SAD). There is a single copy of UL21C in the asymmetric unit. The final

structure, refined against a 2.7-Å native data set, contains residues 279-530 and has R_{work} = 21.77% and R_{free} = 23.43% (Table 4-2).

Table 4-2. Data collection and refinement statistics (UL21C)

	Native	**Selenomethionine**
Data collection[a]		
Space group	$P4_32_12$	$P4_32_12$
Cell dimensions		
a, *b*, *c* (Å)	54.27, 54.27, 180.76	56.06, 56.06, 185.27
α, β, γ (°)	90, 90, 90	90, 90, 90
Resolution (Å)	54.27-2.72 (2.85-2.72)	185.27-2.88 (3.03-2.88)
R_{sym} or R_{merge}	0.066 (0.864)	0.181 (2.481)
$I/\sigma I$	20.5 (2.4)	28.3 (1.9)
Completeness (%)	99.7 (99.7)	100.00 (100.00)
Redundancy	6.7 (7.2)	50.6 (39.8)
Refinement		
Resolution (Å)	51.98-2.72	
No. reflections (free)	7813 (360)	
R_{work} / R_{free}[b]	0.2177/0.2343	
No. atoms		
Protein	1926	
Water	17	
B-factors		
Protein	79.65	
Water	48.89	
RMS[c] deviations		
Bond lengths (Å)	0.003	
Bond angles (°)	0.685	
Ramachandran plot[d]		

Favored (%)	97.6
Allowed (%)	2.4
Outliers (%)	0.0

[a]Values in parentheses are for highest-resolution shell.

[b]R_{work} and R_{free} are defined as $\sum||F_{obs}|-|F_{calc}||/\sum|F_{obs}|$ for the reflections in the working or the test set, respectively.

[c]RMS, root mean square.

[d]As determined using Molprobity (molprobity.biochem.duke.edu) [140].

4.2.3 Structural characterization of UL21N

4.2.3.1 UL21N has a unique fold.

The model of UL21N contains residues 1-198 of UL21, the N-terminal StrepII affinity tag, and the Gly-Ser linker. Residues 38-46, 76-87, and 199-216 were not resolved despite being present in the crystals. Ser0 is the first residue of the β1 strand, and residues QEF of the strep-tag form a 3_{10} helix. UL21N adopts a single-domain structure of an unusual α/β fold that resembles a wind-filled sail with dimensions of 50X30X30Å. A DALI search [141] identified human kinase PLK1 as most likely to have a similar fold to UL21N; however, the similarity score was not much higher than that of a spurious match, and overlaying structures revealed no appreciable overall structural similarity this or any other proteins. UL21N is composed of two clearly defined "halves", segregated by secondary structure: the oblong β bouquet composed of three antiparallel β sheets (β1-β14) and the α crescent composed of 4 α-helices (α1-α4) and two 3_{10} helices (η1-η2) that are arranged along one face of the molecule (Figure 4-5). Short 3_{10} helix η2 and alpha helix α4 look like one long helix broken by completely conserved glycine 190 (Figure 4-3). The β bouquet consists of 3 antiparallel β sheets oriented at an angle to the longest axis of UL21N creating the appearance of a bouquet arranged with the following

topology: an inner seven-stranded β sheet (β2-β3-β4-β5-(β7)-β6-β14), a three-stranded upper outer β sheet (β1-β11-β10), and a four-stranded lower outer β sheet (β8-β9-β12-β13) (Figure 4-5). Strand β5 is shorter than the neighboring strand β6, and is effectively extended by the strand β7 such that both β5 and β7 form hydrogen bonds along the same margin of strand β6 (Figure 4-5A,C).

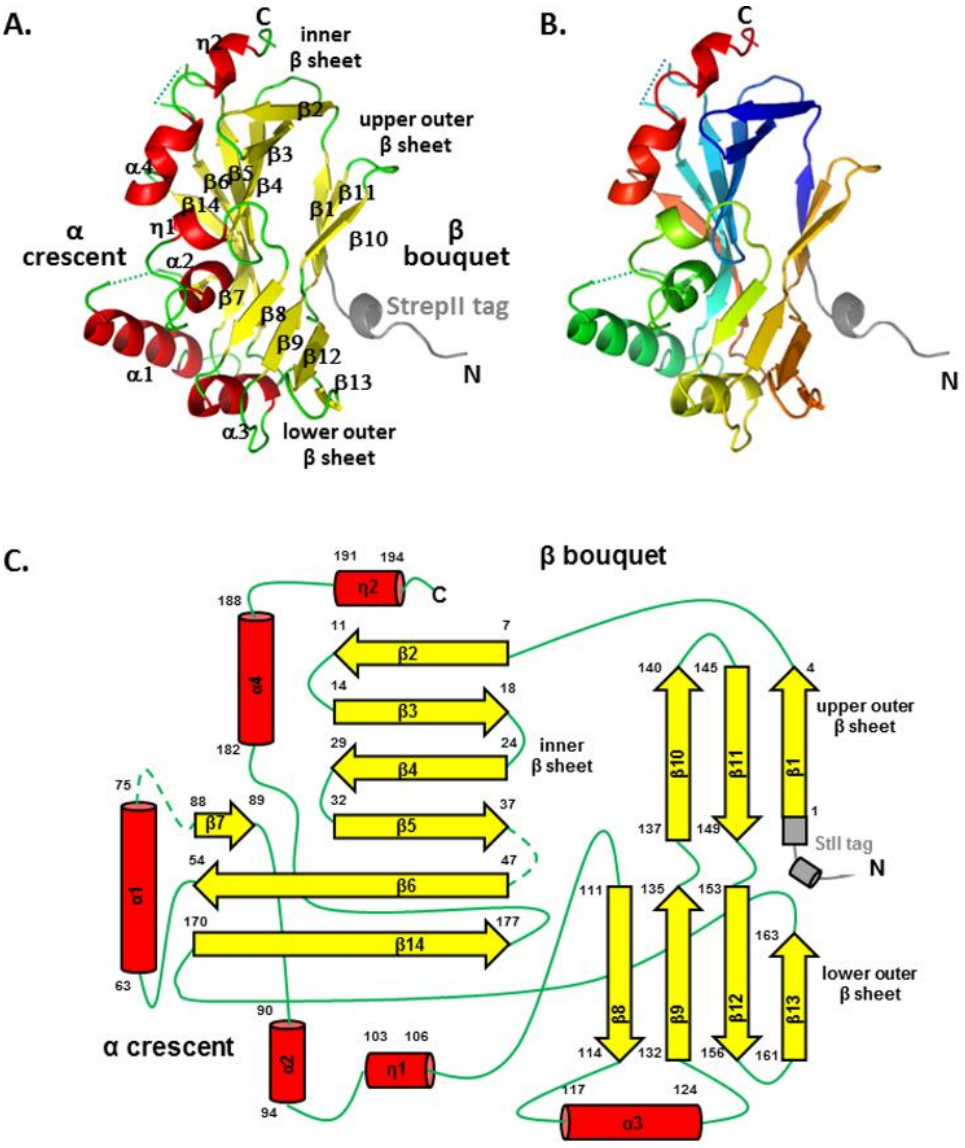


Figure 4-5. UL21N structure
The crystal structure of UL21N is shown in color based on secondary structure (A) and in rainbow coloring (B), from blue (N terminus) to red (C terminus). The StrepII tag is shown in gray. Residues unresolved in the structure are shown as dashed lines. (C) Topology diagram, colored as described for panel A. Secondary structure elements are numbered sequentially, and their amino acid boundaries are given. Helices are show as cylinders, strands as arrows, loops as solid lines, and unresolved loops as dashed lines. All structures were made using Pymol (**http://www.pymol.org**). Reprinted with permission from [53].

4.2.3.2 UL21N is conserved amongst homologs.

UL21 has 5.2% identical residues among 16 alphaherpesviruses, and 21 of 28 identical residues are within UL21N (Figure 4-3,4). 12 are surface exposed (Figure 4-6)

but do not form obvious clusters, which prevented clear assignment of potentially important sites on the surface of UL21N from sequence identity alone. Evolutionary trace analysis (ETA) [142] was performed on the same sequence alignment. ETA generates a phylogenetic tree from a sequence alignment of homologous proteins. At each partition, or branch point, closely related sequences are grouped into classes, trace residues are defined as conserved within specific classes, and these residues are assigned importance scores based on the partition at which they appear. Clustering of important trace residues on the protein surface may indicate regions of functional importance [142]. This method has been used to detect functional sites in a number of proteins [143, 144], including UL37 from pseudorabies virus, another alpha-herpesvirus [74]. ETA on UL21N revealed several surface clusters of trace residues of increasing importance (Figure 4-6B).

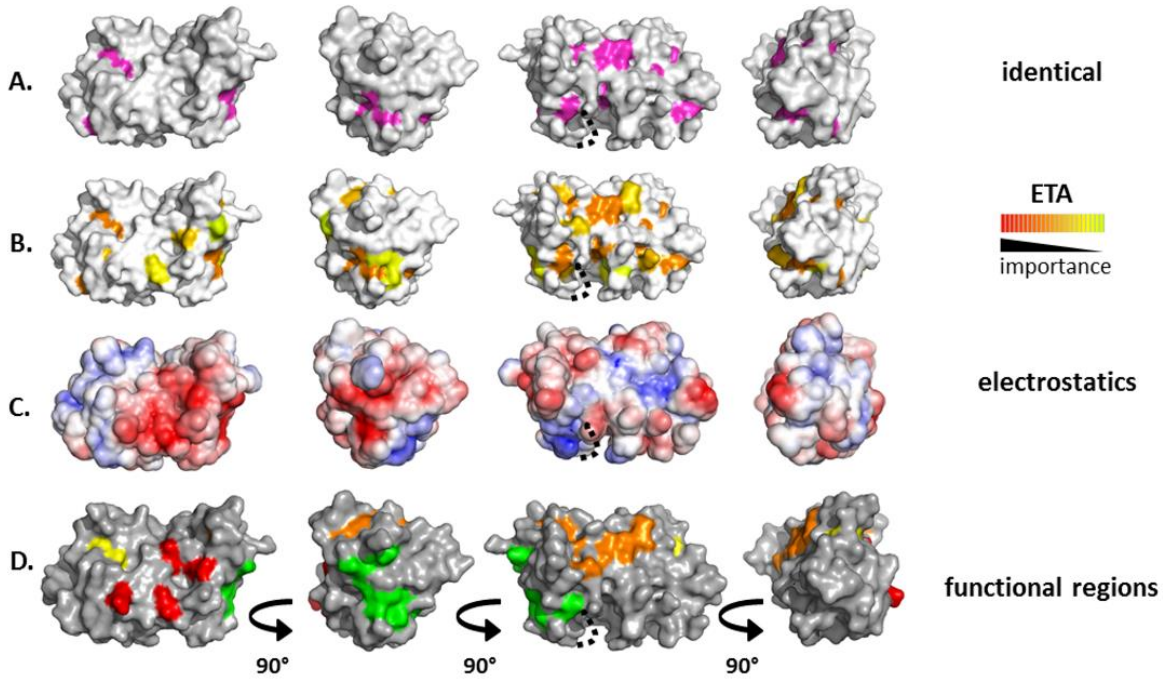


Figure 4-6. Analysis of conservation and charge on the surface of UL21N
A dotted line marks a cavity likely obscured by unresolved residues 76 to 87. (A) Completely conserved residues on the surface of UL21N are shown in magenta. Four orientations based on 90° rotations around the vertical axis are shown. (B) Class-specific residues identified by universal evolutionary trace analysis (**http://mammoth.bcm.tmc.edu/uet/**) are highlighted on the surface of UL21N. The 25% of residues with the highest importance scores are shown. (C) An electrostatic surface potential map of UL21N was generated using the PBEQ Solver function in the Charmm program (**http://www.charmm-gui.org/?doc=input/pbeqsolver**). (D) Potential functional regions assigned on the surface of UL21N. Potential functional regions are composed of the following residues: region 1 (red), D13, D105, D111, and E113; region 2 (orange), M1, E2, R55, N156, Y163, P165, F170, and L172; region 3 (yellow), Y5, Y17; region 4 (green), Y67, R69, S70, E71, D116, E117, and E121. Reprinted with permission from [53].

4.2.3.3 Potential functional regions in UL21N.

Analysis of the electrostatic potential on the surface of UL21N revealed two large negatively charged patches, consistent with its calculated isoelectric point of 5.2, one of which coincides with a shallow depression and the other wraps around the side (Figure 4-6C). A number of the residues that contribute to these charges were also identified as

important trace residues in ETA. In combination, conservation and charge patterns pinpoint several surface regions of potential importance (Figure 4-6D) for functions possibly including but not limited to interacting with the remaining C-terminal residues of UL21 or known binding partner UL16. Region 1 (D13, D105, D111, E113) sits on one flat face of UL21N and is comprised of four aspartate and glutamate residues, two of which are also identified in ETA. Region 2 (M1, E2, R55, N156, Y163, P165, F170, L172) extends to three sides of the protein and contains identical and trace residues. Region 3 (Y5, Y17) is two conserved, surface-accessible tyrosine residues, and large Region 4 (Y67, R69, S70, E71, D116, E117, E121) wraps around one short side and contains residues identified in electrostatic and evolutionary analyses. These regions provide a more educated starting point for mutational analysis in the context of protein biochemistry and viral infection.

4.2.4 Structural characterization of UL21C

4.2.4.1 UL21C has a unique fold.

UL21C is composed of ten α helices and one 3_{10} helix that are arranged into a dragonfly fold with the left and right "wings" formed by α1-α4-η1 and α5-α9, respectively, and the "body" formed by the long helix α10 (Figure 4-7A,B). Helices α1, α3, α4, and α6 are aligned approximately parallel to the helix α10 "body", while helices η1, α2, α5, α7, α8, and α9 lie parallel to each other but perpendicular to helix α10. The boundaries of the helices are mostly in agreement with the secondary structure prediction by PSIPRED [145] except that helices α3, α4, α5, α6, α7, α8, α9, and α10 are between 3 residues shorter and up to 5 residues longer than predicted. 3_{10} helix η1 was not predicted. The α-helical dragonfly fold does not resemble any known protein folds, according to the DALI structural similarity search algorithm [141]. A portion of UL21C

displays a weak structural similarity to 2/2 globins, also known as truncated hemoglobins, which are small proteins from plants and bacteria that share structural and functional similarity with the globin superfamily [146]. Portions of α helices α6, α7, and α10 in UL21C align with three α helices in the heme-binding site with a Z score of 4.7 and an RMSD of 4.5 Å over 87 aligned residues, but helix α5 in UL21C blocks the would-be heme-binding pocket, which rules out heme binding as a possible function for UL21 (Figure 4-7C). Thus, the biological significance of this limited structural similarity, if any, is unclear.

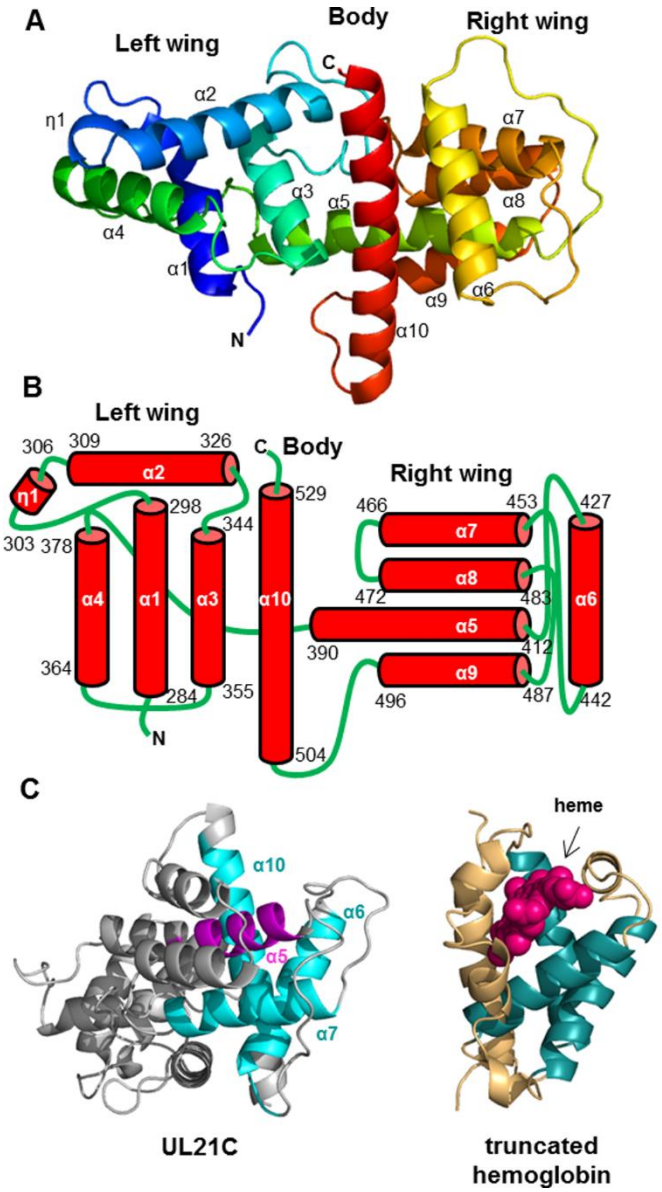


Figure 4-7. The structure of HSV-1 UL21C resembles a dragonfly
(A) Ribbon representation of the crystal structure of UL21C. Blue, N terminus; red, C terminus. Helices are numbered sequentially. (B) Diagram of the topology of UL21C. The amino acid boundaries of helices are marked. (C) UL21C is shown side by side with monomeric truncated hemoglobin (PDB accession no. 4UUR). The structures were aligned using the Dali server. Aligned residues in UL21C and truncated hemoglobin are shown in cyan and teal, respectively (RMSD, 4.5 Å over 87 aligned residues). The heme group in truncated hemoglobin is shown in magenta, and the similarly placed helix (α5) in UL21C is shown in purple. All structures were rendered in PyMOL (**www.pymol.org**). Reprinted with permission from [99].

4.2.4.2 Areas of conservation and diversity on UL21C surface.

HSV-1 UL21 has 28 residues that are identical among 16 alphaherpesviruses, but only 3 of these are within UL21C (Figure 4-8). Additionally, the sequence length of UL21C varies across species, as UL21 homologs from PRV, BHV-1, and BHV-5 contain long insertions at different locations in the sequence. Specifically, PRV UL21C contains a 17-amino-acid insertion between residues equivalent to 385 and 386 within HSV-1 UL21C, while BHV-1 and BHV-5 UL21C each contain a 23-amino-acid insertion between residues equivalent to 425 and 426 and a 28-amino-acid extension after residue 535 within HSV-1 UL21C (Figure 4-4). All three insertions are rich in glycines, aspartates, and glutamates and likely form flexible, negatively charged loops in their native structures (Figure 4-8A). Such virus-specific surface variability may reflect divergent protein functions.

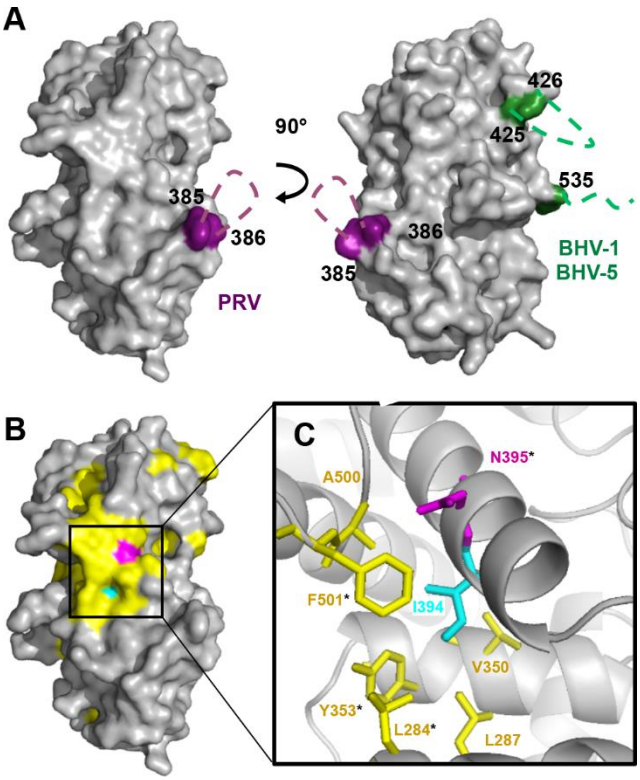


Figure 4-8. Conserved and variable regions on the surface of UL21C
(A) Surface representation of UL21C. Two orientations related by a 90° rotation along the vertical axis are shown. Predicted locations of G/D/E-rich loop insertions in PRV (purple), BHV-1, and BHV-5 (green) are indicated by dashed lines. (B) Surface representation of UL21C in the same orientation as that on the left in panel A. Conserved residue N395 is shown in magenta, and similar residues from the alignment in Figure 4-4 are shown in yellow. Residue I394, which corresponds to PRV UL21 residue V375 (mutated in the Bartha strain), is shown in cyan. (C) Close-up view of residue I394 and its van der Waals interactions within the hydrophobic core with residues L284, L287, V350, Y353, A500, and F501. Surface-accessible residues are marked with asterisks. The color scheme is the same as that in panel B. Reprinted with permission from [99].

The three identical residues in UL21C are Q346, N395, and L521 (Figure 4-4). The side chain of L521 is buried and is likely necessary for structural integrity. In contrast, the side chains of Q346 and N395 are mostly surface accessible and could be functionally important. N395 is noteworthy because it is surrounded by conserved residues, generating a “bullseye” pattern (Figure 4-8B). Three coding mutations—H37R, E355D, and V375A—have been identified in UL21 in the avirulent, spread-deficient

PRV strain Bartha [138]. When these mutations were reverted to the wild type sequence, defects in spread [137] and virulence [138] were also reversed, suggesting that these residues have functional roles. Two of these mutations, E355D and V375A, map to the C-terminal domain of UL21. While E355 in PRV UL21 is located within the G/D/E-rich insertion absent from HSV-1 UL21, V375 aligns with I394 in HSV-1 UL21. The side chain of I394 is mostly buried within the hydrophobic core underlying N395 where it engages in van-der-Waals interactions with the side chains of several conserved residues that surround N395 (Figure 4-8C). The valine-to-alanine mutation at this position in the PRV UL21 homolog would be expected to destabilize the hydrophobic core and thus "deflate" the conserved surface above it, which may disrupt binding of a potential functional partner. Thus, the conserved surface surrounding N395 could be essential for a function related to virulence.

To uncover additional conservation patterns within UL21C, we performed evolutionary trace analysis (ETA), a method of identifying functionally important residues on the basis of their resistance to mutations over evolutionary time [142], as has been done for herpesvirus proteins UL25 [147], UL37N [74], and UL21N [53]. After generating an evolutionary tree (Figure 4-9) from the 16 aligned sequences (Figure 4-4), ETA highlighted surface clusters of evolutionarily conserved residues that form a belt wrapping around the middle of UL21C (Figure 4-10C), parallel to the long central helix α10 (Figure 4-10A).

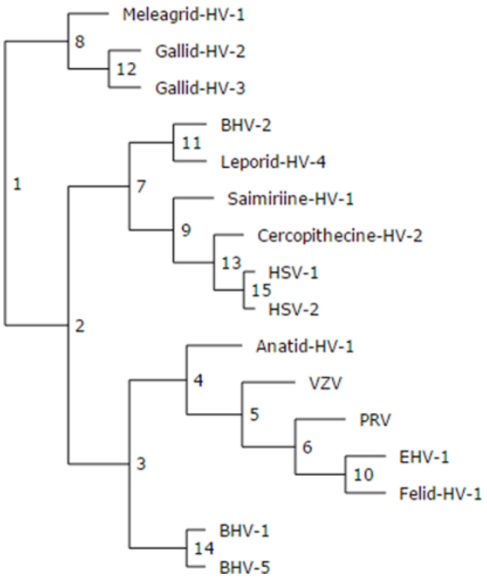


Figure 4-9. Phylogenetic tree of UL21
From universal evolutionary trace analysis (**http://mammoth.bcm.tmc.edu/uet/**) of full-length UL21 homologs from 16 alphaherpesviruses. Partition numbers are marked at each branching point. Reprinted with permission from [99].

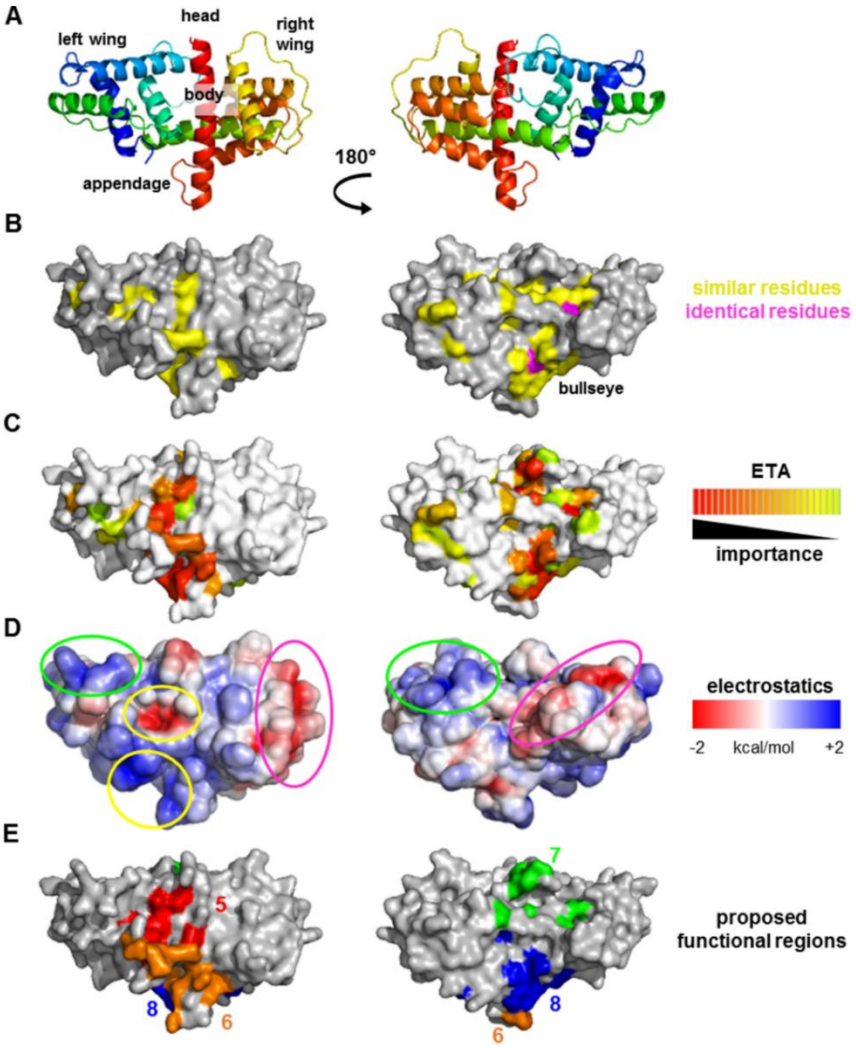


Figure 4-10. Surface analysis of the structure of UL21C
UL21C is shown in two orientations that are related by a 180° rotation around the vertical axis. (A) Ribbon diagram with topological descriptors. (B) Surface representation, with similar residues shown in yellow and identical residues in magenta. (C) Evolutionarily conserved residues identified using universal evolutionary trace analysis (**http://mammoth.bcm.tmc.edu/uet/**) are mapped on the surface of UL21C. Residues with importance scores in the top 25% were assigned colors from red (more conserved) to yellow (less conserved). (D) Electrostatic surface potential was analyzed by composing a map of the surface of UL21C using the PBEQ Solver function in the CHARMM program (**http://www.charmm-gui.org/?doc=input/pbeqsolver**). The two nonconserved acidic valleys wrap around the "wings" of UL21C and are circled in fuchsia; the two nonconserved basic patches sit on the tips of the "wings" and are circled in green; and the conserved acidic patch and the conserved nearby basic pocket are circled in yellow. (E) Potential functional regions on the surface of UL21C. Region 5 (red) includes Q321, Y352, E358, P363, D519, and L526; region 6 (orange), R357, R359, R360, E409, A500, R503, A509, R511, T512, R513, and A515; region 7 (green), D295, G330, P332, R333, Q346, K347, and G469; and region 8 (blue), L284, Y353, D392, N395, R399, D400, D494, S499, F501, and D502. Reprinted with permission from [99].

The surface of UL21C contains acidic and basic patches, some of which are conserved. Analysis of the electrostatic surface potential identified three acidic patches on the surface of HSV-1 UL21C (Figure 4-10D). One acidic patch maps to an area within the evolutionarily conserved belt around the "body" of the protein and may contribute to a common UL21 function. The other two acidic patches, which are located in valleys on either "wing" (Figure 4-10D), are not evolutionarily conserved and could participate in virus-specific functions.

The surface of HSV-1 UL21C has 27 arginines and three lysines (Figure 4-10D, Figure 4-4), which is consistent with the calculated pI of 9.68 (Table 4-3), and some of these residues cluster into three prominent basic patches. Only one basic patch is located within the evolutionarily conserved belt (Figure 4-10D) while the two non-conserved basic patches map to "wing" tips. Interestingly, both the basic pI and twelve arginines are conserved in HSV-2 and the other four most evolutionarily similar viruses that cluster at partition 2 of ETA (Figure 4-9 and Figure 4-10), but are not conserved among more divergent alphaherpesviruses such as PRV and VZV, which have acidic calculated pIs. Such divergence in charge among UL21C homologs is notable compared to the more clustered isoelectric points of both UL21N and the UL21 binding partner, UL16 (Table 4-3). The basic charge could also potentially ascribe acidic membrane binding ability on UL21C. Indeed, the domain associated with MLVS in the cosedimentation assay; however, this was not seen for the full length protein (data not shown).

Table 4-3. Isoelectric points of UL16 and UL21 sequences from 16 α-herpesviruses

	UL21		**UL16**
domain:	*C*	*N*	*full*
Meleagrid-HV-1	4.78	4.94	6.77
Gallid-HV-2	4.62	5.07	7.11
Gallid-HV-3	4.88	5.13	8.89
BHV-2	**9.34**	5.97	N/A
Leporid-HV-4	**8.71**	7.64	7.50
Saimiirine-HV-1	**8.37**	5.41	6.82
Cercopithecine-HV-2	**7.22**	5.96	8.35
HSV-1	**9.68**	5.20	8.06
HSV-2	**8.10**	6.43	7.90
Anatid-HV-1	4.82	5.05	7.95
VZV	5.78	5.23	8.74
PRV	4.92	5.20	8.53
EHV-1	5.12	5.07	8.72
Felid-HV-1	5.33	4.84	6.17
BHV-1	4.81	8.14	9.37
BHV-5	4.77	5.94	9.37

4.2.4.3 Potential functional regions in UL21C.

By combining the results of the aforementioned analyses, we located four conserved regions of potential functional importance on the surface of UL21C (Figure 4-10E), designated regions 5-8 (which join conserved regions 1-4 located within UL21N, (Figure 4-6) [53]). Region 5 surrounds the conserved acidic patch near the center of the "body". Region 6 contains the conserved basic patch on one side of the "appendage". Region 7 lies atop the "head" and includes evolutionarily conserved residues identified

by ETA analysis. Region 8 completes the belt of evolutionarily conserved residues and encompasses the bullseye near the "appendage". Together, these regions wrap around the middle of the UL21C and may mediate conserved functions of UL21C.

4.2.5 Characterization of full length UL21

4.2.5.1 The domains of UL21 do not interact

Since we've found both domains of UL21 to be well folded units connected physically by a long, flexible linker and with surface areas of opposing charge (Figures 21-6, -10), we hypothesized that the domains might interact with each other. We tested this hypothesis by incubating constructs of UL21N and UL21C of varying size and tag placement and assessed their interactions using affinity resin and size exclusion chromatography. Based on available constructs, we first incubated StII-tagged UL21N(1-216) with tag-free GP-UL21C(275-535) and evaluated their interaction on size exclusion chromatography (Figure 4-11A). In this experiment we saw no interaction between the two domains of UL21. Additionally, UL21C (28.4 kDa) runs at a smaller apparent molecular weight than UL21N (25.2 kDa), potentially due to affinity for the chromatography media as we had previously seen that UL21C can bind StrepTactin sepharose without a tag. We then hypothesized that the conserved, predicted strand region in the linker may be important for interaction between the domains. We tested the interaction between StII-UL21C(275-535) with longer constructs of UL21N including just the conservation site [UL21N(1-253)-StII] and the entire linker region [UL21N(1-274)-StII] using size exclusion and saw no interaction between the domains (Figure 4-11B,C). We again saw that UL21C was interacting strongly with the resin because despite similar molecular weights and known globular folds, the peak representing UL21C (29.7 kDa) was well-separated from and retained longer than the peaks

representing UL21N(1-253) (28.4 kDa) (Figure 4-11B) or UL21N(1-275) (30.5 kDa) (Figure 4-11C). We then found that tag-free UL21C did not show a high affinity for glutathione resin (Figure 4-11D). To remove any bias from interaction with chromatography media, the domain interaction was tested using GST-tagged UL21C as bait for glutathione resin after incubation with the UL21N constructs including the linker region (Figure 4-11D). Again, we saw no interaction. Together these results suggest that the two domains of UL21 do not have a strong interaction.

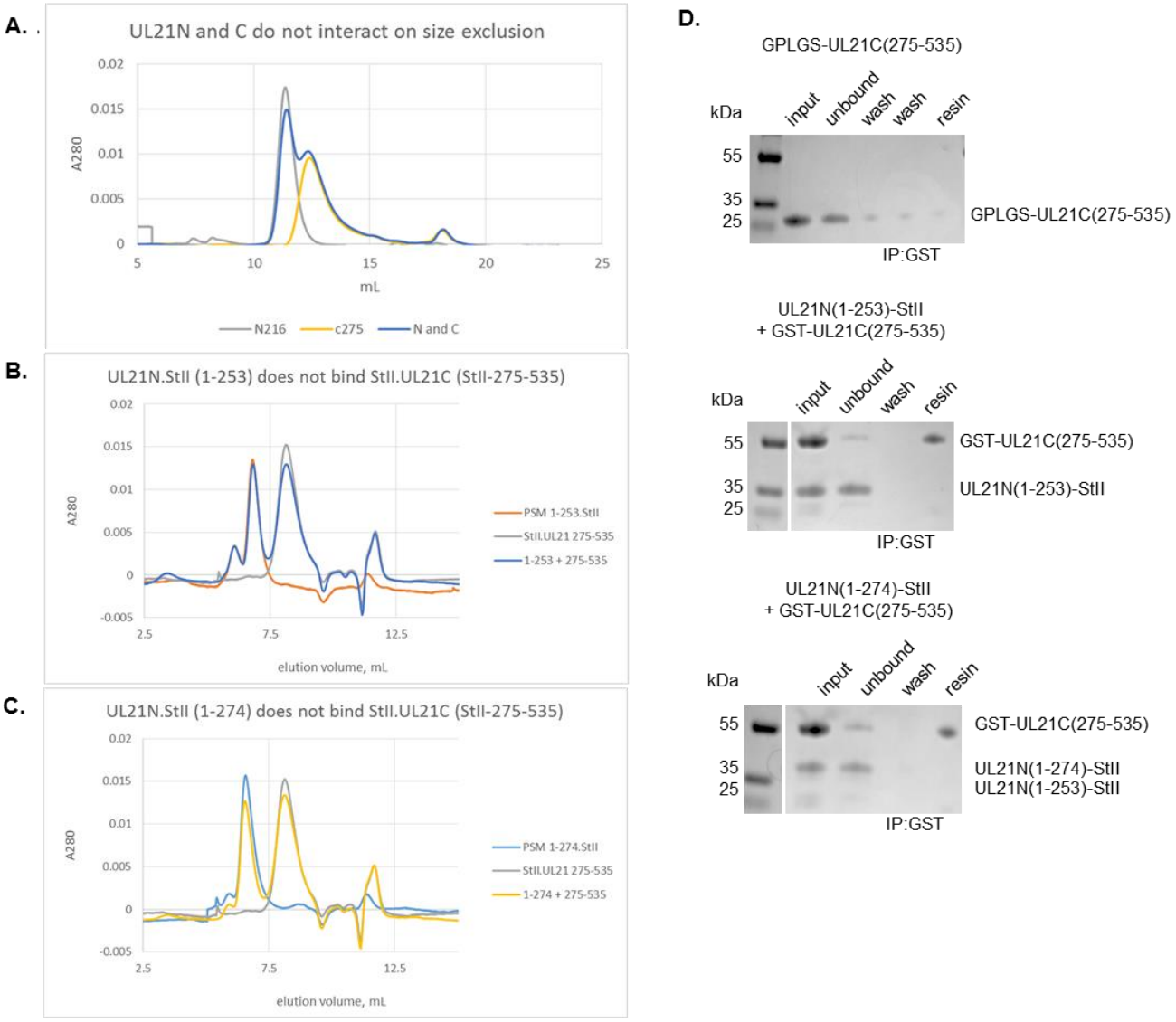


Figure 4-11. Interaction studies of UL21N and UL21C
A-C) Size exclusion chromatograms of UL21N, UL21C, or UL21N and UL21C constructs as depicted in legends. D) Coomassie-stained gels of glutathione resin pull-down using UL21C with or without GST-tag as bait and UL21N as target.

4.2.5.2 Mutations in non-conserved linker region stabilize UL21 against proteolysis

Since the two domains of UL21 do not appear to interact, thereby impeding structural characterization of the full length molecule by studying the domains together, the first hurdle to overcome in characterizing the full length molecule was to stabilize

UL21 against proteolysis. Since the originally identified digestion (Figure 4-2) occurred after large, charged amino acids, we designed UL21 proteolytic site mutant (PSM) including mutations of non-conserved arginines and lysines in the linker region (R216G/R217S/R220S/K234G/R235S). This construct was stable to proteolysis in normal purification (Figure 4-1), but never crystallized with N- or C-terminal StII tags. Based on the intimate interaction between the N-terminal StII tag and UL21N (Figure 4-5), we hypothesized that any affinity tag on UL21 could interfere with domain interactions and aimed to create full length UL21 construct with as few excess residues as possible. Since the very C-terminus of UL21 (QSV) is similar to the N-terminal side of the PreScission cut site (LEV), we designed UL21(1-532)-HRV3C-StII in the PSM background to produce UL21 with only three extra residues on the C-terminus (Figure 4-1). This protein expressed and purified easily and behaved similarly to PSM-StII (Figure 4-12A,C), but the StII tag was not removed as the size of the cleaved protein did not decrease on SDS-PAGE and most of the cleavage reaction mixture was retained on StrepTactin resin (Figure 4-12B). This construct, still including the StII tag, also did not crystallize.

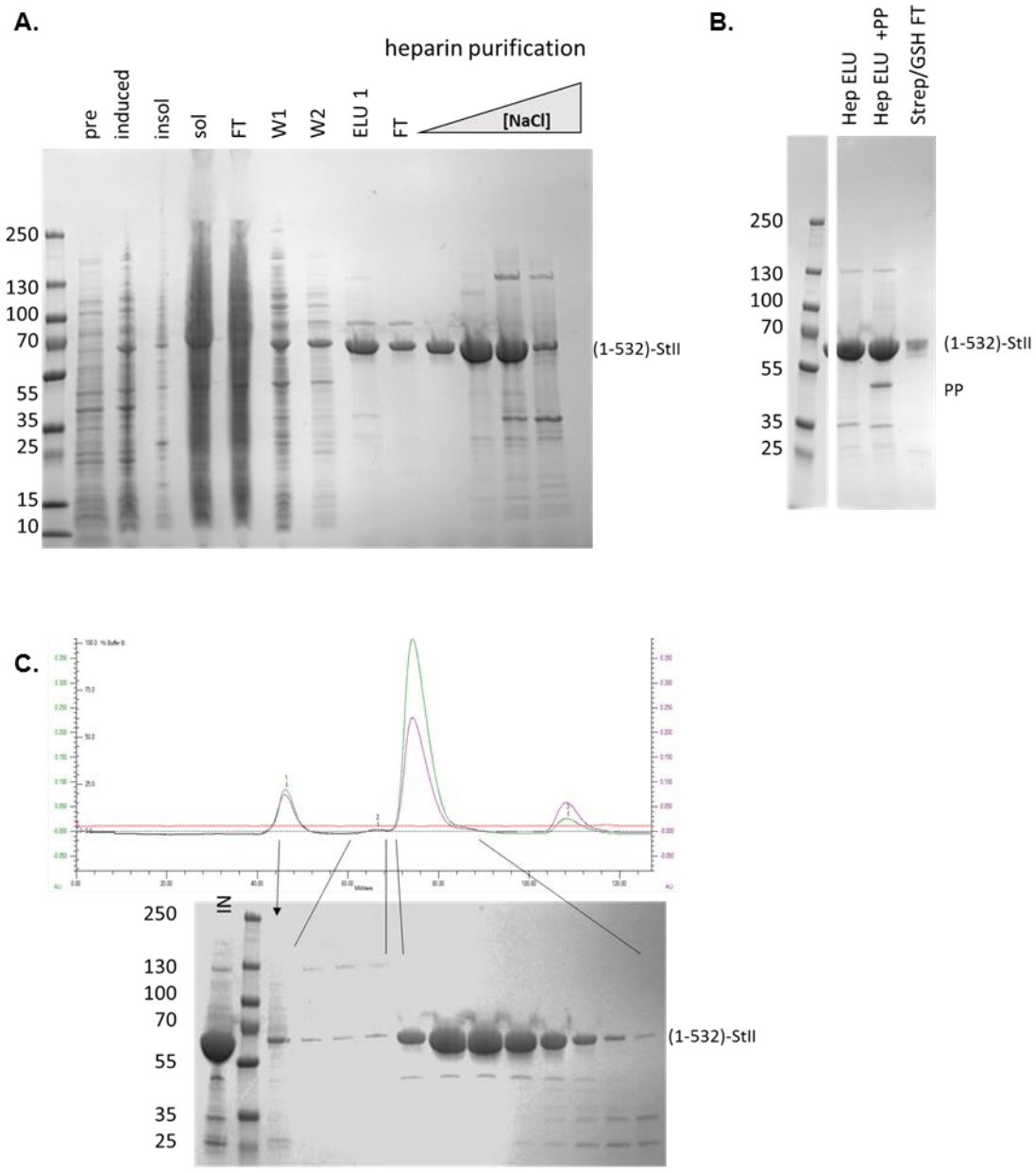


Figure 4-12. PSM(1-532)-HRV3C-StII purification
A) StrepTactin and heparin affinity purification, B) PreScission protease (PP) cleavage, and C) size exclusion chromatogram of PSM(1-532)-HRV3C-StII show that this construct is purified like other UL21 constructs but the tag cannot be removed.

4.2.5.3 Full length UL21 is an elongated molecule.

Although we were able to protect full-length UL21 from proteolytic digestion, no construct ever formed crystals. Since we were able to produce large amounts of pure,

stable UL21, we turned to small-angle X-ray scattering to investigate the full length structure of UL21 in solution. Diffraction data was collected from PSM(1-532)-HRV3C-STII and the resulting subtracted SAXS curve underwent Guinier analysis to identify its Rg of 38.8 Å (Figure 4-13). The data was further analyzed with GNOM to calculate the pairwise distance distribution [p(r)] function which showed a trailing parabola in comparison to a perfectly symmetric bell curve, which suggested that full length UL21 is an extended molecule (Figure 4-13). This p(r) function gave an Rg value of 40.3 Å and a Dmax of 139 Å, in agreement with the "extended" diagnosis for UL21. This can be better appreciated when compared to the SAXS characteristics of bovine serum albumin, a globular protein with a slightly larger molecular weight of 66 kDa (UL21 is 59.1 kDa) that has a smaller Rg of approximately 28 Å and a smaller Dmax of 82 Å (https://www.sasbdb.org/data/SASDBT4/). The Kratky plot of full-length UL21 displays a bell-shaped curve, which suggests that this is a well-folded molecule (Figure 4-13).

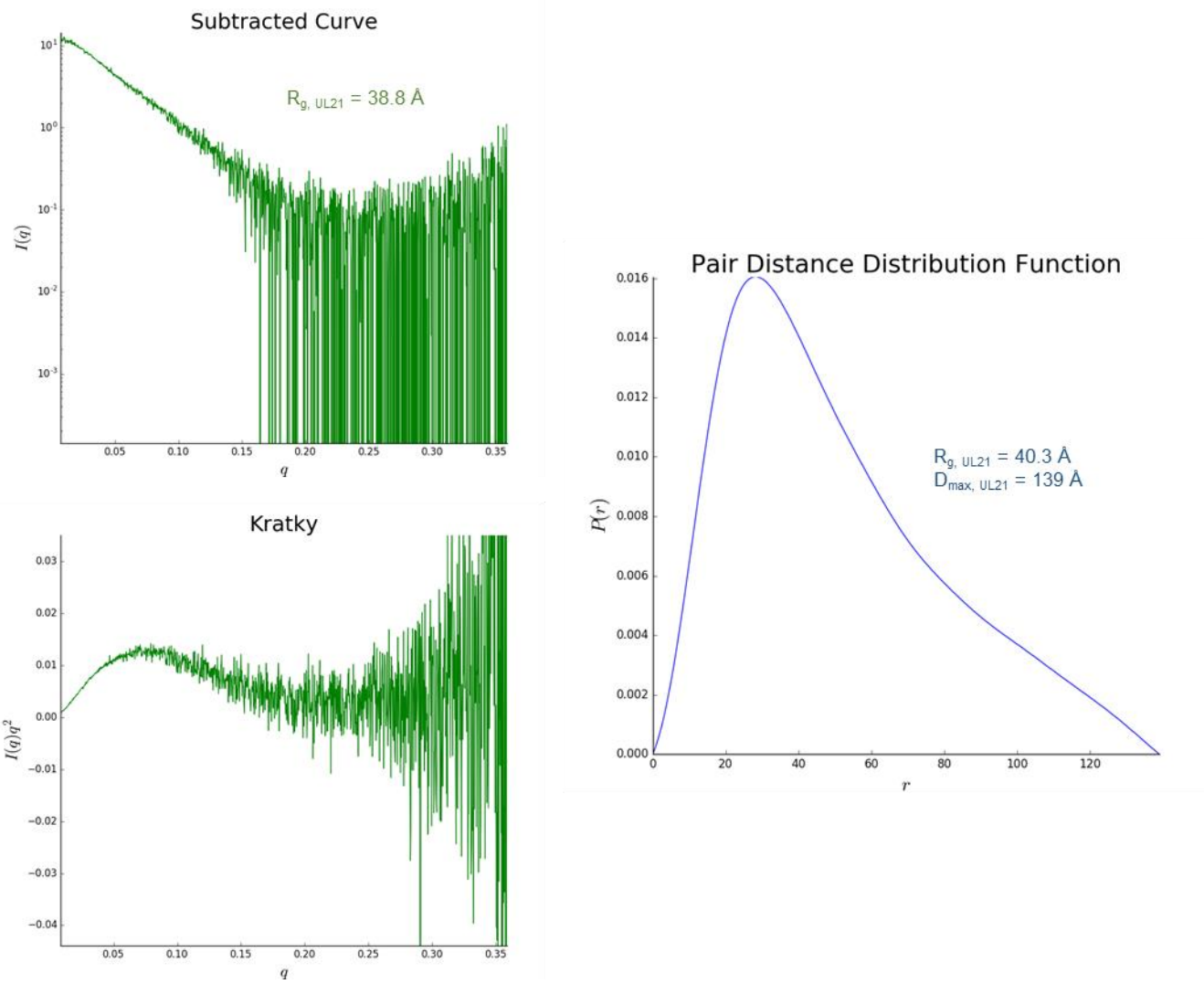


Figure 4-13. SAXS analysis of UL21
Subtracted SAXS curve, Kratky plot, and pair distance distribution function of PSM(1-532)-HRV3C-StII from SEC-SAXS data collected at ambient temperature. Rg and Dmax are shown as ~40 Å and 139 Å, respectively.

The curve was then subjected to *ab initio* rigid body modeling. Since the crystal structures of each domain of UL21 are known [53, 99], modelling was performed with CORAL, which takes these structures into account. Five different instances of CORAL were run. DAMAVER attempted to average these models, but was unable to differentiate between the two domains and often overlaid them, so single models are shown (Figure 4-14). These models are generally elongated (~140 Å), place the two domains close to each

other often end-to-end, and place the flexible linker running along one face of, extending past, and then returning back to UL21N. However, the specific interactions between the domains do not appear to be the same amongst these models.

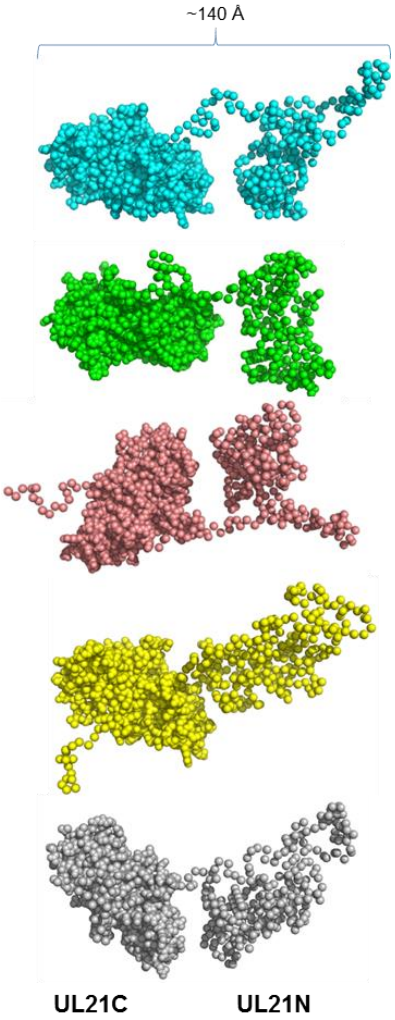
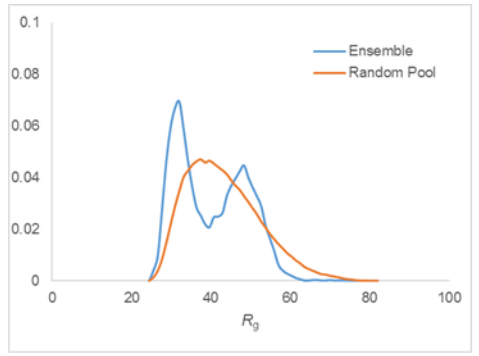


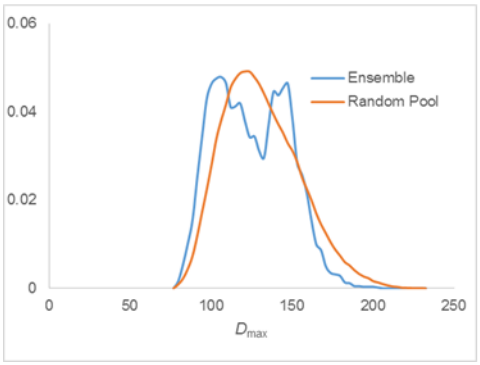
Figure 4-14. Rigid body models of UL21
Five separate rigid body models of UL21 from CORAL using crystal structures of UL21N and UL21C.

4.2.5.4 Full length UL21 is flexible in solution

Since it is likely that UL21 is a flexible molecule based on lack of conservation and predicted secondary structure in the linker region (Figure 4-3,-4) and further backed

up by the observation of multiple orientations in CORAL (Figure 4-14), we also analyzed SAXS data from PSM(1-532)-HRV3C-STII using the Ensemble Optimization Method (EOM). Firstly, this method confirms that UL21 is flexible with an R_{flex} of 83.31 very close to the R_{random} of 87.69, despite the parabolic Kratky plot, though it has been shown that folded domains separated by flexible linkers often show a parabolic Kratky in addition to measures of flexibility. Secondly, this analysis identifies three conformations that UL21 may take in solution: a compact conformation with a radius of about 30 Å and maximum dimension of about 105 Å, an intermediate conformation with a radius of about 45 Å and maximum dimension of about 140 Å, and an extended conformation with a radius of about 57 Å and maximum dimension of about 170 Å (Figure 4-15). Since temperature or the placement of the affinity tag could presumably alter the conformations available to UL21, we collected diffraction data from different full length constructs of UL21/PSM with a StII-tag at either terminus and at either ambient temperature or 4°C to characterize the effects of these factors and found each data set was comparable in EOM and rigid body modeling.





	Rg	Dmax	Percent
compact	31.28	101.09	40.00%
	34.36	116.22	10.00%
average	**31.90**	**104.12**	**50.00%**
intermediate	45.82	135.03	10.00%
	47.2	160.07	10.00%
	48.46	140.92	20.00%
average	**47.49**	**144.24**	**40.00%**
extended	**54.46**	**162.89**	**10.00%**

Figure 4-15. Ensemble modeling of UL21
Dimensions of models found in optimized ensemble from EOM shown as histograms of Rg and Dmax (L) and a summary table of dimensions (R).

4.2.5.5 The conserved region is protected from proteolysis.

The linker region of UL21 contains a stretch of residues from 239-251 that is extremely conserved amongst homologs, in stark comparison to a lack of conservation in residue type or number in the rest of the linker region (Figure 4-4). Furthermore, this stretch is predicted to form a beta strand. For these reasons, we hypothesized that this predicted strand may associate with UL21N despite identification of a fragment from 237-535 in the original proteolysis (Figure 4-2C). UL21N constructs were cloned with the PSM mutations encoding residues 1-253 and residues 1-274 with a C-terminal StII tag (Figure 4-1). These constructs were expressed, purified, and screened for crystallization conditions, but no hits were found. During storage of these proteins, a differential proteolysis pattern emerged; namely, it appeared that UL21N(1-253)-StII was

being cleaved by contaminating proteases to UL21N(1-216) while UL21N(1-274)-StII was first being cleaved to UL21N(1-253) before being further cleaved to UL21N(1-216). Limited proteolysis was performed on UL21N(1-253) and UL21N(1-274) with trypsin and chymotrypsin which showed the same pattern (Figure 4-16). Since UL21N(1-274) was not immediately cleaved to the minimal structured domain of UL21N and instead shows an intermediate cleavage product resembling UL21N(1-253), these results suggest that there is some structural protection from proteolysis in this region.

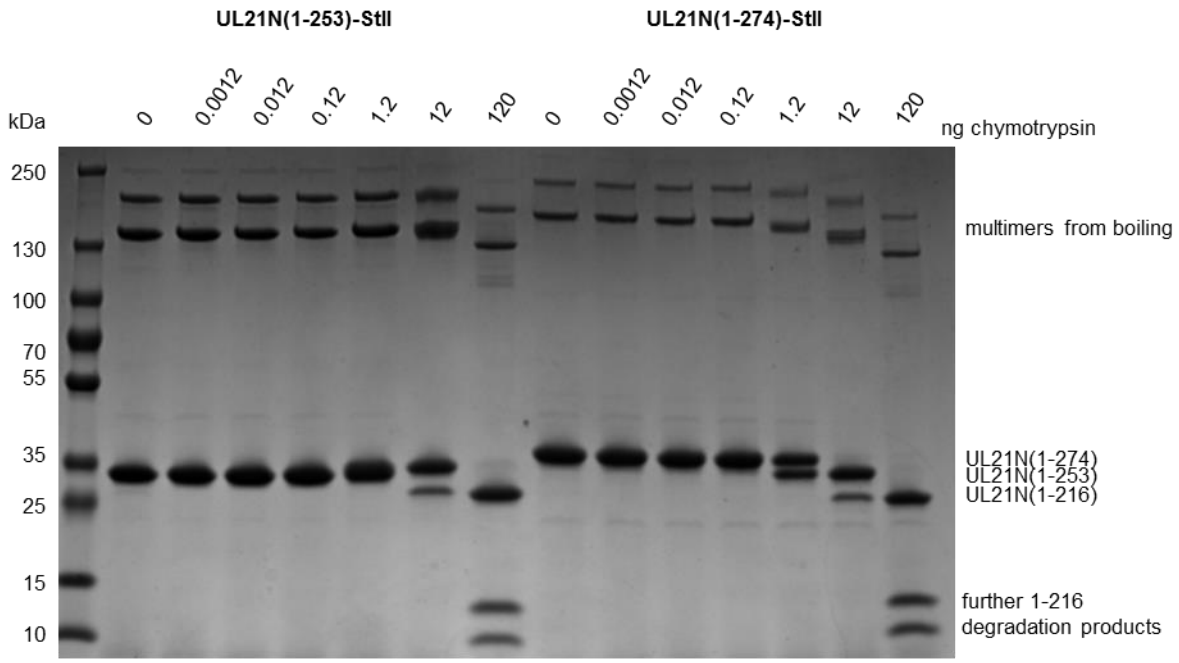


Figure 4-16. Limited proteolysis of UL21N with linker
Coomassie-stained gel of UL21N(1-253)-StII and UL21N(1-274)-StII incubated with increasing amounts of chymotrypsin.

4.2.5.6 UL21 copurifies with RNA after expression in *E. coli*

While purifying UL21, we observed that after StrepTactin affinity and size-exclusion chromatography purification steps, the resulting UL21 sample appeared homogenous as judged by SDS-PAGE (Figure 4-17A) but had a spectroscopic

A260:A280 ratio of 1.6, which is characteristic of a sample that contains approximately 80% protein and 20% nucleic acid (NA) [148]. This suggested that a large amount of *E. coli* NAs copurified with UL21. These NAs could only be separated from UL21 using heparin resin, which binds UL21 (Figure 4-17B) but not NAs. The resulting UL21 sample (Figure 4-17C) had an A260/A280 ratio of 0.5, which is consistent with a homogenous protein sample. Similar results were also observed with UL21C constructs whereas UL21N did not copurify with substantial amounts of NAs. To identify the type of NAs co-purifying with UL21, UL21/NA complex was affinity purified, and bound NAs were precipitated, digested with nucleases, and resolved on ethidium-bromide-stained agarose gels. The NAs were susceptible to digestion by RNase but not DNase (Figure 4-17D). We conclude that UL21 preferentially binds RNA through its C-terminal domain.

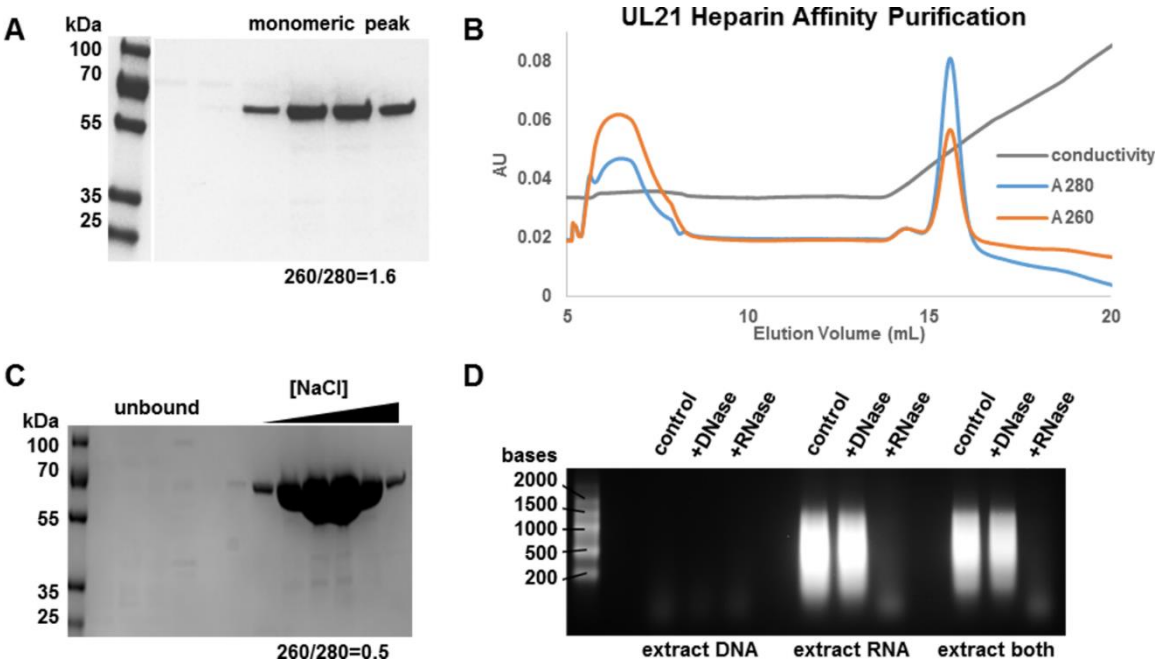


Figure 4-17. UL21 binds *E. coli* RNA
(A) Purified StII-GS-UL21 expressed in *E. coli* appears homogenous as judged by SDS-PAGE and Coomassie staining but has an A_{260}/A_{280} ratio consistent with a large amount of nucleic acid contaminant. (B) UL21 binds heparin Sepharose and is eluted with increasing NaCl concentrations (shown as conductivity). (C) Heparin Sepharose purification produces pure UL21 free of nucleic acid contamination as judged by the A_{260}/A_{280} ratio and SDS-PAGE. (D) The copurified nucleic acids are susceptible to digestion with RNase but not DNase. Samples were resolved on an ethidium bromide-stained formaldehyde agarose gel. Reprinted with permission from [99].

4.3 Discussion

UL21 has been implicated in a number of processes in the viral replication cycle, including nuclear egress, cytoplasmic capsid trafficking, secondary envelopment, and cell-cell spread; yet, the molecular mechanisms by which UL21 enables these and other processes remain unknown. Here we determined and analyzed the crystal structures of the domains of HSV-1 UL21. Together, these structures provide three-dimensional templates for a systematic exploration of the multiple activities of UL21. We also described the full length molecule as a flexible protein capable of taking multiple conformations. Future work will carefully characterize these conformations alone and in

the presence of known binding partners UL16 and RNA to further describe the multiple functions of UL21.

Chapter 5: UL16

5.1 Introduction

HSV-1 UL16 is a 373 amino acid tegument protein conserved among nearly all herpesviruses and its deletion results in a tenfold reduction in virion production [51], plaque size reduced 80%. These results are similar to PRV lacking UL16 [132], but HSV-2 UL16 [79] and the VZV homolog ORF44 [113] are required for replication in cell culture. In particular, a lack of HSV-2 UL16 blocks nuclear egress [79], similar to observations for binding partner HSV-2 UL21 [54]. Beta- and gammaherpesviruses also rely on their UL16 homologs as CMV UL94 [58] and murine gammaherpesvirus-68 ORF33 [149] are required for replication and their absences result in secondary envelopment defects. Like HSV-2, lack of ORF33 also impeded nuclear egress [149] and ORF33 has recently been found on capsids in the nucleus [150]. In transfected cells, HSV-1 UL16 localizes to the cytoplasm, but at early time points in infection UL16 additionally localizes to the nucleus [40], which suggests possible unknown nuclear roles for UL16 in HSV-1 as well, though UL16 is only found on cytoplasmic, not nuclear HSV-1 capsids [40].

UL16 also seems to be part of a signal transduction mechanism that prepares the virus for entry and/or dissociation upon entry because in extracellular virions, capsid-associated UL16 becomes dissociated when the virus binds cell surface receptors [151]. Within the conserved, functional cysteines of UL16 there is a C-X-C…C-X-X-C motif similar to a sequence in bacterial chaperone Hsp33. In Hsp33, this sequence serves as a redox-sensitive switch to turn on or off chaperone activity in oxidative or reducing environments, respectively [56]. While there is not yet evidence that UL16 serves as a chaperone, the authors hypothesize that this sequence serves as a redox-sensitive switch

to affect conformational change in UL16 as well. This hypothesis is strengthened by the observation of inconsistent *in vitro* behavior of UL16 in reducing or oxidizing environments (John Wills, personal communication), the knowledge that HSV-1 infection can generate an oxidative environment in the cell[152], and the extremely conditional binding activities of UL16 that suggest conformational change may be a prerequisite for binding (discussed below).

Based on these observations, UL16 is likely a multifunctional protein and with a role in secondary envelopment and interestingly, HSV-1 infection lacking UL16 resulted in multiple capsids being packaged together [51]. HSV-1 UL16 binds secondary envelopment regulator HSV-1 UL11 through its N-terminal domain and this interaction is conserved amongst homologs; however, in HSV-1, this interaction requires activation by binding the C-terminal domain of alphaherpesvirus-specific tegument protein UL21, mutation of cysteine residues, deletion of the self-regulatory C-terminal domain, or an oxidative environment [56]. Strengthened by these interactions with UL11 and UL21, UL16 binds gE and regulates cell-cell spread [45]. The domains of UL16 interact with each other and the C-terminal domain interacts with mitochondria in HSV-1 infected cells [153]. The mitochondrial interaction requires UL16 binding membranes (a highly conditional interaction requiring extensive deletion or mutation of UL16) and may be correlated with yet undescribed viral assembly or apoptosis [153]. Additionally, HSV-1 UL16 has been shown to directly bind tegument protein VP22/UL49 [51]. Clearly UL16 is an important, complex, and dynamic molecule that makes contact with many host and viral molecules that would benefit from careful biochemical and structural study, but our

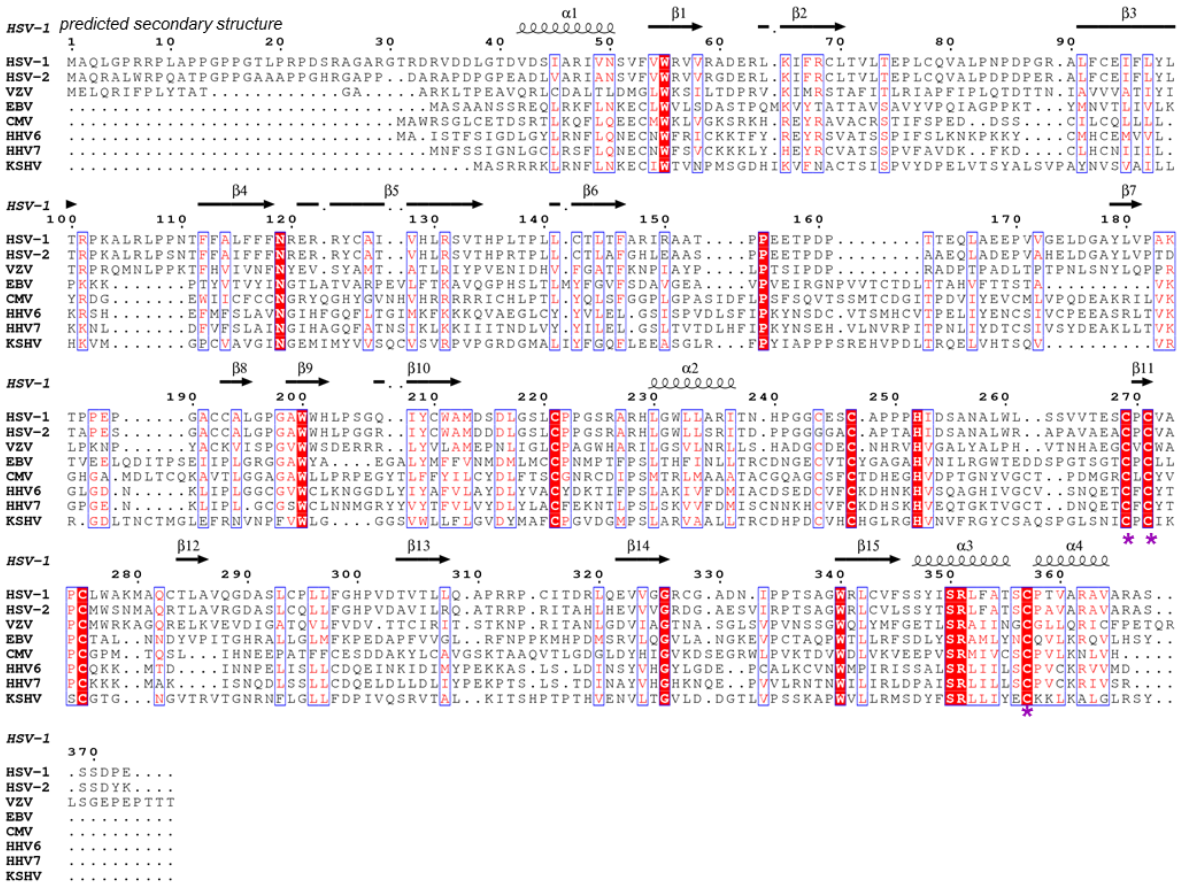
attempts were hampered by the inability to produce soluble protein, not only due to the presence of 20 cysteine residues. Our progress is described below.

5.2 Results

5.2.1 StII-tagged UL16 constructs are not expressed or very soluble

We started characterizing UL16 by expressing StII-tagged full length UL16 or its domains (Figure 5-1) in Rosetta *E. coli* grown in LB with IPTG induction. On a Coomassie-stained gel, only StII-Xa-UL16C(156-373) appeared to express protein and it did so preferentially with expression at 37°C (Figure 5-2A). A western blot probed with StrepTactin-HRP (Figure 5-2B) showed that StII-Xa-UL16N(1-155) was not expressed. StII-Xa-UL16 was weakly expressed, but found in the insoluble fraction after a freeze-thaw solubility experiment in both a buffer with pH 7.5 and 100 mM NaCl and a buffer with pH 5.5 and 500 mM NaCl. In contrast, StII-Xa-UL16C(156-373) appeared 50% soluble in both of these buffers after IPTG induction at 16°C.

A.


HSV-1 predicted secondary structure
HSV-1
HSV-2
VZV
EBV
CMV
HHV6
HHV7
KSHV

B.

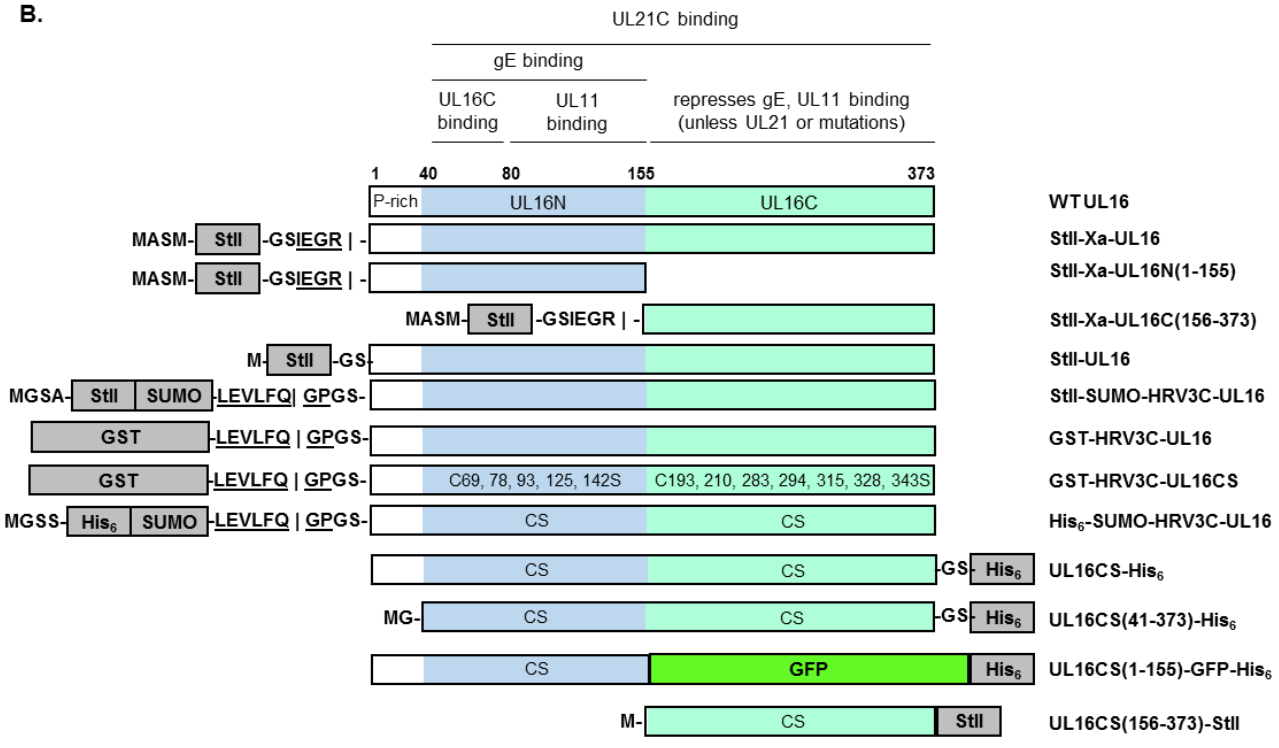

UL21C binding
gE binding
UL16C binding
UL11 binding
represses gE, UL11 binding (unless UL21 or mutations)
1
40
80
155
373
P-rich
UL16N
UL16C
WT UL16
MASM-StII-GSIEGR | -
StII-Xa-UL16
StII-Xa-UL16N(1-155)
StII-Xa-UL16C(156-373)
M-StII-GS-
StII-UL16
MGSA-StII SUMO-LEVLFQ| GPGS-
StII-SUMO-HRV3C-UL16
GST-LEVLFQ | GPGS-
GST-HRV3C-UL16
C69, 78, 93, 125, 142S
C193, 210, 283, 294, 315, 328, 343S
GST-HRV3C-UL16CS
MGSS-His6 SUMO-LEVLFQ | GPGS-
His6-SUMO-HRV3C-UL16
CS
UL16CS-His6
MG-
UL16CS(41-373)-His6
GFP
UL16CS(1-155)-GFP-His6
UL16CS(156-373)-StII

Figure 5-1. Secondary structure of UL16
A) Sequence alignment of UL16 homologs from human herpesviruses. Predicted secondary structure of HSV-1 UL11 is marked above. Cysteine mutations that activate HSV-1 UL16 for binding UL11 are marked with purple asterisks. B) UL16 expression constructs. Boundaries of binding requirements are shown. Protease recognition sites are underlined and cleavage sites are shown with a vertical line. CS, cysteine to serine mutant. All constructs are codon-optimized except for those with a StII-Xa tag.

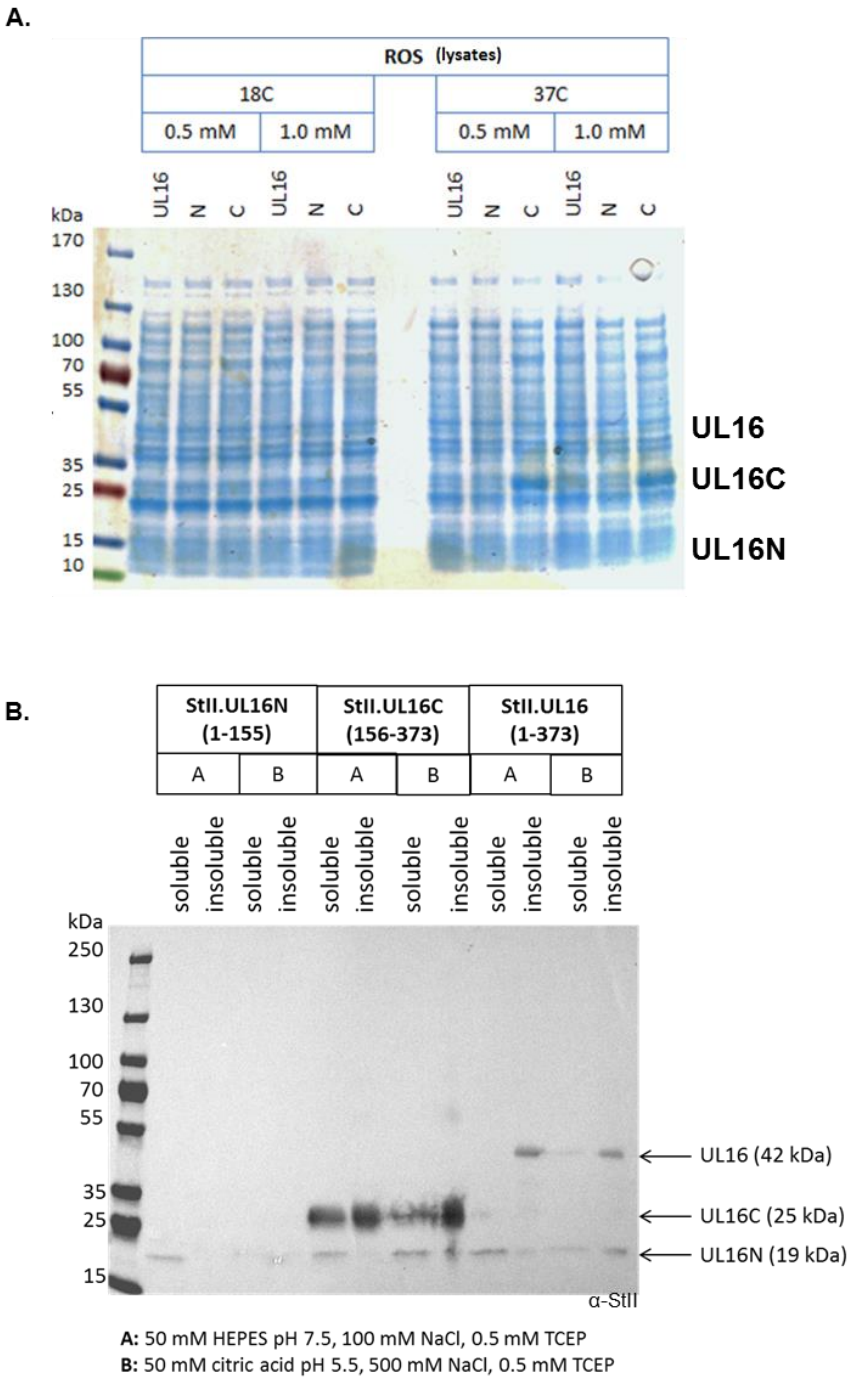


Figure 5-2. Expression and solubility of StII.Xa.UL16 constructs
A) Coomassie-stained gel of UL16 expression in different conditions. B) Anti-StII western of soluble and insoluble fractions of UL16 constructs

StII-Xa-UL16C(156-373) was expressed on larger scale and purified in buffer A over StrepTactin affinity resin, but copurified with a large proportion of a contaminant

around 70 kDa (potentially a chaperone protein) and nucleic acids (NAs), as the average 260/280 of elution fractions was 1.14 (Figure 5-3A) [129]. The elution from affinity purification was purified with size exclusion chromatography and although most of the protein eluted in the void with the 70kDa contaminant, there was a peak containing only a 25 kDa band that eluted later (Figure 5-3A,B). The nucleic acid interaction remained after size exclusion as the 260/280 ratio in the peak for the 25 kDa protein remained above 1. This pool was subjected to heparin purification but the 25 kDa species did not bind, suggesting that this protein may be sticky rather than specifically binding NAs (Figure 5-3C). Furthermore, the 25 kDa band appeared fuzzy on a gel (Figure 5-3A,C) and was not recognized by polyclonal serum to UL16C in a western blot (Figure 5-3D "second peak"), suggesting that the 25 kDa band is a contaminant.

Another purification of StII-Xa-UL16C was performed in buffer A with 1 mM EDTA. Using a high salt (1M NaCl) wash in affinity purification resulted in an elution fraction containing 100% protein (260/280~0.75), but it did not remove the 70 kDa protein contaminant (Figure 5-3E). StII-Xa-UL16C without NAs remained predominantly in the void with the 70 kDa protein in size exclusion, again with a second peak consisting of the fuzzy 25 kDa band. The void was collected and subjected to anion exchange chromatography in an attempt to separate the contaminant from StII-Xa-UL16C, but the separation was inefficient (Figure 5-3F).

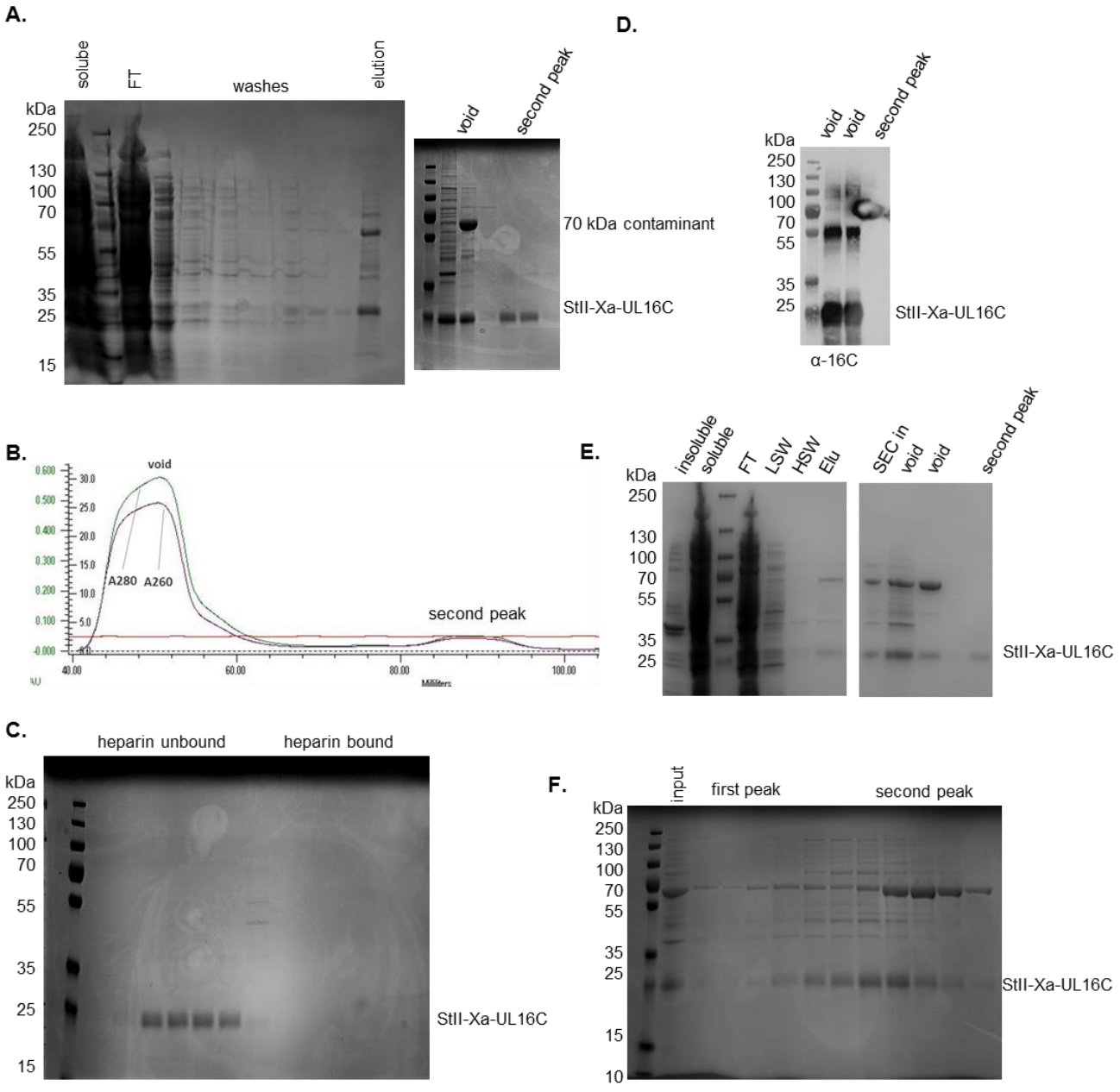


Figure 5-3. Purification of StII-Xa-UL16C
A) Coomassie-stained gels of affinity and size exclusion purification of UL16C. B) Chromatogram from A. C) Coomassie-stained gel of UL16C heparin purification. D) Anti-16C western from A-B. E) Coomassie-stained gel of affinity and size exclusion purification of UL16C including a high-salt wash to decrease nonspecific binding. F) Coomassie-stained gel of UL16C on anion exchange.

5.2.2 GST-tagged, codon-optimized UL16 can be expressed, but remains mostly insoluble

In order to improve expression levels of the full length protein, we obtained codon-optimized DNA, used it to clone constructs StII-SUMO-UL16 and GST-UL16 (Figure 5-1), and tested their expression in Rosetta cells grown in LB with different amounts of IPTG and solubility in buffer A after freeze/thaw lysis. The StII-SUMO construct was very lowly expressed and only at 16°C (Figure 5-4A), but the GST-UL16 construct was expressed highly in all conditions (Figure 5-4B). Unfortunately, all of these constructs were found to be insoluble (Figure 5-4A,B).

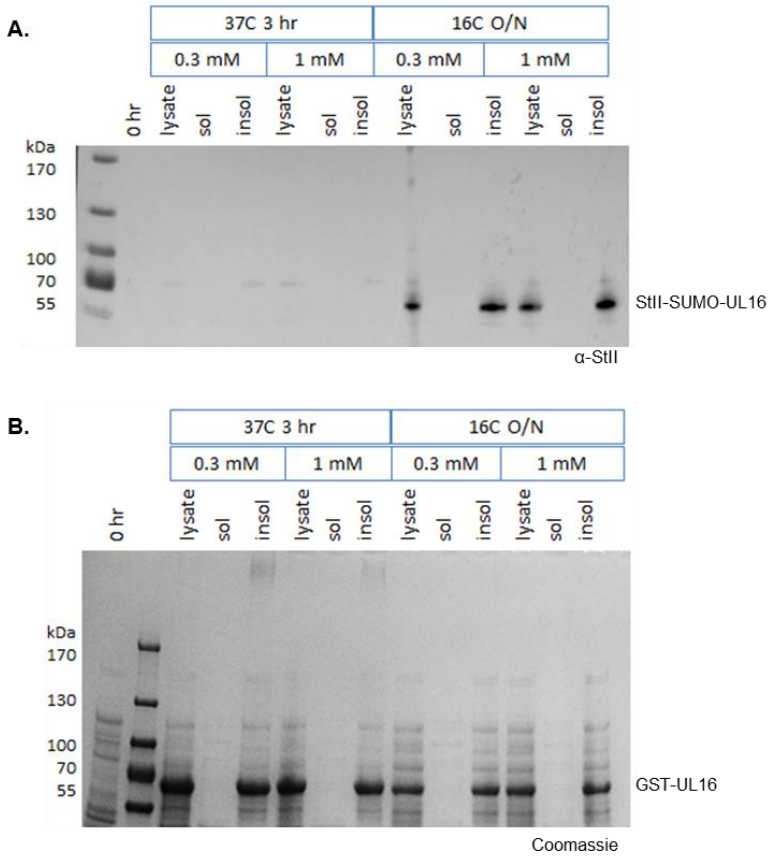


Figure 5-4. Expression and solubility of codon-optimized UL16
A) Anti-StII western blot of soluble and insoluble fractions of StII-SUMO-UL16 and B) Coomassie-stained gel of soluble and insoluble fractions of GST-UL16 expressed under different conditions.

5.2.3 Coexpressing GST-UL16 with StII-UL21 decreases expression of UL16

We also obtained codon optimized DNA for full length UL16 with 12 non-conserved cysteines mutated to serines and subcloned it to create a construct encoding GST-UL16CS (Figure 5-1). As a binding partner can often mediate the behavior of a stubborn protein, UL16 is known to bind UL21, and we have shown UL21 to be very well behaved, we assessed this construct's expression level and solubility alone and in coexpression with StII-UL21. In Origami B cells, we also tested coexpression of StII-UL21 with wild type GST-tagged UL16, since some of the mutated cysteines may be important for folding and Origami cells provide an environment that allows the formation of disulfide bonds. All UL16 constructs were expressed alone in T7 (Figure 5-5A), Rosetta (Figure 5-5B), and Origami (Figure 5-5C) cells, but expression of GST-UL16CS or GST-UL16 was decimated in the presence of StII-UL21 in all cell types (Figure 5-5) (discussed below). Notably, UL21 was expressed alone and in coexpressions in all cell types and was found to be soluble in all cell types, although in Origami cells, soluble UL21 was cleaved into domains (Figure 5-5C), which was not seen in other expression systems (Figure 5-5A, B). Although both UL16 constructs were mostly insoluble after freeze/thaw lysis in reducing buffer A from all cell types, a small portion of GST-UL16CS expressed alone in Rosetta cells was soluble (Figure 5-5B). Since multiple rounds of freezing and thawing successfully lyses cells but can be harsh on proteins and therefore underestimate solubility, we decided to pursue this construct in large scale purification.

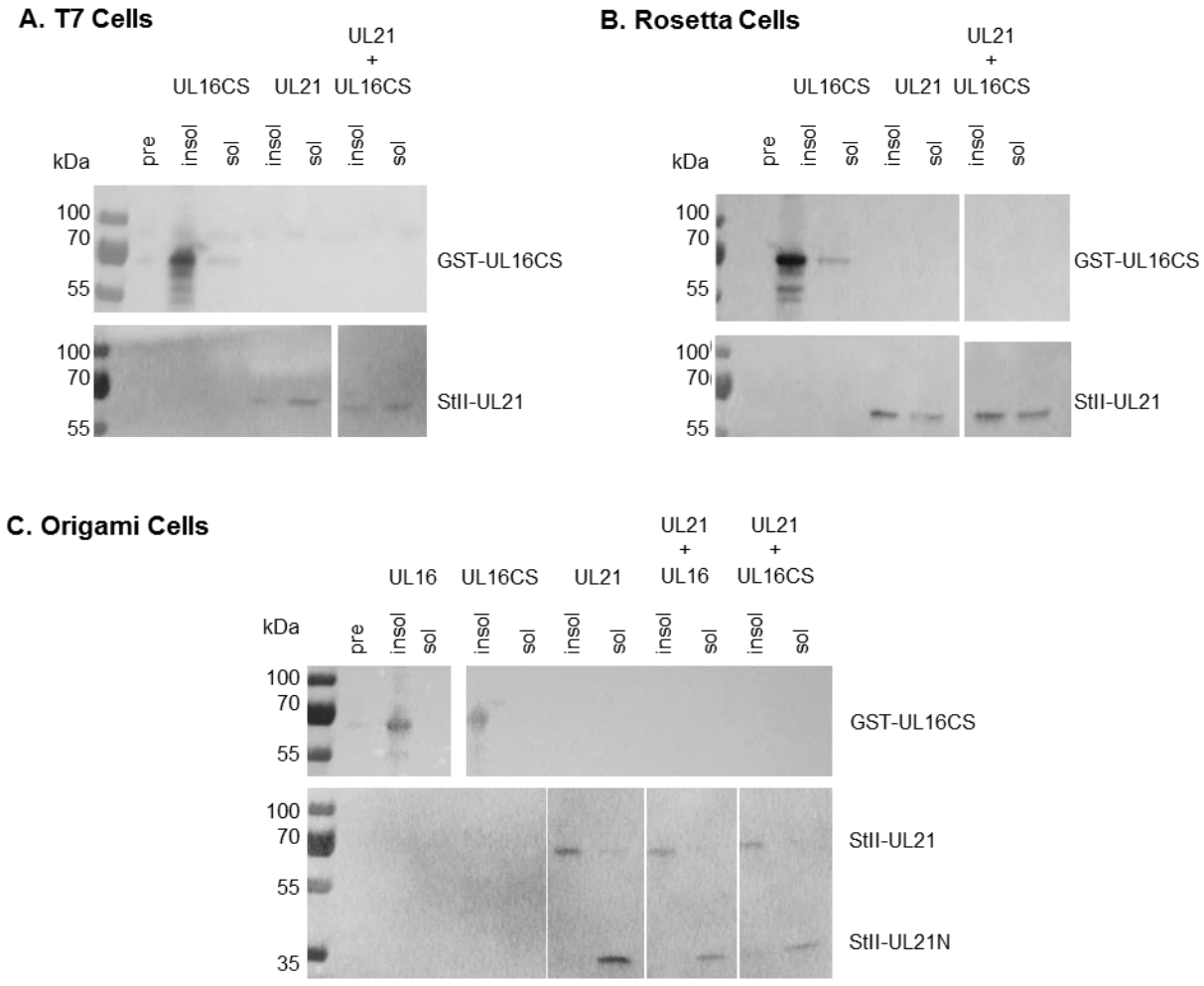


Figure 5-5. Coexpression of UL16 and UL21
(top) Anti-GST and (bottom) anti-StII western blots of soluble and insoluble fractions of GST-UL16, GST-UL16CS, and StII-UL21 expressed singly or together in A) T7, B) Rosetta, or C) Origami B cells.

5.2.4 A cysteine to serine mutant improves solubility but not purity of GST-UL16

GST-UL16CS was expressed in Rosetta cells overnight at 16°C, lysed in buffer A, and purified by glutathione affinity purification. There is a large band consistent with the size of GST-UL1CS in the diluted insoluble fraction, the flow through, and the wash, suggesting that this protein may be mostly insoluble, and what is soluble doesn't bind very well (Figure 5-6A). There was some protein eluted from the glutathione resin that was incubated with PreScission protease and indeed a small amount of cleaved GPLGS-

UL16CS appeared (Figure 5-6C), but there was still a large band around 70 kDa (Figure 5-6A,C), as was previously seen for StII-Xa-UL16C (Figure 5-2). This pool was purified by S200 size exclusion chromatography and four peaks appeared (Figure 5-6B): a void with a large A260 (Figure 5-6B) and no protein (Figure 5-6C), suggesting it to be mostly NAs; a post-void peak that contained the 70 kDa contaminant and the cleaved GPLGS-UL16CS (Figure 5-6C); a peak with low absorbance readings that contained PreScission protease (Figure 5-6C); and the peak with the largest absorbance reading that contained the free GST-tag and prescisson (Figure 5-6C). Since there was so little of the cleaved UL16 product and it remained in the void with apparently the same chaperone that we were unable to remove before, we moved on.

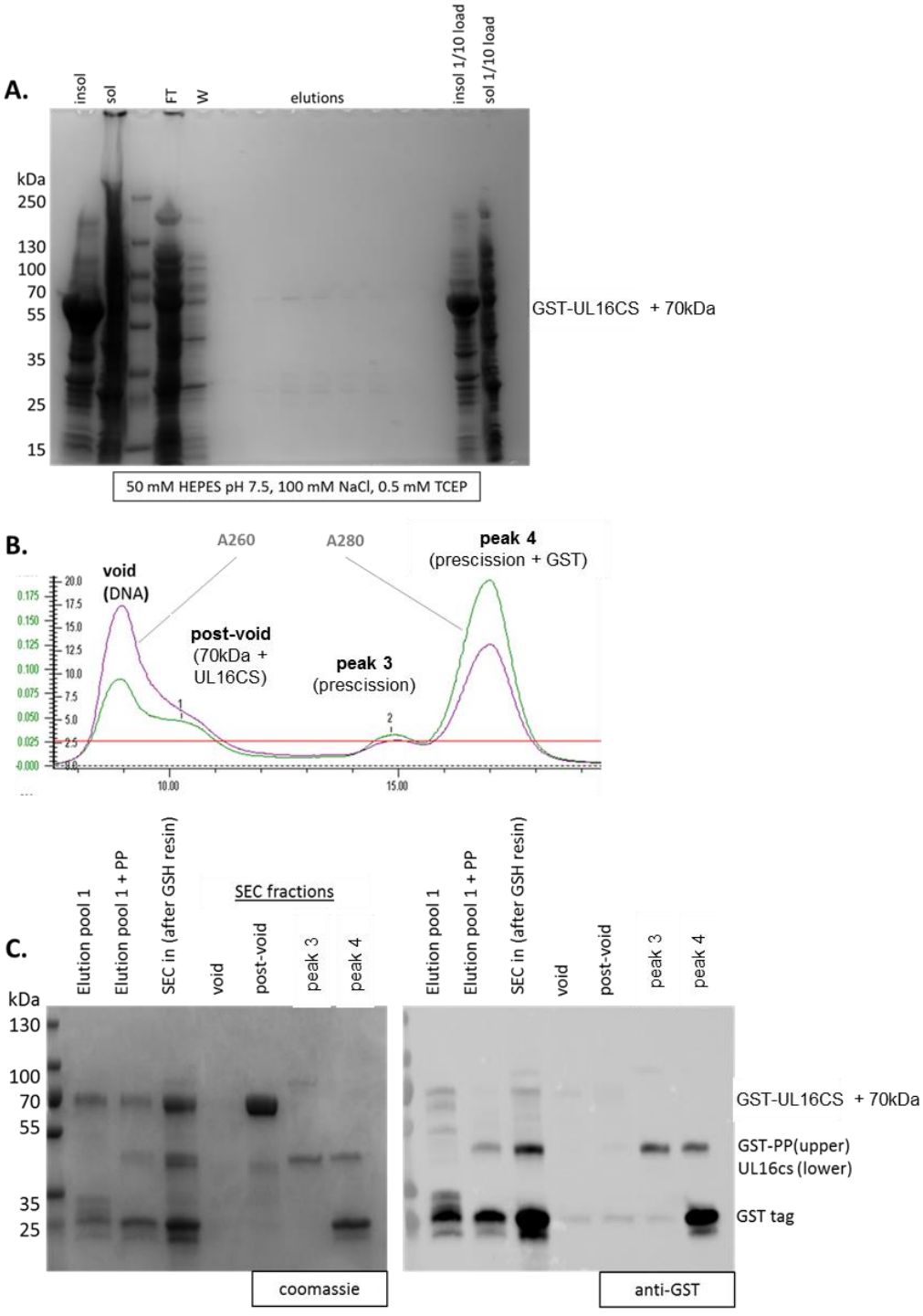


Figure 5-6. Purification of GST-UL16CS
A) Coomassie-stained gel of affinity purification of GST-UL16CS. B) Chromatogram of GPLGS-UL16CS size exclusion purification, 260 nm absorbance in purple, 280 nm absorbance in green. C) (L) Coomassie-stained gel and (R) anti-GST western of B.

5.2.5 StII-UL21 does not solubilize GST-UL16CS when lysates are mixed

Coexpression of StII-UL21 with GST-UL16CS resulted in extremely decreased expression of UL16, but UL21 could still in theory improve the solubility of UL16. Therefore, we asked if lysing cells expressing UL21 together with cells expressing UL16 could solubilize UL16. Binding UL16 has been mapped to the c-terminal domain of UL21, so we mixed harvested cells expressing StII-HRV3C-UL21C(275-535) with harvested cells expressing GST-UL16CS, lysed them together in buffer A, and subjected them to affinity purification over glutathione resin followed by StrepTactin resin. As shown in Figure 5-7, when lysed together, GST-UL16CS remains mostly insoluble whereas StII-HRV3C-UL21C(275-535) is approximately 50% soluble. When subjected to glutathione resin, both proteins are found in the flow through and wash. Only a very small proportion of GST-UL16CS is eluted from the first round of affinity purification and no StII-UL21C comes along. This suggests that GST-UL16CS is mostly insoluble and cannot bind StII-UL21C.

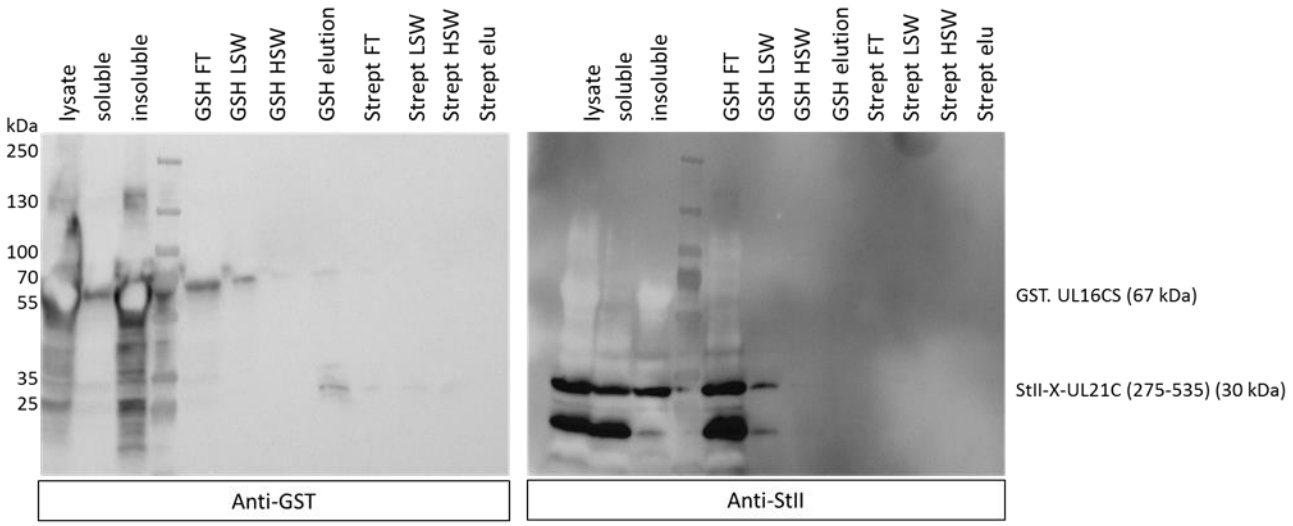


Figure 5-7. Purification of GST-UL16CS lysed with StII-HRV3C-UL21C(275-535) (L) Anti-GST and (R) anti-StII western blots of affinity purification of UL16CS from cells lysed with cells expressing UL21C. GSH, glutathione resin. Strept, StrepTactin resin. FT, flow through. LSW, low salt wash. HSW, high salt wash. Elu, elution.

5.2.6 Coexpression with a pET vector decreases expression levels of pGEX insert

Our collaborators have reported that the interactions between the two domains of UL16 are only seen when the constructs are coexpressed [153], so it is possible that coexpression is necessary. Unfortunately we saw that UL16 expression was severely decreased in the presence of UL21 (Figure 5-5). Since UL21 has been reported to have RNA binding ability [99] and potential roles in gene regulation, we asked whether this was a common effect of coexpressing proteins with UL21. We coexpressed StII-UL21 (pET vector) with GST-UL34 (pGEX vector) and indeed saw that expression of UL34 was severely decreased in the presence of UL21 (Figure 5-8A). As controls, we also tested the coexpression of GST-UL34 (pGEX) with His_6-SUMO-UL31 (pET) and GST-UL16 (pGEX) with His_6-SUMO-UL31 (pET) and found that in general, the expression level of the protein encoded by the pGEX vector was decreased when coexpressed with

pET plasmid (Figure 5-8A). Furthermore, the pET-encoded protein expression level remained unchanged. We confirmed this observation by coexpression GST-UL21 (pGEX) with StII-UL16CS (pET) and were able to express both proteins at once (Figure 5-8B).

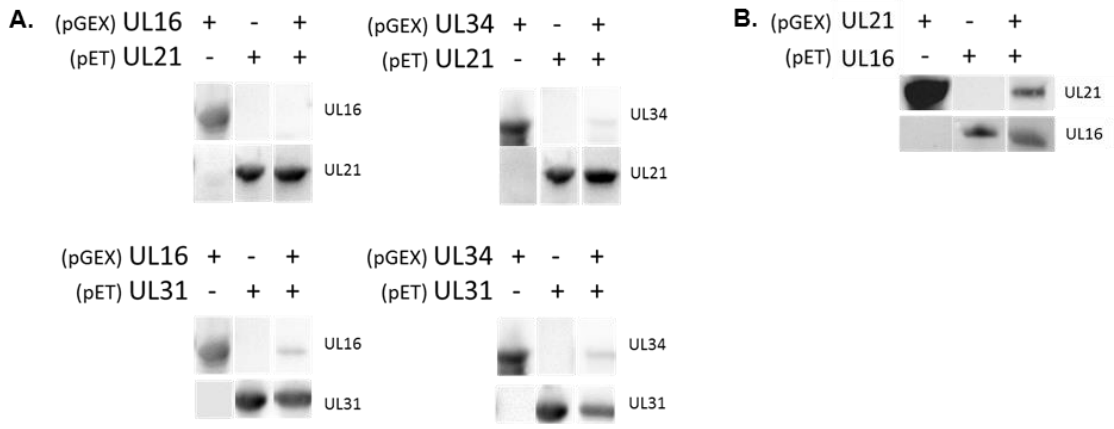


Figure 5-8. Coexpression from pET and pGEX vectors
A) Anti-GST (UL16, UL34) and anti-His (UL31) or anti-StII (UL21) westerns of single or coexpressed proteins. B) Anti-GST (UL21) and anti-StII (UL16) westerns of single or coexpressed GST-UL21 and StII-UL16.

5.2.7 Coexpression of N-terminally tagged UL16CS with binding partners does not improve UL16CS solubility or purity

Since we were finally able to coexpress UL16 and UL21 (Figure 5-8B), we tried to purify GST-UL21 in complex with StII-UL16 with StrepTactin resin or with glutathione resin and evaluated the purification with SDS-PAGE (Figure 5-9A). There were large bands in the insoluble fraction and either flow through fraction consistent with the sizes of GST-UL21 and StII-UL16. Many proteins bound the StrepTactin resin, with slight enrichment of bands around 100 kDa (GST-UL21), 70 kDa, 40 kDa, and 25 kDa. The glutathione resin captured more GST-UL21, but there was only a slight concentration of a band at 40 kDa. Together these observations suggest that most of

GST-UL21 and StII-UL16 are found in the insoluble fraction, what is soluble binds affinity resin only poorly, and the protein that does bind affinity resin does so in a non-stoichometric fashion, so the soluble portion of this complex is likely a soluble aggregate. We were also able to coexpress small amounts of GST-UL16CS (Figure 5-1) with binding partner His_6-UL11 in this system compared to single expressions; however, GST-UL16CS was not soluble (Figure 5-9B), and therefore we did not pursue purification.

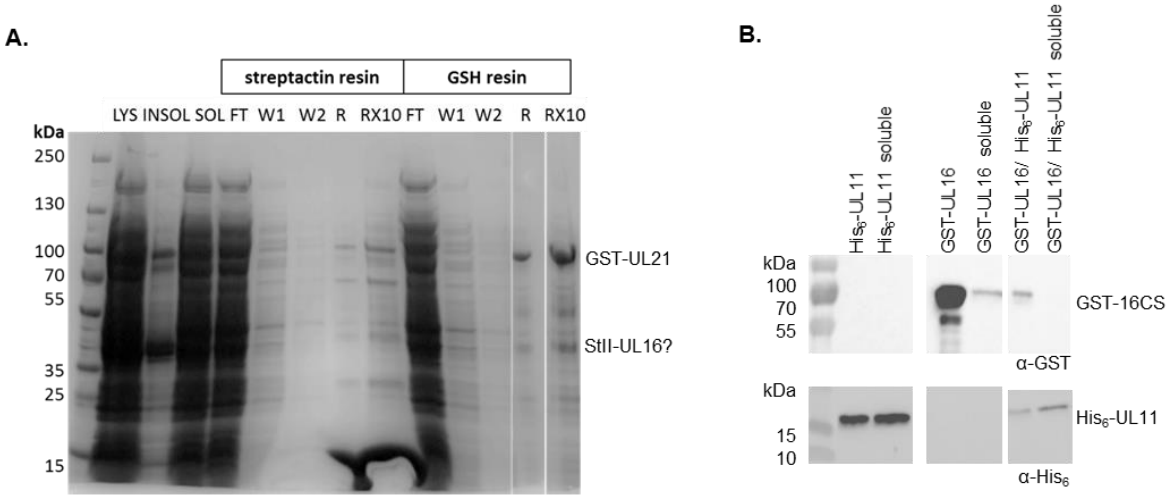


Figure 5-9. Interaction with binding partners *in vitro*
A) Purification of coexpressed StII-UL16 and GST-UL21 over StrepTactin or glutathione (GSH) resin. B) Anti-GST and anti-His western of soluble and insoluble fractions of His6-UL11 and GST-UL16CS expressed singly or together.

5.2.8 Inclusion of metal ions in expression does not improve solubility or purity of GST-UL16CS

As described above, we hypothesize that UL16 may act like Hsp33, a well-studied protein with published expression and purification protocols. The only uncommon step in the production of Hsp33 from bacteria is the inclusion of $ZnCl_2$ in the culture medium to occupy a zinc coordination site in reduced Hsp33. GST-UL16CS (Figure 5-1) was

expressed by autoinduction with and without $ZnCl_2$. Both cultures produced a portion of soluble protein, but this protein did not bind glutathione resin and is likely forming soluble aggregates (Fig 5-10).

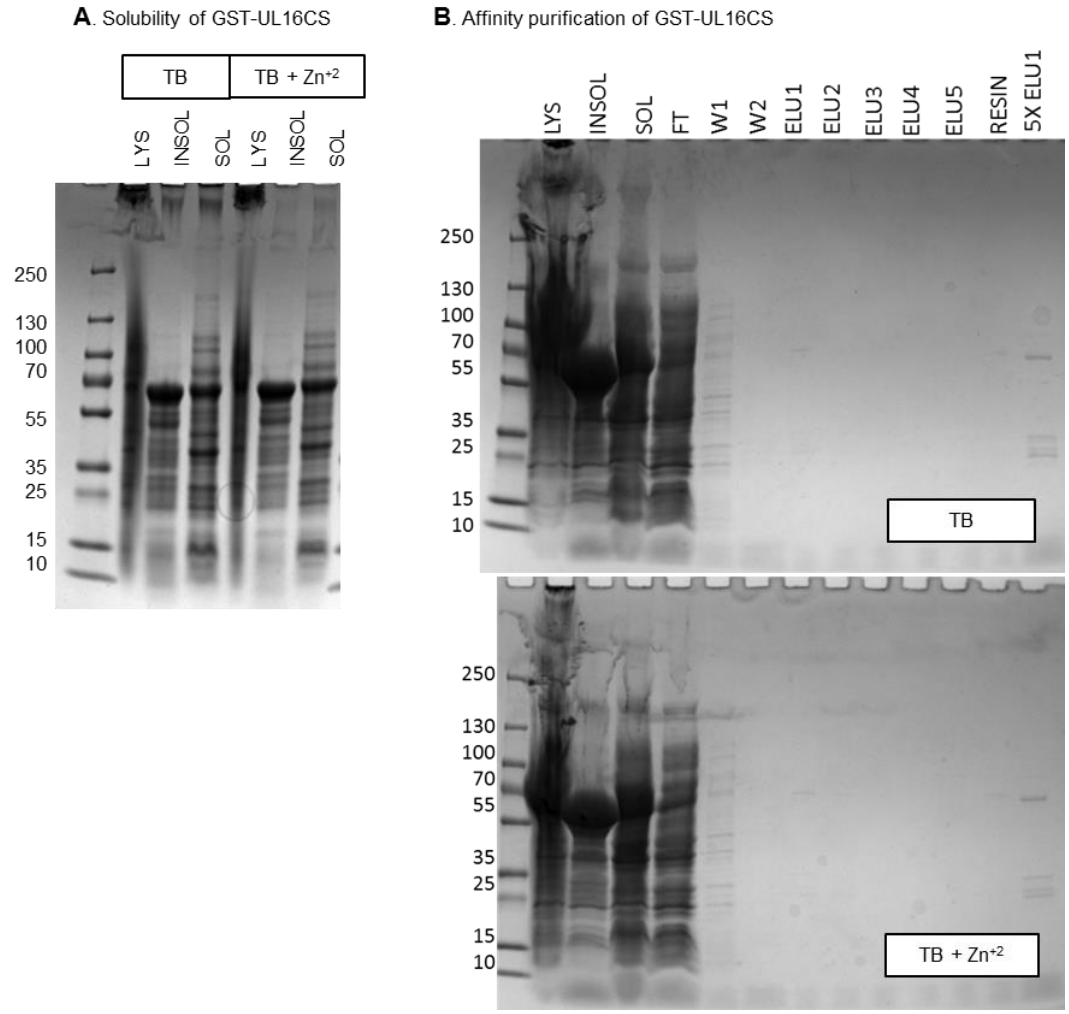


Figure 5-10. Expression and purification of GST-UL16CS with zinc
A) Coomassie-stained gel of total, insoluble, and soluble fractions of GST-UL16CS expressed in the presence of $ZnCl_2$. B) Affinity purification of GST-UL16CS with (top) or without (bottom) zinc.

5.2.9 UL16 with C-terminal affinity tags are expressed in some cell types

None of the UL16 constructs described above behaved well *in vitro* and they all featured an N-terminal affinity tag, either large (GST) or small (StII) (Figure 5-1). Many

of the experiments were performed with GST-UL16CS and it is possible that this construct is inherently flawed for two reasons: 1) When a mutant virus expresses UL16 with large N-terminal tag, the protein is excluded from the virion (Greg Smith, personal communication), suggesting that a large N-terminal tag may result in misfolded or non-functional UL16 and 2) the GST tag is a functional redox enzyme and we predict that UL16 is sensitive to changes in redox environment. Therefore, we moved to new UL16 constructs with these considerations in mind: UL16CS(1-373)-His_6 and UL16CS(41-373)-His_6 (Figure 5-1). The latter construct lacks the first 40 predicted unstructured and not well conserved residues that are also dispensable for binding UL21. The expression of these constructs was tested in four different cell types with different characteristics. T7 cells are run of the mill *E. coli* expression cells. LoBStr (low background strain) cells have been engineered to eliminate major *E. coli* contaminant proteins [87], which may eliminate the major contaminant seen in previous purifications (Figure 5-3, Figure 5-6). SHuffle cells express a chaperone and disulfide bond isomerase that helps correct disulfide bonds to form in the cytoplasm. Origami cells also allow disulfide bond formation in the cytoplasm by engineered mutations in various endogenous reductases. UL16CS-His_6 constructs were expressed from these cell types in LB with IPTG induction at either 16 or 37°C with varying degrees of success (Figure 5-11 A, B). Interestingly, UL16CS(41-373)-His_6 was expressed highly from origami cells at both temperatures (Figure 5-11B). Protein expressed at 37°C was assayed for solubility in the presence and absence of reducing agents TCEP and DTT and was found to be insoluble in both cases (Figure 5-11C). Notably, these solubility assays were performed in the fluidizer, which is directly representative of large scale purification conditions.

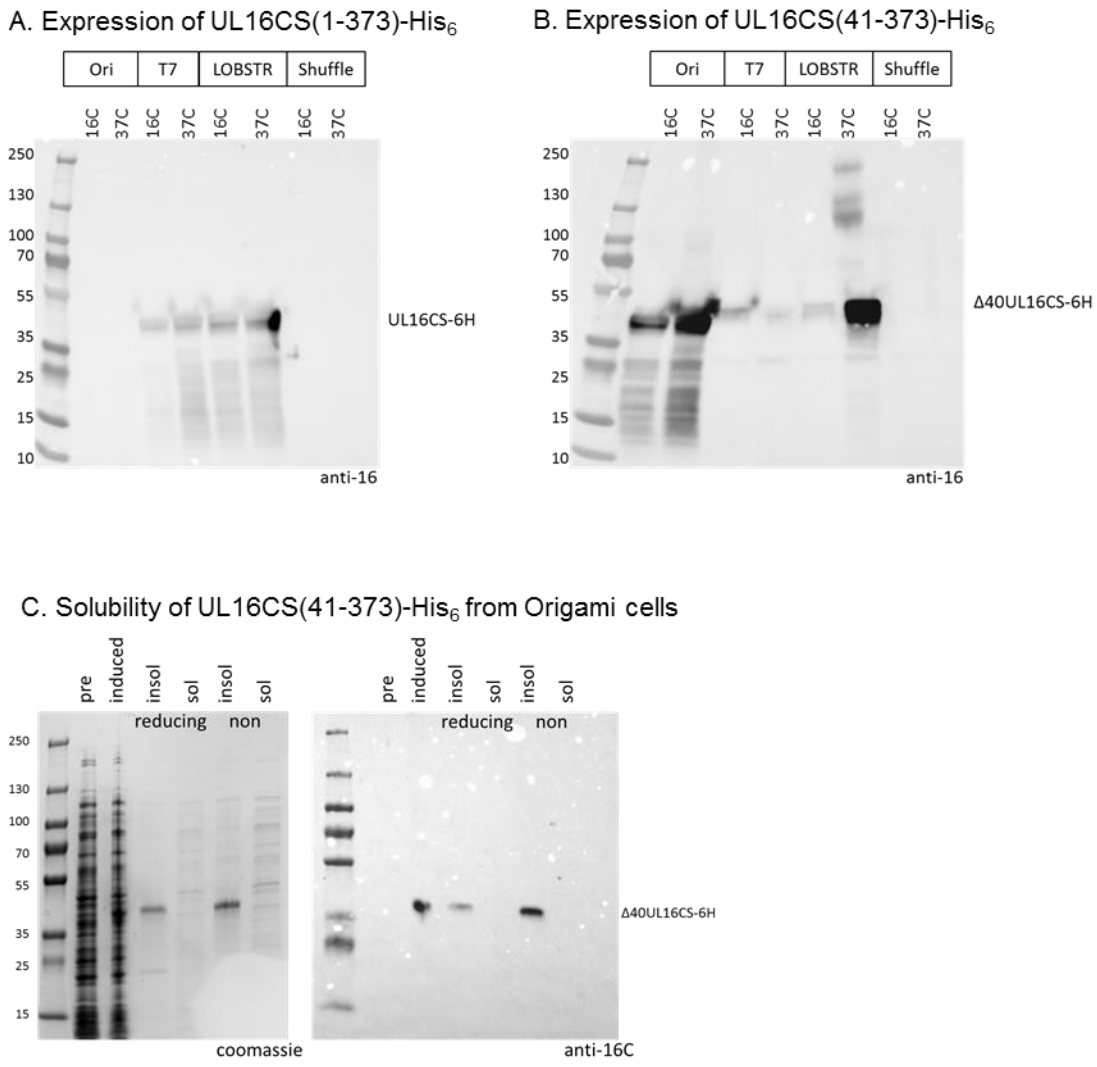


Figure 5-11. Expression and solubility of C-terminally UL16 constructs
Anti-UL16C western blots of A) UL16CS(1-373)-His6 and B) UL16CS(41-373)-His6 expressed under different conditions. C) Coomassie-stained gel and anti-UL16C western of soluble and insoluble fractions of UL16CS(41-373)-His6 in reducing or non-reducing buffer from expression in Origami B cells.

5.2.10 Coexpressing domains of UL16 produces protein in soluble aggregates despite oxidation state

Based on recent cellular experiments suggesting that the two domains of UL16 are stable *in vivo*, have separate functions, and interact [153], we decided to clone new constructs representing the domains two domains of UL16, namely UL16N(1-155)-GFP-

His_6 and UL16C(156-373)-StII, both with the CS mutations. This work suggested that a large tag after residue 155 somehow stabilizes UL16N, UL16C is functional with a small C-terminal tag, and these two domains interact in cells with these types of tags. Expression of these constructs was tested alone (Figure 5-12A, B) and together (Figure 5-12C) in three cell types and three expression conditions. We found that alone, UL16N-GFP-His6 was expressed only in LoBStr cells, but somewhat evenly in all expression conditions (Figure 5-12A). UL16C-StII was also expressed in all conditions in LoBStr cells, as well as in Origami cells grown in LB (Figure 5-12B). Nothing was expressed in Shuffle cells (Figure 5-12AB). When these constructs were coexpressed, the expression patterns changed (Figure 5-12C). While UL16N was still expressed in all cases of LoBStr expression, UL16C only expressed in LoBStr cells in the presence of UL16N when the cultures were grown at 37°C in LB. Additionally, UL16N was expressed in low level in Origami cells grown at 37°C in LB in the presence of UL16C.

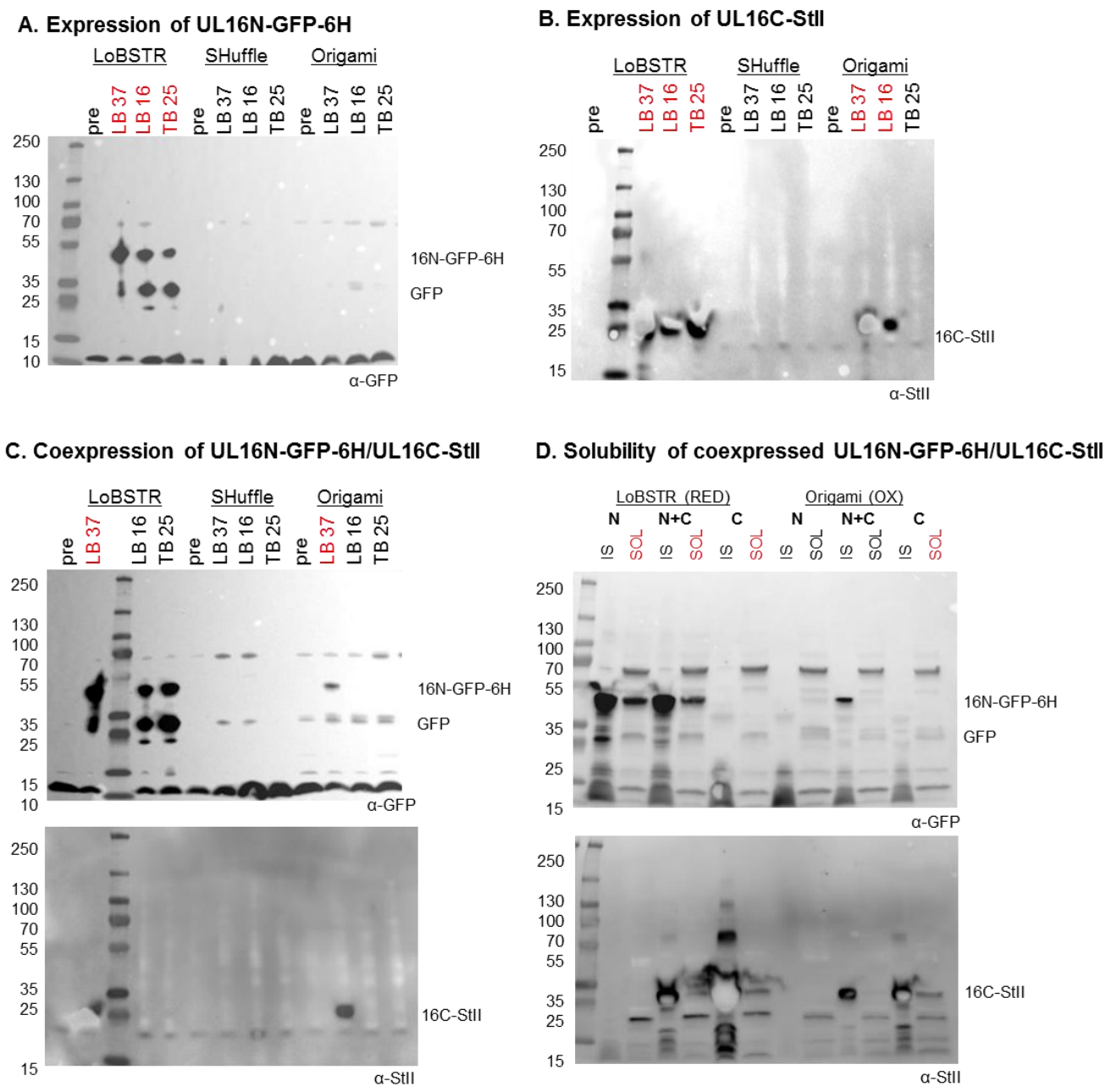


Figure 5-12. Expression and solubility of domains of UL16
Westerns of cells expressing A) UL16(1-155)-GFP-His6, B) UL16(156-373)-StII, and C) both in different expression conditions. D) Solubility of these constructs expressed alone or together. Red, reducing buffer. Ox, non-reducing buffer. Antibody marked at bottom right of blot. Promising conditions marked in red.

We then tested the solubility of these UL16N and UL16C constructs after expression alone and together in LB at 37°C in either LoBStr or Origami cells. To lyse the cells, we used lysozyme. The LoBStr cells were lysed in a reducing buffer while the Origami cells were lysed in a buffer without reducing agent in order to preserve any disulfide bonds. As before, UL16N was expressed alone in LoBStr cell and it was partially soluble (Figure 5-12D). UL16C was highly expressed alone in LoBStr cells but was mostly insoluble (Figure 5-12D). Together, these proteins were both expressed and slightly soluble in reducing buffer when coexpressed in LoBStr cells (Figure 5-12D). UL16C was expressed alone in Origami cells and slightly soluble in non-reducing buffer whereas UL16N was not expressed alone in Origami cells. Both proteins were coexpressed lowly in Origami cells, but neither protein was soluble in the non-reducing buffer (Figure 5-12D).

Although neither UL16N nor UL16C construct appeared soluble in non-reducing buffer when expressed in Origami cells (Figure 5-12D), previous work in the lab has shown that lysozyme lysis can underestimate the solubility of large proteins. Therefore, we attempted to purify these proteins from a larger scale culture. Figure 5-13A shows that while nothing binds or elutes from StrepTactin resin, there are many bands that elute from nickel resin including bands that are consistent with the sizes of the UL16N construct and the UL16C construct. Furthermore, the elution fraction is green consistent with the presence of GFP. However, western blots probing for UL16 domains or affinity tags (Figure 5-13B) show that only GFP exists in the nickel resin elution and these proteins are indeed insoluble in non-reducing buffer when coexpressed in Origami cells.

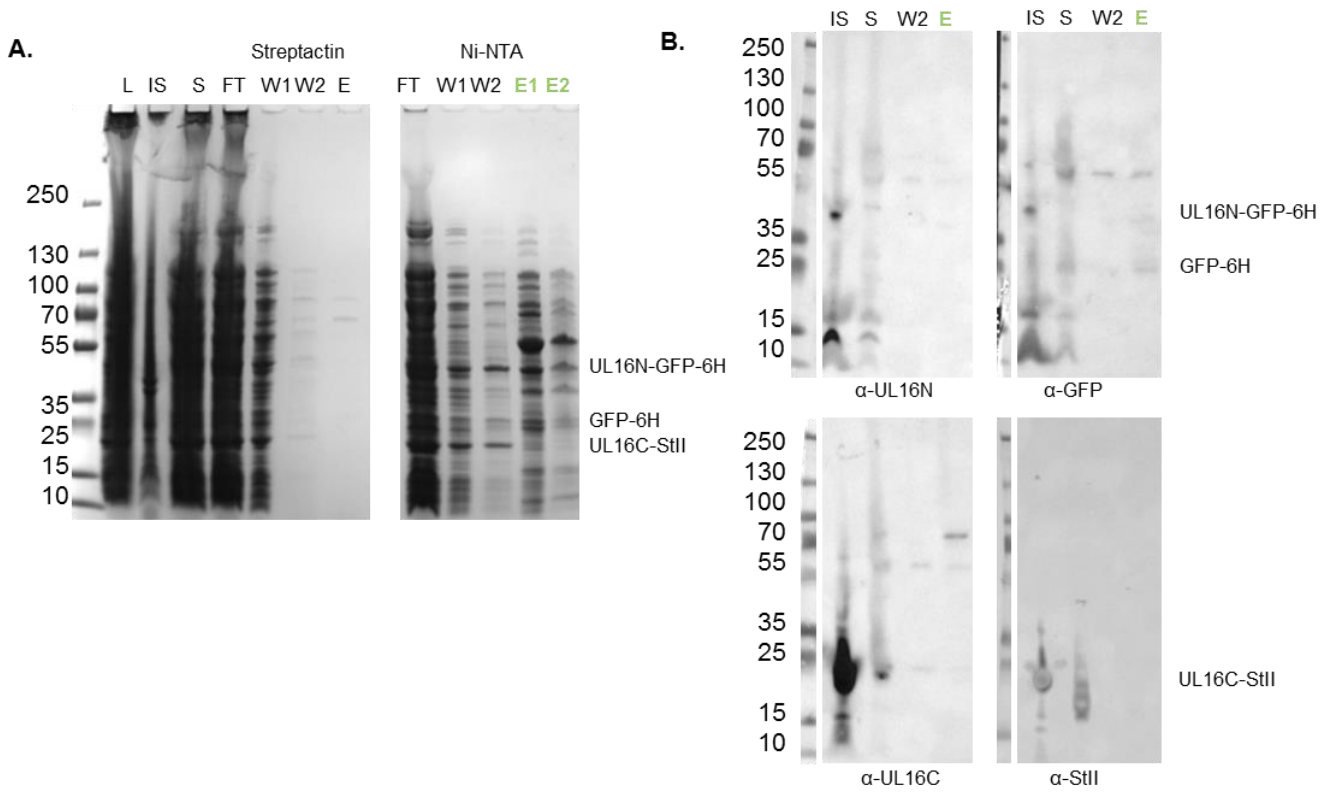


Figure 5-13. Purification of coexpressed UL16N and UL16C in non-reducing conditions
A) Coomassie-stained gel of non-reducing affinity purification of UL16N-GFP-6H coexpressed with UL16C-StII in Origami B cells. Green labels indicate color of solution. B) Western blots of fractions from purification. Antibodies labeled below blots.

Since the UL16N and UL16C constructs were highly coexpressed in LoBStr cells (Figure 5-12D), we attempted to purify the complex from a large scale culture. The cells were mechanically lysed in reducing buffer and passed over nickel resin to capture the His_6 tag. Indeed, bands in the elution fraction appeared to represent UL16N around 40 kDa, UL16C around 25 kDa, and the common UL16 contaminant around 70 kDa (Figure 5-14A). These bands did not bind to StrepTactin resin (Figure 5-14A). Western blots probing for either UL16 domain or affinity tag revealed that while UL16N was partially soluble and bound to nickel resin, only a portion of UL16C was soluble and was not

eluted nickel resin indicating that it was not bound to UL16N (Figure 5-14B). This is consistent with the observation that nothing bound StrepTactin because there was nothing in the input with the correct affinity tag.

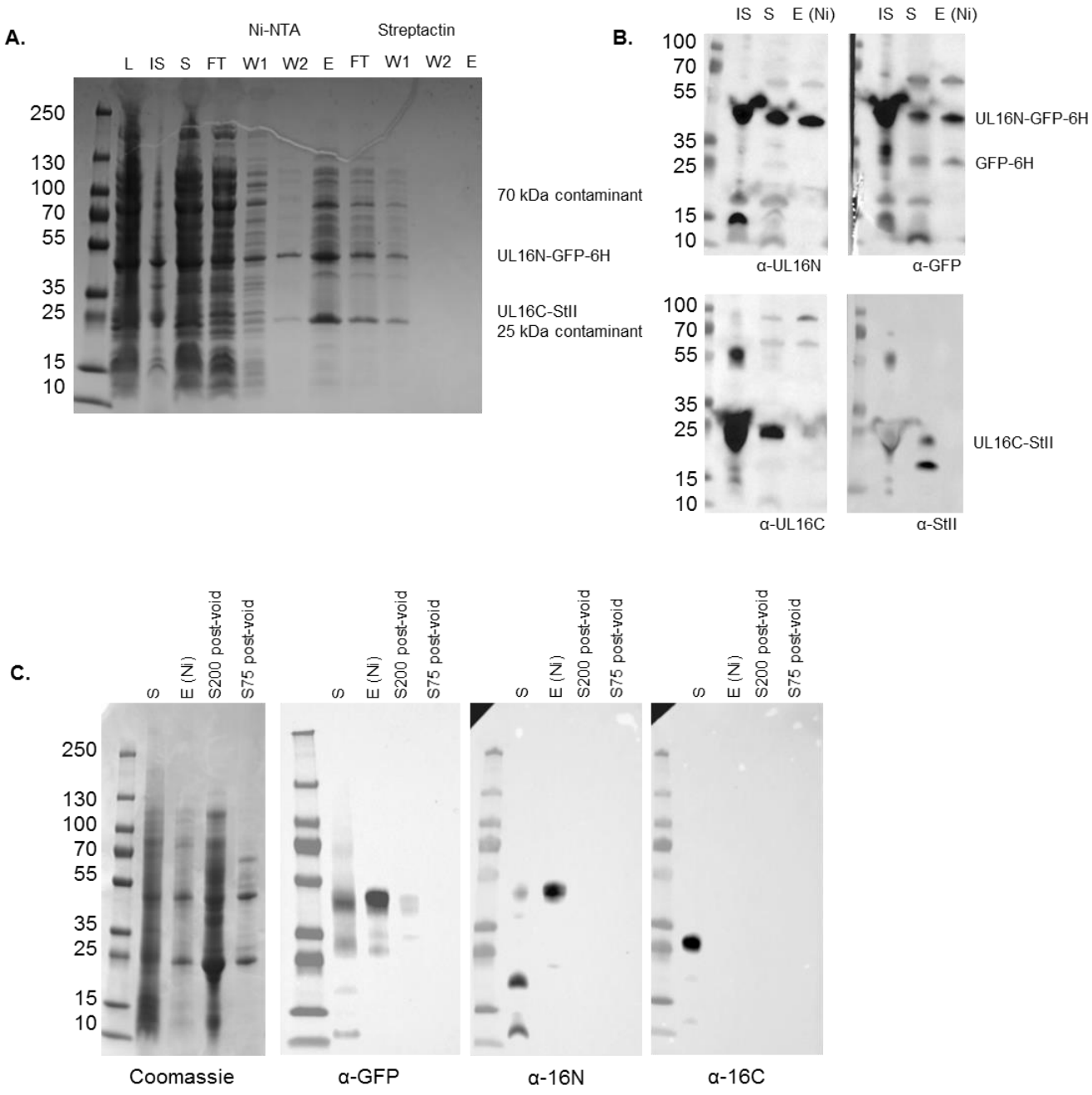


Figure 5-14. Purification of coexpressed UL16N and UL16C in reducing conditions
A) Coomassie-stained gel of reducing affinity purification of UL16N-GFP-6H coexpressed with UL16C-StII in LoBStr cells. Green labels indicate color of solution. Western blots of fractions from B) affinity and C) size exclusion purification. Antibodies labeled below blots.

Since there was abundant, impure UL16N, the elution from nickel resin was subjected to S200 and S75 size exclusion chromatography (Figure 5-14C) to clean it up and further evaluate its solubility. While there was a band below 55 kDa consistent with UL16N in fractions after the void, this band was not recognized with antibodies raised to UL16N or the GFP tag. As expected, the 25 kDa band was not recognized with an antibody for UL16C, consistent with its previous diagnosis as a contaminant (Figure 5-14B). Together we conclude that when UL16N-GFP-His6 is coexpressed with UL16C-StII, both proteins are nominally soluble, but likely exist in a soluble aggregate because UL16C-StII does not bind StrepTactin resin and UL16N-GFP-His6 is not found after the void volume of a size exclusion column. Confusing matters further is the observation that in addition to the common 70 kDa protein, these proteins copurify with contaminants around 25 kDa and 55 kDa that match the size of the UL16C and UL16N constructs, respectively, and these off target proteins were the predominant species purified by size exclusion.

5.3 Discussion

In this chapter we have described that UL16 was extremely hard to solubilize when expressed in *E. coli*, despite the use of distinct and differentially placed affinity tags, truncated constructs, constructs with non-conserved cysteines mutated to serines, and strains of *E. coli* engineered to mediate disulfide bond formation. When able to get any soluble UL16, it was aggregated and didn't bind the affinity resin, or remained in a complex with chaperones and unknown contaminants in the void. Future work will focus on strategies to isolate pure, soluble, UL16, likely in the presence of binding partners, in order to better understand this molecule's complex conformational rearrangement.

Chapter 6: UL51

6.1 Introduction

HSV-1 UL51 is a conserved, 244-amino acid tegument protein that associates with golgi-derived membranes through palmitylation at its N-terminus[65]. UL51 has been shown to associate with a number of other viral proteins including gE [46] and directly bind tegument proteins UL14 and UL7[64, 66, 68]. Through these interactions, UL51 is thought to play a structural role linking the envelope to the capsid. Indeed, infection with mutant HSV-1 lacking UL51, UL7, or UL14, UL51 and UL7, or UL51 and UL14 leads to an accumulation of partially enveloped single capsids in the cytoplasm [64, 68]. However, it is likely that these proteins have independent functions and roles outside of secondary envelopment. Specifically, the UL51/UL7 complex has been shown to maintain uninfected cell-like morphology in infected cells in a gE-independent mechanism[64]. Stabilizing cell-cell contacts may be an important proponent to cell-cell spread, but in HSV-1 infected cells, UL51 has been found in a complex with gE, a major regulator of cell-cell spread, and may also contribute to spread this way [46]. Like many other tegument proteins, it is unclear exactly what functions UL51 performs and how it accomplishes them. To begin to understand this multifunctional protein, we characterized it biochemically and structurally and found it to be an alpha-helical molecule with a structured core, conformational flexibility, and the ability to polymerize.

6.2 Results

6.2.1 Expression and purification of UL51

Our first characterization target was the full length protein and previous work in the lab showed that a His_{10}-SUMO at the N-terminus (Figure 6-1) produced soluble protein for the full length construct. Early work on this protein suggested that it was

prone to aggregation, so initial lysis and purification buffers include high amounts of salt (0.5M) and glycerol. The protein was first purified by nickel resin, to which we have previously seen UL51 take an extreme affinity for, even without a His_{10}-tag. A step gradient of increasing amounts of imidazole was added to improve yield in this first purification step, but little additional protein was gained and cleavage efficiency of PreScission protease, the next step in purification, decreased in the new elution buffer conditions (Figure 6-2). Elution fractions were incubated with PreScission protease overnight and the protease efficiently cleaved the His_{10}-SUMO tag from GPGS-UL51 (Figure 6-1). Magnesium and ATP were also added to loosen common copurified chaperones.

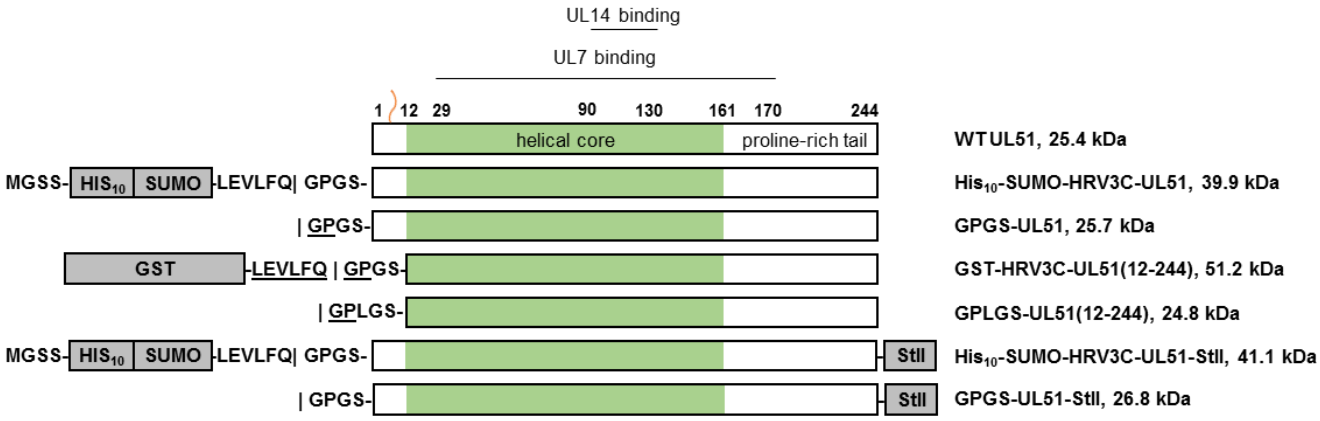


Figure 6-1. UL51 construct map
Expression constructs for UL51. The predicted helical core, proline-rich tail, and binding sites for UL14 and UL17 are marked. The palmitylation site is identified with a wavy vertical line. Protease recognition sites are underlined and cleavage sites are marked with a vertical line.

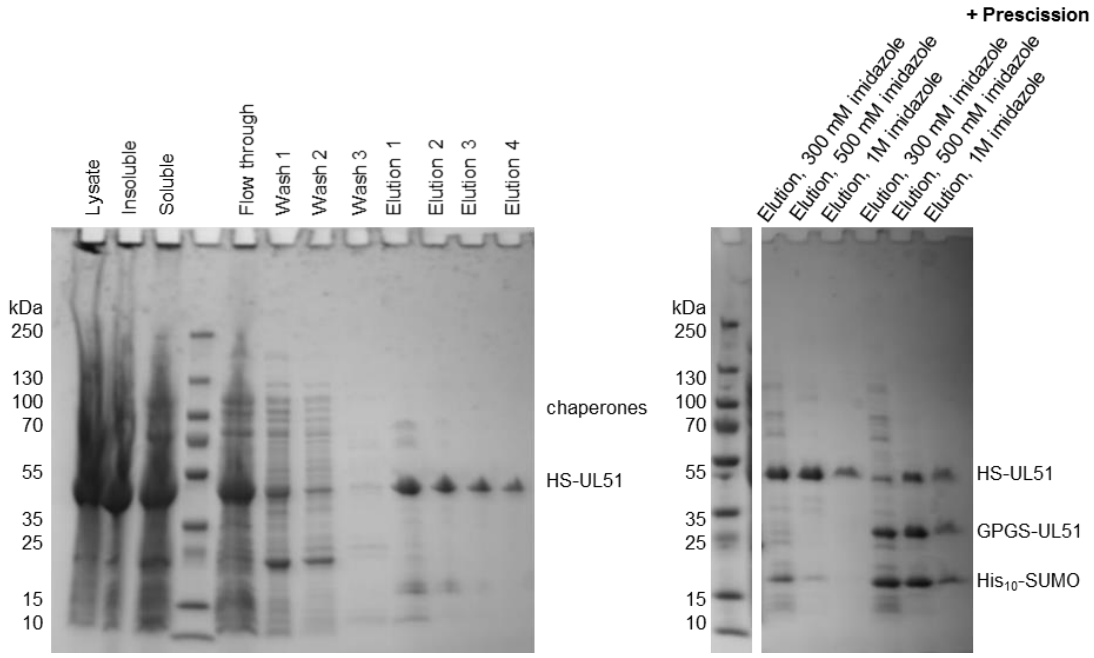


Figure 6-2. Purification of GPLGS-UL51 and its affinity for nickel
Coomassie stained gels of (L) affinity purification, (R) PreScission cleavage, and attempt to liberate UL51 from nickel resin with excess imidazole.

Due to tag-less UL51's affinity for nickel resin, the straightforward methods for His_{10}-SUMO separation of passage over nickel resin before or during size exclusion chromatography were not applicable, and instead GPGS-UL51 was separated from His_{10}-SUMO using size exclusion chromatography on an S75 superdex column. This was also our first indicator that UL51 was either elongated or forming multimers as the peak for GPGS-UL51, a ~25 kDa protein that would be expected to fall well within the included volume of an S75 column, overlapped with the void (Figure 6-3). The GPGS-UL51 protein eluted from the S75 was found to still contain contaminants (Figure 6-3), so the protein was further purified with ion exchange using Q sepharose, an anion exchanger. As the isoelectric point of GPGS-UL51 is calculated to be 5.88 and the purification buffers were pH 7.0, GPGS-UL51 was expected to bind the Q column. Surprisingly, most

of GPGS-UL51 flowed through this resin, but since the offending contaminants bound the resin, this remained an effective purification step (Figure 6-4). After removal of contaminants with ion exchange, aggregation-prone UL51 was purified with S200 size exclusion to separate aggregates from mono-dispersed protein (Figure 6-5A). After all of these purification steps, the concentrated monodispersed peak of GPGS-UL51 still retained some contaminants and was potentially being proteolysed (discussed below). The protein was set up anyway in sitting drops to screen for crystallization conditions and no hits were found. We predicted that impeding crystallization were three main hurdles: lasting impurities, aggregation, and potentially proteolysis. Attempts to maneuver these hurdles are discussed below.

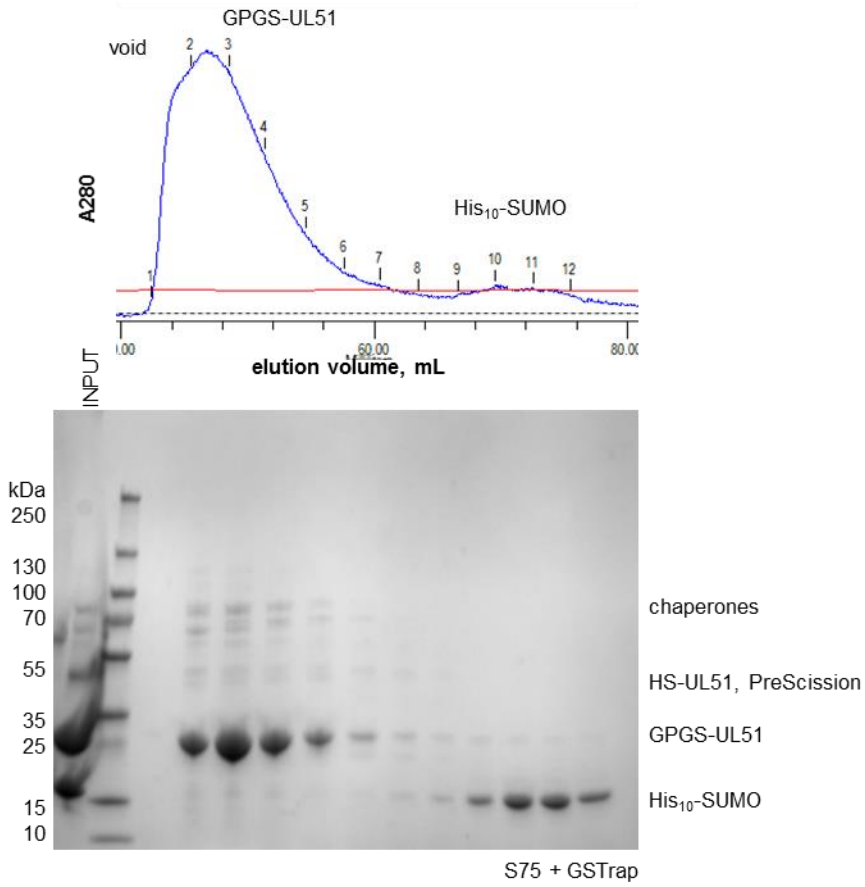


Figure 6-3. S75 size exclusion purification of GPGS-UL51
(Top) chromatogram and (bottom) Coomassie-stained gel of fractions from S75 purification of GPGS-UL51 to separate from HisSUMO.

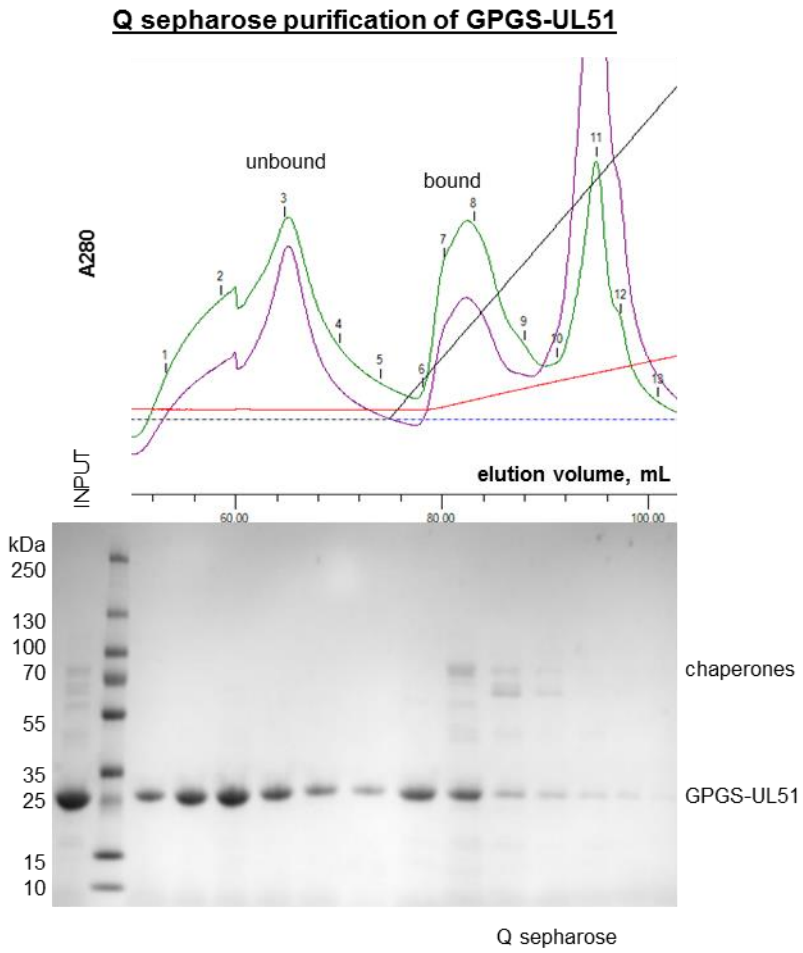


Figure 6-4. Ion exchange purification of GPGS-UL51
(Top) chromatogram and (bottom) Coomassie-stained gel of fractions of purification of GPGS-UL51 over Q sepharose to separate from chaperones. A260 trace in purple, A280 trace in green.

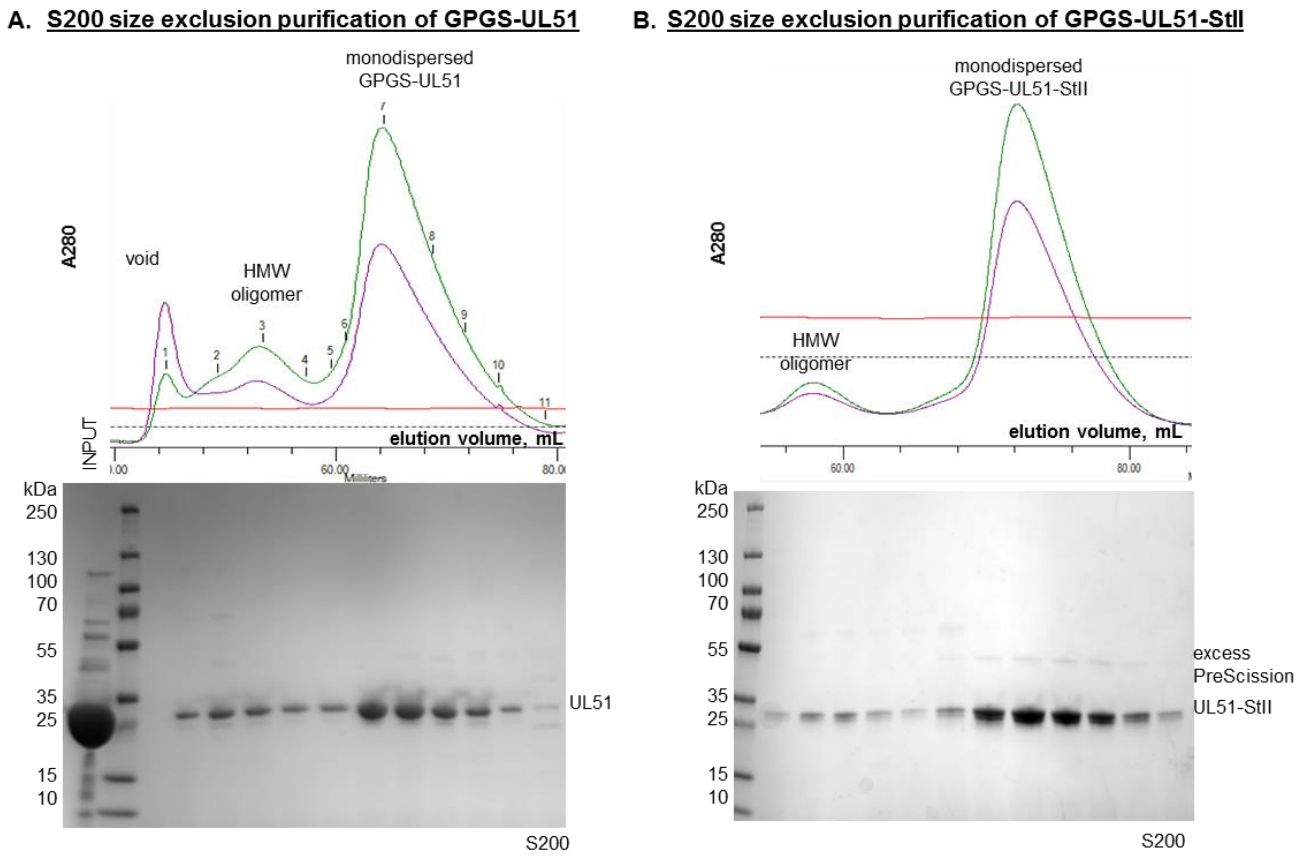


Figure 6-5. Size exclusion purification of GPGS-UL51 and GPGS-UL51-StII
S200 size exclusion purification including (top) chromatogram and (bottom) Coomassie-stained gel of fractions for (A) GPGS-UL51 and (B) GPGS-UL51-StII. A260 trace in purple, A280 trace in green.

In order to improve the purity of full length UL51, a StII-tag was added to C-terminus in addition to the N-terminal His_{10}-SUMO tag (Figure 6-1) based on the observation in our lab that the StII purification system generally produces very clean protein. The StII tag was added to the HisSUMO construct rather than used as an alternative affinity tag because previous efforts to express StII-UL51 or UL51-StII produced insoluble protein, if it was even expressed. The HS-UL51-StII construct was expressed, soluble, and purified by nickel resin like HS-UL51 and then purified by StrepTactin resin, which produced very pure protein, as predicted. This protein was also efficiently cleaved with PreScission protease and separated from the His_{10}-SUMO tag

with S75 size exclusion. Despite its relative purity, GPGS-UL51-StII was purified by ion exchange, which ended up being unnecessary. The protein was then purified by S200 size exclusion and eluted at a similar volume GPGS-UL51, suggesting that the C-terminal tag does not greatly alter the protein's oligomeric conformation (Figure 6-5B).

6.2.2 UL51 is folded and alpha-helical

Our first attempt to decrease the aggregation of UL51 was to identify more optimal buffer conditions using the thermofluor assay. In thermofluor, protein is mixed with a dye in a variety of buffer conditions and subjected to a temperature gradient. The dye fluoresces when it comes in contact with the hydrophobic interior of a protein and by monitoring fluorescence over time, you can identify the melting point of a protein. We interpret an increase in melting temperature as an increase in protein stability and therefore seek those conditions for purification and crystallization. Over multiple attempts, however, all UL51 constructs showed melting curves with an overall negative slope in comparison to sigmoidal curves for well folded UL21C and PreScission protease (Figure 6-6B). These aberrant melting curves are consistent with a protein that is unfolded or disordered in comparison to a folded protein like UL21C that shows a sigmoidal fluorescence signal. In contrast, analyzing GPGS-UL51 with circular dichroism shows that this is a well-folded, mostly alpha helical protein. The predicted secondary structure suggestion that UL51 is 49% alpha helical (122 out of 248 residues for GPGS-UL51, Figure 6-6C) may be an over estimation as CD measures 38% helical content and 14% strand content; however, the total content of secondary structure (161 out of 244 residues, ~66%) is consistent with CD measurements (70%) (Figure 6-6A, Table 6-1). We concluded that thermofluor was not a reasonable assay to perform with

this protein for identification of protein stability, potentially due to the likely flexible C-terminal region “soaking up” the hydrophobic dye, and that GPGS-UL51 has some secondary structure, despite propensity for aggregation.

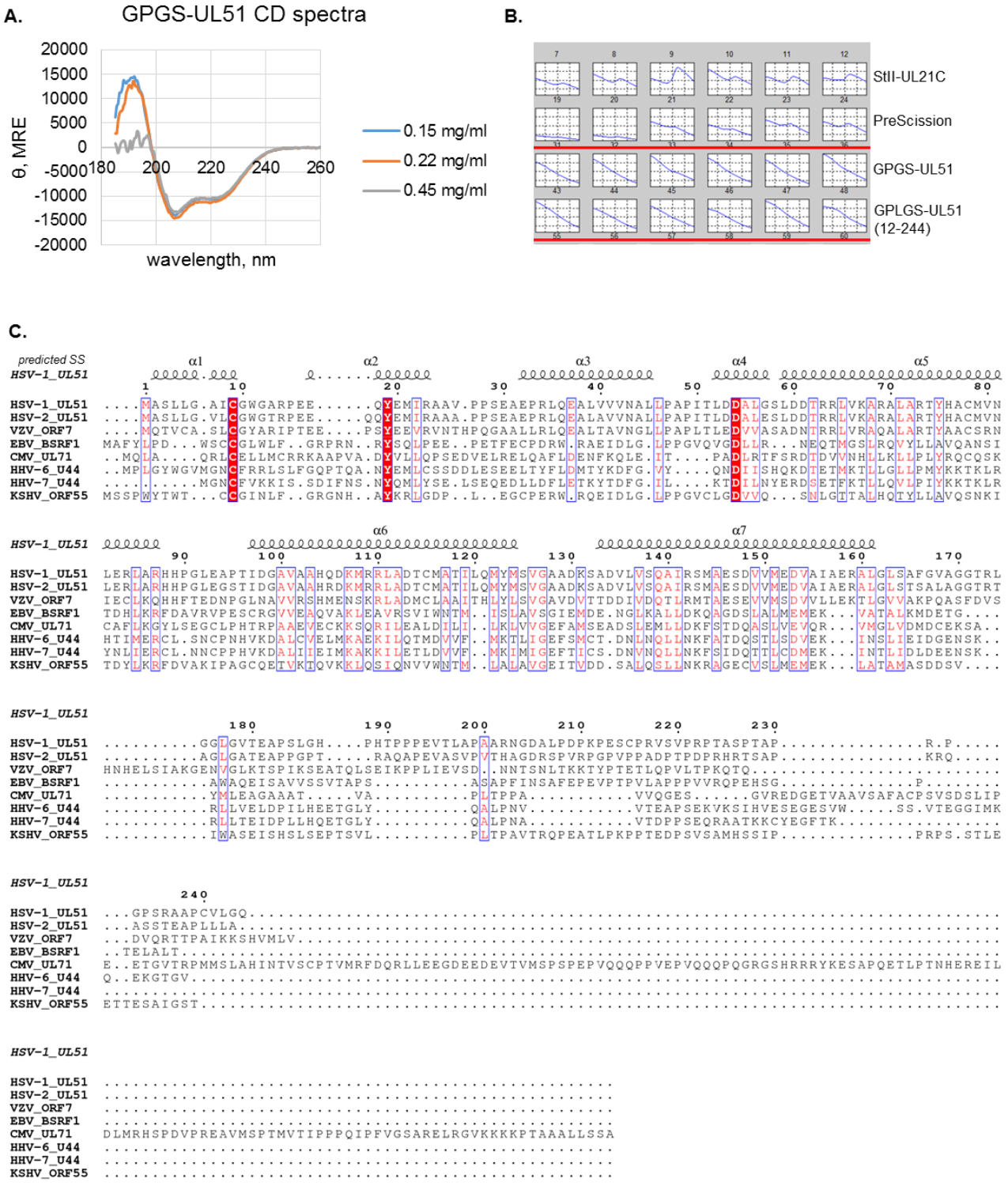


Figure 6-6. Secondary structure of UL51
A) CD spectra, B) thermofluor melting curves, and C) sequence alignment of human herpesvirus UL51 homologs. Secondary structure predicted for HSV-1 UL51 is shown above the alignment.

Table 6-1. Secondary structure of UL51 from CD

Concentration (mg/mL)	Ref. Set	Analysis Program	NRMSD	Helix1	Helix2	Strand1	Strand2	Turns	Coil
0.45	SP175	CDSSTR	0.023	0.24	0.16	0.04	0.05	0.14	0.36
	Set 6	CDSSTR	0.016	0.21	0.14	0.11	0.06	0.17	0.31
	Set 3	CDSSTR	0.015	0.23	0.17	0.11	0.07	0.19	0.23
0.22	SP175	CDSSTR	0.019	0.19	0.16	0.08	0.07	0.15	0.36
	Set 6	CDSSTR	0.023	0.22	0.18	0.06	0.07	0.21	0.26
	Set 3	CDSSTR	0.023	0.25	0.19	0.06	0.07	0.2	0.23
0.15	SP175	CDSSTR	0.019	0.17	0.15	0.1	0.08	0.15	0.36
	Set 6	CDSSTR	0.03	0.2	0.16	0.05	0.06	0.2	0.33
	Set 3	CDSSTR	0.025	0.21	0.19	0.06	0.07	0.22	0.26
			average	0.213	0.167	0.074	0.067	0.181	0.300
				% alpha	0.380	**% beta**	0.141		
						% struct	0.702	**% coil**	0.300

6.2.3 UL51 comprises a folded core with N- and C-terminal extensions

We then sought to define the structured domain boundaries of this protein for further construct design, with the idea that the half of the molecule without predicted regular structure may be promoting aggregation and nonspecific binding. Furthermore, the observation of a very slightly lower band appearing in SDS-PAGE after size exclusion purification of GPGS-UL51-StII and GPGS-UL51 (Figure 6-7A) made us hypothesize that this protein was undergoing proteolysis during purification by proteases contaminating the preparation or our equipment. To test for the presence of domains, limited proteolysis with trypsin or chymotrypsin was performed on purified GPGS-UL51 and the banding pattern that emerged suggested the presence of structured cores of approximately 15 kDa, 10 kDa, and several smaller fragments that are resistant to proteolysis (Figure 6-7C). Digested GPGS-UL51 and contaminated GPGS-UL51-StII were further analyzed by mass spectrometry and N-terminal sequencing to identify the boundaries of these domains. Since the two bands seen on a coomassie gel after

purification of GPGS-UL51-StII were both recognized on a western blot probing for the StII tag (Figure 6-7A), we hypothesized that the contaminating protease was removing a short part of the N-terminus of UL51. Indeed, mass spectrometry of proteolysed GPGS-UL51-StII showed two peaks in the size range of the full length protein (Figure 6-7B). The difference between these peaks is consistent with the removal of the first ten residues of UL51-StII (Figure 6-7B). Furthermore, N-terminal sequencing of chymotrypsinized GPGS-UL51 identified that the largest species began at residue 12, the intermediate band around 15 kDa was predominantly a species that began at residue 12 as well as a fragment that began at residue 20, and the smallest band isolated in this experiment around 10 kDa was a species beginning at residue 167 with traces of a species beginning at residue 164 (Figure 6-7C, D). Taken together, we concluded that the full length UL51 constructs previously purified were being clipped at the N-terminus and therefore we decided to pursue constructs representing UL51 (12-244). Furthermore, there may be a second domain in the C-terminal region that is predicted to be disordered (discussed below).

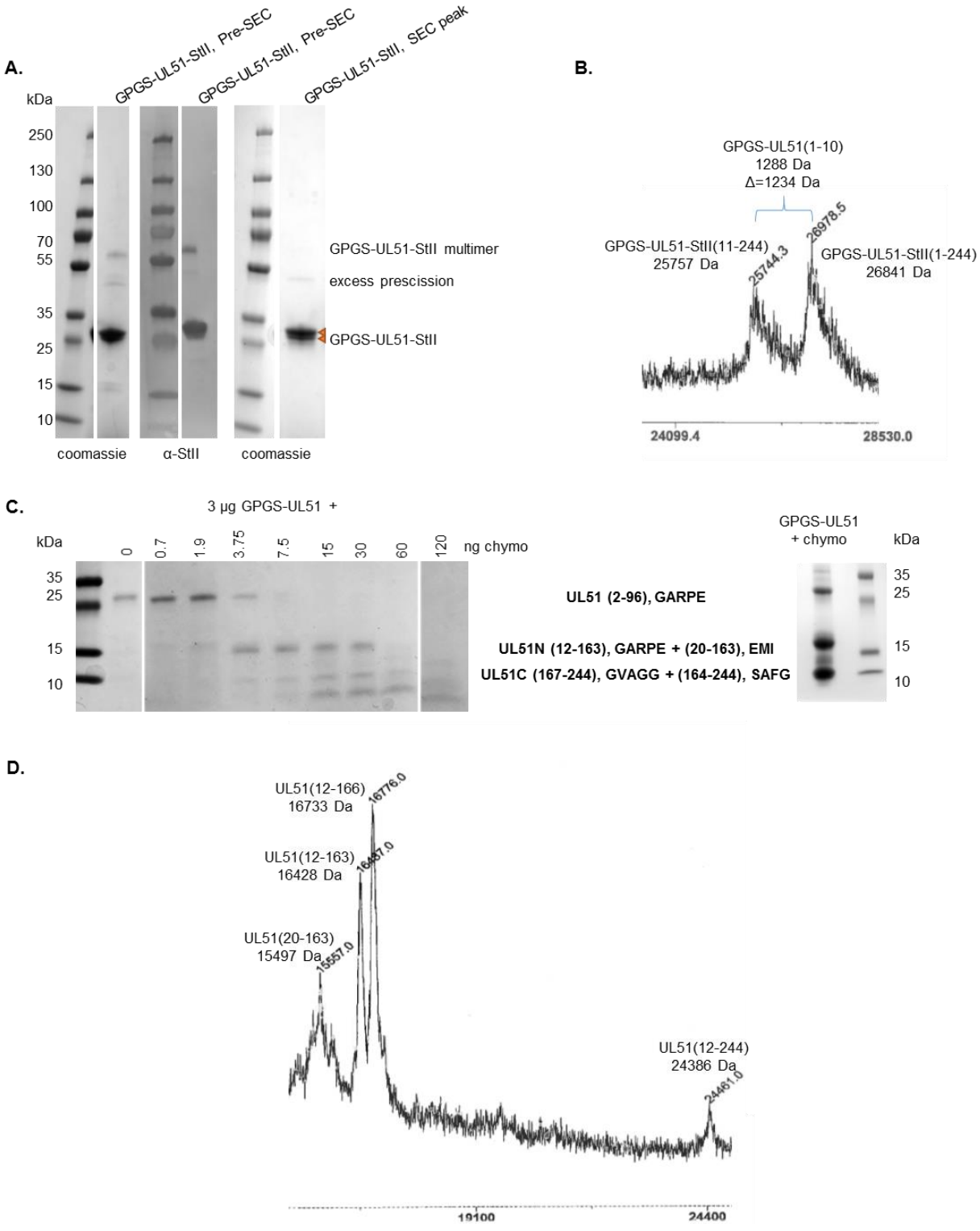


Figure 6-7. Assignment of structural boundaries in UL51
A) N-terminal cleavage of UL51 and UL51-StII in storage. B) Mass spectrum of samples from A. C) Coomassie-stained gels of chymotrypsin limited proteolysis of GPGS-UL51 with assignments from N-terminal sequencing. D) Mass spectrum of samples from C.

The first available N-terminally truncated UL51 construct was GST-UL51(12-244) (Figure 6-1). This construct produced soluble protein in small scale assays, but this construct did not bind glutathione resin very well (Figure 6-8A) and eluted with two contaminants including a chaperone protein (Figure 6-8B). GST-UL51(12-244) that was eluted from glutathione was efficiently removed from the GST-tag with PreScission protease (Figure 6-8B). GPGS-UL51 (12-244) was then subjected to S75 size exclusion purification, wherein it eluted just after the void like full length protein, suggesting that it too is aggregating (potentially with chaperone) or multimerizing (Figure 6-8C).

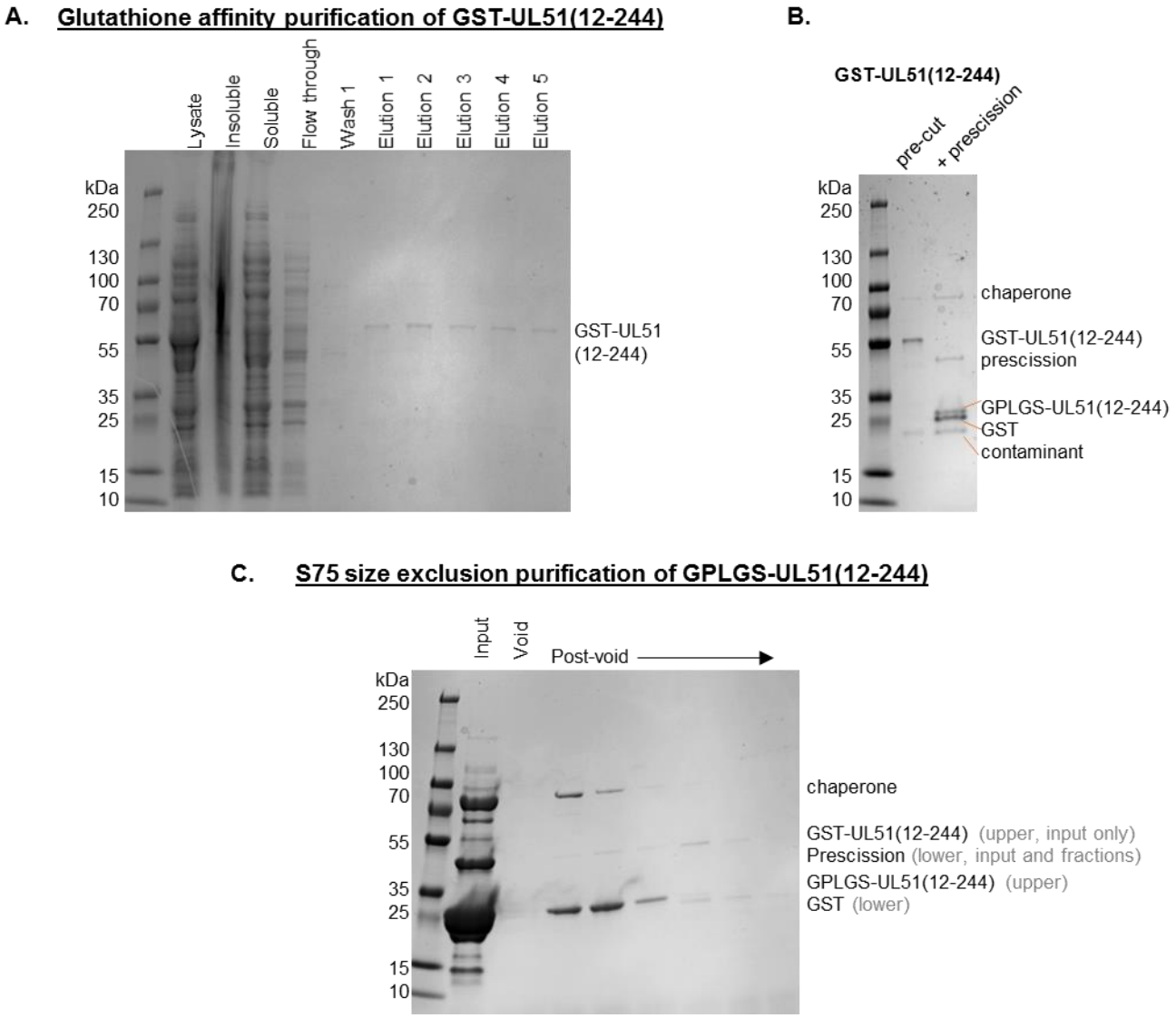


Figure 6-8. Purification of GPLGS-UL51 (12-244)
Coomassie-stained gels of A) glutathione affinity purification, B) tag removal, and C) S75 size exclusion purification of GPLGS-UL51 (12-244).

As mentioned above, limited proteolysis of GPGS-UL51 digested with chymotrypsin also identified major fragments at approximately 15 kDa and 10 kDa on SDS-PAGE. N-teminal sequencing combined with mass spectrometry identifies the 15 kDa band as a mixture of 12-163, 12-166, and 20-163 (Figure 6-7D). There were also trace signals in the mass spectrum (not shown) to suggest that the 10 kDa band beginning

at residue 167 or 164 is a mixture of 164-244, and 167-244. Taken with secondary structure prediction, we conclude that UL51 comprises a flexible N-terminus from 1-11, the structured core of the molecule from 12-163, and a C-terminal tail from 164-244 that is extremely proline-rich and predicted to be disordered, yet resists proteolysis starting at residue 167.

6.2.4 UL51 forms multimers and higher ordered structures

Since no UL51 construct has yet crystallized, we turned to SAXS to characterize this molecule in solution. In particular, we used SEC-SAXS which is critical for UL51 due to its propensity to aggregate. Indeed, after post-purification storage and travel, when previously monodispersed GPGS-UL51 (Figure 6-5A) was run over size exclusion at the beam line, the majority of the protein had shifted to the void (Figure 6-9A). This shift could be the result of random aggregation of UL51 after time and concentration or the formation of higher order multimers. To further investigate the ability of UL51 to form ordered assemblies, chemical crosslinking was performed with GPGS-UL51-StII from the monodispersed size exclusion peak. Even without crosslinker, the potentially dimeric species is seen, but with increasing amounts of crosslinker, larger multimers appear first above the 250 kDa marker and then large enough to be excluded from the gel (Figure 6-10). This result suggests that UL51 is forming ordered multimers in solution.

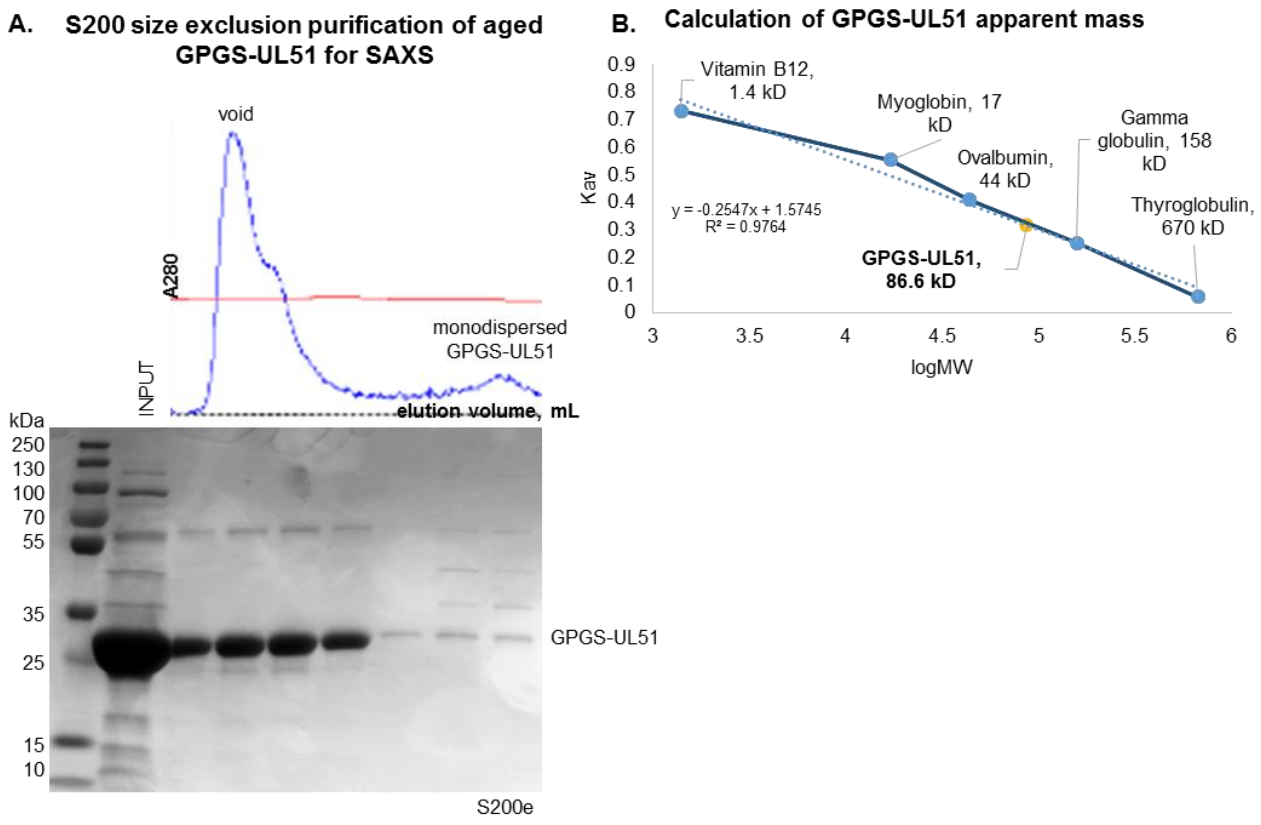


Figure 6-9. Characterization of GPGS-UL51 aggregation over time
A) (top) Chromatogram and (bottom) Coomassie-stained gel of S200 size exclusion purification fractions of aged GPGS-UL51. B) Calculation of GPGS-UL51 apparent mass, for protein that has not shifted to the void.

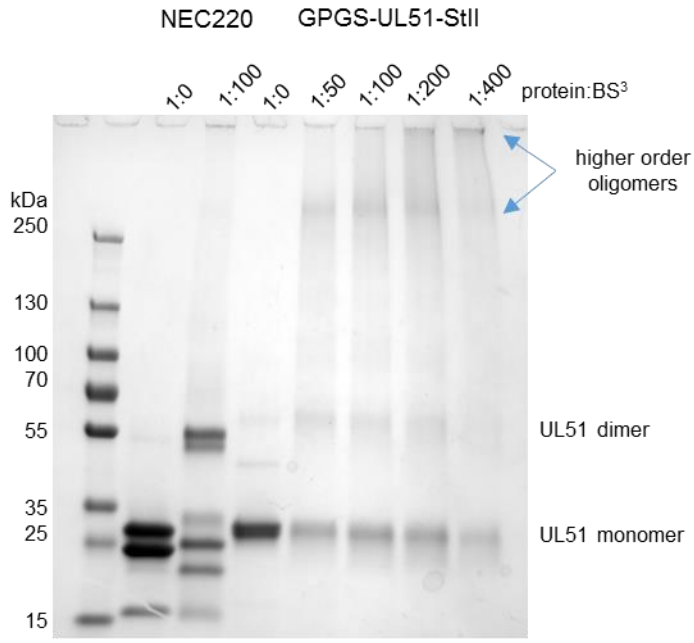


Figure 6-10. Crosslinking UL51
Coomassie-stained gel of GPGS-UL51-StII incubated with increasing amounts of homobifunctional amine-reactive BS^3 crosslinker. NEC220 is shown as a positive control.

6.2.5 UL51 is dimeric, elongated, and flexible

Despite the shift of the majority of UL51 to higher order multimers in SEC-SAXS, there was a small amount of diffraction data from monodispersed protein suitable for preliminary analysis (Figure 6-9A). Firstly, the apparent molecular weight of monodispersed material was calculated against calibration standards and found to be 86.6 kDa (Figure 6-9B), which for this ~25 kDa protein could represent a trimer or an elongated dimer. Frames representing monodispersed GPGS-UL51 from SEC-SAXS with as consistent as possible Rg values were selected and averaged and buffer frames averaged prior were subtracted. Guinier analysis of the resulting SAXS curve gave an Rg of 35.1 Å (Figure 6-11A). The distance distribution function [p(r)] gave GPGS-UL51 an Rg 38.8 Å and a Dmax of 123 Å (Figure 6-11B). In comparison, a monomer of bovine

serum albumin (BSA), a globular protein of 66 kDa, has an Rg of approximately 28 Å from either method and a Dmax of 82 Å (https://www.sasbdb.org/data/SASDBT4/). Furthermore, the shape of the p(r) function for GPGS-UL51 was particularly extended and does not take on the symmetrical bell shape that is typical of compact, globular proteins (Figure 6-11B). Together, these features of the SAXS data suggest that UL51, which we expect to be multimeric, is also an elongated molecule. These conclusions literally took shape in *ab initio* modelling of the p(r) function in DAMMIN, which output elongated models (Figure 6-11C). When averaged, although DAMAVER had a hard time superimposing these somewhat divergent models, these models have two clear lobes of density which suggests that based on this method UL51 is dimeric (Figure 6-11C). When the averaged model is filtered to remove loosely associated beads, dimeric UL51 appears as two crescents associated end to end longways (Figure 6-11C).

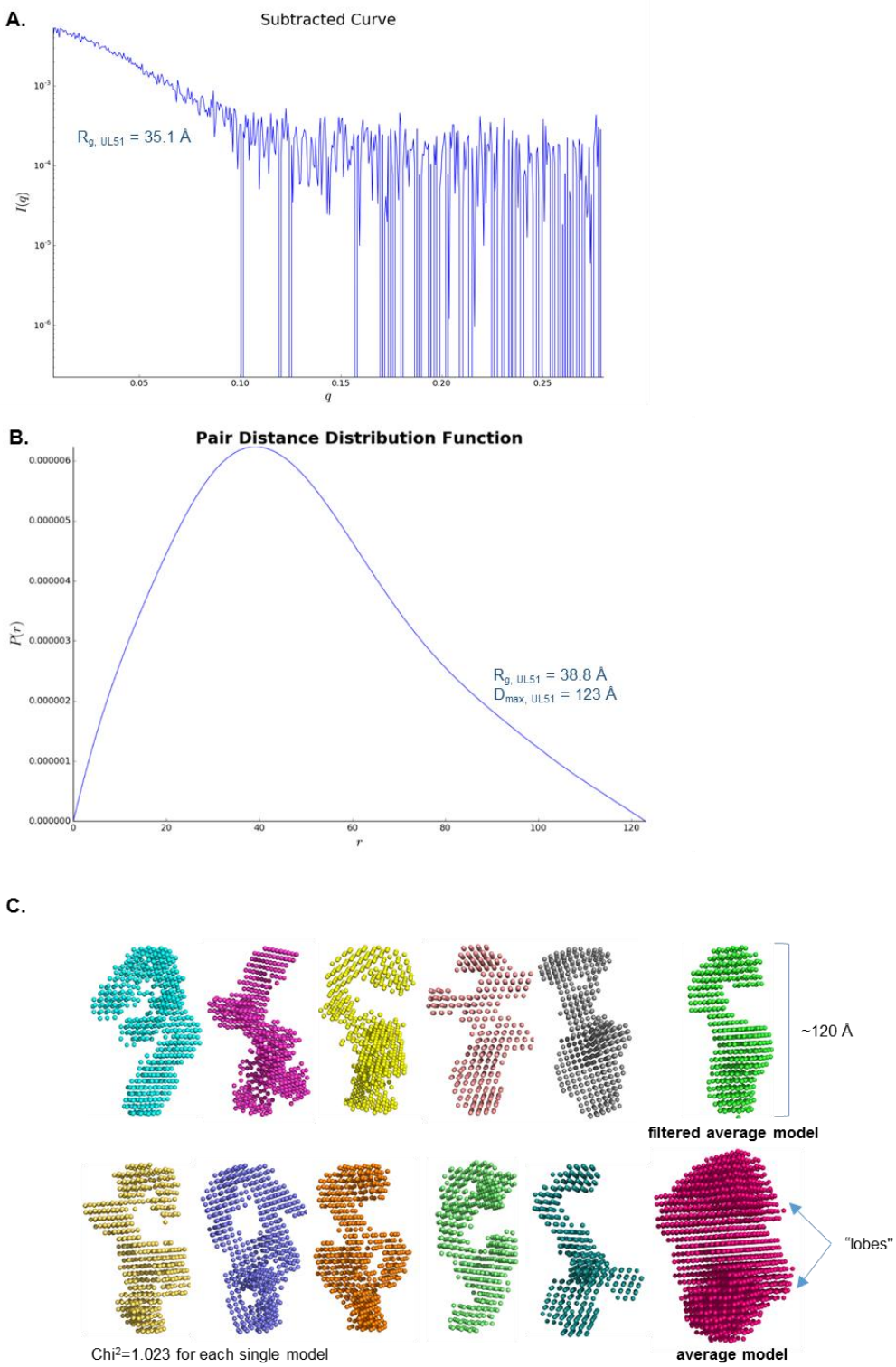
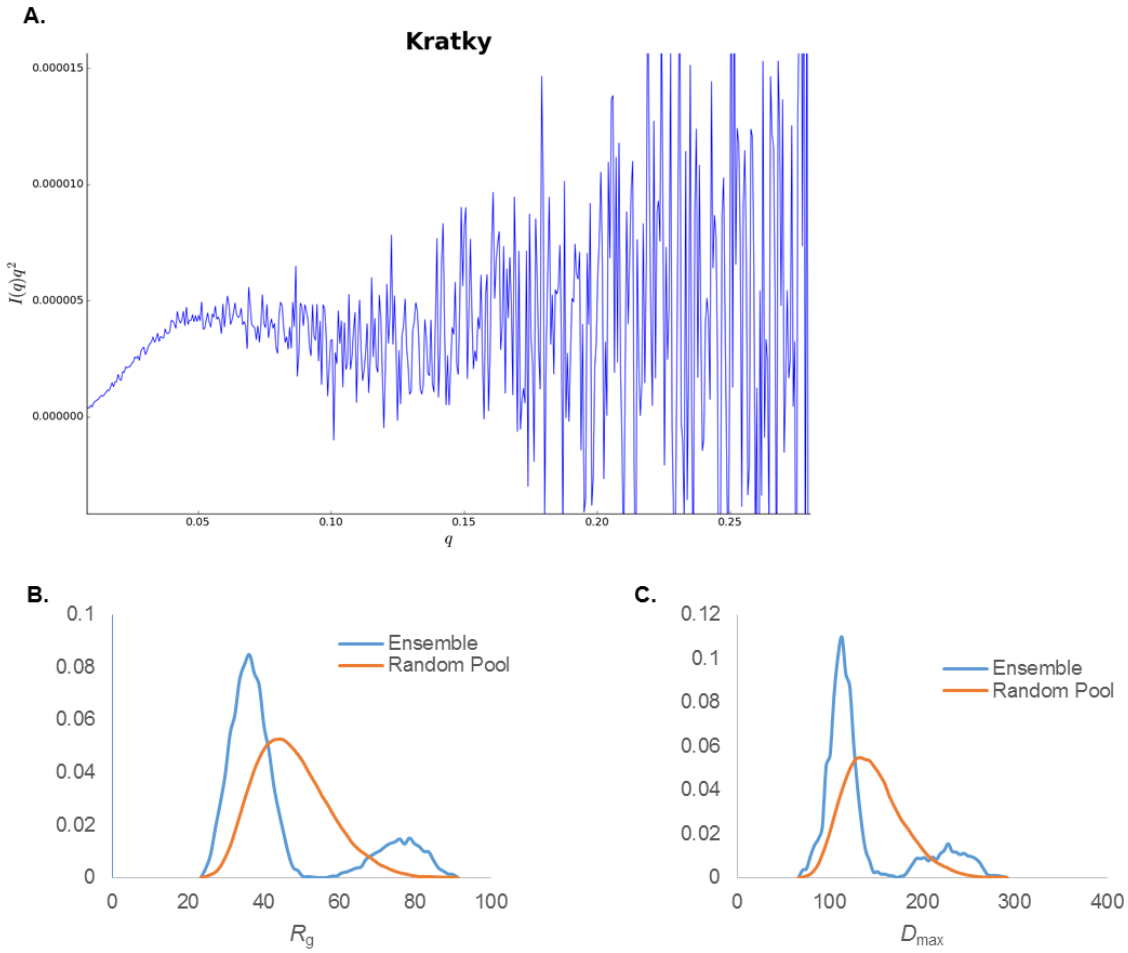


Figure 6-11. SAXS characterization and rigid body modeling of GPGS-UL51
A) Subtracted SAXS curve including Guinier Rg value, B) pair distance distribution function including Rg and Dmax values, and C) 10 individual DAMMIN rigid body models including average models of GPGS-UL51.

Since secondary structure prediction suggests that a large portion of the C-terminal region of UL51 is disordered (Figure 6-6C), it is likely that this molecule is flexible. One of the first indicators of flexibility of a molecule via SAXS is the shape of its Kratky plot, where a flexible molecule will often plateau at a maximum value or increase infinitely in contrast to the bell shape typical of a rigid molecule. Molecules with both characteristics tend to have additive Kratky plots. The signal from the small amount of diffraction data representing the smallest unit size of GPGS-UL51 was very weak and therefore it is difficult to evaluate the shape of the Kratky plot; however, you can see a hint of a bell shape indicating that there is some rigidity to this molecule (Figure 6-12A). To further evaluate the flexibility of UL51, the SAXS curve was analyzed using Ensemble Optimization Modeling (EOM), as was performed for UL11 and UL21. EOM using this preliminary SAXS data suggests that there are again two pools of GPGS-UL51 (Figure 6-12B, C). The majority of the sample appears to be in compact conformation with a small portion of the sample in an extended conformation (Figure 6-12B, C, D). Looking more closely at the distribution of representative models that make up this ensemble, it appears that there may be some flexibility within the compact conformation as half the models in the compact peak have an Rg of 30.4 Å and a Dmax of 94.4 Å and the other half have slightly larger dimensions of Rg approximately 40 Å and a Dmax approximately 115 Å (Figure 6-12D) These larger dimensions agree well with the dimensions of the averaged bead model. Although the remainder of the models appear to represent a higher order multimer of UL51, potentially a tetramer since the dimensions are approximately doubled, this diffraction data came from one species on size exclusion and a tetramer would likely elute separately. We conclude instead that the majority of

UL51 is in equilibrium between two dimensionally similar, compact, dimeric states and there is a portion in an extremely extended conformation.



D.

	R_g	D_{max}	Curves (out of 10)
compact	**30.40**	**94.43**	**4 (40%)**
Intermediate	39.28	114.51	1
	43.61	116.48	3
average	**42.53**	**115.99**	**4 (40%)**
extended	**78.94**	**254.08**	**2 (20%)**

Figure 6-12. Ensemble modeling of GPGS-UL51 SAXS data
A) Kratky plot and dimensions of models found in optimized ensemble from EOM displayed as B,C) histograms and D) summary table.

6.2.6 UL51 does not associate with membranes without modification

Finally, UL51 is palmitylated at its N-terminus which allows it to associate with cellular membranes[62]; however, some membrane-associated proteins are able to interact with membranes even when their transmembrane domains or lipid anchors are removed [38]. To test whether UL51 could interact with membranes, an experiment was performed in which purified recombinant GPGS-UL51 was incubated with multilamellar vesicles (MLVs) of varying lipid composition and binding was assessed by UL51's ability to pellet with the lipids. In stark comparison with positive control NEC220 which strongly binds acidic MLVs [38], neither UL51 nor UL51 (12-244) was able to associate with lipids without palmitoylation (Figure 6-13).

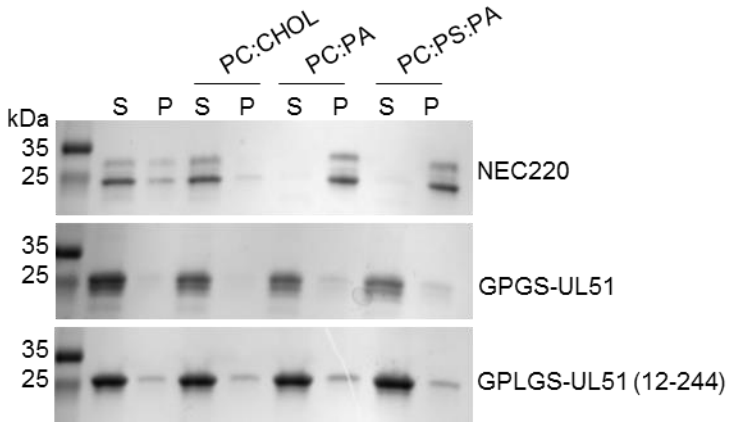


Figure 6-13. Membrane binding assay with UL51
Coomassie-stained gels of soluble and insoluble fractions of GPGS-UL51 or GPLGS-UL51(12-244) incubated with MLVs. NEC220 was used as a positive control. PC, phosphatidylcholine; chol, cholesterol; PA, phosphatidic acid; PS, phosphatidylserine.

6.3 Discussion

The results presented in this chapter characterize UL51 as an elongated protein with an alpha-helical core and a protease-resistant, proline-rich flexible tail. This protein may exist as a dimer in solution with compact and extended conformations and forms larger ordered assemblies over time. The implications of this polymerization are exciting

as UL51 is known to interact with membranes and has poorly understood roles in secondary envelopment, so it is tempting to imagine UL51 polymerizing on the membrane to form a scaffold of sorts for the maturing virion. Future work will focus on characterizing the multiple structures of UL51 alone, in assemblies, and with binding partners including membranes, UL7, and UL14 to deconvolute its roles in the viral replication cycle.

Chapter 7: Discussion

The results presented in this thesis describe UL11, UL21, UL51, and to some extent UL16 as flexible proteins. UL21 likely uses its flexibility to increase the pool of binding partners and therefore number of functions, and this is probably a general mechanism exploited by many tegument proteins including UL11, UL51, and UL16. Work in our lab is building evidence that UL37 and UL25 may function this way as well. It is likely that UL11 additionally uses its flexibility to organize the tegument during secondary envelopment. UL51 likely uses its flexibility to modulate its polymerization with an effect on cellular membranes in infected cells. These hypotheses and their implications on viral replication and assembly are discussed in depth in this chapter.

7.1 UL51 is a flexible, polymerizing protein

7.1.1 UL51 and the ESCRT pathway

The results presented in chapter 6 characterize UL51 as an elongated protein with an alpha-helical core and a protease-resistant, proline-rich flexible tail that predominantly exists as a dimer in solution with compact and extended conformations. This protein also forms larger ordered assemblies over time. Specifically, our results describe UL51 as comprising flexible N-terminal residues 1-11, structured core 12-163, and flexible C-terminal tail 164-244. Since 167-244 is predicted to lack structure that would protect it from proteolysis, it is striking that this fragment is so well represented on SDS-PAGE after digestion with chymotrypsin, as it contains a number of leucines. This protection suggests that despite predicted disorder, this fragment might be protected from proteolysis by interacting with the core and/or by forming a structure not predicted *in silico*, such as a polyproline helix, as this region is rich in prolines (Figure 6-6).

Combining the characteristics of UL51 described in this chapter with the observations made about this protein during infection brings ESCRT (endosomal sorting complexes required for transport) proteins to mind. These proteins are necessary for "reverse topology" scission of membrane vesicles away from the cytoplasm[154, 155]—the direction in which herpesviral nucleocapsids must bud in order to gain their envelope. In CMV, some secondary envelopment is thought to occur at multivesicular bodies and indeed, deletion the CMV homolog (UL71) of cytoplasmic-budding regulator UL51, results in large multivesicular bodies full of unenveloped nucleocapsids in close proximity to the membrane [156]. Since multivesicular body formation is usually controlled by ESCRT proteins in the host cell, the authors propose that UL71 may manipulate host ESCRT functions. Indeed, CMV UL71 features a proline-rich, low sequence complexity region at its C-terminus (Figure 6-6) that could contain signals known to recruit ESCRT machinery [157]. Despite early assertions that these proteins were dissimilar based on sequence divergence, ESCRT recruitment could be conserved as HSV-1 UL51 and the remainder of the human herpesvirus UL51 homologs also feature these proline rich regions at their C-termini (Figure 6-6) and many UL51 homologs have been shown important for secondary envelopment. Interestingly, HSV-1 UL51 binding partners and envelopment effectors UL7 and UL14 also have proline-rich C-termini (data not shown), so it will be interesting to see how these molecules play into the ESCRT pathway as well. In support of homology amongst UL51 homologs, CMV UL71 has also been shown to form dimers and higher ordered structures, like HSV-1 UL51 (this work), and this multimerization was found to be required for efficient secondary envelopment [158].

In contrast, it has been shown that knockout of ESCRT proteins does not inhibit secondary envelopment or virion production in CMV [159]. It is already known that herpesviruses encode their own proteins to mediate membrane budding and scission at the nuclear membrane [35], so it is not out of the question that they also encode ESCRT-like proteins to manipulate cytoplasmic membranes. The characteristics of UL51 presented in this chapter support the hypothesis that this protein may act like an ESCRT-III protein. ESCRT-III proteins are approximately 200-amino acid domains that are largely alpha-helical, with a globular N-terminus and a flexible C-terminus, and we have shown in this chapter that UL51 has these qualities (Figure 6-6). The flexible C-terminus of ESCRT-III proteins also features a regulatory helix that associates with the globular N-terminal domain, and although no helix is predicted in the c-terminal domain of UL51 (Figure 6-6), our limited proteolysis data exhibits proteolytic protection in this region consistent with the tail interacting with the core and/or transient structure formation (Figure 6-7). ESCRT-III molecules have both compact and extended conformations, like we have seen with UL51 (Figure 6-11, -12), and transition to the extended conformation can require much of the previously flexible C-terminus to refold into a helix. This may be possible based on the C-terminal sequence of UL51 (Figure 6-6). Finally, these proteins polymerize to deform and promote scission of cellular membranes and we have shown that HSV-1 forms large molecular weight complexes (Figures 51-9, -10). Finally, ESCRT-III proteins are found as soluble monomers in the cytoplasm in the absence of signals that recruit them to membranes. Similarly, UL51 is associated with membranes via palmitylation (a reversible modification). We have shown that the unacylated protein does not associate with MLVS (Figure 6-13), suggesting that it too may have reversible

membrane association capabilities *in vivo*. Since the UL51-mediated defects to secondary envelopment also occur when only binding partners UL7 or UL14 are missing, it will be interesting to see how these molecules play into the ESCRT pathway as well.

While these hypotheses are exciting, future structural, biochemical, and biological experiments are necessary to fully describe how UL51 and its binding partners fit into the ESCRT membrane remodeling pathway to affect herpesvirus replication. In particular, stabilization of the compact and extended conformations of UL51, either chemically, by the inclusion of binding partners, or truncation, followed by structural characterization will allow for comparison with the known structures of ESCRT-III proteins or other proteins of known structure and function. Interaction studies including co-immunoprecipitation or BioID from transfected and infected cells will help identify host factors that are known to be part of the ESCRT machinery or other pathways in which UL51 has been implicated. Most excitingly, fully describing how this protein may polymerize and performing microscopic studies of the polymerized form of UL51, in particular in the presence of membranes and its binding partners UL7 and UL14, will help characterize this protein's role in membrane deformation and scission and eventually, viral secondary envelopment and replication.

7.2 Recombinant UL16 is insoluble and hard to purify

7.2.1 Solubilizing UL16

The work presented in chapter 5 describe recombinant UL16 as a stubborn, insoluble molecule. In line with the central theme of this thesis, it is possible that these solubility issues come from flexibility of UL16 and its inability to "pick" a conformation *in vitro*. This makes intuitive sense based on the number of cysteines in this model

available for spurious disulfide bonding that could lead to aggregation (Figure 5-1). Supporting flexibility and "misbehavior" in UL16 are many notes from literature including the observation that it reversibly associates with the capsid in extracellular virions [40], the idea that its conformation may be controlled by a redox-sensitive switch[56], and its requirement for activators (UL21 and UL11) to take part in binding interactions with UL11 and gE[45]—all models hypothesizing that UL16 undergoes conformational change. Furthermore, it is unlikely that aggregation and impurity are indicative of UL16 function as the 70 kDa and 25 kDa contaminants are likely to be a common *E. coli* chaperone and metal-binding protein, respectively, and while the C-terminal domain of UL16 has been described to self-associate, and we also see that in Figure 5-3D, the N-terminal domain has not [153].

In order to produce protein to study UL16 *in vitro*, we attempted many common strategies to improve expression and solubility including optimizing codons, changing affinity tag identity and location, expressing the protein in different systems (cells and media), removing hypothetically non-essential cysteines, truncating unstructured regions, and splitting up domains, but were not successful in acquiring soluble, pure UL16 (or any piece of it). We also tried to coexpress UL16 with stabilizing binding partners, but were unsuccessful there as well due to vector incompatibility.

Coexpression is still the most sensible way forward. Future work should focus on coexpressing UL16 or its domains with direct binding partners UL21 and/or UL11 using pET-duet or a polycistronic vector. If full length, UL16 should have a small affinity tag at its C-terminus. Care must be taken in designing this system though, as only a limited number of affinity tags exist, we hypothesize that inclusion of GST could be problematic

for UL16, His_6 tags are problematic for UL11 (because it might bind metal, not shown) and UL21 (because it has a precipitates in the presence of imidazole, not shown), and capturing an MBP tag on amylose resin may also capture UL21C. Since we can't put an StII tag on everything, an almost-ideal strategy is to coexpress UL21(PSM)-HRV3C-UL11-StII with UL16(CS)-His_6 in a pET-duet vector, purify over StrepTactin and nickel resin to isolate the complex associated through a UL16-UL21C interaction, then cleave with PreScission protease to free UL11 and form the tripartite complex. This should overcome previously seen expression issues seen with vector incompatibility as well as low expression of UL11 with something large at its C-terminus. We also know that UL21 expression is improved with C-terminal extensions and UL16(CS)-His_6 is expressed in some cell types. The major risks here are using the His_6 tag (described above), we have previously seen UL16 pull otherwise soluble molecules into an aggregate, and we do not know whether UL16 can bind UL21C with a C-terminal extension. Two other common strategies that have not yet been attempted are to refold UL16 from insoluble protein and to try a different strain of Origami cells.

Although it is also possible that there are some undescribed post-translational modifications necessary for proper folding of UL16, our collaborators at Penn State have had similar trouble isolating soluble protein from baculovirus-infected insect cells. It is tempting to try another homolog as HSV-1 UL16 has the most cysteines, but our CS construct did not produce soluble protein either. *In vivo* work suggests that separate domains, in particular UL16N(1-155), are stable and behave well [153] but we did not see that to be the case *in vitro* (Figure 5-14). If we can isolate pure, soluble UL16 it will

be interesting to map structural domain boundaries with limited proteolysis, n-terminal sequencing, and mass spectrometry to see if these domains have been correctly assigned.

7.3 UL21 is a bimodal, RNA-binding protein

7.3.1 Structures of UL21 domains lead to functional hypotheses

The constructs used to determine the structures of UL21N and UL21C were designed based on the products of accidental proteolysis that occurred during purification. The proteolytic susceptibility of this protein *in vitro* has been observed previously [53, 136, 139], typically with affinity-tagged protein constructs expressed in *E. coli.* One study observed a potential proteolytically processed UL21 species in immunoprecipitates from PRV-infected cells [133]; however, at 40 kDa, this species is larger than any stable cleavage product of UL21 predicted from the structures of the two domains. Therefore, there is little evidence that proteolysis occurs during infection. In combination with secondary structure prediction, proteolytic instability of full-length UL21 *in vitro* along with the relative stability of UL21N and UL21C supports the model in which UL21 is composed of two distinct domains connected by a flexible linker, reminiscent of two balls connected by a string. We originally predicted that the globular domains of UL21N and UL21C could interact through electrostatic interactions because HSV-1 UL21N is largely acidic [53] whereas HSV-1 UL21C is basic [99] and that this potential interaction could feasibly be blocked by binding partners of UL21, e.g., UL16; however, the two domains were never seen to interact (discussed below).

The basic patches on the surface of HSV-1 UL21C (Figure 4-10) may explain why UL21 is found in the nucleus in transfected and infected cells [45, 55, 133, 136, 139]. Although the UL21 sequence lacks an obvious linear classical nuclear localization

signal (cNLS), one of the positively charged surface areas could structurally mimic an NLS or provide a nonclassical NLS. For example, monomers of STAT1 are unable to bind importin α for nuclear import, but through dimerization, an NLS is formed out of residues from each monomer that allows the dimer to bind importin α [160]. Since evolutionarily distant UL21 from HSV-1 and PRV are both found in the nucleus [45, 55, 133, 136, 139], we could potentially identify a putative basic NLS on the surface of UL21 by mapping the few arginines and lysines conserved between HSV-1 and PRV sequences; however, these are spread around the surface of the molecule and therefore there is no obvious cluster that could represent a basic nuclear localization signal. Alternatively, UL21 could localize to the nucleus by piggybacking on a cellular protein containing an NLS. Confusingly, either domain of HSV-1 UL21 localizes to the nucleus of transfected cells [136], despite the acidic nature of UL21N. This may be due to the fact that either domain is appropriately sized for passive entry into the nucleus [136].

Both UL21N and UL21C have unique folds (Figures 4-5,-7). Whether this property enables functions unique to alphaherpesviruses, or provides a new way to perform a function mediated by other, unrelated protein structures is unknown. UL21 is thought to participate in different processes during the viral replication cycle. To provide a starting point for mutational analysis, we have designated four surface regions within either domain that could have functional roles. Based on sequence similarity, the roles of UL21N are likely largely conserved, whereas UL21C likely has some conserved and some virus-specific functions. If UL21C interacts with UL21N, one potential conserved function could be binding to UL21N, which has a higher degree of sequence conservation than UL21C. Another conserved function could be interaction with UL16 [118, 132,

136]; however, UL16 sequences could have evolved in parallel with UL21C sequences such that while the complex formation is conserved, the binding interface is not. UL21C binds capsid [55, 133] and as this has been seen in both HSV-1 and PRV, direct or indirect capsid binding is likely a conserved function.

In contrast to the more conserved UL21N (Figure 4-3), to which no binding functions have been ascribed, the surface of UL21C shows a lot of variation among homologs as is evident from both sequence similarity and evolutionary conservation (Figures 4-4, -10). Hence, we hypothesize that while UL21N may mediate conserved functions, UL21C additionally facilitates functions that adapt the virus to its specifically preferred host. These virus-specific characteristics of UL21C are also apparent in its electrostatic properties. Firstly, there is a stark division in the calculated isoelectric points of UL21C: the 5 sequences most closely related to HSV-1 UL21C are basic while the rest of the sequences are acidic. These differences can be localized to three basic regions on the surface of HSV-1 UL21C that appear specific to six basic UL21 homologs, including HSV-1, but not to the ten acidic UL21 homologs (Table 4-3).

Sequence analysis revealed acidic glycine-rich insertions in the sequences of PRV, BHV-1 and BHV-5 UL21 that are absent from other UL21 homologs (Figure 4-4). The insertion in PRV UL21, which likely forms a flexible loop, was previously hypothesized to serve as a hinge and site of proteolytic processing [133, 139], based on the inherent flexibility of glycine residues and the similarity between this sequence and the proteolytic processing sequence found in HSV-1 ribonucleotide reductase [161]. While this loop or the similar loops in BHV-1 and BHV-5 may be cleaved in the cell, based on the structure of HSV-1 UL21 presented here, it is unlikely that this cleavage

would result in separation of the domain as it would expose many hydrophobic residues (Figure 4-8).

7.3.2 Structural characteristics of full length UL21

We have shown that the two recombinant domains of UL21 do not interact *in vitro*, with or without the flexible linker and conserved region (Figure 4-11), despite featuring oppositely charged patches on either domain (Figures 4-6, -10). Although it is possible that purification tags remaining on either domain could interfere with the interaction, these studies were performed with multiple constructs and multiple tags and no interaction was seen. Since it is unlikely that this protein gets cleaved during infection (discussed above), and all homologs have retained the flexible linker and C-terminal folded domain through evolution (Figure 4-4), it is possible that the virus just needs the domains of UL21 close to each other, potentially as a scaffold. Alternately there may be an interaction with UL21 that requires an interface created by both domains. No such binding partner has yet been described, but the only direct binding partners of UL21 described so far are UL16 and RNA, which both only require the C-terminal domain.

This work shows that UL21 comprises two folded domains that do not strongly interact and are connected by a linker that is predicted to lack structure. These observations are in line with our results that no full length construct crystallized and that UL21 is flexible in solution (Figure 4-15). Intriguingly, ensemble modeling showed two main populations of conformations—compact, like cherries (see below), and extended, like two separate balls on a string (Figure 7-1A)—instead of a range of conformations that would be associated with the infinite flexibility of two non-interacting proteins connected by a linker. Combining this observation with results from bead modeling

(Figure 4-14), which places the two domains close to each other and limited proteolysis (Figure 4-16), which shows that the conserved region is protected from cleavage, we have developed two potential model for full length UL21 (Figure 7-1B). In the "strand-to-sheet" model, the conserved region forms a β-strand to reversibly associate with a β-sheet in UL21N (Figure 4-5). This could pull UL21C into close proximity with UL21N, forming a semi-rigid, elongated molecule (similar to what was modeled with CORAL in Figure 4-14), without the need for the two molecules to interact. In the "hairpin" model, the conserved region could form an isolated β-hairpin, leaving the domains of UL21 in proximity but not directly attached, reminiscent of a pair of cherries (Figure 7-1). Future biochemical work will attempt to stabilize the closed conformation of full length UL21, either with chemical crosslinking or with binding partners, to better understand the structure of the full length protein. With better understanding of what factors control the conformational state of UL21, mutations can be made in the context of infection to describe what functions the conformational states and overall flexibility of UL21 provide for the virus.

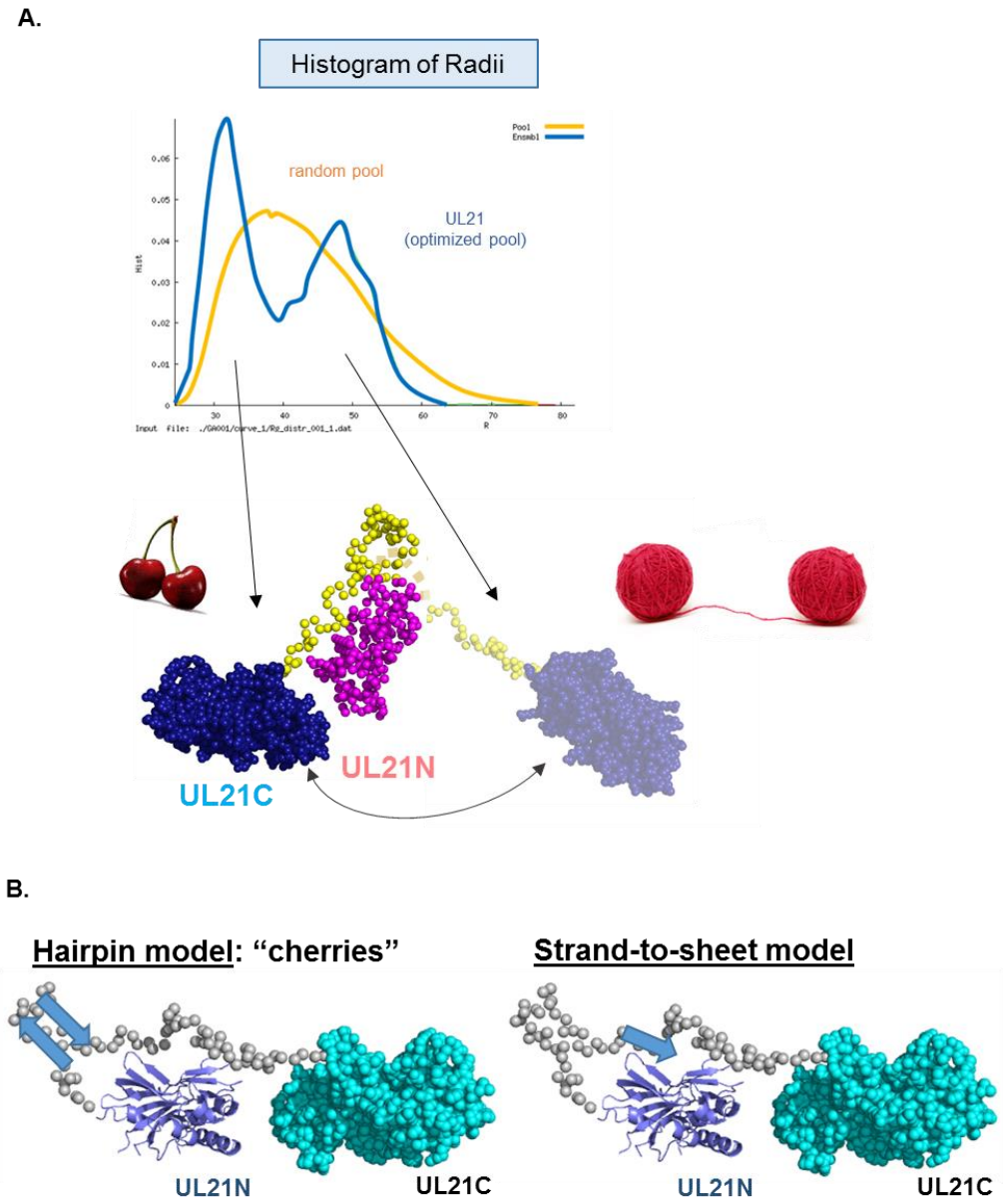


Figure 7-1. Multiple structures of full length UL21
A) (Top) Histogram of Dmax from ensemble modeling of UL21 SAXS data. (Bottom) Potential compact and extended conformations of full length UL21. B) Two potential models for the compact conformation of UL21.

7.3.3 UL21 binds RNA through the C-terminal domain

The ability of UL21 to bind RNA through its C-terminal domain (Figure 4-17) was unexpected, based on the previous observation that there are four HSV-1 proteins

that bind total RNA from infected cells in a Northwestern assay: US11, UL49, UL47, and an unidentified 110 kDa protein that is too large to represent any known species of UL21 [162]. This conclusion, however, does not rule out UL21 as an RNA-binding protein. Since UL47 and UL49 are among the most abundant proteins in the tegument [163] and would therefore be easily detected, it is possible that the amount of UL21 in the virion is too low to detect by this analysis. Alternately, the UL21-RNA binding interaction may require specific structures that would have been sensitive to the multiple denaturation and renaturation steps required in this technique. While the precise location of the RNA binding site awaits experimental validation, being a negatively charged polymer, RNA could, in principle, bind to any of the three prominent basic patches in HSV-1 UL21 (Figure 4-10). Whether RNA binding is conserved among all UL21 homologs or limited to the species with basic isoelectric points is yet unknown (Table 4-3).

So far, this phenomenon has only been observed with *E. coli* RNA, no RNA sequence information has been collected, and further studies are necessary to identify the biologically relevant RNA target of UL21 and to fully characterize the UL21/RNA interaction. Interestingly, in the absence of UL21 in HSV-1 [80] and HSV-2 [54], there is a delay in the production of viral RNA and proteins. It is tempting to speculate that this delay in viral RNA and protein production could reflect a role for UL21 in transcription and/or translation. Moreover, UL21C appears to contribute to virulence in PRV because infection with a UL21C deletion mutant increased time to death in infected mice approximately threefold [133]. In contrast, UL21 RNA binding may serve a different purpose, such as packaging RNA into the virion like UL47, and the delay in viral

production could be indicative of a separate function of UL21. In either case, UL21 clearly plays additional non-structural roles in viral replication.

7.4 UL11 is a flexible protein with an ordered core and an IDR tail

7.4.1 UL11 flexibility is maintained *in vivo*

We found that UL11 without acylation is a soluble, flexible protein in solution. These characteristics are in agreement with previous reports suggesting that without myristylation or palmitylation, UL11 acts as a soluble protein[120]. Furthermore, UL11 and homologs are relatively unstable in transfected and infected cells when expressed without lipid modification or binding partner UL16, potentially due to rapid turnover [52, 58, 76, 116]. Since inherently flexible proteins are more available to proteolytic digestion, and association with a binding partner such as UL16 or a membrane would confer proteolytic protection, this instability suggests that UL11 remains flexible in the cell. In support of *in vivo* flexibility, UL11 homologs isolated from transfected and infected cells run at an elevated molecular weight on SDS-PAGE, a phenomenon occurs in proteins with IDRs [124]. Though this feature has been commonly ascribed to the protein's post-translational modifications, we make the same observation with recombinant unmodified UL11. This also supports the idea that despite sequence divergence, the flexibility of UL11 is conserved amongst homologs. The N-terminal core of UL11 may also be more conserved than previously appreciated, as this region is predicted to be structured and acylated, contains the characterized acidic clusters, and binds UL16 in each homolog.

7.4.2 The IDR in UL11 contributes to RNA binding

We additionally report that UL11 has the ability to copurify with heterologous RNA from E. coli, like we have described for HSV-1 UL21 [99]. Unlike UL21C, which

interacts with RNA through its structured C-terminal domain, this function of UL11 maps to its basic, unstructured tail. Binding nucleic acids often aligns with intrinsic disorder because a concentration of positively (or negatively) charged residues in low complexity sequences is thought to contribute to "unstructure" [124]. However, this should not be confused with irrelevance because these promiscuous interactions between charged low complexity sequences and nucleic acids often function to form functional complexes and cellular substructures (discussed below).

UL11 copurified specifically with ribosomal RNA. One deletion study reports delays in virion production without UL11[76], so it is possible that the interaction with ribosomal RNA points to a role in translation regulation, as we have proposed for UL21C, though UL11 has not yet been described to regulate translation. Despite the observation that one species of RNA was identified, it is also possible that this is a nonspecific interaction with the most prominent RNA species, as UL11 has been described as "sticky" [34]. Alternately, UL11, UL16, and UL21 may interact with RNA and function together as UL21 binds RNA [99] and UL16 has been reported to interact with nucleic acids, potentially through a putative leucine zipper motif. Further studies are necessary to describe the function and specificity of UL11 binding RNA alone or in complex with UL16 and UL21 in the context of infection.

7.4.3 Transient structure in UL11

Although we characterized UL11 as having a folded, β-rich core using CD, ensemble modeling suggests that UL11 samples three conformations in solution: compact, intermediate, and extended (Figure 7-2). The extended conformation likely represents a fully unstructured UL11 as the Rg is close to that calculated for a denatured

protein of the same size. This conformation, like a piece of string, could allow UL11 to take on many functions. The rigid body model is consistent with the dimensions of the intermediate form and looks like a dancing poodle, where the "front paws" could potentially represent the folded, β-rich core, with the "head" and "hind legs" representing the flexible N- and C-terminal extensions, respectively (Figure 3-4, Figure 7-2). This conformation has a structured core and an unstructured portion, so it can be likened to a partially unraveled ball of yarn that can also take on a number of conformations and therefore multiple functions. Finally, it is tempting to interpret the compact pool as a closed conformation with the tail in contact with the core (a sleeping poodle). Like a ball of yarn, there is likely only one conformation in the closed group, likely associated with one or just a couple of functions (Figure 7-2). One can imagine that the positively charged tail might interact with the acidic cluster to compact the structure of UL11. Based on the proline-rich sequence (Figure 3-2), it is also possible that the dimensions are condensed if the tail forms a transient structure such as a polyproline helix. In support of the former case, our data suggests that the N- and C-termini of UL11-StII come close enough together to form an intramolecular bond in the presence of a chemical crosslinker (Figure 3-5). It is also possible that both scenarios occur. It is likely that binding partners influence the conformational state of UL11. The CD data also suggests that the secondary structure of UL11 is concentration dependent, so clustering of UL11, such as on lipid rafts[65], may be another mechanism by which UL11 selects a conformational state. Future work will address potential for different induced structures in the presence of membranes and UL16 and delineate which structures correlate with each binding partner to affect UL11's multiple functions.

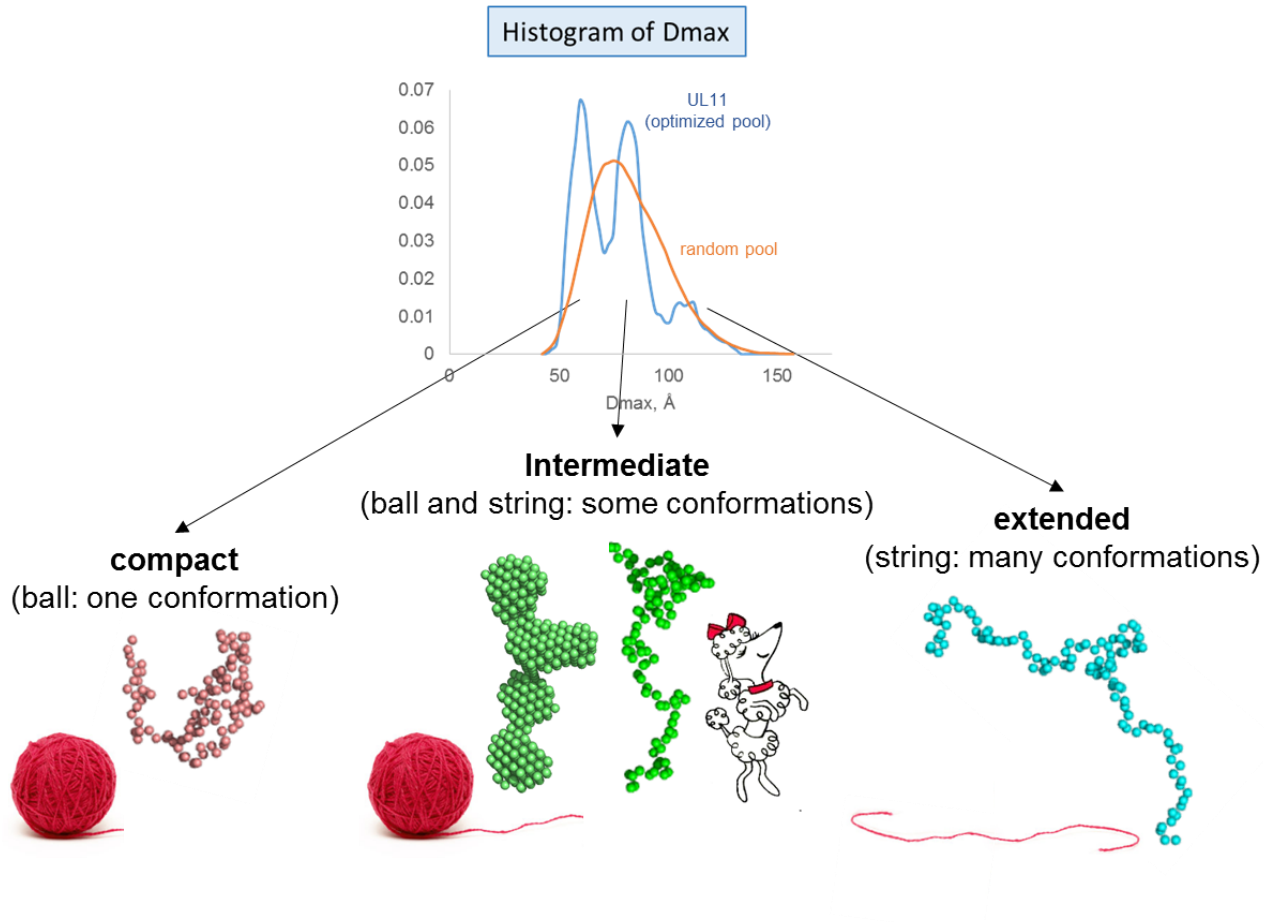


Figure 7-2. Multiple structures of UL11
(Top) Histogram of Dmax from ensemble modeling of UL11 SAXS data. (Bottom) Potential conformations of UL11. Dancing poodle illustration from Charles Schulz.

7.4.4 The tegument as a phase separated droplet

Another indicator of UL11's flexibility was its ability to phase separate *in vitro*. Proteins that undergo this phase transition often contain long intrinsically disordered regions (IDRs). Many of these proteins retain the ability to phase separate *in vivo* [123] and form "biological condensates" that often function like organelles without a membrane. Interestingly, UL11 is often described as localizing to "punctate" structures [56, 118] and even "nuclear speckles"[81]. We propose that UL11 remains flexible *in vivo* (see above) and may phase separate *in vivo* that manifests as these puncta or

speckles. If this is the case, one might also envision the herpesviral outer tegument as being phase separated or a biological condensate. These dynamic assemblages always contain flexible proteins with IDRs [164]. In addition to UL11, UL21 is flexible, and several other outer tegument proteins are predicted to have large IDRs , e.g., VP22/UL19 (our own analysis) and outer tegument "hub" VP16/UL48[165]. These condensates are also formed by many multivalent, sometimes nonspecific interactions[164], characteristics that may correlate with the loosely structured outer tegument. Indeed, outer tegument proteins not only make many redundant specific interactions (reviewed in [16]), but also UL11, VP22/UL19, and VP16/UL48 have been described as "sticky" and make many nonspecific interactions that may be important for tegument formation [34]. Also, RNA and cellular proteins can be incorporated into the outer tegument in the absence of some tegument proteins, potentially acting as "space fillers". Finally, these droplets are stable to loss of components[164], which is in line with the striking observation that the independent deletion of most outer tegument proteins from HSV-1 hinders, but does not abolish replication. Considering an ordered protein network as a brick wall and a phase separated droplet as a bowl of noodles provides a helpful analogy. While removal of only one or two bricks can send the wall crumbling down, a bowl of noodles remains a cohesive entity if several noodles are removed [123]. The outer tegument forming a multiphase state rather than a strictly ordered protein network could be a general tactic for subverting host defenses, allowing for productive replication even when the availability of one or more outer tegument proteins is decreased by the host.

Furthermore, based on their liquid-like characteristics, phase separated droplets are inherently spherical. Driven by membrane associated tegument proteins, one could

imagine the energy of this phase separation being enough to pinch off a spherical bud to drive secondary envelopment, much like a lava lamp. This is consistent with the observation that during normal HSV-1 infection, dense bodies/L-particles are formed that lack nucleocapsids, which suggests that a nucleocapsid-tegument-envelope bridge is not necessary to drive secondary envelopment. Though this idea doesn't rule out the contribution of host factors or undiscovered viral mediators of membrane deformation and scission, it could be an insurance mechanism used by the virus to ensure envelopment when these mediators have been compromised by the host. Continued investigation is necessary to fully characterize the flexibility of the tegument as a whole and how this characteristic contributes to multiple mechanisms driving herpesviral replication.

Chapter 8: Bibliography

1. Flint, S.J., Enquist, L. W., Racaniello, V. R., and Skalka, A. M., *Principles of Virology*. 3 ed. 2009: American Society for Microbiology.
2. Taubenberger, J.K. and D.M. Morens, *1918 Influenza: the mother of all pandemics.* Emerg Infect Dis, 2006. **12**(1): p. 15-22.
3. Liu, F. and Z.H. Zhou, *Comparative virion structures of human herpesviruses*, in *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis*, A. Arvin, et al., Editors. 2007: Cambridge.
4. Coen, D.M. and P.A. Schaffer, *Antiherpesvirus drugs: a promising spectrum of new drugs and drug targets.* Nat Rev Drug Discov, 2003. **2**(4): p. 278-88.
5. Becker, S., et al., *Interactions of Marburg virus nucleocapsid proteins.* Virology, 1998. **249**(2): p. 406-17.
6. Muhlberger, E., et al., *Three of the four nucleocapsid proteins of Marburg virus, NP, VP35, and L, are sufficient to mediate replication and transcription of Marburg virus-specific monocistronic minigenomes.* J Virol, 1998. **72**(11): p. 8756-64.
7. Muhlberger, E., et al., *Comparison of the transcription and replication strategies of marburg virus and Ebola virus by using artificial replication systems.* J Virol, 1999. **73**(3): p. 2333-42.
8. Cardenas, W.B., et al., *Ebola virus VP35 protein binds double-stranded RNA and inhibits alpha/beta interferon production induced by RIG-I signaling.* J Virol, 2006. **80**(11): p. 5168-78.
9. Hale, B.G., et al., *The multifunctional NS1 protein of influenza A viruses.* J Gen Virol, 2008. **89**(Pt 10): p. 2359-76.
10. Davison, A.J., et al., *A novel class of herpesvirus with bivalve hosts.* J Gen Virol, 2005. **86**(Pt 1): p. 41-53.
11. Farooq, A.V. and D. Shukla, *Herpes simplex epithelial and stromal keratitis: an epidemiologic update.* Surv Ophthalmol, 2012. **57**(5): p. 448-62.
12. Ayoade, F. and W.G. Gossman, *Varicella (Chickenpox), Zoster*, in *StatPearls*. 2017: Treasure Island (FL).
13. Nair, P.A. and S.S. Bhimji, *Herpes, Zoster (Shingles)*, in *StatPearls*. 2017: Treasure Island (FL).
14. Mettenleiter, T.C., B.G. Klupp, and H. Granzow, *Herpesvirus assembly: an update.* Virus Res, 2009. **143**(2): p. 222-34.
15. McGeoch, D.J., F.J. Rixon, and A.J. Davison, *Topics in herpesvirus genomics and evolution.* Virus Res, 2006. **117**(1): p. 90-104.
16. Owen, D.J., C.M. Crump, and S.C. Graham, *Tegument Assembly and Secondary Envelopment of Alphaherpesviruses.* Viruses, 2015. **7**(9): p. 5084-114.
17. Cooper, R.S. and E.E. Heldwein, *Herpesvirus gB: A Finely Tuned Fusion Machine.* Viruses, 2015. **7**(12): p. 6552-69.
18. Lippe, R., *Deciphering novel host-herpesvirus interactions by virion proteomics.* Front Microbiol, 2012. **3**: p. 181.
19. Amen, M.A. and A. Griffiths, *Packaging of Non-Coding RNAs into Herpesvirus Virions: Comparisons to Coding RNAs.* Front Genet, 2011. **2**: p. 81.
20. Harrison, S.C., *Viral membrane fusion.* Virology, 2015. **479-480**: p. 498-507.
21. Eisenberg, R.J., et al., *Herpes virus fusion and entry: a story with many characters.* Viruses, 2012. **4**(5): p. 800-32.
22. Kelly, B.J., et al., *Functional roles of the tegument proteins of herpes simplex virus type 1.* Virus Res, 2009. **145**(2): p. 173-86.
23. Smiley, J.R., M.M. Elgadi, and H.A. Saffran, *Herpes simplex virus vhs protein.* Methods Enzymol, 2001. **342**: p. 440-51.

24. Granzow, H., B.G. Klupp, and T.C. Mettenleiter, *Entry of pseudorabies virus: an immunogold-labeling study.* J Virol, 2005. **79**(5): p. 3200-5.
25. Luxton, G.W., et al., *Targeting of herpesvirus capsid transport in axons is coupled to association with specific sets of tegument proteins.* Proc Natl Acad Sci U S A, 2005. **102**(16): p. 5832-7.
26. Copeland, A.M., W.W. Newcomb, and J.C. Brown, *Herpes simplex virus replication: roles of viral proteins and nucleoporins in capsid-nucleus attachment.* J Virol, 2009. **83**(4): p. 1660-8.
27. Dohner, K., et al., *Function of dynein and dynactin in herpes simplex virus capsid transport.* Mol Biol Cell, 2002. **13**(8): p. 2795-809.
28. Wolfstein, A., et al., *The inner tegument promotes herpes simplex virus capsid motility along microtubules in vitro.* Traffic, 2006. **7**(2): p. 227-37.
29. Douglas, M.W., et al., *Herpes simplex virus type 1 capsid protein VP26 interacts with dynein light chains RP3 and Tctex1 and plays a role in retrograde cellular transport.* J Biol Chem, 2004. **279**(27): p. 28522-30.
30. Zaichick, S.V., et al., *The herpesvirus VP1/2 protein is an effector of dynein-mediated capsid transport and neuroinvasion.* Cell Host Microbe, 2013. **13**(2): p. 193-203.
31. Yamauchi, Y., et al., *The UL14 tegument protein of herpes simplex virus type 1 is required for efficient nuclear transport of the alpha transinducing factor VP16 and viral capsids.* J Virol, 2008. **82**(3): p. 1094-106.
32. Yamauchi, Y., et al., *The UL14 protein of herpes simplex virus type 2 translocates the minor capsid protein VP26 and the DNA cleavage and packaging UL33 protein into the nucleus of coexpressing cells.* J Gen Virol, 2001. **82**(Pt 2): p. 321-30.
33. Baines, J.D., *Herpes simplex virus capsid assembly and DNA packaging: a present and future antiviral drug target.* Trends Microbiol, 2011. **19**(12): p. 606-13.
34. Johnson, D.C. and J.D. Baines, *Herpesviruses remodel host membranes for virus egress.* Nat Rev Microbiol, 2011. **9**(5): p. 382-94.
35. Bigalke, J.M. and E.E. Heldwein, *Nuclear Exodus: Herpesviruses Lead the Way.* Annu Rev Virol, 2016. **3**(1): p. 387-409.
36. Fuchs, W., et al., *The interacting UL31 and UL34 gene products of pseudorabies virus are involved in egress from the host-cell nucleus and represent components of primary enveloped but not mature virions.* J Virol, 2002. **76**(1): p. 364-78.
37. Klupp, B.G., et al., *Vesicle formation from the nuclear membrane is induced by coexpression of two conserved herpesvirus proteins.* Proc Natl Acad Sci U S A, 2007. **104**(17): p. 7241-6.
38. Bigalke, J.M., et al., *Membrane deformation and scission by the HSV-1 nuclear egress complex.* Nat Commun, 2014. **5**: p. 4131.
39. Meckes, D.G., Jr., J.A. Marsh, and J.W. Wills, *Complex mechanisms for the packaging of the UL16 tegument protein into herpes simplex virus.* Virology, 2010. **398**(2): p. 208-13.
40. Meckes, D.G., Jr. and J.W. Wills, *Dynamic interactions of the UL16 tegument protein with the capsid of herpes simplex virus.* J Virol, 2007. **81**(23): p. 13028-36.
41. Lyman, M.G. and L.W. Enquist, *Herpesvirus interactions with the host cytoskeleton.* J Virol, 2009. **83**(5): p. 2058-66.
42. Pawliczek, T. and C.M. Crump, *Herpes simplex virus type 1 production requires a functional ESCRT-III complex but is independent of TSG101 and ALIX expression.* J Virol, 2009. **83**(21): p. 11254-64.
43. Tandon, R., D.P. AuCoin, and E.S. Mocarski, *Human cytomegalovirus exploits ESCRT machinery in the process of virion maturation.* J Virol, 2009. **83**(20): p. 10797-807.
44. Dingwell, K.S., et al., *Herpes simplex virus glycoproteins E and I facilitate cell-to-cell spread in vivo and across junctions of cultured cells.* J Virol, 1994. **68**(2): p. 834-45.

45. Han, J., et al., *Function of glycoprotein E of herpes simplex virus requires coordinated assembly of three tegument proteins on its cytoplasmic tail.* Proc Natl Acad Sci U S A, 2012. **109**(48): p. 19798-803.
46. Roller, R.J., et al., *The herpes simplex virus 1 UL51 gene product has cell type-specific functions in cell-to-cell spread.* J Virol, 2014. **88**(8): p. 4058-68.
47. Ambrosini, A.E.a.E., L. W., *Cell-fusion events induced by α-herpesviruses.* Future Virol, 2015. **10**(2).
48. Lee, J.H., et al., *Identification of structural protein-protein interactions of herpes simplex virus type 1.* Virology, 2008. **378**(2): p. 347-54.
49. Yeh, P.C., D.G. Meckes, Jr., and J.W. Wills, *Analysis of the interaction between the UL11 and UL16 tegument proteins of herpes simplex virus.* J Virol, 2008. **82**(21): p. 10693-700.
50. MacLean, C.A., B. Clark, and D.J. McGeoch, *Gene UL11 of herpes simplex virus type 1 encodes a virion protein which is myristylated.* J Gen Virol, 1989. **70 (Pt 12)**: p. 3147-57.
51. Starkey, J.L., et al., *Elucidation of the block to herpes simplex virus egress in the absence of tegument protein UL16 reveals a novel interaction with VP22.* J Virol, 2014. **88**(1): p. 110-9.
52. Han, J., et al., *Interaction and interdependent packaging of tegument protein UL11 and glycoprotein e of herpes simplex virus.* J Virol, 2011. **85**(18): p. 9437-46.
53. Metrick, C.M., P. Chadha, and E.E. Heldwein, *The unusual fold of herpes simplex virus 1 UL21, a multifunctional tegument protein.* J Virol, 2015. **89**(5): p. 2979-84.
54. Le Sage, V., et al., *The herpes simplex virus 2 UL21 protein is essential for virus propagation.* J Virol, 2013. **87**(10): p. 5904-15.
55. Takakuwa, H., et al., *Herpes simplex virus encodes a virion-associated protein which promotes long cellular processes in over-expressing cells.* Genes Cells, 2001. **6**(11): p. 955-66.
56. Chadha, P., et al., *Regulated interaction of tegument proteins UL16 and UL11 from herpes simplex virus.* J Virol, 2012. **86**(21): p. 11886-98.
57. Silva, M.C., et al., *Human cytomegalovirus UL99-encoded pp28 is required for the cytoplasmic envelopment of tegument-associated capsids.* J Virol, 2003. **77**(19): p. 10594-605.
58. Phillips, S.L. and W.A. Bresnahan, *The human cytomegalovirus (HCMV) tegument protein UL94 is essential for secondary envelopment of HCMV virions.* J Virol, 2012. **86**(5): p. 2523-32.
59. Wu, J.J., et al., *ORF33 and ORF38 of Kaposi's Sarcoma-Associated Herpesvirus Interact and Are Required for Optimal Production of Infectious Progeny Viruses.* J Virol, 2015. **90**(4): p. 1741-56.
60. Leege, T., et al., *Effects of simultaneous deletion of pUL11 and glycoprotein M on virion maturation of herpes simplex virus type 1.* J Virol, 2009. **83**(Pt 2): p. 896-907.
61. Sarfo, A., et al., *The UL21 tegument protein of herpes simplex virus type-1 is differentially required for the syncytial phenotype.* J Virol, 2017.
62. Nozawa, N., et al., *Subcellular localization of herpes simplex virus type 1 UL51 protein and role of palmitoylation in Golgi apparatus targeting.* J Virol, 2003. **77**(5): p. 3204-16.
63. Klupp, B.G., et al., *Functional analysis of the pseudorabies virus UL51 protein.* J Virol, 2005. **79**(6): p. 3831-40.
64. Albecka, A., et al., *Dual Function of the pUL7-pUL51 Tegument Protein Complex in Herpes Simplex Virus 1 Infection.* J Virol, 2017. **91**(2).
65. Koshizuka, T., et al., *Herpes simplex virus protein UL11 but not UL51 is associated with lipid rafts.* Virus Genes, 2007. **35**(3): p. 571-5.

66. Roller, R.J. and R. Fetters, *The herpes simplex virus 1 UL51 protein interacts with the UL7 protein and plays a role in its recruitment into the virion.* J Virol, 2015. **89**(6): p. 3112-22.
67. Fuchs, W., et al., *The UL7 gene of pseudorabies virus encodes a nonessential structural protein which is involved in virion formation and egress.* J Virol, 2005. **79**(17): p. 11291-9.
68. Oda, S., et al., *The Interaction between Herpes Simplex Virus 1 Tegument Proteins UL51 and UL14 and Its Role in Virion Morphogenesis.* J Virol, 2016. **90**(19): p. 8754-67.
69. Kattenhorn, L.M., et al., *A deubiquitinating enzyme encoded by HSV-1 belongs to a family of cysteine proteases that is conserved across the family Herpesviridae.* Mol Cell, 2005. **19**(4): p. 547-57.
70. Dong, X., et al., *The herpes simplex virus 1 UL36USP deubiquitinase suppresses DNA repair in host cells via deubiquitination of proliferating cell nuclear antigen.* J Biol Chem, 2017. **292**(20): p. 8472-8483.
71. Wang, S., et al., *Herpes simplex virus 1 ubiquitin-specific protease UL36 inhibits beta interferon production by deubiquitinating TRAF3.* J Virol, 2013. **87**(21): p. 11851-60.
72. Jambunathan, N., et al., *Herpes simplex virus 1 protein UL37 interacts with viral glycoprotein gK and membrane protein UL20 and functions in cytoplasmic virion envelopment.* J Virol, 2014. **88**(11): p. 5927-35.
73. Zhao, J., et al., *A Viral Deamidase Targets the Helicase Domain of RIG-I to Block RNA-Induced Activation.* Cell Host Microbe, 2016. **20**(6): p. 770-784.
74. Pitts, J.D., et al., *Crystal structure of the herpesvirus inner tegument protein UL37 supports its essential role in control of viral trafficking.* J Virol, 2014. **88**(10): p. 5462-73.
75. Koenigsberg, A.L. and E.E. Heldwein, *Crystal structure of the N-terminal half of the traffic controller UL37 from Herpes Simplex virus Type 1.* J Virol, 2017.
76. Baird, N.L., et al., *Myristylation and palmitylation of HSV-1 UL11 are not essential for its function.* Virology, 2010. **397**(1): p. 80-8.
77. Chouljenko, D.V., et al., *Functional hierarchy of herpes simplex virus 1 viral glycoproteins in cytoplasmic virion envelopment and egress.* J Virol, 2012. **86**(8): p. 4262-70.
78. Baines, J.D. and B. Roizman, *The UL11 gene of herpes simplex virus 1 encodes a function that facilitates nucleocapsid envelopment and egress from cells.* J Virol, 1992. **66**(8): p. 5168-74.
79. Gao, J., T.J.M. Hay, and B.W. Banfield, *The Product of the Herpes Simplex Virus 2 UL16 Gene Is Critical for the Egress of Capsids from the Nuclei of Infected Cells.* J Virol, 2017. **91**(10).
80. Mbong, E.F., et al., *Deletion of UL21 causes a delay in the early stages of the herpes simplex virus 1 replication cycle.* J Virol, 2012. **86**(12): p. 7003-7.
81. Baines, J.D., et al., *The herpes simplex virus 1 UL11 proteins are associated with cytoplasmic and nuclear membranes and with nuclear bodies of infected cells.* J Virol, 1995. **69**(2): p. 825-33.
82. Leege, T., et al., *Effects of simultaneous deletion of pUL11 and glycoprotein M on virion maturation of herpes simplex virus type 1.* J Virol, 2009. **83**(2): p. 896-907.
83. Yamauchi, Y., et al., *Herpes simplex virus type 2 UL14 gene product has heat shock protein (HSP)-like functions.* J Cell Sci, 2002. **115**(Pt 12): p. 2517-27.
84. Yamauchi, Y., et al., *Herpes simplex virus UL14 protein blocks apoptosis.* Microbiol Immunol, 2003. **47**(9): p. 685-9.
85. Tanaka, M., T. Sata, and Y. Kawaguchi, *The product of the Herpes simplex virus 1 UL7 gene interacts with a mitochondrial protein, adenine nucleotide translocator 2.* Virol J, 2008. **5**: p. 125.

86. Bigalke, J.M. and E.E. Heldwein, *Expression, Purification and Crystallization of the Herpesvirus Nuclear Egress Complex (NEC).* Bio Protoc, 2016. **6**(14).
87. Andersen, K.R., N.C. Leksa, and T.U. Schwartz, *Optimized E. coli expression strain LOBSTR eliminates common contaminants from His-tag purification.* Proteins, 2013. **81**(11): p. 1857-61.
88. Silverman, J.L., et al., *Membrane requirement for folding of the herpes simplex virus 1 gB cytodomain suggests a unique mechanism of fusion regulation.* J Virol, 2012. **86**(15): p. 8171-84.
89. Sreerama, N. and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: comparison of CONTIN, SELCON, and CDSSTR methods with an expanded reference set.* Anal Biochem, 2000. **287**(2): p. 252-60.
90. Sreerama, N., S.Y. Venyaminov, and R.W. Woody, *Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis.* Anal Biochem, 2000. **287**(2): p. 243-51.
91. Lees, J.G., et al., *A reference database for circular dichroism spectroscopy covering fold and secondary structure space.* Bioinformatics, 2006. **22**(16): p. 1955-62.
92. Whitmore, L. and B.A. Wallace, *Protein secondary structure analyses from circular dichroism spectroscopy: methods and reference databases.* Biopolymers, 2008. **89**(5): p. 392-400.
93. Nielsen, S.S., et al., *BioXTAS RAW, a software program for high-throughput automated small-angle X-ray scattering data reduction and preliminary analysis.* Journal of Applied Crystallography, 2009. **42**: p. 959-964.
94. Franke, D., et al., *ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions.* J Appl Crystallogr, 2017. **50**(Pt 4): p. 1212-1225.
95. Petoukhov, M.V., et al., *ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis.* Journal of Applied Crystallography, 2007. **40**: p. S223-S228.
96. Svergun, D.I., *Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing.* Biophys J, 1999. **76**(6): p. 2879-86.
97. Volkov, V.V. and D.I. Svergun, *Uniqueness of ab initio shape determination in small-angle scattering.* Journal of Applied Crystallography, 2003. **36**: p. 860-864.
98. Tria, G., et al., *Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering.* IUCrJ, 2015. **2**(Pt 2): p. 207-17.
99. Metrick, C.M. and E.E. Heldwein, *Novel Structure and Unexpected RNA-Binding Ability of the C-Terminal Domain of Herpes Simplex Virus 1 Tegument Protein UL21*. J Virol, 2016. **90**(12): p. 5759-69.
100. Heckman, K.L. and L.R. Pease, *Gene splicing and mutagenesis by PCR-driven overlap extension.* Nat Protoc, 2007. **2**(4): p. 924-32.
101. Otwinowski, Z. and W. Minor, *Processing of X-ray diffraction data collected in oscillation mode.* Methods Enzymol, 1997. **276**: p. 307-26.
102. de La Fortelle, E. and G. Bricogne, *[27] Maximum-likelihood heavy-atom parameter refinement for multiple isomorphous replacement and multiwavelength anomalous diffraction methods.* Methods Enzymol, 1997. **276**: p. 472-494.
103. Emsley, P. and K. Cowtan, *Coot: model-building tools for molecular graphics.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2126-32.
104. Adams, P.D., et al., *PHENIX: a comprehensive Python-based system for macromolecular structure solution.* Acta Crystallogr D Biol Crystallogr, 2010. **66**(Pt 2): p. 213-21.
105. Sheldrick, G.M., *A short history of SHELX.* Acta Crystallogr A, 2008. **64**(Pt 1): p. 112-22.
106. Konarev, P.V., et al., *PRIMUS: a Windows PC-based system for small-angle scattering data analysis.* Journal of Applied Crystallography, 2003. **36**: p. 1277-1282.

107. Svergun, D.I., *Determination of the Regularization Parameter in Indirect-Transform Methods Using Perceptual Criteria.* Journal of Applied Crystallography, 1992. **25**: p. 495-503.
108. Franke, D. and D.I. Svergun, *DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering.* J Appl Crystallogr, 2009. **42**(Pt 2): p. 342-346.
109. Petoukhov, M.V., et al., *New developments in the ATSAS program package for small-angle scattering data analysis.* J Appl Crystallogr, 2012. **45**(Pt 2): p. 342-350.
110. Phillips, K. and A.H. de la Pena, *The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization.* Curr Protoc Mol Biol, 2011. **Chapter 10**: p. Unit10 28.
111. MacLean, C.A., et al., *The myristylated virion proteins of herpes simplex virus type 1: investigation of their role in the virus life cycle.* J Gen Virol, 1992. **73 (Pt 3)**: p. 539-47.
112. Fulmer, P.A., et al., *UL20 protein functions precede and are required for the UL11 functions of herpes simplex virus type 1 cytoplasmic virion envelopment.* J Virol, 2007. **81**(7): p. 3097-108.
113. Sadaoka, T., et al., *Varicella-zoster virus ORF49 functions in the efficient production of progeny virus through its interaction with essential tegument protein ORF44.* J Virol, 2014. **88**(1): p. 188-201.
114. Bowzard, J.B., et al., *Membrane targeting properties of a herpesvirus tegument protein-retrovirus Gag chimera.* J Virol, 2000. **74**(18): p. 8692-9.
115. Koshizuka, T., et al., *Association of two membrane proteins encoded by herpes simplex virus type 2, UL11 and UL56.* Virus Genes, 2006. **32**(2): p. 153-63.
116. Chiu, Y.F., et al., *Characterization and intracellular trafficking of Epstein-Barr virus BBLF1, a protein involved in virion maturation.* J Virol, 2012. **86**(18): p. 9647-55.
117. Sanchez, V., E. Sztul, and W.J. Britt, *Human cytomegalovirus pp28 (UL99) localizes to a cytoplasmic compartment which overlaps the endoplasmic reticulum-golgi-intermediate compartment.* J Virol, 2000. **74**(8): p. 3842-51.
118. Loomis, J.S., R.J. Courtney, and J.W. Wills, *Binding partners for the UL11 tegument protein of herpes simplex virus type 1.* J Virol, 2003. **77**(21): p. 11417-24.
119. Sadaoka, T., et al., *Deletion in open reading frame 49 of varicella-zoster virus reduces virus growth in human malignant melanoma cells but not in human embryonic fibroblasts.* J Virol, 2007. **81**(22): p. 12654-65.
120. Loomis, J.S., et al., *Intracellular trafficking of the UL11 tegument protein of herpes simplex virus type 1.* J Virol, 2001. **75**(24): p. 12209-19.
121. Baird, N.L., et al., *Sequences in the UL11 tegument protein of herpes simplex virus that control association with detergent-resistant membranes.* Virology, 2008. **374**(2): p. 315-21.
122. Gluck, J.M., et al., *Single vector system for efficient N-myristoylation of recombinant proteins in E. coli.* PLoS One, 2010. **5**(4): p. e10081.
123. Uversky, V.N., *Protein intrinsic disorder-based liquid-liquid phase transitions in biological systems: Complex coacervates and membrane-less organelles.* Advances in Colloid and Interface Science, 2017. **239**: p. 97-114.
124. Tompa, P., *Intrinsically unstructured proteins.* Trends Biochem Sci, 2002. **27**(10): p. 527-33.
125. Uversky, V.N., *Natively unfolded proteins: a point where biology waits for physics.* Protein Sci, 2002. **11**(4): p. 739-56.
126. Wilkins, D.K., et al., *Hydrodynamic radii of native and denatured proteins measured by pulse field gradient NMR techniques.* Biochemistry, 1999. **38**(50): p. 16424-31.
127. Bernado, P. and D.I. Svergun, *Structural analysis of intrinsically disordered proteins by small-angle X-ray scattering.* Mol Biosyst, 2012. **8**(1): p. 151-67.

128. Bernado, P., *Effect of interdomain dynamics on the structure determination of modular proteins by small-angle scattering.* Eur Biophys J, 2010. **39**(5): p. 769-80.
129. Research, J.P.L.o.t.L.I.f.C. and a.W.a.E.H.I.o.M. Research, *Measuring protein concentration in the presence of nucleic acids by a280/a260: the method of warburg and christian.* CSH Protoc, 2006. **2006**(1).
130. Muto, Y., et al., *Generation and Characterization of UL21-Null Herpes Simplex Virus Type 1.* Front Microbiol, 2012. **3**: p. 394.
131. Klopfleisch, R., et al., *Influence of pseudorabies virus proteins on neuroinvasion and neurovirulence in mice.* J Virol, 2006. **80**(11): p. 5571-6.
132. Klupp, B.G., et al., *Complex formation between the UL16 and UL21 tegument proteins of pseudorabies virus.* J Virol, 2005. **79**(3): p. 1510-22.
133. de Wind, N., et al., *The pseudorabies virus homology of the herpes simplex virus UL21 gene product is a capsid protein which is involved in capsid maturation.* J Virol, 1992. **66**(12): p. 7096-103.
134. Frans Wagenaar, J.M.A.P., Niels DeWind, and T.G. Kimman, *Deletion of the UL21 gene in Pseudorabies virus results in the formation of DNA-deprived capsids: an electron microscopy study.* Vet Res, 2001. **32**: p. 47-54.
135. Michael, K., et al., *Efficient incorporation of tegument proteins pUL46, pUL49, and pUS3 into pseudorabies virus particles depends on the presence of pUL21.* J Virol, 2007. **81**(2): p. 1048-51.
136. Harper, A.L., et al., *Interaction domains of the UL16 and UL21 tegument proteins of herpes simplex virus.* J Virol, 2010. **84**(6): p. 2963-71.
137. Curanovic, D., et al., *Repair of the UL21 locus in pseudorabies virus Bartha enhances the kinetics of retrograde, transneuronal infection in vitro and in vivo.* J Virol, 2009. **83**(3): p. 1173-83.
138. Klupp, B.G., et al., *Mutations affecting the UL21 gene contribute to avirulence of pseudorabies virus vaccine strain Bartha.* Virology, 1995. **212**(2): p. 466-73.
139. Baines, J.D., et al., *The UL21 gene products of herpes simplex virus 1 are dispensable for growth in cultured cells.* J Virol, 1994. **68**(5): p. 2929-36.
140. Davis, I.W., et al., *MolProbity: all-atom contacts and structure validation for proteins and nucleic acids.* Nucleic Acids Res, 2007. **35**(Web Server issue): p. W375-83.
141. Holm, L. and P. Rosenstrom, *Dali server: conservation mapping in 3D.* Nucleic Acids Res, 2010. **38**(Web Server issue): p. W545-9.
142. Lichtarge, O., H.R. Bourne, and F.E. Cohen, *An evolutionary trace method defines binding surfaces common to protein families.* J Mol Biol, 1996. **257**(2): p. 342-58.
143. Sowa, M.E., et al., *Prediction and confirmation of a site critical for effector regulation of RGS domain activity.* Nat Struct Biol, 2001. **8**(3): p. 234-7.
144. Chakravarty, S., et al., *Evolutionary trace residues in noroviruses: importance in receptor binding, antigenicity, virion assembly, and strain diversity.* J Virol, 2005. **79**(1): p. 554-68.
145. Buchan, D.W., et al., *Scalable web services for the PSIPRED Protein Analysis Workbench.* Nucleic Acids Res, 2013. **41**(Web Server issue): p. W349-57.
146. Pesce, A., M. Bolognesi, and M. Nardini, *The diversity of 2/2 (truncated) globins.* Adv Microb Physiol, 2013. **63**: p. 49-78.
147. Bowman, B.R., et al., *Structural characterization of the UL25 DNA-packaging protein from herpes simplex virus type 1.* J Virol, 2006. **80**(5): p. 2309-17.
148. Joint Proteomic, S.L.o.t.L.I.f.C.R., Walter, and M.A. Eliza Hall Institute of Medical Research, *Measuring protein concentration in the presence of nucleic acids by a280/a260: the method of warburg and christian.* CSH Protoc, 2006. **2006**(1).

149. Guo, H., et al., *Open reading frame 33 of a gammaherpesvirus encodes a tegument protein essential for virion morphogenesis and egress.* J Virol, 2009. **83**(20): p. 10582-95.
150. Shen, S., et al., *Gammaherpesvirus Tegument Protein ORF33 Is Associated With Intranuclear Capsids at an Early Stage of the Tegumentation Process.* J Virol, 2015. **89**(10): p. 5288-97.
151. Meckes, D.G., Jr. and J.W. Wills, *Structural rearrangement within an enveloped virus upon binding to the host cell.* J Virol, 2008. **82**(21): p. 10429-35.
152. Fraternale, A., et al., *GSH and analogs in antiviral therapy.* Mol Aspects Med, 2009. **30**(1-2): p. 99-110.
153. Chadha, P., et al., *Domain Interaction Studies of Herpes Simplex Virus 1 Tegument Protein UL16 Reveal Its Interaction with Mitochondria.* J Virol, 2017. **91**(2).
154. Schoneberg, J., et al., *Reverse-topology membrane scission by the ESCRT proteins.* Nat Rev Mol Cell Biol, 2017. **18**(1): p. 5-17.
155. Henne, W.M., H. Stenmark, and S.D. Emr, *Molecular mechanisms of the membrane sculpting ESCRT pathway.* Cold Spring Harb Perspect Biol, 2013. **5**(9).
156. Schauflinger, M., et al., *The tegument protein UL71 of human cytomegalovirus is involved in late envelopment and affects multivesicular bodies.* J Virol, 2011. **85**(8): p. 3821-32.
157. Ren, X. and J.H. Hurley, *Proline-rich regions and motifs in trafficking: from ESCRT interaction to viral exploitation.* Traffic, 2011. **12**(10): p. 1282-90.
158. Meissner, C.S., et al., *A leucine zipper motif of a tegument protein triggers final envelopment of human cytomegalovirus.* J Virol, 2012. **86**(6): p. 3370-82.
159. Fraile-Ramos, A., et al., *The ESCRT machinery is not required for human cytomegalovirus envelopment.* Cell Microbiol, 2007. **9**(12): p. 2955-67.
160. Fagerlund, R., et al., *Arginine/lysine-rich nuclear localization signals mediate interactions between dimeric STATs and importin alpha 5.* J Biol Chem, 2002. **277**(33): p. 30072-8.
161. Ingemarson, R. and H. Lankinen, *The herpes simplex virus type 1 ribonucleotide reductase is a tight complex of the type alpha 2 beta 2 composed of 40K and 140K proteins, of which the latter shows multiple forms due to proteolysis.* Virology, 1987. **156**(2): p. 417-22.
162. Donnelly, M., J. Verhagen, and G. Elliott, *RNA binding by the herpes simplex virus type 1 nucleocytoplasmic shuttling protein UL47 is mediated by an N-terminal arginine-rich domain that also functions as its nuclear localization signal.* J Virol, 2007. **81**(5): p. 2283-96.
163. Newcomb, W.W., et al., *Role of a reducing environment in disassembly of the herpesvirus tegument.* Virology, 2012. **431**(1-2): p. 71-9.
164. Mitrea, D.M. and R.W. Kriwacki, *Phase separation in biology; functional organization of a higher order.* Cell Commun Signal, 2016. **14**: p. 1.
165. Liu, Y., et al., *Crystal structure of the conserved core of the herpes simplex virus transcriptional regulatory protein VP16.* Genes Dev, 1999. **13**(13): p. 1692-703.