The Role of Novel Gene UL79 Encoded by Human Cytomegalovirus in Viral Replication

Yi-Chieh Perng

Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/etd

Recommended Citation


This Dissertation is brought to you for free and open access by Washington University Open Scholarship. It has been accepted for inclusion in All Theses and Dissertations (ETDs) by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.
The Role of Novel Gene UL79 Encoded by Human Cytomegalovirus in Viral Replication

by

Yi-Chieh “EJ” Perng

A dissertation presented to the
Graduate School of Arts and Sciences
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

August 2014

Saint Louis, Missouri
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>List of Figures</td>
<td>iv</td>
</tr>
<tr>
<td>List of Tables</td>
<td>vi</td>
</tr>
<tr>
<td>List of Abbreviations</td>
<td>vii</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>ix</td>
</tr>
<tr>
<td>Abstract</td>
<td>xi</td>
</tr>
<tr>
<td><strong>Chapter 1 – Introduction</strong></td>
<td>1</td>
</tr>
<tr>
<td>Introduction</td>
<td>2</td>
</tr>
<tr>
<td>Figures</td>
<td>10</td>
</tr>
<tr>
<td>Tables</td>
<td>16</td>
</tr>
<tr>
<td>References</td>
<td>17</td>
</tr>
<tr>
<td><strong>Chapter 2 - Characterization of HCMV novel essential protein pUL79 and determine its roles in viral late gene expression</strong></td>
<td>21</td>
</tr>
<tr>
<td>Abstract</td>
<td>22</td>
</tr>
<tr>
<td>Introduction</td>
<td>23</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>26</td>
</tr>
<tr>
<td>Results</td>
<td>32</td>
</tr>
<tr>
<td>Discussion</td>
<td>42</td>
</tr>
<tr>
<td>Figures</td>
<td>46</td>
</tr>
<tr>
<td>Tables</td>
<td>60</td>
</tr>
<tr>
<td>References</td>
<td>61</td>
</tr>
<tr>
<td><strong>Chapter 3 - Identification of pUL79 protein partners and the mechanism of pUL79 activity</strong></td>
<td>65</td>
</tr>
<tr>
<td>Abstract</td>
<td>66</td>
</tr>
<tr>
<td>Introduction</td>
<td>67</td>
</tr>
<tr>
<td>Materials and Methods</td>
<td>71</td>
</tr>
</tbody>
</table>
Chapter 4 Summary and Future Directions

Summary

Future Directions

Figures

References
# LIST OF FIGURES

## Chapter 1

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Lytic infection cycle of HCMV</td>
<td>10</td>
</tr>
<tr>
<td>1.2</td>
<td>Functional map of the HCMV Genome</td>
<td>11</td>
</tr>
<tr>
<td>1.3</td>
<td>Using ddFKBP-Shield-1 system to conditionally control protein stability</td>
<td>13</td>
</tr>
<tr>
<td>1.4</td>
<td>Growth kinetics of the recombinant HCMV in the presence or absence of Shield-1</td>
<td>14</td>
</tr>
</tbody>
</table>

## Chapter 2

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1</td>
<td>(A) Schematic diagram for creating recombinant virus ADddUL79 and applying a destabilization domain (ddFKBP)-mediated protein stability-based approach to regulate gene function. (B) Dose-dependent regulation of ADddUL79 replication by Shield-1. (C) HFFs overexpressing the hemagglutinin (HA)-tagged UL79 complemented the growth of ADddUL79 in the absence of Shield-1</td>
<td>46</td>
</tr>
<tr>
<td>2.2</td>
<td>UL79 is essential for HCMV replication at late stages of infection</td>
<td>48</td>
</tr>
<tr>
<td>2.3</td>
<td>UL79 is required for the efficient accumulation of late viral proteins during HCMV infection</td>
<td>50</td>
</tr>
<tr>
<td>2.4</td>
<td>UL79 is required for the efficient accumulation of late viral transcripts during HCMV infection</td>
<td>51</td>
</tr>
<tr>
<td>2.5</td>
<td>Viral DNA replication is not altered in the absence of UL79 during infection</td>
<td>52</td>
</tr>
<tr>
<td>2.6</td>
<td>UL79 is transcribed with early-late kinetics and is regulated by a positive-feedback loop</td>
<td>53</td>
</tr>
<tr>
<td>2.7</td>
<td>A mutant virus in which the UL79 ORF has been abrogated by a stop codon mutation has the same defect as ADddUL79</td>
<td>54</td>
</tr>
<tr>
<td>2.8</td>
<td>UL79 proteins are primarily nuclear proteins that localize within viral replication compartments during infection</td>
<td>56</td>
</tr>
<tr>
<td>2.9</td>
<td>The UL79 protein is not detected in HCMV virions</td>
<td>58</td>
</tr>
<tr>
<td>2.10</td>
<td>Immunoblot analysis of viral ddFKBP fusion proteins</td>
<td>59</td>
</tr>
</tbody>
</table>
Chapter 3
3.1 – Identification of pUL79 interacting proteins.................................................................91
3.2 – pUL79 interacts with the RNAP II protein complex.......................................................93
3.3 – pUL79 does not alter protein accumulations of RNAP II.................................................95
3.4 – pUL79 alters RNAP II occupancy at viral loci.................................................................96
3.5 – pUL79 does not alter a particular phosphorylated form of the RNAP II CTD domain......97
3.6 – pUL79 alters the rate of transcriptional elongation at viral loci......................................99
3.7 – Potential role of pUL79 in RNAP II-mediated viral transcription.................................101

Chapter 4
4.1 – Both N- and C- terminuses of pUL79 are required for pUL79-RNAP II interaction........121
4.2 – pUL79 potentially contains RPB1 CTD-binding domains.............................................122
4.3 – Deletion of pUL79’s CID-like domains reduces viral late protein accumulation and viral progeny production........................................................................................................123
4.4 – Identification of residues critical to pUL79-RPB1 mediated viral replication..............124
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Chapter 1</strong></td>
<td>1.1 – Selected Uncharacterized HCMV Essential and Augmenting Genes</td>
<td>16</td>
</tr>
<tr>
<td><strong>Chapter 2</strong></td>
<td>2.1 – Primers and probes used for RT-qPCR in Chapter 2</td>
<td>60</td>
</tr>
<tr>
<td><strong>Chapter 3</strong></td>
<td>3.1 – pUL79 protein partners identified by mass spectrometry</td>
<td>103</td>
</tr>
<tr>
<td></td>
<td>3.2 – Antibodies used in chromatin immunoprecipitation assays</td>
<td>104</td>
</tr>
<tr>
<td></td>
<td>3.3 – Primers and probes used in ChIP and RT-qPCR analysis</td>
<td>105</td>
</tr>
</tbody>
</table>
### LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>BAC</td>
<td>bacterial artificial chromosomes</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>CDV</td>
<td>cidofovir</td>
</tr>
<tr>
<td>ChIP</td>
<td>chromatin immunoprecipitation</td>
</tr>
<tr>
<td>CTD</td>
<td>carboxyl terminal domain</td>
</tr>
<tr>
<td>DD</td>
<td>destabilizing domain</td>
</tr>
<tr>
<td>ddfFKBP</td>
<td>unstable variant of the FKBP12 protein</td>
</tr>
<tr>
<td>DMEM</td>
<td>Dulbecco modified Eagle medium</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr virus</td>
</tr>
<tr>
<td>Flp</td>
<td>Flippase</td>
</tr>
<tr>
<td>FOS</td>
<td>forscarnet</td>
</tr>
<tr>
<td>FTICR</td>
<td>Fourier transform ion cyclotron resonance</td>
</tr>
<tr>
<td>GAPDH</td>
<td>glyceraldehyde-3-phosphate dehydrogenase</td>
</tr>
<tr>
<td>GCV</td>
<td>Ganciclovir</td>
</tr>
<tr>
<td>GFP</td>
<td>green fluorescent protein</td>
</tr>
<tr>
<td>GTF</td>
<td>general transcription factor</td>
</tr>
<tr>
<td>H3Ac</td>
<td>histone 3 acetylation</td>
</tr>
<tr>
<td>H3K4me2</td>
<td>histone 3 with lysine 4 methylation</td>
</tr>
<tr>
<td>H3K9</td>
<td>histone 3 lysine 9</td>
</tr>
<tr>
<td>HA</td>
<td>hemagglutinin</td>
</tr>
<tr>
<td>HCMV</td>
<td>human cytomegalovirus</td>
</tr>
<tr>
<td>HFF</td>
<td>human foreskin fibroblast</td>
</tr>
<tr>
<td>HRP</td>
<td>horseradish peroxidase</td>
</tr>
<tr>
<td>HSV</td>
<td>herpes simplex virus</td>
</tr>
<tr>
<td>IE1</td>
<td>HCMV gene UL123</td>
</tr>
<tr>
<td>IE2</td>
<td>HCMV gene UL122</td>
</tr>
<tr>
<td>LC-MS/MS</td>
<td>liquid chromatography-tandem mass spectrometry</td>
</tr>
<tr>
<td>MCMV</td>
<td>murine cytomegalovirus</td>
</tr>
<tr>
<td>MCP</td>
<td>major capsid protein</td>
</tr>
</tbody>
</table>

vii
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHV-68</td>
<td>murine gammaherpesvirus 68</td>
</tr>
<tr>
<td>MIE HCMV</td>
<td>major immediate-early genes</td>
</tr>
<tr>
<td>MOI</td>
<td>multiplicity of infection</td>
</tr>
<tr>
<td>MRC5</td>
<td>embryonic lung fibroblast</td>
</tr>
<tr>
<td>NELFs</td>
<td>negative elongation factors</td>
</tr>
<tr>
<td>NRO</td>
<td>nuclear run-on assay</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>OriLyt</td>
<td>origin of viral DNA synthesis</td>
</tr>
<tr>
<td>PAA</td>
<td>peroxycetic acid</td>
</tr>
<tr>
<td>PBS</td>
<td>phosphate-buffered saline</td>
</tr>
<tr>
<td>PCNA</td>
<td>proliferating cell nuclear antigen</td>
</tr>
<tr>
<td>PEI</td>
<td>polyethylenimine</td>
</tr>
<tr>
<td>PIC</td>
<td>pre-initiation complex</td>
</tr>
<tr>
<td>POI</td>
<td>protein of interest</td>
</tr>
<tr>
<td>PVDF</td>
<td>polyvinylidene difluoride</td>
</tr>
<tr>
<td>qPCR</td>
<td>quantitative real-time PCR</td>
</tr>
<tr>
<td>RNAP I</td>
<td>cellular RNA polymerase I</td>
</tr>
<tr>
<td>RNAP II</td>
<td>cellular RNA polymerase II</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>reverse transcription-coupled qPCR</td>
</tr>
<tr>
<td>SDS</td>
<td>sodium dodecyl sulfate</td>
</tr>
<tr>
<td>Shld1</td>
<td>Shield-1</td>
</tr>
<tr>
<td>SV40</td>
<td>simian virus 40</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA-box binding protein</td>
</tr>
<tr>
<td>TCID&lt;sub&gt;50&lt;/sub&gt;</td>
<td>50% tissue culture infectious dose assay</td>
</tr>
<tr>
<td>TRS/IRS</td>
<td>terminal repeats/internal repeats</td>
</tr>
<tr>
<td>UL</td>
<td>unique long region</td>
</tr>
<tr>
<td>US</td>
<td>unique short region</td>
</tr>
<tr>
<td>VGCV</td>
<td>Valganciclovir</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

This thesis is done with the help and support of many people. First I would like to thank my thesis advisors, Dong and Debbie. Dong allowed me to pursue my Ph.D. research on HCMV viral pathogenesis in his lab. He gave me the freedom to pursue the topics I am interested in but prioritized what to do to make projects on track. Debbie picked me up after Dong’s leave and shaped me the thoughts how pUL79 interacts with RNA polymerase II. I also want to thank all my committee members’ supports. Without their inspirational suggestions and inputs, my projects cannot move any further. Especially, I want to thank Henry. The brainstorming with him was an interesting adventure and created a lot of ideas worthy to pursue.

I want to thank all the members of both Yu and Lenschow labs. Zhikang and Baoqin provided excellent technical supports. Anne and Tony taught me how to write scientific papers in English. Jesse and Mari helped me solve the problems of both ChIP and NRO assays. With their warm supports, I can have a chance to know the beauty of science.

I also want to Michael Nevels (University of Regensburg, Germany), Gregory Pari (University of Nevada), Efrem Lim (Wang Lab), and Stacy Gish (Doering lab), for their technical assistance with chromatin immunoprecipitation. Members of Lodge lab, Kimberly Gerik, Rajendra Upadhya, and Woei Lam, helped me setup the nuclear run-on assay and made the experiments feasible.

Most importantly, I want to thank my family. My father is the reason I want to study viral pathogenesis. I will never forget I am a son of a patient who eager to live to see how wonderful the world is. My mother supported the whole family after my father passed away. The supports of her and my sister’s family help me overcame all the difficulties. My wife, Chia-Li, and my
daughter, Sabrina, are the biggest supports behind my in the past years. Without their sacrifice, there is nothing I can do.

In the end, I want to thank God. I am blessed and surrounded by angels he sent. Being a scientist and having chances to know how amazing he created is the biggest blessing in my life.
ABSTRACT OF THE DISSERTATION

The Role of Novel Gene UL79 Encoded by Human Cytomegalovirus in Viral Replication

by

Yi-Chieh Perng

Doctor of Philosophy in Biology and Biomedical Sciences
Molecular Microbiology and Microbial Pathogenesis
Washington University in St. Louis, 2014
Professor Deborah J Lenschow, Chairperson

Human cytomegalovirus (HCMV) is a ubiquitous pathogen that infects the majority of the world’s population. Even though HCMV infection is usually asymptomatic, it acts as an opportunistic pathogen and is the major cause of morbidity and mortality in immunocompromised individuals, including transplant recipients and AIDS/HIV patients. Understanding HCMV biology is critical to the development of HCMV therapeutics. However, our current knowledge of HCMV biology is limited by the fact that only less than half of the HCMV genes have been characterized, especially viral essential and augmenting genes which play critical roles in viral replication.

To study the function of HCMV essential and augmenting genes, we devised a conditional approach to facilitate the analysis. In this approach, we constructed recombinant virus where the viral open reading frame (ORF) of interest is tagged with the destabilization domain FKBP (ddFKBP) which targets the fusion protein for rapid degradation. However, the fusion protein can be stabilized by the synthetic ligand Shield-1 (Shld-1). This system allows us to monitor the
effects of the viral protein on the HCMV life cycle simply by infecting human fibroblasts in the presence or absence of Shld-1.

We adopted this conditional protein genetic approach to characterize the role of the human cytomegalovirus (HCMV) gene UL79 during virus infection. We constructed ADddUL79, a recombinant HCMV in which the annotated UL79 open reading frame (ORF) was tagged with ddkBP. ADddUL79 failed to replicate without Shield-1, but it grew at wild-type levels with Shield-1 or in human foreskin fibroblasts overexpressing hemagglutinin (HA)-tagged UL79 (HF-UL79HA cells), indicating an essential role of UL79 and the effectiveness of this approach. Without Shield-1, representative immediate-early and early viral proteins as well as viral DNA accumulated normally, but late transcripts and proteins were markedly reduced. UL79 was transcribed with early-late kinetics, and was also regulated via a positive-feedback loop. We also found that the UL79 protein localized to viral replication compartments during HCMV infection. Moreover, we created a second UL79 mutant virus (ADinUL79stop) in which the UL79 ORF was disrupted by a stop codon mutation and found that ADinUL79stop phenocopied ADddUL79 under the destabilizing condition. Taking together, we showed that UL79 acts after viral DNA replication to promote the accumulation of late viral transcripts.

To test how pUL79 regulates viral late gene transcription, we analyzed the proteome associated with pUL79 during virus infection by mass spectrometry. We identified both cellular transcriptional factors, including multiple RNA polymerase II (RNAP II) subunits, and novel viral transactivators, including pUL87, pUL92, and pUL95, as protein binding partners of pUL79. Co-immunoprecipitation (co-IP) followed by immunoblot analysis confirmed the pUL79-RNAP II interaction, and this interaction was independent of any other viral proteins. We showed that this interaction did not alter the total levels of RNAP II or its recruitment to viral
late promoters. Furthermore, pUL79 did not alter the phosphorylation profiles of the RNAP II C-terminal domain, which is critical for transcriptional regulation. Rather, nuclear run-on assay indicated that, in the absence of pUL79, RNAP II failed to elongate and stalled on viral DNA. Surprisingly, pUL79-dependent RNAP II elongation was required for transcription from all three kinetic classes of viral genes (i.e. immediate-early, early, and late) at late times during virus infection. In contrast, host gene transcription during HCMV infection was independent of pUL79. In summary, we have identified a novel viral mechanism by which pUL79 regulates the rate of RNAP II transcription machinery on viral transcription during late stages of HCMV infection.

Together these data provide important insight into how HCMV uses pUL79 to promote viral transcription specifically during late stages of viral infection. Understanding the mechanisms by which pUL79 regulates RNAP II elongation as well as its association with other viral factors will aid in the development of future therapeutics against HCMV infection.
Chapter 1:

Introduction
INTRODUCTION

**HCMV as human pathogen.** Human cytomegalovirus (HCMV), the prototypic β-herpesvirus, is a ubiquitous pathogen that infects the majority of the world’s population. Upon primary infection, HCMV establishes a lifelong persistent and latent/recurrent infection in a host(12). Even though HCMV infection is usually asymptomatic, it acts as an opportunistic pathogen and is the major cause of morbidity and mortality in immunocompromised individuals, including transplant recipients and AIDS/HIV patients(13). In addition, HCMV is the leading viral cause of birth defects leading to mental retardation and blindness. Each year, HCMV infects approximately 30,000 newborns, causing an estimated 400 deaths and 5,000 disabilities in the United States alone(1). The health and economic impact of HCMV infection is tremendous. The healthcare cost associated with HCMV-related disease in the United States is estimated at 4 billion dollars per year(22). Moreover, there is evidence for HCMV to act as a risk factor in the development of vascular diseases, such as atherosclerosis, transplant vascular sclerosis, and coronary restenosis after angioplasty surgery(19, 26, 32, 36, 48, 52, 59). Recently, a role for HCMV has also been implicated in multiple forms of human cancers, where it may contribute to oncogenic transformation, onco-modulation, and tumor cell immune evasion(6, 16, 23, 42).

**HCMV therapeutics.** To date, a limited number of drugs are licensed for the treatment of HCMV infection and no vaccine is available(1). Current standard therapy for HCMV relies on oral/intravenous ganciclovir (GCV) or its oral prodrug, valganciclovir (VGCV). Although efficacious, GCV treatment suffers from dose-related toxicities. Forscanet (FOS) and cidofovir (CDV), the two commonly used second-line treatments for HCMV, are also associated with significant toxicities, including renal toxicity. In addition, during
prolonged/repeated application, HCMV can become resistant to GCV and lead to treatment failure in immunocompromised patients. Of concern, GCV resistance in HCMV infections has increased during recent years, and the emergence of cross-resistance to either or both second-line agents (FOS & CDV) is frequently encountered because all licensed drugs share a common target molecule, the viral DNA polymerase, pUL54(28). Moreover, all current HCMV antivirals are teratogenic, mutagenic, and potential carcinogens(4, 24, 27, 29). The dosing limitations due to toxicities further facilitate the emergence of drug-resistance. Currently no HCMV vaccine is available(1). Therefore, there is an urgent need to develop new, safe, and efficacious antiviral drugs. A better understanding of HCMV biology, its interactions with the host, and pathogenesis is critical for us to develop novel and more effective antiviral therapeutics to control this globally important pathogen.

**HCMV biology and reverse genetics.** HCMV is the largest known human virus with a genome of ~240 kb encoding at least 166 annotated viral open reading frames (ORFs)(37, 38) and several miRNAs(15, 21). The viral genome is composed of a unique long (UL) and a unique short (US) region with repeated segments at the end of each region, namely TRL/IRL and TRS/IRS, respectively. ORFs are numbered sequentially in each region. HCMV displays a lifecycle very similar to other herpesviruses but with a slow replication kinetics, as the virus does not produce progeny virions until 48-72 hours after infection in primary human fibroblast cultures *in vitro*. After binding and internalization, the viral DNA is transported to the nucleus after which the immediate-early transcripts are expressed. Viral DNA replication occurs between 24 and 48 hours after infection followed by late gene expression and capsid assembly. Capsids are then translocated out of the nucleus and acquire their envelope in the cytoplasm. Once virions are formed they exit the cell beginning around 48 hours post
infection and continuing for several days until the cell dies(17). The lytic infection cycle of HCMV is also diagramed in Figure 1.1.

HCMV exhibits strict host specificity as it can only infect human cells. On the other hand, HCMV can infect a wide range of cell types in its natural host, including epithelial cells, endothelial cells, smooth muscle cells, fibroblasts, neurons, and macrophages. The virus spreads systemically via a leukocyte-associated viremia, and the primary infection can continue for months in adults and years in children. Following the prolonged acute infection the virus enters a lifelong latent state in lineage-committed myeloid cells and possibly endothelial cells. The latent virus can then reactivate following differentiation of latently infected cells into cells permissive for HCMV lytic infection, the reduction in immune pressure, or due to other unknown determinants(17).

Laboratory-adapted strains of HCMV, such as AD169 and Towne, have been widely used to study HCMV biology. The laboratory-adapted strains were created by passaging the virus in fibroblasts for > 100 times in an effort to create a live attenuated vaccine. These strains replicate efficiently in fibroblasts but inefficiently in other cell types, such as endothelial cells, epithelial cells, and macrophages(17), which are readily infected by clinical strains of HCMV. In AD169, this loss in cell tropism can be partially attributed to mutation in the UL131 genes, as an AD169 variant which has the UL131 gene repaired can efficiently infect both epithelial and endothelial cells(55, 56). We and others have cloned laboratory strains (AD169 and Towne)(58), low-passage strain (Toledo), as well as several clinical isolates (TR and Fix) as infectious bacterial artificial chromosomes (BAC) and developed BAC-based reverse genetic system for HCMV. The advent of the HCMV reverse genetics greatly enhances our ability to dissect the molecular mechanisms underlining its biology and pathogenesis(39).
**Expression kinetics of HCMV genes.** During lytic infection, HCMV genes are expressed in a highly ordered temporal cascade(3, 34, 50, 57). Viral transcripts accumulate with three kinetic classes, namely immediate-early, early, and late. The HCMV major IE (MIE) genes UL123 (IE1) and UL122 (IE2) play critical roles in predisposing the cellular environment to infection and also act as transactivators to induce early gene transcription. Many early genes encode proteins required for viral DNA synthesis(18, 20, 51). The transcript accumulation of early genes is independent of viral DNA synthesis; however, the continued accumulation of a subset of these genes at late times is enhanced by the onset of viral DNA synthesis(49).

Following viral DNA synthesis, late viral genes start to transcribe, which mostly encode structural proteins required for virus assembly and egress, ultimately leading to the release of infectious particles. Previous studies have shown that the activation of both beta- and gamma-herpesvirus late gene promoters is dependent on the origin of viral DNA synthesis (OriLyt) in cis(2, 14, 35). This further supports the notion that late gene transcription is tightly coupled to viral DNA synthesis. However, whether viral late gene expression is subjected to additional viral regulation remains poorly defined.

**RNA polymerase II (RNAP II) transcribe HCMV genes.** During cytomegalovirus infection, viral genes are transcribed by cellular RNA polymerase II (RNAP II), a large multi-subunit enzyme. Rpb1, the largest subunit of RNAP II, has a carboxy terminal domain (CTD) that contains repeats of the heptapeptide sequence of Tyr1-Ser2-Pro3-Thr4-Ser5-Pro6-Ser7(11). The mammalian CTD is compromised of 52 copies of heptapeptides(8). This unstructured domain is strategically located adjacent to the mRNA exit channel of the enzyme. This helps the CTD act as a scaffold to coordinate interactions with proteins involved in different phases of transcription and allow for the coupling of transcription with other nuclear processes, such as mRNA maturation and chromatin modification(40, 46).
amino acids within the CTD repeats are targets for post-translational modifications, including phosphorylation, glycosylation, ubiquitination, and cis-trans isomerization(10). Different combinations of CTD modifications, known as the CTD codes, orchestrate the sequential recruitment of numerous factors during the transcription cycle(7).

RNAP II activity is tightly regulated by phosphorylation of its CTD. RNAP II is recruited to preinitiation complexes (PIC) in an unphosphorylated state(30). Following the binding of RNAP II to a promoter, serine 5 residue of the CTD (Ser5) is rapidly phosphorylated by cdk7, the kinase subunit of the transcription factor TFII-H. This facilitates the dissociation of RNAP II from the PIC(47), and also promotes the recruitment of capping and splicing factors as well as other histone modification complexes(40). After dissociating from PIC, RNAP II proceeds to intrinsic pausing sites where it is halted by negative elongation factors (NELFs). The onset of productive elongation requires the positive transcription elongation factor P-TEFb composed of cdk9 and cyclin T, which phosphorylates serine 2 residue (Ser2) of the CTD to produce stable elongation complexes(31). Ser2 phosphorylation also couples RNA synthesis with RNA processing by promoting the recruitment of splicing and polyadenylation factors(41). At the 3’ end of the coding region, RNAP II dissociates from the DNA template and RNA transcript prior to transcript polyadenylation. Specific phosphatases, Ssu72 and Fcp1, also dephosphorylate the CTD. Thus, RNAP II is recycled as an unphosphorylated, initiation-competent form for another round of transcription(9, 25).

HCMV utilizes RNAP II and the accompanying host machinery for transcription of viral genes. During early times of viral infection, RNAP II and other transcription machinery are recruited to early replication sites to drive viral IE and early gene expression(53). The protein levels of RNAP II, including hyper-phosphorylated forms, increase as infection progresses(53, 54). Treatment of infected cells with cdk inhibitors inhibits viral gene
expression as well as viral replication(43). During late stages of viral infection, cdk kinase and RNAP II-associated transcriptional machinery proteins continue to accumulate and relocate into the peri-replication center(54). However, how RNAP II transcription machinery remains active on viral loci during late infection requires further investigation.

**Functional profile of HCMV genes.** Our current knowledge of HCMV biology is limited by the fact that only less than half of the HCMV genes have been characterized. With the development of the BAC-based genetic system for HCMV, our group and others have used global mutagenesis approaches to delineate the functions of the genes encoded by HCMV(58). These studies led to the creation of a functional genomic map for HCMV strains AD169 and Towne. In AD169, we identified 3 different classes of genes based on how they are required for virus growth properties in human foreskin fibroblasts (HFFs) (Figure 1.2). There were 39 essential gene (mutants that produce no infectious particles), 100 nonessential genes (mutants that had no growth defect in human fibroblasts), and 27 augmenting genes (mutants that produced small plaques or had a >10 fold in replication on HFFs(58). Among 66 essential and augmenting viral genes, 25 of them remain uncharacterized and their roles in HCMV infection are completely unknown. This is because complementing cells that support growth of null mutants are difficult to obtain for β- and γ- herpesviruses, such as HCMV, and therefore it is often difficult, if not impossible, to generate recombinant HCMV viruses, especially the ones defective in essential genes. Currently, only a very few complementing cell lines have been documented for HCMV mutant viruses(33, 44, 45). As essential and augmenting genes are critical for efficient virus growth, insight into functions of these genes is a prerequisite for the rational development of antiviral compounds and vaccine vectors.
**ddFKBP complement system.** To study the function of HCMV essential and augmenting genes, an efficient genetic system is the prerequisite tool. Null mutants of essential genes produce no progeny and the ones of augmenting genes may have severe growth defects and produce low viral titers, thus impeding the functional characterization of these genes. The conventional complementing approach, namely the use of cells expressing mutated gene products to support growth of null mutants, have been reported for only a very few HCMV genes because it is difficult to obtain such complementing cells for slowing replicating β-herpesviruses (33, 44, 45). Therefore, development of an alternative and efficient complementation system is necessary for functional studies of HCMV essential and augmenting genes.

To overcome this hurdle, we devised a conditional approach to facilitate the analysis of viral proteins critical to viral growth and pathogenesis. A new approach to conditionally regulate proteins has been established by tagging a protein with a 102-aa unstable variant of the FKBP12 protein (ddFKBP) (5). In this approach, we construct recombinant virus where the viral open reading frame (ORF) of interest is tagged with the destabilization domain FKBP (ddFKBP) which targets the fusion protein for rapid degradation (5). However, the fusion protein can be stabilized by the synthetic ligand Shield-1 (Shld-1). This system allows us to monitor the effects of the viral protein on the HCMV life cycle simply by infecting human fibroblasts in the presence or absence of Shld-1 (Figure 1.3).

**Create a library of recombinant HCMV viruses where accumulation of novel essential and augmenting viral proteins can be regulated by ddFKBP-based protein genetic approach.** Twelve essential genes and 13 augmenting genes for which null mutants have severe growth defects have not been characterized because of the lack of efficient complementing system (Table 1.1). We created recombinant HCMVs in which these 25 viral
genes were tagged with ddFKBP. The 25 genes that were chosen for this study fit two criteria: 1) these genes are either essential or augmenting for HCMV replication in fibroblasts; 2) these genes and their homologous in other herpesviruses have not previously been characterized. Using linear recombination, we constructed 24/25 ddFKBP fusions to create conditional recombinant viruses for further analysis. (Table 1.1) Recombinant BAC clones for 10 essential genes and 9 augmenting genes had been electroporated into HFFs to reconstitute recombinant viruses in the presence of Shld-1. Among these 19 recombinant viruses, the ones for UL71(ADddUL71), UL77(ADddUL77), UL79(ADddUL79), UL91(ADddUL91), and UL103(ADddUL103), were efficiently regulated by Shld-1 (Table.1.1). In the presence of Shld-1, these recombinant viruses replicated at wild-type levels while in the absence of Shld-1 recombinant viruses reconstituted from BAC transfection were not detected (Growth curves of representative genes shown in Figure 1.4).

**Goal of the dissertation.** The ddFKBP-based viral genetic system allow us to propagate recombinant viruses for multiple candidate viral genes, and a prioritization is necessary for us to select one viral gene for further analysis. Among the five genes that can be regulated by ddFKBP, UL79 was shown to be required for late stages of viral infection in our preliminary analysis (Data not shown). HCMV late gene expression is critical to viral replication, however, how and what viral factors regulate it was completely unknown. Therefore, we chose UL79 for further analysis. In this dissertation, we characterized the function of UL79 during viral infection and elucidated the mechanism of how it facilitates HCMV gene expression and viral growth.
Figure 1.1. Lytic infection cycle of HCMV.
Figure 1.2. **Functional Map of the HCMV Genome.** ORFs are color coded according to whether they are essential (blue) or nonessential (red), or augment (green) replication in human fibroblasts (HFFs). For essential genes, the electroporated BACs of recombinant HCMV generated GFP-positive cells 2-3 days after electroporation but failed to produce virus. For augmenting genes, the BACs produced small plaques and/or generated \( \geq 10 \)-fold
reduced yields of cell-free virus. For nonessential genes, the electroporated BACs produced normal-sized plaques and virus stocks with wild-type yields(58)
Figure 1.3. Using ddFKBP-Shield-1 System to Conditionally Control Protein Stability.

Fusion of a destabilizing domain ddFKBP (DD) to a protein of interest (POI) results in degradation of the entire fusion. Addition of a Shield-1, a ligand for the destabilizing domain, protects the fusion from degradation(5).
Figure 1.4. Growth kinetics of the recombinant HCMV in the presence or absence of Shield-1. Fibroblasts infected with the ADddUL71, ADddUL77, and ADddUL79 recombinant viruses were cultivated with or without 1 µM Shield-1. At the time points indicated, supernatants were collected and analyzed for the presence of virus by TCID$_{50}$ assays on fibroblasts maintained in 1 µM Shield-1 (Shld1/Shld1). To test for leakiness of the system, supernatant of infected cells grown in the absence of Shield-1 was assayed by TCID$_{50}$ on cells receiving the Shield-1 ligand (nonShld-1/Shld1). To test the occurrence of
revertant viruses, both the growth curve and the TCID\textsubscript{50} assays were performed in the absence of Shield-1 (nonShld1/nonShld1).
<table>
<thead>
<tr>
<th>gene</th>
<th>type</th>
<th>size</th>
<th>ddFKBP-tagged BAC</th>
<th>BAC DNA transfected</th>
<th>Regulated by Shld-1</th>
<th>Grow like wt without Shld-1</th>
<th>Virus do not grow with Shld-1</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL103</td>
<td>augmenting</td>
<td>28 kDa</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL71</td>
<td>augmenting</td>
<td>40 kDa</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL91</td>
<td>essential</td>
<td>12 kDa</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL77</td>
<td>essential</td>
<td>71 kDa</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL79</td>
<td>essential</td>
<td>34 kDa</td>
<td>√</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
</tr>
<tr>
<td>US26</td>
<td>augmenting</td>
<td>66 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL27</td>
<td>augmenting</td>
<td>70 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL72</td>
<td>augmenting</td>
<td>40 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL76</td>
<td>augmenting</td>
<td>36 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL124</td>
<td>augmenting</td>
<td>17 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL35</td>
<td>augmenting</td>
<td>71 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL32</td>
<td>essential</td>
<td>115 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL49</td>
<td>essential</td>
<td>63 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL57</td>
<td>essential</td>
<td>136 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td></td>
<td>√</td>
</tr>
<tr>
<td>UL92</td>
<td>essential</td>
<td>33 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL95</td>
<td>essential</td>
<td>59 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL34</td>
<td>essential</td>
<td>45 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL87</td>
<td>essential</td>
<td>104 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL48</td>
<td>augmenting</td>
<td>247 kDa</td>
<td>√</td>
<td>√</td>
<td></td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>UL96</td>
<td>essential</td>
<td>14 kDa</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL94</td>
<td>essential</td>
<td>38 kDa</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL93</td>
<td>essential</td>
<td>65 kDa</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL114</td>
<td>augmenting</td>
<td>28 kDa</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL47</td>
<td>augmenting</td>
<td>108 kDa</td>
<td>√</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>UL30</td>
<td>augmenting</td>
<td>12 kDa</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1.1. Selected Uncharacterized HCMV Essential and Augmenting Genes
REFERENCES


33. **Mocarski, E. S., G. W. Kemble, J. M. Lyle, and R. F. Greaves.** 1996. A deletion mutant in the human cytomegalovirus gene encoding IE1(491aa) is replication


Chapter 2:

Characterization of HCMV novel essential protein pUL79 and determine its roles in viral late gene expression

This chapter contains data published in the following publication

ABSTRACT

In this chapter, we adopted a conditional protein genetic approach to characterize the role of the human cytomegalovirus (HCMV) gene UL79 during virus infection. We constructed ADddUL79, a recombinant HCMV in which the annotated UL79 open reading frame (ORF) was tagged with the destabilization domain of a highly unstable variant of the human FKBP12 protein (ddFKBP). The ddFKBP domain targets the tagged protein for rapid proteasomal degradation, but the synthetic ligand Shield-1 can stabilize ddFKBP, allowing accumulation of the tagged protein. ADddUL79 failed to replicate without Shield-1, but it grew at wild-type levels with Shield-1 or in human foreskin fibroblasts overexpressing hemagglutinin (HA)-tagged UL79 (HF-UL79HA cells), indicating an essential role of UL79 and the effectiveness of this approach. Without Shield-1, representative immediate-early and early viral proteins as well as viral DNA accumulated normally, but late transcripts and proteins were markedly reduced. UL79 was transcribed with early-late kinetics, which was also regulated via a positive-feedback loop. Using HF-UL79HA cells, we found that the UL79 protein localized to viral replication compartments during HCMV infection. Finally, we created a second UL79 mutant virus (ADinUL79 stop) in which the UL79 ORF was disrupted by a stop codon mutation and found that ADinUL79 stop phenocopied ADddUL79 under the destabilizing condition. Taking these results together, we conclude that UL79 acts after viral DNA replication to promote the accumulation of late viral transcripts. Importantly, the comparative analysis of ADddUL79 and ADinUL79 stop viruses provide additional proof for the power of the protein stability-based conditional approach to dissect the role of viral factors in HCMV biology.
INTRODUCTION

Human cytomegalovirus (HCMV) is the prototypical betaherpesvirus and a ubiquitous opportunistic pathogen that infects the majority of the world's population. Upon primary infection, HCMV establishes a lifelong persistent and latent/recurrent infection in a host(8). Even though HCMV infection is usually asymptomatic in healthy individuals, it is a significant source of severe disease in immunocompromised adults, such as AIDS patients and transplant recipients. Importantly, HCMV is the leading infectious cause of birth defects in newborns. Additionally, there is evidence that HCMV is a possible risk factor in the development of vascular disease, such as atherosclerosis, transplant vascular sclerosis, and coronary restenosis after angioplasty surgery(15, 24, 31, 35, 40, 42, 57). To acquire a comprehensive understanding of HCMV biology and facilitate the effort to develop effective therapeutics to combat disease caused by HCMV, it is imperative to dissect the roles of previously uncharacterized viral genes in both acute and latent infections of this virus.

During lytic infection, HCMV genes are expressed in a highly ordered temporal cascade. Viral transcripts accumulate in three different kinetic classes, namely, immediate early (IE), early, and late. The HCMV major IE (MIE) genes UL123 (IE1) and UL122 (IE2) play a critical role in predisposing the cellular environment to infection, and they act as transactivators to induce transcription of early genes. Many early genes encode proteins involved in viral DNA replication or predisposing the cellular environment to infection(17, 18, 29, 30). The transcript accumulation of early genes is independent of viral DNA replication; however, the continued accumulation of a subset of genes (i.e., early-late) is enhanced by the onset of viral DNA replication(41). Following viral DNA replication, late viral genes that mainly encode structural proteins start to transcribe, ultimately leading to the assembly and release of infectious particles. Although late gene transcription is tightly coupled to viral DNA replication, the underlying mechanism is poorly understood.

23
HCMV contains a 240-kb double-stranded DNA genome that contains at least 166 putative open reading frames (ORFs) and several microRNAs (7, 15, 17-19, 29, 30). Previously we and others have used genome-wide mutagenic approaches to classify the entire set of HCMV genes that encode annotated ORFs into three functional categories (9, 54), and we found that about 40 genes are essential for the HCMV laboratory strain AD169 to replicate in human fibroblasts. Many of the HCMV genes have not been experimentally characterized and lack homologues with known functions in other herpesviruses (33), and their functions therefore remain elusive. With the advent of the infectious bacterial artificial chromosome clone-based system for HCMV (BAC-HCMV) (4, 55), the functions of many HCMV genes have started to be elucidated. However, defining the role of essential genes remains challenging because of a paucity of a reliable system to propagate null mutant viruses. The conventional complementation approach, namely, the use of cells expressing a viral gene in trans to support the growth of the null mutant, has been reported for only a few essential HCMV genes (33). To overcome this technical hurdle, we and others have recently adopted a conditional approach (3) to facilitate the analysis of proteins critical to viral growth and pathogenesis (14, 37). This approach takes advantage of a mutant variant of the human FKBP12 protein that is highly unstable and rapidly degraded when expressed in mammalian cells (3). Using this method, a recombinant virus in which a viral ORF of interest is tagged with the destabilization domain of this FKBP variant (ddFKBP) that confers rapid instability can be constructed. The addition of the cell-permeable, synthetic ligand Shield-1 stabilizes the fusion protein, allowing propagation of the recombinant virus. Upon withdrawal of Shield-1, ddFKBP targets the fusion protein for rapid degradation, thus allowing the study of the virus infection in the absence of the protein of interest. ddFKBP tagging offers a robust genetic tool to study the role of viral genes, particularly the ones that are essential for HCMV infection.
In this chapter, we characterized the role of viral gene UL79 in HCMV replication. UL79 was previously classified as being essential for HCMV replication in fibroblasts by two large-scale mutagenic analyses (9, 54). To define its function during virus replication, we engineered a recombinant HCMV virus (AD\textit{dd}UL79) in which the annotated UL79 ORF was tagged with \textit{ddFKBP}. Using the \textit{ddFKBP} approach, we found that UL79 was critical for the accumulation of late viral transcripts but not viral DNA synthesis. This phenotype was further validated by our analysis of a second UL79 mutant virus in which the UL79 ORF was disrupted by a stop codon mutation. Furthermore, UL79 was expressed with early-late kinetics, was regulated by a positive-feedback loop, and primarily localized to viral nuclear replication compartments during infection.
MATERIALS AND METHODS

Plasmids, antibodies, and chemicals. pYD-C169 is a retroviral vector derived from pRetro-EBNA(23). It was created by inserting a PCR fragment containing the sequence of the UL79 ORF along with a C-terminal hemagglutinin (HA) tag into the multiple cloning site of pRetro-EBNA. pYD-C630 was derived from pGalK(47) and carried the 110-amino-acid destabilization domain of the human FKBP12 protein variant that contained the F36V and L106P destabilizing mutations (ddFKBP)(3). In this vector, ddFKBP was followed by a GalK/kanamycin dual expression cassette that was flanked by the Flp recognition target (FRT) sequence.

The primary antibodies used in this study included anti-β-actin (clone AC15; Abcam), anti-pUL44 (clone 10D8; Virusys), anti-IE2 (MAB8140; Chemicon), anti-HA (H6908; Sigma), anti-PCNA (clone F-2; Santa Cruz), and anti-FKBP12 (8/FKBP12; BD Biosciences). Other primary antibodies used in this study were anti-IE1(43), anti-pp28(39), anti-UL38(39), anti-UL69(25), anti-pp71(6), and anti-pp150 (all generous gifts from Thomas Shenk, Princeton University). Secondary antibodies used for immunoblotting were horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG and HRP-conjugated goat anti-rabbit IgG (Jackson Laboratory). The secondary antibodies used for immunofluorescence were Alexa Fluor 568-conjugated goat anti-rabbit IgG and Alexa Fluor 488-conjugated goat anti-mouse IgG (Invitrogen-Molecular Probes).

The synthetic chemical ligand Shield-1 (Shld1) that was used to regulate the stability of ddFKBP-tagged proteins was purchased from Cheminpharma (Farmington, CT).

Cells and viruses. Primary human newborn foreskin fibroblasts (HFFs) and embryonic lung fibroblasts (MRC5) were propagated in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal calf serum, nonessential amino acids, sodium pyruvate, and
penicillin-streptomycin. To create HFFs expressing the HA-tagged UL79 (HF-UL79HA cells) or empty vector (HF-vector cells), retrovirus stocks were prepared from Phoenix-Ampho cells(23) that were transfected with the retroviral vector pYD-C169 or pRetroEBNA, respectively. HFFs were transduced with retrovirus three times and then allowed to recover for 72 h to generate a pool of cells expressing the protein of interest.

Various BAC-HCMV clones were constructed to reconstitute recombinant HCMV viruses. Two BAC-HCMV clones, pAD/Cre and pAD-GFP, were used as the parental clones to produce wild-type viruses ADwt.2 and ADwt, respectively. pAD/Cre carries the full-length genome of HCMV strain AD169, and pAD-GFP is identical to pAD/Cre except that it contains a simian virus 40 (SV40) early promoter-driven green fluorescent protein (GFP) gene in place of the viral US4-US6 region(43, 55). All other recombinant BAC clones were constructed by using a linear recombination protocol in the bacterial strain SW105 that contained an arabinose-inducible Flp gene for the transient expression of Flp recombinase(47). To create a BAC clone in which the UL79 ORF was tagged with the \textit{ddFKBP} sequence at its N terminus, a cassette that contained the \textit{ddFKBP} sequence followed by the GalK/kanamycin dual marker was amplified by PCR from pYD-C630 with a pair of 70-bp primers. The 5’-terminal 50-bp sequences of these primers were homologous to the viral sequence immediately upstream or downstream of the N terminus of the UL79 ORF (5’-TCGTCACATCGTATTGTCGTCACCGTCGCTACCCGCTCACCGAGCGAACGATGGGAGTGCAGGTGGAAACCATC-3’ and 5’-TTGCCCGTGCGGACCCGCGGGACGGCGGGTTCTCTCTCGTCGGGGCCATGCTGGAGCCACCGGGAAGTTC-3’). The cassette was recombined into pAD-GFP at the N terminus of the UL79 ORF by linear recombination, and the resulting transformants were selected on kanamycin-containing LB plates to identify clones carrying the marker cassette. As the GalK/kanamycin marker was also flanked by FRT sites, it was then removed from the
BAC by adding arabinose to a fresh culture to induce Flp recombinase expression and plating for isolation of single colonies lacking kanamycin resistance. Therefore, the final clone (ADddUL79) contained the ddFKBP sequence along with a small FRT site fused in frame at the N terminus of UL79 (see Figure 2.1A). To create a BAC clone in which the UL79 ORF was prematurely terminated by a stop codon, a cassette that contained a stop codon followed by the FRT-bracketed GalK/kanamycin dual marker was amplified by PCR from pYD-C649 with a pair of 70-bp primers. The 5'-terminal 50-bp sequences of these primers were homologous to the viral sequence immediately upstream or downstream of the 5th amino acid of the UL79 ORF (5’-GTCGTCACCGTGCCTACCCGCTACCCGAGCGAACGATGATGGCCCGA

CTGATCACTATAGGGCGAATTGGGTGA-3’ and 5’-GTAAAGGAGAATTTGGCTGATCC-3’). The cassette was recombined into the UL79 ORF of pAD-GFP, and the GalK/kanamycin marker was then removed by Flp/FRT recombination as described above (see Figure 2.7A). The final clone (ADinUL79stop) contained a stop codon along with a small FRT site inserted in frame at the 6th amino acid codon of the UL79 ORF. All recombinant BACs were verified by restriction digestion, PCR, and direct sequencing analysis.

To reconstitute virus, 2 µg of the BAC-HCMV DNA and 1 µg of the pp71 expression plasmid were transfected into HFFs or HF-UL79HA cells by electroporation(55), and the culture medium was changed 24 h later. For reconstitution of ddFKBP-tagged virus, the synthetic chemical ligand Shield-1 was added every 48 h to maintain the concentration at 1 µM. The recombinant virus was harvested by collecting cell-free culture supernatant when the entire monolayer of infected cells was lysed. Virus was amplified by collecting cell-free culture supernatant from HFFs infected at a multiplicity of infection (MOI) of 0.01 in the presence of Shield-1. To remove any residual Shield-1, virus-containing culture supernatants
were then purified by ultracentrifugation through a 20% D-sorbitol cushion at an average relative centrifugal force of 53,000 \( \times \) g for 1 h, resuspended in DMEM with 10% fetal calf serum, and saved as viral stocks. HCMV titers were determined by 50% tissue culture infectious dose (TCID\(_{50}\)) assay. For \( ddFKBP \)-tagged virus, the titer was determined in duplicate in HFFs in the presence of 1 \( \mu \)M Shield-1 (unless indicated otherwise). For \( ADinUL79_{\text{stop}} \) virus, the titer was determined in duplicate in HF-UL79HA cells.

**Analysis of viral growth kinetics.** HFFs, HF-vector cells, or HF-UL79HA cells were seeded in 12-well plates overnight to produce a subconfluent monolayer. Cells were then inoculated with recombinant HCMV viruses for 1 h at an MOI of 0.1 for multistep growth analysis or an MOI of 3 for single-step growth analysis. The inoculum was removed, the infected monolayers were rinsed with phosphate-buffered saline, and infected cells were cultured in medium in the presence or absence of Shield-1 (1 \( \mu \)M). To infect cells with virus in the presence of ligand Shield-1, the ligand was added every 48 h to maintain its concentration. At various times postinfection, cell-free virus was collected by harvesting the medium from infected cultures, and the titers of the virus were determined by TCID\(_{50}\) assay.

**DNA and RNA analysis.** Intracellular viral DNA was measured by quantitative PCR (qPCR) as previously described(38). HCMV-infected cells were collected at various times postinfection, resuspended in lysis buffer (200 mM NaCl, 20 mM Tris [pH 8.0], 50 mM EDTA, 0.2 mg/ml proteinase K, 1% sodium dodecyl sulfate [SDS]), and incubated at 55°C overnight. DNA was extracted with phenol-chloroform, treated with RNase A (100 \( \mu \)g/ml), extracted again with phenol-chloroform, precipitated with ethanol, and resuspended in nuclease-free water (Ambion). Viral DNA was quantified by qPCR as previously described using a TaqMan probe (Applied Biosystems) and primers specific for the HCMV UL54
gene(38). Cellular DNA was quantified with SYBR PCR master mix (Clontech) and a primer pair specific for the human β-actin gene as previously described(38). The accumulation of viral DNA was normalized by dividing the number of UL54 gene equivalents by the number of β-actin gene equivalents. The accumulation of wild-type viral DNA at 4 h postinfection (hpi) was arbitrarily set at 1.

Transcript accumulation was analyzed by reverse transcription coupled to qPCR (RT-qPCR) as previously described(53). Total RNA was extracted using the Trizol reagent (Invitrogen) and treated with the TURBO DNA-free reagent (Ambion) to remove genomic DNA contaminants. cDNA was reverse transcribed with random hexamer primers using the high-capacity cDNA reverse transcription kit (Applied Biosystems). For IE1, cDNA was quantified by qPCR using Maxima probe/Rox 2× qPCR master mix (Fermentas), and a TaqMan probe (Applied Biosystems) and a primer pair specific for IE1. Alternatively, cDNA was quantified using SYBR Advantage qPCR premix (Clontech) and primer pairs specific for viral genes UL32, UL79, UL82, UL93-99, the ddFKBP sequence of ddFKBP-UL79, or the GAPDH (glyceraldehyde-3-phosphate dehydrogenase) cellular gene(10, 32, 48). All of the primer pairs and TaqMan probes are listed in Table 2.1. cDNA from infected cells was used to generate a standard curve for each gene examined. The standard curve was then used to calculate the relative amount of a specific transcript present in a sample. The amounts of IE1, UL32, UL79, UL82, and UL93-99 were normalized using GAPDH as an internal control.

Protein analysis. Protein accumulation and subcellular localization were determined by immunoblotting and immunofluorescence, respectively. For immunoblotting, cells were washed with phosphate-buffered saline and lysed in the sodium dodecyl sulfate (SDS)-containing sample buffer. Virion proteins were prepared by purifying cell-free virions from
culture medium of infected cells by ultracentrifugation through a sorbitol cushion and resuspending the cells in SDS-containing sample buffer. Proteins were resolved by electrophoresis on an SDS-containing polyacrylamide gel, transferred to a polyvinylidene difluoride (PVDF) membrane, hybridized with primary antibodies, reacted with the HRP-conjugated secondary antibody, and visualized by SuperSignal West Pico chemiluminescent substrate (Thermo Scientific)(53).

For immunofluorescence, cells grown on glass coverslips were fixed in 2% paraformaldehyde, permeabilized with 0.1% Triton X-100 for 15 min, incubated with a primary antibody, and subsequently labeled with an Alexa Fluor 488- or 568-conjugated secondary antibody (Invitrogen-Molecular Probes). Labeled cells were also counterstained with TO-PRO-3 (Invitrogen-Molecular Probes) to visualize the nuclei. Finally, labeled cells were mounted on slides with Prolong Gold antifade reagent (Invitrogen-Molecular Probes). Images were captured using Zeiss LSM Image software with a Zeiss LSM 510 META confocal laser scanning microscope.
RESULTS

Construction of a BAC-HCMV clone carrying the \textit{ddFKBP}-tagged UL79 and propagation of recombinant virus. The HCMV UL79 ORF has been reported to be essential for HCMV replication in fibroblasts\cite{9, 54}. To study the function of UL79 in virus infection, we adopted a protein-based viral genetic approach\cite{3, 14, 37}. We constructed a recombinant BAC-HCMV clone, pAD\textit{dd}UL79, in which the annotated UL79 coding sequence was tagged with a 110-amino-acid destabilization domain derived from a mutant variant of the human FKBP protein (\textit{ddFKBP}) at its N terminus (Figure 2.1A). This recombinant clone was derived from the parental BAC clone (pAD-GFP) which was also used to reconstitute wild-type virus (AD\textit{wt}) in this study\cite{55}. In cells infected with recombinant virus (AD\textit{dd}UL79) that was derived from pAD\textit{dd}UL79, the \textit{ddFKBP}-tagged UL79 protein was anticipated to be expressed but directed for rapid degradation by the \textit{ddFKBP} tag. However, the \textit{ddFKBP} tag should be stabilized in the presence of the synthetic chemical ligand Shield-1, and therefore, the tagged protein should retain its activity. This \textit{ddFKBP} tagging approach offers some unique advantages for creating mutant viruses. It abrogates protein products of essential genes during virus infection without the need for creating null mutant virus and complementing cell lines. As only one virus stock is needed for the comparative study of HCMV infection in the presence or absence of a viral protein product, it eliminates the need for creating marker rescued virus and avoids the potential complication resulting from an altered particle-to-PFU ratio in mutant virus. Finally, it provides evidence that it is the protein encoded by the tagged ORF and not other potential products derived from the UL79 locus that is important for viral replication.

Transfection of pAD\textit{dd}UL79 into human foreskin fibroblasts (HFFs) did not result in any detectable productive viral infection when the cells were surveyed for 30 days. However, when Shield-1 was added at a concentration of 1 µM, a cytopathic effect (CPE) was
observed, similar to that in transfection with wild-type pAD-GFP. Recombinant virus ADddUL79 produced from pADddUL79 transfection was then amplified by infecting fresh HFFs in the presence of Shield-1, purified by ultracentrifugation, and resuspended in fresh medium to generate a virus stock. To confirm that viral growth was dependent on the presence of Shield-1, HFFs were infected with ADddUL79 at an MOI of 3 with various concentrations of Shield-1. Infectious culture supernatants were collected at 96 h postinfection (hpi), and the titer of cell-free virus was determined by TCID_{50} assay (Figure 2.1B). The rescue of ADddUL79 replication by Shield-1 was dose dependent. While ADddUL79 failed to produce detectable progeny without Shield-1, its replication and CPE were indistinguishable from those of the wild-type virus with 1 µM Shield-1 (Figure 2.1B and C). Our results suggest that Shield-1 efficiently regulates the replication of ADddUL79 recombinant virus.

To provide additional proof that Shield-1 regulated ADddUL79 replication by controlling the stability of the ddFKBP-tagged UL79 protein, we wanted to examine the expression of the tagged protein in fibroblasts infected with ADddUL79. Surprisingly, we could not detect the tagged protein during infection regardless of Shield-1 treatment using an anti-FKBP12 antibody. This was not due to the inability of the antibody to recognize the ddFKBP domain, because we have used the same antibody to detect other ddFKBP-tagged viral proteins (Figure 2.10). Moreover, we were able to readily detect the ddUL79 protein from cells transfected with a plasmid overexpressing ddUL79 in the presence of Shield-1 using the anti-FKBP12 antibody (data not shown). Therefore, our inability to detect the tagged protein is likely the result of the low-level accumulation of the UL79 protein during HCMV infection. As an alternative approach to confirm that the UL79 ORF encoded a protein essential for HCMV growth, we tested the growth of ADddUL79 in UL79-overexpressing HFFs (HF-UL79HA cells) that were made by transduction of retrovirus carrying the C-terminally HA-
tagged UL79 ORF. HF-UL79HA cells expressed the tagged protein at the predicted size (35 kDa) (Figure 2.1C). In normal HFFs, ADddUL79 grew like wild-type virus did in the presence of Shield-1 but could not replicate without it. In contrast, ADddUL79 virus could replicate in HF-UL79HA cells in the absence of Shield-1, producing virus titers similar to those of ADwt. Therefore, the growth defect of ADddUL79 in the absence of Shield-1 is the direct result of the abrogation of UL79.

**UL79 is essential for HCMV replication.** To characterize the role of UL79 in HCMV replication, we first analyzed the growth kinetics of the ADddUL79 virus in HFFs. For progeny virus collected at each time point, we determined the titer of virus in the presence and absence of Shield-1 to determine whether there were additional escape mutations acquired during the course of infection. Under the multistep growth condition, ADddUL79 replicated indistinguishably from ADwt in the presence of Shield-1. However, without Shield-1, the production of ADddUL79 was almost undetectable by 12 days postinfection (dpi), whereas ADwt growth peaked and the titers reached $1 \times 10^7$ TCID$_{50}$ units/ml (Figure 2.1A). It was noted that at 15 dpi in the absence of Shield-1, ADddUL79 started to produce very low levels of progeny that could be determined even in the absence of Shield-1. We interpreted these results to mean that ADddUL79 might acquire spontaneous escape mutations after a prolonged period under nonpermissive conditions.

Next, we examined the single-step growth kinetics of ADddUL79 to determine whether the defect was dependent on the MOI. At an MOI of 3 and in the absence of Shield-1, ADddUL79 produced almost no virus by 4 dpi and very low levels of progeny ($<1 \times 10^3$ TCID$_{50}$ units/ml) by 6 dpi, while ADwt reached peak titers of $1 \times 10^7$ TCID$_{50}$ units/ml (Figure 2.1A). In contrast, Shield-1 treatment restored the growth of ADddUL79 to wild-type levels. The very low levels of ADddUL79 produced after 5 dpi in the absence of Shield-
1 might be due to the slight leakiness of Shield-1 regulation of ddFKBP degradation. It was unlikely that this low level of virus growth represented escape mutants because they could not establish a second round of infection without Shield-1, their titers did not increase over time, and their titers could not be measured without Shield-1.

Taken together, these data indicate that UL79 is essential for HCMV replication in cultured fibroblasts and that ddFKBP-mediated Shield-1 regulation allows for tight control of UL79 function during the virus infection cycle.

**UL79 is required for viral replication at late stages of infection.** To determine when UL79 is required during the viral replication cycle, we stabilized UL79 protein accumulation by Shield-1 at various time points and for different lengths of time in infected cells and then determined the impact of such temporal accumulation of UL79 proteins on virus production. To determine whether UL79 was required at early times, Shield-1 was added at the onset of virus infection but was removed from the medium at 24 or 48 hpi. To determine whether UL79 was required at late times during infection, Shield-1 was added at 24 or 48 hpi. In both experiments, the titers of virus in the supernatants of infected cells were determined at 120 hpi (Figure 2.2B). When we added Shield-1 as late as 48 hpi, ADdd/UL79 replication was indistinguishable from virus grown in the continuous presence of Shield-1. In contrast, when Shield-1 was present for only the first 48 h during infection, almost no viral progeny were produced, similar to ADdd/UL79 infection in the complete absence of Shield-1. Taken together, these data indicate that UL79 is required for viral replication during the late stages of viral infection.

**UL79 is required for late viral gene expression.** To determine which step of the viral replication cycle was compromised in the absence of UL79, we first examined the viral
protein accumulation profile in the absence of Shield-1. The accumulation of viral immediate-early proteins (IE1-72 and IE2-86) or early/early-late proteins (UL38, UL44, UL69, and pp65) in cells infected with AD<em>dd</em>UL79 was not affected (Figure 2.3 and data not shown). In contrast, the true late proteins, pp28, pp71, and pp150, accumulated at markedly reduced levels.

To determine whether the reduced late protein accumulation in AD<em>dd</em>UL79 infection was the result of reduced viral late transcript accumulation, we analyzed the IE1 transcript (control) and late UL32 (pp150), UL82 (pp71), and coterminated UL93-99 transcripts by RT-qPCR analysis (Figure 2.4)(10, 32, 48). All transcripts detected were specific and were not the result of genomic DNA contamination, as mock-infected cells and qPCR reactions done in the absence of reverse transcriptase failed to produce any products (data not shown). At 24 hpi, late transcripts were barely detectable, whereas the IE1 transcript accumulated to a high level (data not shown). Importantly, stabilization of UL79 by Shield-1 had no substantial effects on viral transcript accumulation at this time point (Figure 2.4), consistent with the observation that UL79 was not required at early times of infection (Figure 2.2B). At 72 hpi, even though degradation of UL79 had only a small effect on IE1 transcript accumulation, as the viral DNA synthesis inhibitor phosphonoacetic acid (PAA) did(10), it reduced the accumulation of late transcripts almost as markedly as PAA did. The levels of late UL32, UL82, and UL93-99 transcripts when UL79 was absent were 7-, 27-, and 13-fold lower than the levels when UL79 was present, respectively (Figure 2.4). Together, our results indicate that UL79 has minimal effect on viral immediate-early and early gene expression but is critical for the accumulation of late viral transcripts and proteins.

**UL79 is dispensable for viral DNA replication.** As viral late gene expression is dependent on viral DNA replication, it was possible that the defect in late gene expression in the
absence of UL79 was due to a defect in viral DNA replication. Thus, we examined viral DNA synthesis in the absence of Shield-1 (Figure 2.5A). Surprisingly, the kinetics of viral DNA replication during ADddUL79 infection in cells in the absence of Shield-1 was indistinguishable from that of infection of cells in the presence of Shield-1. To confirm that Shield-1 withdrawal was effective, supernatants from infected culture were collected at 72 hpi and measured for virus production (Figure 2.5B). Infected cells in the absence of Shield-1 produced no detectable virus, indicating that UL79 function was indeed abrogated in these cells. Therefore, we conclude that UL79 acts after viral DNA replication and is specifically required for the accumulation of late viral transcripts.

The UL79 transcript is expressed with early-late kinetics and is regulated by a positive-feedback loop. To determine the kinetic class of UL79 expression, we analyzed the UL79 transcript accumulation in fibroblasts infected with ADddUL79 by RT-qPCR (Figure 2.6A). When the UL79 protein was stabilized, the UL79 transcript accumulated abundantly at late times of infection, and its level at 72 hpi was 5-fold higher than the level at 24 hpi. The UL79 transcript level was markedly reduced at 72 hpi when infected cells were treated with PAA, indicating that its transcript accumulation was augmented by viral DNA replication. However, in the presence of PAA, the accumulation of UL79 transcript was not completely inhibited, as the UL79 transcript levels increased almost 3-fold from 24 to 72 hpi. Therefore, as the accumulation of UL79 transcript is markedly enhanced by, but not completely dependent on viral DNA synthesis, our data suggest that UL79 is an early-late HCMV gene. However, in the present study, we cannot rule out the possibility that more than one transcript emanates from the UL79 locus and that our result reflects the combined expression profile of these transcripts rather than individual species.
Surprisingly, when UL79 proteins were destabilized, the level of UL79 transcript was drastically reduced in infected cells at 72 hpi (Figure 2.6A). This suggests a positive-feedback loop for UL79 expression which in turn amplifies the expression of late viral genes. To provide additional evidence for this autoregulation of UL79 transcription by its protein products, we tested whether we could enhance endogenous UL79 transcription during infection of UL79 recombinant virus by providing the UL79 protein in trans. We infected HFFs expressing the HA-tagged UL79 (HF-UL79HA cells) or empty vector (HF-vector cells) with ADddUL79 in the absence of Shield-1. This experimental system allowed us to uncouple the UL79 protein accumulation (provided by HF-UL79HA cells) and endogenous UL79 transcript accumulation (from the viral genome). The endogenous UL79 transcription was analyzed by RT-qPCR using a primer pair specific to the ddFKBP sequence. We found that the accumulation of ddFKBP-UL79 transcripts in infected HF-UL79HA cells markedly increased at 72 hpi relative to that in infected HF-vector cells. This indicates that, as anticipated, the UL79 protein enhances the transcription from the UL79 gene locus (Figure 2.6B). Together, these results suggest that efficient transcription of UL79 is positively regulated by both viral DNA replication and its own protein products.

A UL79 stop codon mutant virus has the same defect as ADddUL79. To provide additional proof that the defect of ADddUL79 was the direct result of the loss of UL79 function, we created a second UL79 mutant virus, ADinUL79<sub>stop</sub>. In this recombinant virus, we inserted a stop codon at residue Glu6 of the UL79 ORF to abrogate the expression of UL79 protein products (Figure 2.7A). The ADinUL79<sub>stop</sub> virus could be reconstituted only from transfection of the BAC clone in HF-UL79HA cells, not by transfection in normal HFFs or HF-vector cells. The titer of reconstituted mutant virus was comparable to that of wild-type virus (data not shown), and it replicated efficiently in HF-UL79HA cells but completely
failed to grow in HF-vector cells (Figure 2.7B). These results indicated that loss of UL79 function was solely responsible for the growth defect of ADinUL79stop. Importantly, cells infected with ADinUL79stop synthesized viral DNA at wild-type levels (Figure 2.7C) and accumulated the products of IE genes (IE1-72 and IE2-86) and early genes (pUL38 and pUL44) efficiently but had a marked defect in accumulation of viral late proteins (pp71 and pp28) during infection (Figure 2.7D). Therefore, the defect of ADinUL79stop phenocopied that of ADddUL79 under Shield-1 withdrawal, and our results provide further validation for the power of the protein stability-based conditional approach to dissect the role of viral factors in HCMV biology.

**The HA-tagged UL79 protein is primarily nuclear and localizes in replication compartments during virus infection.** As UL79 is required for the accumulation of late viral transcripts, it is possible that it is a nuclear protein involved in the regulation of viral late gene transcription. We first examined the intracellular distribution of UL79 expressed alone to determine its ability to localize into the nucleus in the absence of any other viral factors. We used HF-UL79HA cells because there was no antibody specific to UL79 available and because the tagged UL79 was fully functional to complement UL79-deficient virus (Figure 2.1C). While no appreciable HA staining was present in normal HFFs, the staining was readily observed in HF-UL79HA cells, predominantly localizing within the nuclei, suggesting that UL79 proteins are primarily nuclear (Figure 2.8A). Next we determined whether the UL79 proteins localized to any virus-induced intracellular structures during infection, particularly replication compartments. These organized intranuclear viral structures are sites where sets of viral proteins (e.g., pUL44, pUL57, pUL117, and IE2)(1, 36, 38) and cellular proteins (e.g., p53, Nbs1, and Rad50)(12, 26) are recruited to, and viral activities, including viral DNA synthesis, late gene transcription, and DNA packaging, take
place. To test our hypothesis, we examined the localization of the HA-tagged UL79 proteins expressed from HF-UL79HA cells that were infected with HCMV, as we were not able to detect tagged UL79 proteins expressed from the endogenous locus of the viral genome during infection (data not shown). The tagged UL79 proteins formed large intranuclear domains which colocalized with pUL44, the virus-encoded DNA polymerase accessory protein that has been used as a marker for replication compartments (36, 38) (Figure 2.8A), indicating that a major fraction of HA-tagged UL79 proteins is localized in replication compartments. Importantly, the formation of UL79-containing large intracellular domains appeared to occur at late times during infection (Figure 2.8B), providing additional evidence that UL79 functions at late times to regulate gene expression.

We noted that the anti-HA antibody also appeared to generate low-level cytoplasmic staining in some normal HFFs that were infected with HCMV, even though this staining intensified in infected cells that expressed the tagged UL79 protein (Figure 2.8A). This cytoplasmic staining colocalized with pp28-containing viral assembly centers (data not shown). Nonetheless, we decided not to pursue this further, as we could not rule out the possibility that the cytoplasmic HA staining represents the nonspecific recognition of assembly centers by anti-HA antibody.

The UL79 protein is not detected in HCMV virions. A proteomics study has previously shown that a single UL79-derived peptide can be detected in the HCMV virions only by the sensitive Fourier transform ion cyclotron resonance (FTICR) mass spectrometry, not by the more traditional liquid chromatography-tandem mass spectrometry (LC-MS-MS) analysis, suggesting that this peptide may represent a nonspecific contaminant (46). To determine whether the UL79 protein is a tegument protein, we infected HF-UL79HA cells or HF-vector cells with AD*UL79 in the presence or absence of Shield-1. We then harvested
supernatants from infected cells at 96 hpi and partially purified cell-free virions by ultracentrifugation through a sorbitol cushion. These virion samples appeared to be largely free of contamination of cellular debris, as they did not contain detectable amounts of the viral IE1 protein or the cellular PCNA protein (Figure 2.9). On the other hand, the major capsid protein (MCP) was readily detected, and importantly, pUL69, a protein that is present in the tegument and dense bodies in very small amounts (46, 50), could also be detected in the virion samples (Figure 2.9). Under these conditions, however, we could not detect the HA-tagged UL79 protein in the virion samples, even though it was readily detectable in total lysate of infected HF-UL79HA cells. Thus, the UL79 protein is unlikely to be a major component of the HCMV virion. This was consistent with the result that UL79 appeared to be required only at late stages (i.e., after 48 hpi) (Figure 2.2B) and suggested the important role of the newly synthesized UL79 protein in promoting the HCMV lytic infection cycle. Collectively, our data indicate that UL79 plays a key role at late times after viral DNA replication to regulate the accumulation of late viral transcripts.
DISCUSSION

In this study, we applied the *ddFKBP*-mediated, Shield-1-regulated protein genetic approach\(^3\) to investigate the role of viral gene UL79 during HCMV replication. Moreover, we also compared the effectiveness of this approach to the more traditional viral genetic approach by creating and analyzing the second UL79 mutant virus in which the UL79 ORF was disrupted by a stop codon mutation (Figure 7). This comparative analysis provides additional proof for the power of the protein stability-based conditional approach to dissect the role of viral factors in HCMV biology. Our work represents one of the few studies that take advantage of this powerful approach to define viral gene functions in HCMV biology\(^{14, 37}\). Shield-1 regulation of *ddFKBP*-tagged UL79 is robust, reversible, and precisely controlled and validates the role of UL79 as an essential HCMV gene (Figures 2.1 and 2.2). Interestingly, we could not detect *ddFKBP*-tagged UL79 protein in cells infected with AD*ddUL79* even with Shield-1 using an anti-FKBP12 antibody. Nonetheless, the full complementation of AD*ddUL79* in the absence of Shield-1 by HFFs overexpressing UL79 provided additional evidence that the defect of AD*ddUL79* was the direct result of UL79 depletion (Figure 2.1C). Our inability to detect the endogenous UL79 proteins during infection is likely due to either their extremely low levels of expression or their unstable nature, similar to other HCMV proteins such as pUL21a\(^{11}\). Notably, the accumulation of the HA-tagged UL79 protein overexpressed from transduced HFFs was also reduced during infection relative to that in mock-infected cells, suggesting that the turnover of UL79 proteins may be enhanced during infection (Figure 2.1C). A more sensitive antibody or assay will need to be developed in order to detect and characterize endogenous UL79 proteins during HCMV infection.

Our systematic dissection of AD*ddUL79* infection in the absence of Shield-1 allowed us to define the stage of the infection cycle where UL79 functioned. Shield-1-dependent rapid
degradation/stabilization of ddFKBP-tagged proteins(3) gave us a tool to define the requirement of UL79 in HCMV infection in a nearly real-time manner (Figure 2.2B). Treating infected cells with Shield-1 for the first 48 hpi could not rescue the growth of recombinant virus, whereas adding Shield-1 at late times, even at 48 hpi, fully restored its growth (Figure 2.2), suggesting that UL79 was required at late stages of HCMV infection. Consistent with the timing of its involvement in virus infection, UL79 was expressed with early-late kinetics (Figure 2.6) and localized into viral replication compartments where viral DNA replication and late gene transcription took place (Figure 2.8). These spatial and temporal characteristics of UL79 support its role in regulating the accumulation of viral late gene products at transcript levels (Figure 2.3, 2.4, and 2.7).

Depletion of UL79 during HCMV infection was able to uncouple viral DNA synthesis and viral late gene expression, indicating that the former is not sufficient to promote the latter (Figure 2.4 and 2.5). For many DNA viruses, the temporal regulation of viral gene expression is a common feature of their lytic infection. For simian virus 40, the large T antigen plays an essential role in the activation of viral promoters(5, 21), and viral DNA replication attenuates in trans the repressor of viral late promoters(49, 58). For adenovirus, both viral trans-acting factors(16, 34) and viral DNA replication(44, 45) are required for its late gene expression. Viral DNA replication facilitates late gene expression directly(44) or indirectly by promoting the expression of viral trans-acting factors(16) or recruitment of cellular transcription factors to late promoters(45).

For herpesviruses, viral late gene expression has been studied mostly with herpes simplex virus (HSV) and murine gammaherpesvirus 68 (MHV-68). In HSV, DNA replication is required in cis for activity of late promoters(20, 28). HSV proteins, including ICP4, ICP8, and ICP27, are necessary for efficient expression of late genes by interacting with the general transcription machinery(13, 22, 46), and they facilitate the assembly of transcription
preinitiation complexes(46, 56). In MHV-68, both viral DNA replication (12) and four viral proteins, ORF18, ORF24, ORF30, and ORF34, are required for late gene expression(2, 51, 52). HCMV UL79, UL87, and UL95 share sequence homology with ORF18, ORF24, and ORF34, respectively, while ORF30 is conserved only in gamma-herpesviruses(2, 51, 52). Both UL79 and ORF18 are dispensable for viral DNA replication but are required for the accumulation of viral late gene products (Figures 2.4 and 2.5)(2), and UL79 also localizes to viral replication compartments (Figure 2.8). However, UL79 shares only 28% amino acid identity to ORF18(2) and is transcribed with early-late kinetics (Figure 2.6), whereas ORF18 was identified as an early gene(27). This body of evidence suggests that a general regulatory principle of late viral gene expression is conserved across the herpesvirus family, but the precise mechanism may be unique to each herpesvirus.

How does UL79 regulate late viral gene expression? Three possible mechanisms are responsible for its activity. It is possible that UL79 may modulate the initiation of late viral gene transcription(2). It may act as a transcription activator, a function reminiscent of ORF18. However, as in silico analysis did not reveal any significant homology of UL79 to known transcription factors or conserved DNA binding motif, UL79 may also stimulate late gene transcription by activating other transcriptional factors or dissociating repressors from late gene promoters. Alternatively, UL79 may be involved in regulation of chromatin structures of replicating viral genomes, which may in turn facilitate late viral gene expression. Finally, it is also possible that UL79 may promote the accumulation of late transcripts by maintaining their stability. In any event, the temporal expression (i.e., early-late kinetics) of UL79 may be partially responsible for its involvement in the accumulation of only the late transcripts and not the early transcripts. The precise mechanism for how UL79 selectively promotes viral late transcripts will be a key question to address in future study.
How is the expression of UL79 regulated during HCMV infection? UL79 is expressed with early-late kinetics; its transcripts accumulate even when viral DNA synthesis is inhibited, but the accumulation is markedly enhanced upon viral DNA synthesis (Figure 2.6). Importantly, the efficient accumulation of UL79 transcripts also requires the accumulation of UL79 proteins themselves and is regulated by a positive-feedback loop (Figure 2.6). Therefore, low-level UL79 expression appears to occur independently of viral DNA replication. Viral DNA replication may increase the copy number of DNA templates for UL79 transcription. Perhaps more importantly, any stimulatory effect of viral DNA replication on UL79 expression is further amplified via this positive-feedback loop. Therefore, it is sensible to speculate that viral DNA replication acts in cis and UL79 proteins function in trans to promote their own expression.

In conclusion, we have identified UL79 as a key viral factor bridging viral DNA replication and late gene expression during HCMV infection. We are investigating potential interactions of UL79 with cellular proteins or other viral proteins to define its mechanism during HCMV infection.
**Figure 2.1** (A) Schematic diagram for creating recombinant virus ADddUL79 and applying a destabilization domain (ddFKBP)-mediated protein stability-based approach to regulate gene function. ADddUL79 carried the UL79 ORF that was tagged at its N terminus with ddFKBP (indicated by gray box) and expressed the ddFKBP-UL79 fusion protein. Without the small-molecule ligand Shield-1 (Shld1), ddFKBP should direct the entire fusion protein for rapid degradation, thus effectively depleting UL79 proteins during infection. However, Shield-1 should bind to ddFKBP, protect the fusion protein from degradation, and maintain the accumulation of the UL79 proteins, thus allowing productive virus replication. BAC, bacterial artificial chromosome. (B) Dose-dependent regulation of ADddUL79 replication by Shield-1. Human foreskin fibroblasts (HFFs) were infected with ADddUL79 at a multiplicity of infection (MOI) of 3 in the presence of various concentrations of Shield-1 (as indicated). At 96 h postinfection (hpi), supernatants of infected cells were collected and analyzed for the...
presence of virus by TCID\textsubscript{50} assay in HFFs maintained in 1 \( \mu \text{M} \) Shield-1. (C) HFFs overexpressing the hemagglutinin (HA)-tagged UL79 complemented the growth of AD\textit{dd}UL79 in the absence of Shield-1. Normal HFFs (HF) or HFFs overexpressing the HA-tagged UL79 (HF-UL79HA) were infected with the AD\textit{wt} (wild-type virus control) or AD\textit{dd}UL79 virus at an MOI of 3 with Shield-1 or without Shield-1. At 120 hpi, the titers of cell-free virus were determined as described above for panel B (left panel), and total cell lysates were analyzed for the expression of the tagged UL79 by immunoblotting using the anti-HA antibody (\( \alpha \)-HA) (right panel). The antibody to actin was used as a loading control. The detection limit of the TCID\textsubscript{50} assay is indicated by the dashed line.
Figure 2.2. **UL79 is essential for HCMV replication at late stages of infection.** (A) Growth kinetic analysis of ADddUL79 in the presence or absence of Shield-1. HFFs were infected with ADwt or ADddUL79 at an MOI of 3 for single-step growth analysis or at an MOI of 0.1 for multistep growth analysis. Infected cells were cultivated in the presence (+) or absence (−) of 1 μM Shield-1, and supernatants were collected at the indicated times and analyzed for the presence of virus by TCID_{50} assay in HFFs maintained in 1 μM Shield-1 (−/+) or (−/−). To test the potential occurrence of spontaneous escape mutants after prolonged growth of recombinant virus, supernatants from cells infected with ADddUL79 in the absence of Shield-1 were also analyzed by TCID_{50} assay in the absence of Shield-1 (−/−). (B) UL79 is required at late stages of HCMV infection. HFFs were infected with ADddUL79 at an MOI of 0.1, Shield-1 (Shld1) was added at the indicated times postinfection, and infected cells were cultivated with Shield-1 for various time periods (indicated by the white boxes in
the schematic drawing in the left panel). Supernatants were collected at 120 hpi and analyzed for the presence of virus by the TCID$_{50}$ assay in HFFs maintained in 1 µM Shield-1 (right panel). The number of hours of Shield-1 treatment are shown in parentheses to the right of the graph. The detection limits of the TCID$_{50}$ assay are indicated by dashed lines in the graphs.
Figure 2.3. UL79 is required for the efficient accumulation of late viral proteins during HCMV infection. HFFs were infected with AD<sup>dd</sup>UL79 at an MOI of 3 in the presence (+) or absence (−) of 1 μM Shield-1. Cells were harvested at different times post-infection. The accumulation of immediate-early proteins (IE1-72 and IE2-86), early proteins (pUL38 and pUL44), and late proteins (pp71, pp150, and pp28) were determined by immunoblot analysis.
Figure 2.4. **UL79 is required for the efficient accumulation of late viral transcripts during HCMV infection.** HFFs were infected with ADddUL79 at an MOI of 3 in the presence (+) or absence (-) of Shield-1 (Shld1) (1 µM) and with or without the viral DNA synthesis inhibitor phosphonoacetic acid (PAA) (100 µg/ml). Total RNA was isolated at 24 and 72 hpi, and the amounts of selected viral transcripts were then measured by reverse transcription coupled to quantitative PCR (RT-qPCR) and normalized to that of GAPDH. The normalized amount of viral transcript at 24 hpi in the presence of Shield-1 but without PAA was set at 1.
Figure 2.5. **Viral DNA replication is not altered in the absence of UL79 during infection.** HFFs were infected with AD<sup>dd</sup>UL79 in the presence or absence of 1 µM Shield-1. (A) Total DNA was isolated at different times postinfection, and the accumulation of viral DNA was determined by qPCR using primers specific for the HCMV UL54 gene. (B) Supernatants of infected cells with or without Shield-1 treatment were also collected at 72 hpi and analyzed for the presence of virus by the TCID<sub>50</sub> assay in HFFs maintained in 1 µM Shield-1. The detection limit of the TCID<sub>50</sub> assay is indicated by the dashed line.
Figure 2.6. **UL79 is transcribed with early-late kinetics and is regulated by a positive-feedback loop.** (A) HFFs were infected with AD<sup>dd</sup>UL79 at an MOI of 3 in the presence or absence of Shield-1 (1 μM) and with or without the viral DNA synthesis inhibitor PAA (100 μg/ml). Total RNA was isolated at 24 and 72 hpi, and the amount of UL79 transcript was measured by RT-qPCR analysis and normalized to that of GAPDH. The normalized amount of UL79 transcript at 24 hpi in the presence of Shield-1 but without PAA was set at 1. (B) HFFs expressing the HA-tagged UL79 (HF-UL79HA) or empty vector (HF-vector) were infected with AD<sup>dd</sup>UL79 at an MOI of 3 in the absence of Shield-1 and with or without PAA (100 μg/ml). Total RNA was isolated at 24 and 72 hpi, and the amount of <i>dd</i>FKBP-tagged UL79 transcript expressed from the viral genome was measured by RT-qPCR analysis using the primers specific to the <i>dd</i>FKBP sequence and normalized to that of GAPDH. The normalized amount of <i>dd</i>FKBP-tagged UL79 transcript at 24 hpi without PAA in infected HF-UL79HA cells was set at 1.
Figure 2.7. A mutant virus in which the UL79 ORF has been abrogated by a stop codon mutation has the same defect as ADddUL79. (A) Schematic diagram for creating recombinant virus ADinUL79stop. A cassette that contained a stop codon followed by the FRT-bracketed GalK/kanamycin dual marker was amplified and recombined into the UL79 ORF of the wild-type HCMV BAC clone. The GalK/kanamycin marker was then removed by Flp/FRT recombination. The final clone, ADinUL79stop, contained a stop codon along with a small FRT site inserted in frame at the 6th amino acid (E6X) of the UL79 ORF. (B) Single-step growth analysis of ADinUL79stop. Complementing HF-UL79HA cells or control HF-vector cells were infected with ADwt or ADinUL79stop at an MOI of 3. Supernatants were collected at the indicated days postinfection and analyzed for the presence of virus by TCID50 assay in HF-UL79HA cells. The detection limit of the TCID50 assay is indicated by...
the dashed line. (C) Viral DNA replication of ADinUL79stop. HF-UL79HA cells or HF-vector cells were infected with ADinUL79stop at an MOI of 3. Total DNA was isolated at different times postinfection, and the accumulation of viral DNA was determined by qPCR using primers specific for the HCMV UL54 gene. (D) Viral gene expression of ADinUL79stop. HF-UL79HA cells or HF-vector cells were infected with ADinUL79stop at an MOI of 3. Cells were harvested at different times postinfection. The accumulation of immediate-early proteins (IE1-72 and IE2-86), early proteins (pUL38 and pUL44), and late proteins (pp71 and pp28) was determined by immunoblot analysis.
Figure 2.8. UL79 proteins are primarily nuclear proteins that localize within viral replication compartments during infection. (A) Normal HFFs or HF-UL79HA cells were either mock infected or infected with the GFP-less wild-type HCMV virus AD<sub>wt</sub> 2 at an MOI of 0.5. At 72 hpi, the cells were harvested and examined by confocal immunofluorescence analysis for the localization of UL79 using anti-HA antibody and viral replication compartments using anti-pUL44 antibody. The cells were also counterstained with TO-PRO3 to visualize the nuclei. Mock-infected HFFs or HFFs infected with AD<sub>wt</sub>.2 were used as a negative control for the detection of tagged UL79 proteins by anti-HA antibody. (B) HF-UL79HA cells were either mock infected or infected with AD<sub>wt</sub>.2 at an MOI of 0.5. At different times postinfection, cells were examined by confocal fluorescence analysis for viral replication compartments using anti-pUL44 antibody and UL79 localization using anti-HA antibody. The cells were also counterstained with TO-PRO3 to visualize the nuclei. The white arrows indicate viral replication compartments, and the yellow arrows indicate the localization of a portion of HA staining to viral assembly center-like cytoplasmic structures. Bars, 20 μm.
either mock infected or infected with the GFP-less wild-type HCMV virus AD\textsubscript{wt.2} at an MOI of 0.5. At 72 hpi, the cells were harvested and examined by confocal immunofluorescence analysis for the localization of UL79 using anti-HA antibody and viral replication compartments using anti-pUL44 antibody. The cells were also counterstained with TO-PRO3 to visualize the nuclei. Mock-infected HFFs or HFFs infected with AD\textsubscript{wt.2} were used as a negative control for the detection of tagged UL79 proteins by anti-HA antibody. (B) HF-UL79HA cells were either mock infected or infected with AD\textsubscript{wt.2} at an MOI of 0.5. At different times postinfection, cells were examined by confocal fluorescence analysis for viral replication compartments using anti-pUL44 antibody and UL79 localization using anti-HA antibody. The cells were also counterstained with TO-PRO3 to visualize the nuclei. The white arrows indicate viral replication compartments, and the yellow arrows indicate the localization of a portion of HA staining to viral assembly center-like cytoplasmic structures. Bars, 20 µm.
in the experiment, and the protein was analyzed using anti-pUL69, and cellular PCNA.

Figure 2.9. The UL79 protein is not detected in HCMV virions. A total of 1X10^7 HF-UL79HA or HF-vector cells were infected with ADddUL79 in the presence or absence of Shield-1 at an MOI of 5. Total cell lysates and cell-free virions were collected at 96 hpi. Cell-free virions were then partially purified by ultracentrifugation through a sorbitol cushion. Lysates of 1X10^5 cells or 20% of total virion samples harvested were loaded onto each lane of an SDS-polyacrylamide gel and analyzed by immunoblotting for MCP, IE1-72, pUL69, and cellular PCNA. For the UL79 protein, twice as many virion samples were used in the experiment, and the protein was analyzed using anti-HA antibody.
Figure 2.10. Immunoblot Analysis of viral ddFKBP Fusion Proteins. Human fibroblast cells were infected with ADddUL71(A) and ADddUL77(B) recombinant viruses at MOI 3.0 in the presence and absence of 1µM Shield-1. The cells were harvested at 48 and 96 hours post infection. The accumulations of various ddFKBP-fusion proteins were analyzed by immunoblot analysis using anti-FKBP12 antibody.
TABLES

| Transcript | qPCR reaction | Primer sequence | 6FAM-TAMRA TaqMan probe sequence
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>IE1</td>
<td>TaqMan</td>
<td>5′-CAAGTGACCAGGATGCAA-3′</td>
<td>5′-TCCTGGGAGAACTGCAAAAC-3′</td>
</tr>
<tr>
<td>UL32</td>
<td>SYBR green</td>
<td>5′-GGTTTCTGGGCTGAGATGCG-3′</td>
<td>5′-CACCAAAACGCGTCCGGATTAC-3′</td>
</tr>
<tr>
<td>UL79</td>
<td>SYBR green</td>
<td>5′-CCGACACGGGAAAATTCTCTC-3′</td>
<td>5′-TGATCGACACACCCCCAGGTGT-3′</td>
</tr>
<tr>
<td>UL82</td>
<td>SYBR green</td>
<td>5′-TGCTGTATGCTGCGGATGAC-3′</td>
<td>5′-CGGACTGATCTGTGACCG-3′</td>
</tr>
<tr>
<td>UL93-99</td>
<td>SYBR green</td>
<td>5′-GTGCCCATCCCGAGCCTCG-3′</td>
<td>5′-TCCACAGTCCACCCACC-3′</td>
</tr>
<tr>
<td>ddFKBP-UL79</td>
<td>SYBR green</td>
<td>5′-TGAAAACCATCTCCCCAGAGA-3′</td>
<td>5′-TCACCTCTGCTGTAGCTAGCAT-3′</td>
</tr>
<tr>
<td>GAPDH</td>
<td>SYBR green</td>
<td>5′-CTGTGGCTGTAGCGAAATTCGT-3′</td>
<td>5′-ACCCACTCCCTCACCTTGAC-3′</td>
</tr>
</tbody>
</table>

*6FAM, 6-carboxyfluorescein; TAMRA, 6-carboxytetramethylrhodamine; NA, not applicable.

Table 2.1 Primers and probes used for RT-qPCR in Chapter 2.
REFERENCES


Chapter 3:

Identification of pUL79 protein partners and the mechanism of pUL79 activity

This chapter contains data submitted to *PLoS Pathogens*

ABSTRACT

In this chapter, we identified a unique mechanism in which human cytomegalovirus (HCMV) protein pUL79 acts as an elongation factor to direct cellular RNA polymerase II for viral transcription during late times of infection. We and others previous reported that pUL79 and its homologues were required for viral transcript accumulation after viral DNA synthesis (28, 52). We hypothesized that pUL79 represented a unique mechanism to regulate viral transcription at late times during HCMV infection. To test this hypothesis, we analyzed the proteome associated with pUL79 during virus infection by mass spectrometry. We identified both cellular transcriptional factors, including multiple RNA polymerase II (RNAP II) subunits, and novel viral transactivators, including pUL87, pUL92, and pUL95, as protein binding partners of pUL79. Co-immunoprecipitation (co-IP) followed by immunoblot analysis confirmed the pUL79-RNAP II interaction, and this interaction was independent of any other viral proteins. Using a recombinant HCMV virus where pUL79 protein is conditionally regulated by a protein destabilization domain dddFKBP, we showed that this interaction did not alter the total levels of RNAP II or its recruitment to viral late promoters. Furthermore, pUL79 did not alter the phosphorylation profiles of the RNAP II C-terminal domain, which was critical for transcriptional regulation. Rather, nuclear run-on assay indicated that, in the absence of pUL79, RNAP II failed to elongate and stalled on viral DNA. pUL79-dependent RNAP II elongation was required for transcription from all three kinetic classes of viral genes (i.e. immediate-early, early, and late) at late times during virus infection. In contrast, host gene transcription during HCMV infection was independent of pUL79. In summary, we have identified a novel viral mechanism by which pUL79, and potentially other viral factors, regulates the rate of RNAP II transcription machinery on viral transcription during late stages of HCMV infection.
INTRODUCTION

Human cytomegalovirus (HCMV) is the prototypical beta-herpesvirus and a ubiquitous pathogen in the human population. Upon primary infection, HCMV establishes a lifelong persistent and latent/recurrent infection in a host (15). Even though HCMV infection is usually asymptomatic, it acts as an opportunistic pathogen and is the major cause of morbidity and mortality in immunocompromised individuals, including transplant recipients and AIDS/HIV patients (16). Importantly, HCMV is the leading infectious cause of birth defects in newborns (6). Furthermore, there is evidence for HCMV to act as a risk factor in the development of vascular diseases, such as atherosclerosis, transplant vascular sclerosis, and coronary restenosis after angioplasty surgery (23, 36, 42, 45, 60, 66, 79). Recently, a role for HCMV has also been implicated in multiple forms of human cancers, where it may contribute to oncogenic transformation, onco-modulation, and tumor cell immune evasion (5, 19, 31, 54).

During lytic infection, HCMV genes are expressed in a highly ordered temporal cascade (reviewed in (3, 43, 62, 72)). Viral transcripts accumulate with three kinetic classes, namely immediate-early, early, and late. The HCMV major IE (MIE) genes UL123 (IE1) and UL122 (IE2) play critical roles in predisposing the cellular environment to infection and also act as transactivators to induce early gene transcription. Many early genes encode proteins required for viral DNA synthesis (22, 24, 63). The transcript accumulation of early genes is independent of viral DNA synthesis; however, the continued accumulation of a subset of these genes at late times is enhanced by the onset of viral DNA synthesis (61). Following viral DNA synthesis, late viral genes start to transcribe, which mostly encode structural proteins required for virus assembly and egress, ultimately leading to the release of infectious particles. Previous studies have shown that the activation of both beta- and
gamma-herpesvirus late gene promoters is dependent on the origin of viral DNA synthesis (OriLyt) \textit{in cis} (2, 17, 44). This further supports the notion that late gene transcription is tightly coupled to viral DNA synthesis. However, whether viral late gene expression is subjected to additional viral regulation remains poorly defined.

Recently, we and others have demonstrated that HCMV encodes five essential proteins, UL79, UL87, UL91, UL92, and UL95, which are required for the expression of viral late genes after viral DNA synthesis (28, 48, 52). Murine cytomegalovirus (MCMV) M79 and M92, homologs of HCMV UL79 and M92, respectively, are also required for late gene expression (9, 10). Homologs of UL79, UL87, UL91, UL92, and UL95 are found in murine gammaherpesvirus 68 (MHV-68) (ORF18, ORF24, ORF30, ORF31, and ORF34, respectively), which have been shown to have similar functions (4, 30, 74, 75). Epstein-Barr virus (EBV) BcRF1, a UL87 homolog, is a novel viral TATA-box binding protein with greater specificity for a non-classical TATA-box sequence (25, 76). Intriguingly, these factors are conserved only in beta- and gamma-herpesviruses and have no known homologues in herpes simplex virus (HSV) (43, 48), suggesting a unique viral regulatory mechanism shared by these two herpesviral subfamilies. However, the underlying mechanisms of how these viral factors regulate late gene expression are incompletely understood.

During cytomegalovirus infection, viral genes are transcribed by cellular RNA polymerase II (RNAP II), a large multi-subunit enzyme. Rpb1, the largest subunit of RNAP II, has a carboxy terminal domain (CTD) that contains repeats of the heptapeptide sequence of \text{\text{\text{\text{\text{Tyr}}}^{1}\text{\text{\text{\text{\text{-}}}\text{\text{\text{\text{\text{Ser}}}^{2}\text{\text{\text{-}}}\text{\text{\text{\text{\text{Pro}}}^{3}\text{\text{\text{-}}}\text{\text{\text{\text{\text{Thr}}}^{4}\text{\text{\text{-}}}\text{\text{\text{\text{\text{Ser}}}^{5}\text{\text{\text{-}}}\text{\text{\text{\text{\text{Pro}}}^{6}\text{\text{\text{-}}}\text{\text{\text{\text{\text{Ser}}}^{7}\text{}}}}}}}}}}}}}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{}}\text{
interactions with proteins involved in different phases of transcription and allow for the coupling of transcription with other nuclear processes, such as mRNA maturation and chromatin modification (51, 58). The amino acids within the CTD repeats are targets for post-translational modifications, including phosphorylation, glycosylation, ubiquitination, and cis-trans isomerization (13). Different combinations of CTD modifications, known as the CTD codes, orchestrate the sequential recruitment of numerous factors during the transcription cycle (8).

RNAP II activity is tightly regulated by phosphorylation of its CTD. RNAP II is recruited to preinitiation complexes (PIC) in an unphosphorylated state (38). Following the binding of RNAP II to a promoter, serine 5 residue of the CTD (Ser5) is rapidly phosphorylated by cdk7, the kinase subunit of the transcription factor TFII-H. This facilitates the dissociation of RNAP II from the PIC (59), and also promotes the recruitment of capping and splicing factors as well as other histone modification complexes (51). After dissociating from PIC, RNAP II proceeds to intrinsic pausing sites where it is halted by negative elongation factors (NELFs). The onset of productive elongation requires the positive transcription elongation factor P-TEFb composed of cdk9 and cyclin T, which phosphorylates serine 2 residue (Ser2) of the CTD to produce stable elongation complexes (40). Serine 2 phosphorylation also couples RNA synthesis with RNA processing by promoting the recruitment of splicing and polyadenylation factors (53). At the 3’ end of the coding region, RNAP II dissociates from the DNA template and RNA transcript prior to transcript polyadenylation. Specific phosphatases, Ssu72 and Fcp1, also dephosphorylate the CTD. Thus, RNAP II is recycled as an unphosphorylated, initiation-competent form for another round of transcription (12, 35).

HCMV utilizes RNAP II and the accompanying host machinery for transcription of
viral genes. During early times of viral infection, RNAP II and other transcription machinery are recruited to early replication sites to drive viral IE and early gene expression (67). The protein levels of RNAP II, including hyper-phosphorylated forms, increase as infection progresses (67, 69). Treatment of infected cells with cdk inhibitors inhibits viral gene expression as well as viral replication (55). During late stages of viral infection, cdk kinase and RNAP II-associated transcriptional machinery proteins continue to accumulate and relocate into the peri-replication center (69). However, how RNAP II transcription machinery remains active on viral loci during late infection requires further investigation.

In this chapter, we dissected the mechanism of CMV late gene expression by investigating the proteins that were associated with late transcription regulator pUL79 during HCMV infection. We found that pUL79 interacted with a panel of viral and host proteins, including RNAP II, other novel late transcription regulators pUL87, pUL92, and pUL95, as well as components of the viral DNA replication complex. We delineated the pUL79-RNAP II interaction and found that pUL79 bound to RNAP II in the nucleus independent of additional viral factors. Mechanistically, pUL79 did not alter RNAP II protein levels or the phosphorylation profile of its CTD domain. Instead, in the absence of pUL79, RNAP II stalled on viral DNA loci, including those of viral immediate-early, early, and late genes, but not those of host genes, during late times of infection. This resulted in a significantly diminished elongation rate of RNAP II-driven transcription on viral loci. We conclude that during late times of infection HCMV induces the formation of a unique transcriptional machinery in which pUL79 acts as an elongation factor to specifically drive RNAP II-mediated transcription on the viral genome.
MATERIALS AND METHODS

Plasmids and reagents. pYD-C755 (i.e. pLKO) was a pLKO-based lentiviral vector (also referred as pLKO.DCMV.TetO.mcs in (20), a generous gift from Roger Everett, University of Glasgow Centre for Viral Research). pYD-C751 (i.e. pLKO-HA-pUL79) was created by cloning a PCR fragment containing the UL79 coding sequence along with a N-terminal hemagglutinin (HA) tag into the multiple cloning site of pYD-C755. pYD-C744 was derived from pGalK (71), and carried a cassette in which 3×FLAG tag was followed by a GalK/kanamycin dual expression cassette flanked by the Flp recognition target (FRT) sequence (49).

The synthetic chemical ligand Shield-1 (Shld1) used to regulate the stability of ddFKBP-tagged proteins was purchased from Cheminpharma (Farmington, CT). Benzonase was purchased from EMD Millipore. The following primary antibodies were used in this study: anti-beta actin (clone AC15, Abcam); anti-FLAG (clone M2, Sigma-Aldrich); anti-HA (clone 16B12, Covance; clone 3F10, Roche); anti FKBP12 (clone 8/FKBP12, BD Biosciences); anti-Rpb1 (clone N-20 from Santa Cruz to detect total Rpb1; or clone 8WG16 from Abcam to detect both total Rpb1 and unphosphorylated CTD form of Rbp1); anti-Rpb2 (S-20, Santa Cruz); anti-Rpb1 phospho-CTD Ser5/Ser2 (clone H-14, Covance); anti-Rpb1 phospho-CTD Ser5 (clone 3E8, Millipore); anti-Rpb1 phospho-CTD Ser2 (ab5095, Abcam); anti-CDK9 (clone H-169, Santa Cruz); anti-cyclin T1 (clone H-245, Santa Cruz); anti-pUL44 (clone 10D8, Virusys); anti-IE1, anti-pp28, and anti-pp71 (generous gifts from Thomas Shenk, Princeton University).

Cells and Viruses. Primary human newborn foreskin fibroblasts (HFFs) and HEK-293T cells were propagated in Dulbecco modified Eagle medium (DMEM) supplemented with 10% fetal calf serum, nonessential amino acids, sodium pyruvate, and penicillin-
streptomycin.

Three HCMV recombinant viruses, ADwt, ADDdUL79, and ADflagUL79, were used in this study. The wildtype virus ADwt was reconstituted from the BAC-HCMV clone pADwt (also referred as pAD-GFP in the previous study (52)). pADwt carries the full-length genome of HCMV strain AD169, with the exception that it contains a simian virus 40 (SV40) early promoter-driven green fluorescent protein (GFP) gene in place of the viral US4-US6 region that are dispensable for viral replication in HFFs (68, 77). ADDdUL79 was derived from ADwt using BAC recombinase engineering, where the pUL79 coding sequence was fused to that of destabilizing domain ddFKBP (52).

ADflagUL79 was reconstituted from the BAC clone pADflagUL79. This BAC clone was derived from pADwt, and was constructed by using a linear recombination approach in the bacterial strain SW105 that contained an arabinose-inducible Flp gene for the transient expression of Flp recombinase (49). Briefly, the cassette that carried 3×FLAG followed by the GalK/kanamycin dual marker was first generated by PCR from pYD-C744 with a pair of 70-bp primers, so that the PCR-generated cassette was also flanked by 50-bp viral sequences immediately upstream or downstream of the 5’-end of the UL79 coding sequence. The cassette was recombined into pADwt at the 5’-end of the UL79 coding sequence by using linear recombination. The GalK/kanamycin marker was subsequently removed by Flp-FRT recombination (49). The final clone pADflagUL79 contained the 3×FLAG sequence along with a small FRT site fused in frame at the 5’-terminus of the UL79 coding sequence (Figure 3.1A).

To reconstitute virus, 2µg of the BAC-HCMV DNA and 1µg of the pp71 expression plasmid were transfected into HFF cells by electroporation (77), and the culture medium was changed 24 hours later. For reconstitution of ADDdUL79 virus, Shld1 was added every 48
hours to maintain the concentration at 1 µM. Reconstituted virus was harvested by collecting cell-free culture supernatant when the entire monolayer of cells was lysed. To produce virus stocks, cell-free culture supernatants were collected from HFFs infected at an MOI of 0.01. Viruses were pelleted by ultracentrifugation through a 20% D-sorbitol cushion at an average relative centrifugal force of 53,000 × g for 1 hour, resuspended in DMEM with 10% tissue fetal calf serum, and saved as viral stocks. HCMV titers were determined by 50% culture infectious dose (TCID₅₀) assay in HFFs (52).

**Transient transfection.** Four µg of plasmid DNA and 12 µl polyethylenimine (PEI) (1mg/ml, Polysciences) were mixed with 100 µl OPTI-MEM (Invitrogen) and incubated at room temperature for 10 minutes. The mixture was then added to 900 µl complete medium containing 10% fetal calf serum, and applied to 5×10⁶ HEK-293T cells that were seeded one day before. Cells were incubated for 4 hours before medium was changed.

**Analysis of Immunoprecipitation, Mass Spectrometry, and Immunoblot.** For total cell lysates, immunoprecipitation was performed using a protocol modified from previous studies (56, 64, 65). In brief, HFF cells (5×10⁷) were infected with HCMV ADflagUL79 or ADwt at a multiplicity of infection (MOI) of 3. At 72 hpi, cells were collected, rinsed twice with cold phosphate-buffered saline (PBS), and lysed in 2ml EBC2 buffer (50 mM Tris [pH 8.0], 300 mM NaCl, 0.5% NP40) supplemented with protease and phosphatase inhibitors. Cell lysates were then supplemented with 250 unit (U) Benzonase nuclease (Millipore), incubated at 4°C for 15 minutes. One aliquot of cell lysates was saved as the input control and boiled in the LDS sample buffer in the presence of sample reducing agent (Novex). The remainder was clarified by centrifugation at 10,000 × g at 4°C for 15 minutes. The supernatant was incubated with protein A-dynabeads (Novex) conjugated with antibody to FLAG (M2) or Rpb1 (N-20) together with an additional 250U of Benzonase at
4°C overnight. In addition, to confirm the nuclease activity of Benzonase, an aliquot of the supernatant was analyzed on a 0.8% agarose gel containing 100 µg/ml ethidium bromide for the detection of DNA/RNA. The following day the beads were washed three times with 1 ml EBC2 buffer and once with EBC2 buffer without NP40. The immuneprecipitants were eluted by boiling in reducing sample buffer for 5 minutes. For nuclear extracts, immunoprecipitation was performed using the Nuclear Complex Co-IP kit according to the manufacturer’s instructions (Active Motif).

For mass spectrometry analysis, proteins precipitated by anti-FLAG antibody was resolved on a NuPAGE 4-12% gradient gel (Novex) and subsequently stained using a ProteoSilver Silver Stain kit (Sigma-Aldrich) according to the manufacturer’s instruction. Protein bands unique to ADflagUL79-infected sample were extracted. In addition, gel bands from the ADwt-infected sample with migrating positions corresponding to those of ADflagUL79-specific bands were also extracted as negative controls. Extracted gel samples were submitted to the Keck Mass Spectrometry and Proteomics Facility (School of Medicine, Yale University) for liquid chromatography (LC)-mass spectrometry analysis for protein identification.

Protein amounts were determined by immunoblot analysis as previously described (52). In brief, proteins were resolved on an SDS polyacrylamide gel, transferred to a polyvinylidene difluoride (PVDF) membrane, hybridized with a primary antibody, reacted with the horseradish peroxidase-conjugated secondary antibody, and visualized using chemiluminescent substrate (Thermo Scientific).

**Chromatin immunoprecipitation (ChIP).** The ChIP was performed using the MAGnify chromatin-immunoprecipitation system (Life Technologies) and reagents provided in the kit according to the manufacturer’s protocol with modifications. $2 \times 10^6$ HFFs were
infected with ADddUL79 at an MOI of 3.0 in the presence or absence of Shld1. At 72 hours, infected cells were washed twice with PBS, trypsinized, and crosslinked with 1% formaldehyde at room temperature with mixing for 10 minutes. Glycine was added to the final concentration of 125 mM and incubated at room temperature for 5 minutes to stop the cross-linking reaction. Cells were collected by centrifugation at 4°C, 200×g for 10 minutes, washed twice in ice-cold PBS, and lysed in 100 µl lysis buffer with protease inhibitors. Chromatin was shared into 200-500 bp fragments by cup-horn sonication in ice water at 30-second pulse and 60% output with 40-second interval for 70 times (Branson Sonifier 450). Samples were gently vortexed every five sonication cycles and allowed to cool in ice water for additional 2 minutes. Lysates were cleared by centrifugation (20,000×g, 15 minutes; 4°C) and stored as 20-µl aliquots. To confirm the size of sheared chromatin fragments, one 20-µl aliquot was treated with RNase A at 37°C for 1 hour and de-crosslinked by protease K treatment overnight. DNA was purified and analyzed by agarose gel electrophoresis (data not shown).

To immunoprecipitate protein-bound chromatin fragments, each 20-µl aliquot was diluted in dilution buffer with protease inhibitors, and first incubated with 40 µl BSA-preblocked protein A/G Dynabeads to pre-clean for 2 hours. Beads were removed, and one tenth volume of the supernatant was saved as the input sample. The remainder of the supernatant was incubated with appropriated antibodies to generate protein-antibody complexes or with IgG (negative control) (Table 3.2) at 4°C for 16 hours. Forty µl BSA-preblocked protein A/G Dynabeads (Invitrogen) was added to the samples and incubated at 4°C for another 1.5 hours to immunoprecipitate the complexes. Beads were collected, washed twice with IP Buffer 1 and three times with IP Buffer 2. Protein-antibody complexes were eluted from Dynabeads by incubation with reverse crosslinking buffer with proteinase K.
K at 55°C for 15 minutes. Dynabeads were removed, and crosslinking of protein-antibody complexes in the supernatant were reversed by incubation at 65°C for 15 minutes. In addition, the input sample was also treated with the reverse crosslinking buffer in the same procedure to reverse crosslinking. Both input and immunoprecipitated DNAs were isolated by DNA purification on magnetic beads. DNA fragments were quantified by quantitative real-time PCR (qPCR) using SYBR Select Mix (Invitrogen) kit or Taqman Fast Advanced Master Mix kit (Invitrogen). The sequences of primers and Taqman probes are listed in Table 3.3.

**Nuclear run-on assay.** The protocol of nuclear run-on assay was adapted from previous studies with modifications (18, 50, 57). 1×10^7 HFFs were infected with ADddUL79 at an MOI of 3 in the presence or absence of Shld1. At 72 hpi, cells were washed twice with PBS, trypsinized, collected by centrifugation (4°C, 270×g), and washed twice with cold PBS again to remove residual calcium and magnesium. To extract nuclei, cell pellets were resuspended in 4 mL cell lysis buffer (10 mM Tris-HCl, pH 7.2, 3 mM MgCl₂, 10 mM NaCl, 150 mM sucrose, and 0.5% NP40) for 5 minutes on ice. Extracted nuclei were collected by centrifugation (4°C, 170×g) and gently washed with cell lysis buffer to remove NP40. Pellets were resuspended in 300 µl freezing buffer (50 mM Tris-HCl, pH 8.3, 40% glycerol, 5 mM MgCl₂, and 0.1 mM EDTA), washed once with 1× run-on reaction buffer (20 mM Tris-HCl, pH 7.5, 10 mM MgCl₂, 150 mM KCl, and 20% (v/v) glycerol). To perform run-on assay, 10^7 nuclei were incubated in 100 µl 1× run-on reaction buffer with ATP, CTP, GTP (0.5 mM each), and 0.2 mM biotin-16-UTP (Invitrogen) at 29°C for 30 minutes. The reaction was stopped by snap freezing in liquid nitrogen. As negative controls, run-on reactions were also performed with UTP instead of biotin-16-UTP. To isolate biotin-labeled run-on transcripts, streptavidin-coated Dynabeads (Dynabeads MyOne Streptavidin C1, Invitrogen) were
resuspended in binding buffer (10 mM Tris-HCl, pH 7.5, 1 mM EDTA, and 2 M NaCl), and mixed with an equal volume of run-on transcripts. The samples were incubated at 42 °C for 20 minutes and then at room temperature for 1.5 hours. Beads were collected, and washed twice with 15% formamide and three times with 2× standard saline citrate (Invitrogen). Biotinylated RNAs on the beads were reverse transcribed to generate cDNA using Supercript VILO cDNA Synthesis Kit (Invitrogen), and quantified by reverse transcription-coupled qPCR (RT-qPCR) analysis. The relative transcript amounts were normalized to those of 18S rRNA (that is transcribed by RNA polymerase I (RNAP I) so is an unbiased internal control for RNAP II activity). In addition, total RNA of infected cells was also isolated separately by TRIzol extraction (Invitrogen) and the amounts were determined by RT-qPCR analysis (see Table 3.3 for primer sequences).
RESULTS

Identification of pUL79-interacting proteins

To investigate proteins associated with pUL79, we first generated a recombinant HCMV in which the UL79 coding sequence was tagged with the 3×FLAG sequence (AD\textit{flag}UL79) so that protein complexes containing pUL79 in infected cell lysate could be isolated by immunoprecipitation (IP) with an anti-FLAG antibody (Figure 3.1A). Both growth and protein expression profile (Figures 3.1B and 3.1C) of AD\textit{flag}UL79 was indistinguishable from those of wildtype AD169 strain (AD\textit{wt}) in human foreskin fibroblasts cells (HFFs). These results indicate that the addition of 3×FLAG tag to the N-terminus of the UL79 coding sequence does not compromise the function of pUL79.

To identify proteins that interacted with FLAG-pUL79, lysates from HFF cells infected with virus AD\textit{flag}UL79 or AD\textit{wt} (negative control) were collected at 72 hours post infection (hpi) and immunoprecipitated with the anti-FLAG antibody. Immunoprecipitated proteins were resolved by SDS-PAGE and visualized by silver staining (Figure 3.1D). Protein bands unique to AD\textit{flag}UL79 were extracted and their identities were determined by mass spectrometry. For the negative control, we also extracted gel bands from the AD\textit{wt} sample with migrating positions corresponding to those of AD\textit{flag}UL79-specific protein bands as negative controls for mass-spectrometry analysis. The full set of proteins that were identified by this approach and unique to AD\textit{flag}UL79 is listed in Table 3.1.

These pUL79-interacting proteins could be categorized into several functional groups. Most notably, four out of twelve core subunits of human RNA polymerase II (RNAP II), namely Rpb1, Rpb2, Rpb3, and Rpb5, were identified (Table 3.1). Rpb1 is the largest subunit of RNAP II and its C-terminal domain (CTD) plays a critical role in transcription regulation by interacting with various transcriptional factors. Second, several viral proteins...
that are conserved among beta- and gamma- herpesviruses, including pUL87, pUL95, pUL49, and pUL92, were found in the pUL79-protein complexes. pUL87, pUL92, and pUL95 (shown in Table 3.1), together with pUL79, were required for viral late gene expression and reported to be recruited to the viral pre-replication complexes (10, 28, 52). These data together suggest that pUL79 interacts with other viral regulatory proteins involved in late gene expression during HCMV infection. Third, proteins that were involved in HCMV DNA synthesis and shown to bind to viral lytic origin of replication (OriLyt) (32), including pUL44, pIRS1, and pUL112/113, were also found in pUL79 protein complexes. Copurification of pUL79 and viral DNA replication factors suggests that pUL79 may have a role in coordinating viral DNA synthesis and late gene expression. Finally, several cellular proteins involved in protein translation, such as ribosomal protein subunits and elongation factor 1-alpha1, were co-purified with pUL79.

In the chapter, we focused on the interaction between pUL79 and RNAP II subunits. As RNAP II transcribes viral genes during infection, we hypothesized that pUL79 interacted with RNAP II to modify and promote its activity for viral transcription during late times of infection.

**pUL79 interacts with the RNAP II complex**

To further investigate the association of the RNAP II complex with pUL79, we first validated this interaction by immunoprecipitation analysis. HFFs were infected with either AD*flag*UL79 or ADwt (negative control), cell lysates were collected at 72 hpi, and proteins were immunoprecipitated by using antibodies against RNAP II or FLAG, followed by immunoblot analysis (Figure 3.2). Two RNAP II subunits, Rpb1 and Rpb2, were co-
immunoprecipitated with FLAG-pUL79 but not with ADwt-infected samples (Figure 3.2A). In a reciprocal experiment, an anti-Rpb1 antibody co-immunoprecipitated not only the RNAP II complex (indicated by Rpb1 and Rpb2) in both ADflagUL79- and ADwt-infected samples, but also FLAG-pUL79 in ADflagUL79-infected samples (Figure 3.2B). Taken together, these results indicate that pUL79 is associated with the RNAP II complex during viral infection.

The RNAP II complex binds to both DNA and RNA fragments. It was possible that the observed interaction of pUL79 with RNAP II was indirect, the result of the association of both proteins with the same DNA or RNA fragment. To determine if nucleic acids were required for the pUL79-RNAP II interaction, cell lysates were treated with a nonspecific nuclease (Benzonase) prior to immunoprecipitation (65). Benzonase treatment was effective, reducing RNA/DNA levels to undetectable levels in ethidium bromide-stained agarose gel electrophoresis analysis (Figure 3.2A and Figure 3.2B). In the presence of nuclease, pUL79, Rpb1, and Rpb2 remained co-immunoprecipitated in ADflagUL79-infected lysates (Figure 3.2A and 3.2B). Taken together, these results indicate that pUL79 and RNAP II associate with one another, and that this association is not mediated by nucleic acids.

We then determined whether the pUL79-RNAP II interaction could form independent of additional viral factors. To achieve this, we transfected HEK-293T cells with a plasmid expressing HA-tagged pUL79 or an empty vector plasmid. pUL79 contains a PY-nuclear localization signal directing it into the nucleus (70) and is located in viral replication compartments during infection (28, 52). Therefore, we extracted nuclear lysates of transfected cells, and performed coimmunoprecipitation analysis to examine the pUL79-RNAP II interaction using either an anti-HA antibody or anti-Rpb1 antibody in the presence of nuclease. As anticipated, HA-pUL79 was present in the nuclear extracts (Figure 3.2C and
Anti-HA antibody immunoprecipitated HA-pUL79 together with Rpb1, particularly the Rpb1 CTD phosphorylated at Serine 2 (pSer2-CTD) (Figure 3.2C). As pSer2-CTD is a marker of RNAP II undergoing transcriptional elongation, this result suggests that pUL79 may interact with RNAP II during the transcription cycle to modulate its elongation.

Reciprocal coimmunoprecipitation using an anti-Rpb1 further confirmed the association of RNAP II with pUL79 (Figure 3.2D). Together, these results indicate that pUL79 interacts with RNAP II independent of other viral factors. The presence of pSer2-CTD in the pUL79-RNAP II complex also suggests that pUL79 may regulate the elongation activity of RNAP II.

**pUL79 does not alter protein accumulations of RNAP II**

A previous study found that HCMV promoted the accumulation of RNA polymerase II at late times during infection (69). Various isoforms of phosphorylated RNAP II, including pSer2-CTD and pSer5-CTD (i.e. CTD phosphorylated at Serine 5, a hallmark of successful transcription initiation) also accumulate at these late times (67, 69). However, the mechanism of how HCMV regulates these RNAP II-mediated transcriptional events is not clear.

To test if the pUL79-RNAP II association could stabilize the RNAP II complex to increase its protein levels, we measured RNAP II protein accumulation during HCMV infection in the presence or absence of pUL79 protein. We have previously constructed a recombinant HCMV virus ADddUL79 in which the UL79 coding sequence was tagged with the highly unstable dFKBP domain (52). This allowed us to abrogate pUL79 function by targeting it for rapid degradation, or maintain its function by stabilizing the protein with the synthetic ligand Shield-1 (Shld1) (52). Here, we infected HFF cells with ADddUL79 in the presence or absence of Shld1, and analyzed infected cell lysates by immunoblotting at
various times post infection. In the presence of Shld1, ddFKBP-pUL79 was detected at 72 hpi using the antibody recognizing the ddFKBP epitope, whereas in the absence of Shld1, no protein was detected (Figure 3.3). To confirm this regulation of pUL79 activity, we also examined expression profiles of representative viral immediate-early (IE1), early (pUL44), and late (pp71) proteins. In the presence of pUL79, all three classes of viral proteins were accumulated with expected kinetics (Figure 3.3). In the absence of pUL79, immediate-early and early proteins accumulated normally but the accumulation of the late protein was dramatically reduced (Figure 3.3). These results were consistent with the previous study (52), and validated the effectiveness of Shld1-mediated regulation of pUL79 activity in this study. Importantly, the protein levels of Rbp2 and Rpb1 (both total Rpb1 and various CTD-phosphor isoforms) increased as expected when infection progressed (67, 69), but the accumulations were independent of the presence or absence of pUL79 (Figure 3.3). Together, these results indicate that total RNAP II as well as its CTD modified forms accumulate during viral infection in a pUL79-independent manner.

**pUL79 alters RNAP II occupancy at viral loci**

A previous study showed that MHV-68 ORF30 and ORF34, homologues of HCMV UL91 and UL95, respectively, were required for the recruitment of RNAP II to the viral late promoters (75). Like ORF30 and ORF34, both UL91 and UL95 were reported to be essential for late gene expression (28, 48). In this study, we identified pUL95 as a protein partner of pUL79 (Table 3.1). Therefore, we hypothesized that pUL79 formed a complex with pUL95 and other binding partners to recruit RNAP II to promote assembly of the transcription initiation complex at viral late promoters.
To test this, we determined the occupancy of RNAP II on viral late promoters with or without pUL79 during infection using chromatin immunoprecipitation (ChIP) assay. HFFs were infected with ADddUL79 in the presence or absence of Shld1, chromatin fractions from infected cells was collected at 72 hpi and analyzed by ChIP assay using a rabbit anti-RNAP II antibody. The amounts of input and output (immunoprecipitated) DNA were measured by quantitative real-time PCR (qPCR) analysis using primers specific to the promoter or transcript regions of viral genes or cellular housekeeping gene GAPDH (Table 3.3). The results were presented as relative output-to-input ratios to account for the percentages of host/viral genomes occupied by RNAP II during viral infection (Figure 3.4). The levels of viral and cellular DNA immunoprecipitated by Rbp1 antibody were readily detectable whereas DNA immunoprecipitated by control IgG was minimal, indicating the specific binding of Rbp1 to the DNA sequences detected in this assay. However, to our surprise, the occupancy of Rpb1 at promoter or transcript regions of viral genes was not reduced in the absence of pUL79 (Figure 3.4). Instead, without pUL79, Rpb1 levels on viral DNA were ~2-2.5 fold higher than those with pUL79. Importantly, during late times of infection (72 hpi), elevated Rpb1 accumulation occurred not only on the loci of viral late genes (UL32 and UL75), it also occurred on those of viral immediate-early genes (MIE) and early genes (UL54) (Figure 3.4). Moreover, this increased association of RNAP II with viral DNA occurred at both promoter regions and transcript regions. By comparison, Rpb1 occupancy on the host gene GAPDH was not altered by pUL79. Taken together, these results indicate that pUL79 regulates the occupancy of RNAP II, but not its recruitment, onto viral loci during late times of viral infection.
**pUL79 does not alter a particular phosphorylated form of the RNAP II CTD domain**

Next we wanted to determine how dysregulated elevation in the occupancy of RNAP II contributed to its diminished ability to transcribe viral genes when pUL79 was abrogated. Specifically, we wanted to determine which stage of the RNAP II transcription cycle (i.e. initiation, elongation, or termination) was altered by pUL79 by ChIP analysis using antibodies recognizing various forms of RNAP II CTD modifications. In a transcription cycle, Ser5 of RNAP II CTD is rapidly phosphorylated (pSer5-CTD) to facilitate the dissociation of RNAP II from the promoter and recruitment of RNA capping and splicing factors. After that, pSer5 CTD levels decrease with concomitant Ser2 phosphorylation (pSer2-CTD) to facilitate efficient transcription elongation. At 72 hpi, we found that both pSer5-CTD and pSer2-CTD levels significantly increased on viral loci in the absence of pUL79 compared to those in the presence of pUL79 (Figure 3.5A). However, the increase of unphosphorylated CTDs on viral loci also paralleled that of phosphorylated CTD (Figure 3.5A). Therefore pUL79 abrogation appeared to elevate all forms of CTD modifications tested at viral loci.

To more specifically determine whether the elevated accumulation of RNAP II on viral DNA arose from a specific CTD modification in the absence of pUL79, we normalized the ChIP occupancy values of pSer5-CTD, pSer2-CTD, and unphosphorylated CTD to that of total RNAP II. Occupancies of various CTD modifications were proportional to that of total RNAP II, and we found no evidence for the preferential occupancy of a particular CTD modification on any viral locus examined (Figure 3.5B). Therefore, elevated RNAP II occupancy in the absence of pUL79 was unlikely due to the dysregulation of CTD phosphorylation. Consistently, protein levels of CTD kinases (Cyclin T1 and CDK9) and CTD phospho-isoforms (pSer2-CTD, pSer5-CTD, pSer5/pSer2-CTD) were not altered by the
presence or absence of pUL79 (Figure 3.3). These results together indicate that pUL79 is not involved in phosphorylation of RNAP II CTD, and suggest that without pUL79, RNAP II simply stalls during the transcription cycle, resulting in its elevated accumulation at viral loci.

pUL79 alters the rate of transcriptional elongation at viral loci

Based on the above results, we hypothesized that pUL79 was required for efficient elongation of RNAP II-driven transcription at viral loci. To test this, we determined RNAP II elongation activity using nuclear run-on (NRO) assay. The NRO assay allowed us to monitored the contribution of RNAP II transcriptional activity to transcript levels independent of the effect of RNA stability (57). To do this, HFF cells were infected with ADddUL79 in the presence or absence of Shld1, and the nuclei of infected cells were isolated and analyzed by NRO assay. Additionally, total RNA was also harvested to monitor the total transcript accumulation.

We found that in the absence of pUL79, the run-on RNA levels of both MIE and late genes (UL99 and UL32) were reduced at 72 hpi to approximately 40% of those in the presence of pUL79 (Figures 3.6A, 3.6D, and 3.6E). The run-on RNA levels of early genes (UL44 and UL54) without pUL79 were also reduced to approximately 60% of those with pUL79 (Figures 3.6B and 3.6C). As RNAP II transcribes at the rate of 1.3-4.0 kb/minute (39), our NRO assay was performed for 30 minutes, long enough for RNAP II to transcribe all the viral genes tested. However, without pUL79, RNAP II still failed to transcribe viral genes at the levels comparable to those in pUL79-containing controls. Considering the increased accumulation of RNAP II at the viral loci in the absence of pUL79, we concluded that RNAP II requires pUL79 to efficiently elongate on the viral DNA.
As a control, we also examined the run-on RNA levels of host genes GAPDH and RPL30 (which encodes a 60S ribosomal protein). Both genes possess a pattern of histone modifications typical of permissive chromatin, similar to those associated with most CMV viral loci during late times of infection (47). In contrast to viral genes, the run-on RNA levels of both GAPDH and RPL30 were not altered by pUL79 (Figures 3.6F-3.6G). This is consistent with ChIP analysis where the occupancy of RNAP II at GAPDH was found unaltered in the absence of pUL79 (Figure 3.4), and indicates that RNAP II does not stall at host genomic loci even without pUL79. Therefore, pUL79 is specifically required for efficient transcription of viral genes but not host genes.

Taken together, our results from NRO assay provide definitive evidence that pUL79 positively regulates the transcription rates of viral genes but not those of host genes. In the absence of pUL79, RNAP II may still elongate at viral loci but does so at a much slower pace at late times, and ultimately fails to support productive viral gene transcription and viral progeny production.
DISCUSSION

In this chapter, we discovered a novel regulatory mechanism of viral transcription mediated by HCMV protein pUL79. We identified cellular RNA polymerase II (RNAP II) as a key factor that interacted with pUL79. This interaction did not alter the overall accumulation of total RNAP II or its various phospho-isoforms during viral infection. Rather, our data suggest that this interaction allowed pUL79 to act as a virus-encoded elongation factor to stimulate transcriptional elongation activity of RNAP II on viral loci. Without pUL79, RNAP II elongation failed to proceed efficiently and stalled on the viral genome. This caused slow turnover and excessive amount of RNAP II accumulation on viral loci. Ultimately this led to the failure to productive viral late transcription and progeny production.

This pUL79-mediated regulation occurred at viral loci of all three kinetic classes (immediate-early, early, and late) at late times during infection. Previously, when total viral transcripts were analyzed at early times during infection, both immediate-early and early transcripts accumulated efficiently before pUL79, which was a late protein, could be expressed (28, 52). Some of these transcripts could be stable and persist until late times of infection. When overall transcript accumulations were analyzed, the presence of pre-existing transcripts rendered it difficult to reveal the effect of pUL79 on transcription of immediate-early and early genes at late times during infection. NRO assay could measure relative transcription elongation rates at specific gene locus at defined times post infection. It allowed us to show that in fact pUL79 had the potential of driving transcription of many more viral genes than previously expected during late times of viral infection. A systemic, global run-on analysis of virally infected cells will further define the scope of viral gene transcriptions regulated by pUL79.
How does pUL79 modulate the elongation rate of RNAP II? It is possible that pUL79 enhances promoter clearance, a step in which RNAP II transfers from the initiation state to the elongation state (Figure 3.7A). During the transcription cycle, RNAP II is recruited to promoters by cellular TATA-box binding protein (TBP) and other general transcription factors (GTFs) to form the pre-initiation complexes (PIC). The PIC places RNAP II at transcription start sites, denatures DNA, and positions DNA into the RNAP II active site for transcription (7). Once transcription initiates, RNAP II dissociates from PIC and recruits elongation factors for efficient transcription. The dissociation of RNAP II from PIC is mediated by TFIIH and other cellular kinases to facilitate exchange between initiation factors and elongation factors (1, 78). Inefficient dissociation from PIC reduces the rate of RNAP II elongation, resulting in the failure to transcribe genes (1). Several herpesviral proteins have been reported to act as viral transcription initiation factors to form a unique viral PIC. For example, the homologues of HCMV UL87 in gamma-herpesviruses were reported to encode viral TBPs and regulate late transcript accumulation (2, 25). MHV68 ORF30 and ORF34, homologues of HCMV UL91 and UL95, were shown to be required for RNAP II recruitment to viral late promoters (75). EBV Rta, a homolog of CMV UL112/113, was found to associate with viral late promoters during late times of viral infection (26). These viral factors might participate in viral PIC assembly by mediating RNAP II promoter positioning. During HCMV infection, the recruitment of RNAP II to viral promoters was not reduced in the absence pUL79, suggesting that pUL79 was not required for transcription initiation (Figure 3.4). However, the elongation rate of RNAP II at viral loci was reduced drastically, suggesting that pUL79 was essential for a transcription step downstream of initiation (Figure 3.6). Strikingly, pUL79 co-purified with pUL87 and pUL95, two viral factors potentially involved in viral PIC assembly (Table 3.1). As the viral PIC complex may not be recognized by host dissociation factors, it is possible that pUL79 plays a role in the release of RNAP II
from viral PIC prior to elongation (Figure 3.7A). To test this, further analysis is required to determine the composition of RNAP II/viral PIC as well as their distribution on the viral DNA.

It is also possible that pUL79 plays a role in epigenetic regulation to modulate viral transcription (Figure 3.7B). During HCMV infection, viral DNA is chromatinized and undergoes histone modifications to facilitate gene expression (46). In particular, upon the onset of viral DNA replication, newly synthesized viral DNA is wrapped with histone 3 with lysine 4 methylation (H3K4me2), a modification that favors active transcription, suggesting the potential involvement of epigenetic regulation in viral late transcription (47). Even though pUL79 was not required for methylating H3K4 (47), the possibility remains that pUL79 may act as an epigenetic reader to recognize histone modifications unique to viral DNA, and unwrap viral DNA packaged by histones to facilitate RNAP II elongation (Figure 3.7B).

How does pUL79 specifically regulate transcription of viral loci? In this chapter, we showed that pUL79-mediated transcriptional regulation was limited to viral but not host genes (i.e. GAPDH and RPL30). This specificity may be partially due to the localization of pUL79 during infection as pUL79 is enriched in viral replication compartments where late viral transcription occurs (52). In addition, late promoters of beta- and gamma-herpesviruses contain a non-canonical TATA box sequence (25). EBV BcRF1, the homologue of HCMV pUL87, is a viral TATA-box binding protein which preferentially binds to this non-canonical TATA box over the canonical sequence. This suggest that viral transcription machinery directs RNAP II to viral late promoters during late stages of viral infection (25). In HCMV, several characterized viral late promoters also contain the same non-canonical TATA sequences (21, 29, 33, 37, 41, 73). Therefore, pUL79 may also act as viral specific TATA-
box binding protein. However, in this chapter, we observed an overall decrease in transcription rates among all three kinetic classes of viral loci during late times of infection (Figure 3.6). Further analysis is needed to understand how pUL79 can regulate the rate of viral transcription regardless of the structures of gene promoters.

In this chapter, we found that pUL79 also co-purified with other viral regulators of HCMV late gene expression, suggesting that pUL79 may interact with these regulators to form complexes during viral infection (Table 3.1). It is not known whether these viral regulators use similar mechanisms to regulate viral transcription. For example, pUL91 and pUL92 were shown to specifically regulate only true late genes (48). It is possible that these regulators have conserved functions and yet still possess different specificities. In addition, pUL79 also co-purified with viral DNA replication factors (Table 3.1). Previously we have shown that pUL79-mediated viral transcription requires the onset of viral DNA synthesis (52). Expression of neither pUL79 alone nor the combination of all known late gene regulators alters the expression kinetics of viral genes, especially viral late genes (48, 52). Therefore, it is also possible that viral DNA synthesis events predispose viral DNA to late transcription via interactions between replication factors and pUL79.

In conclusion, we have used a systematic proteomic approach to elucidate the mechanism underlying the activity of HCMV late gene expression regulator pUL79. pUL79 interacts with RNAP II to modulate its transcription rate at viral loci during late times of viral infection. This unique viral mechanism is potentially conserved among beta- and gamma-herpesviruses, and provides insight into the design of novel antivirals targeting steps after viral DNA synthesis.
Figure 3.1. Identification of pUL79 interacting proteins. (A) Schematic diagram for creating pADflagUL79, the recombinant HCMV BAC clone used to produce virus ADflagUL79. A cassette that contained a 3×FLAG tag followed by the FRT-bracketed GalK/kanamycin dual selection marker was amplified by PCR and recombined into the wildtype HCMV BAC clone (pADwt) at the 5' terminus of the UL79 coding sequence. The selection marker was then removed by Flp/FRT recombination. The final clone, pADflagUL79, carried the UL79 coding sequence tagged at its 5' terminus with 3×FLAG. (B) Single step viral growth analysis. HFF cells were infected with HCMV recombinant
virus ADflagUL79 (derived from pADflagUL79) or ADwt (derived from pADwt) at an MOI of 3. Infected culture supernatants were collected at indicated days post infection and virus titers were determined by TCID\textsubscript{50} assay. The mean virus titers were derived from two independent experiments and two technical replicates. Standard deviations were presented. The detection limit is indicated by the dashed line. (C) Viral protein expression profile. HFFs were infected as described in (B), and harvested at indicates times post infection. Accumulations of host and viral proteins were determined by immunoblot analysis. FLAG-tagged pUL79 was detected by anti-FLAG antibody. Actin was used as a loading control. Representative results from three independent experiments are shown. (D) Polyacrylamide gel electrophoresis to resolve pUL79 protein complexes. HFFs were infected as described in (B), and at 72 hpi, cell lysates were prepared for immunoprecipitation using an anti-FLAG antibody. Immunoprecipitated proteins were resolved on a gradient polyacrylamide gel and silver stained. Protein bands containing RNAP II subunits identified by mass spectrometry are indicated. Molecular size markers (in kilodaltons) are shown.
Figure 3.2. pUL79 interacts with the RNAP II protein complex. In (A-B), HFFs were infected as described in Figure 3.1, and at 72 hpi cell lysates were immunoprecipitated using either an anti-FLAG antibody (A) or anti-Rpb1 antibody N-20 (B). Immunoprecipitated proteins and lysate inputs were analyzed by immunoblotting. To examine the efficiency of nuclease digestion, the immunoprecipitated samples were also analyzed on an ethidium bromide (EtBr)-stained agarose gel. In (C-D), nuclear lysates from HEK-293T cells transiently expressing HA-tagged pUL79 or empty vector control were prepared at 72 hours post transfection. Lysates were immunoprecipitated using either anti-HA antibody (C) or anti-Rpb1 antibody 8WG16 (D). Immunoprecipitated proteins and lysate inputs were
analyzed by immunoblotting. The clone names of antibodies used in immunoblot analysis are shown. Representative results from three independent experiments are presented.
**Figure 3.3. pUL79 does not alter protein accumulations of RNAP II.** HFFs were infected with ADddUL79 at an MOI of 3 in the presence or absence of 1 µM Shield-1 (Shld1). Cells were harvested at different times post infection and protein accumulation was analyzed by immunoblot analysis with antibodies recognizing various subunits and isoforms of RNAP II, cellular CTD kinases (cyclin T1, CDK9), or viral proteins (immediate-early protein IE1, early-late protein pUL44, late protein pp71). The protein accumulation of the ddFKBP tagged pUL79 were monitored by an antibody recognizing the FKBP-epitope. Representative results from three independent experiments are shown.
Figure 3.4. pUL79 alters RNAP II occupancy at viral loci. HFFs cells were infected with ADddUL79 at an MOI of 3 in the presence or absence of 1 μM Shld1. Cell extracts were prepared at 72 hpi and analyzed by ChIP assay using rabbit anti-RNAP II antibody N-20. Normal rabbit IgG was included as a control for non-specific immunoprecipitation. Amounts of input and precipitated (output) DNAs were quantified by qPCR with primers specific for indicated viral loci or human GAPDH. The output-to-input DNA ratios were determined from four independent ChIP experiments with standard deviations calculated by Prism 6 software. Statistical analysis was performed using Student’s t test (**, P < 0.01; ***, P < 0.005; ****, P < 0.0001; NS, not significant).
**Figure 3.5.** pUL79 does not alter a particular phosphorylated form of the RNAP II CTD domain. HFFs cells were infected with ADdUL79 at an MOI of 3 in the presence or absence of 1 µM Shld1. Cell extracts were harvested at 72 hpi and analyzed by ChIP assays. Rabbit antibody to pSer2 CTD, rat antibody to pSer5-CTD, and mouse antibody to non-phosphorylated CTD (8WG16) were used in ChIP assays. Normal rabbit, rat, and mouse IgGs were included as controls for non-specific precipitation, respectively. Immunoprecipitated DNAs were analyzed as described in Figure 3.4 and the output-to-input DNA ratios were presented in (A). In addition, the immunoprecipitated amount of each phosphor-isoform of RNAP II CTD relative to that of total RNAP II (immunoprecipitated with antibody N-20) was also calculated and presented in (B). Data from four independent
experiments were collected with standard deviations calculated by Prism 6 software.
Statistical analysis was performed using Student’s t test (*, P < 0.05; **, P < 0.01; ***, P < 0.005).
Figure 3.6. **pUL79 alters the rate of transcriptional elongation at viral loci.** HFFs were infected with AD*dd*UL79 at an MOI of 3 in the presence or absence of 1 μM Shld1 (indicated by "+" or "-" sign, respectively). Nuclear extracts were prepared at 72 hpi and analyzed by nuclear run-on (NRO) assays. Transcription elongation was allowed to resume for 30 minutes in the presence of biotin-labeled UTP, labeled RNA was isolated, and their amounts were determined by RT-qPCR. In addition, accumulations of total RNAs were also
determined by RT-qPCR. The normalized amounts of viral run-on transcripts or total transcripts in the presence of Shld1 were set at 1 for run-on assay or total transcript accumulation analysis, respectively. Relative amounts of total and run-on transcripts for viral genes MIE (A), UL44 (B), UL54 (C), UL99 (D), UL32 (E), cellular genes GAPDH (F), and RPL30 (G) were presented. Data from three independent experiments were collected and standard deviations were calculated by Prism 6 software. Statistical analysis was performed using Student’s t test (**, P < 0.01; ***, P < 0.005; NS, not significant).
Figure 3.7. Potential role of pUL79 in RNAP II-mediated viral transcription. During late times of viral infection where pUL79 is expressed, we propose two models where pUL79 may act as an elongation factor to facilitate viral transcription. (A) In the “promoter clearance” model, pUL87, pUL92, pUL95, and potentially other viral factors (shown as red dashed circles) form viral protein pre-initiation complexes (vPIC) to recruit RNAP II to viral promoters. Once transcription initiates, pUL79 interacts with the vPIC to release RNAP II from the vPIC for efficient elongation. In the absence of pUL79, RNAP II is unable to dissociate from vPIC and fails to recruit elongation factors for continued transcription. (B) In the "epigenetic reading" model, pUL79 acts as an epigenetic reader to recognize chromatin modification(s) to facilitate RNAP II elongation. During late times of infection, newly synthesized viral DNA is wrapped with specific histone modifications (shown as purple dashed ovals). pUL79 recognizes these modifications, and then dissociates viral DNA from chromatin binding, with or without other cellular/viral factors, to facilitate RNAP II
elongation. In the absence of pUL79, RNAP II is unable to pass through the unopened chromatin, resulting in transcriptional stalling on viral loci.
### TABLES

<table>
<thead>
<tr>
<th>Protein ID</th>
<th>Description</th>
<th>Size (kDa)</th>
<th>Expectationa</th>
<th>Peptide count</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>HCMV transactivators</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B8YEB2</td>
<td>pUL87</td>
<td>104.7</td>
<td>0</td>
<td>44</td>
</tr>
<tr>
<td>B8YEB9</td>
<td>pUL95</td>
<td>57.2</td>
<td>9.4×10E-37</td>
<td>26</td>
</tr>
<tr>
<td>D2K4K7</td>
<td>pUL92</td>
<td>22.5</td>
<td>6.60×10E-24</td>
<td>6</td>
</tr>
<tr>
<td>A8T7F6</td>
<td>pUL79</td>
<td>33.8</td>
<td>1.5×10E-40</td>
<td>12</td>
</tr>
<tr>
<td><strong>HCMV DNA synthesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Q69214</td>
<td>UL112/113 (pp34)</td>
<td>28.3</td>
<td>2.7×10E-14</td>
<td>7</td>
</tr>
<tr>
<td>D2K4M1</td>
<td>UL112 (pp84)</td>
<td>70.2</td>
<td>3.2×10E-15</td>
<td>6</td>
</tr>
<tr>
<td>D2K5E9</td>
<td>pUL44 (DNA pol. processivity factor)</td>
<td>46.2</td>
<td>8.0×10E-6</td>
<td>2</td>
</tr>
<tr>
<td>C8CFZ3</td>
<td>IRS1 (tegument protein)</td>
<td>91.7</td>
<td>9.6×10E-3</td>
<td>1</td>
</tr>
<tr>
<td><strong>Other HCMV proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>D2K3R0</td>
<td>pUL104 (capsid portal protein)</td>
<td>78.4</td>
<td>2.5×10E-35</td>
<td>10</td>
</tr>
<tr>
<td>D2K4X8</td>
<td>pUL85 (capsid triplex subunit)</td>
<td>34.5</td>
<td>2.8×10E-16</td>
<td>7</td>
</tr>
<tr>
<td>D2K4H6</td>
<td>pUL150 (nuclear egress protein)</td>
<td>43.1</td>
<td>1.1×10E-05</td>
<td>2</td>
</tr>
<tr>
<td>Q1KQ04</td>
<td>pUL49</td>
<td>63.8</td>
<td>0.01</td>
<td>1</td>
</tr>
<tr>
<td><strong>Cellular RNA polymerase II</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RPB2_Human</td>
<td>Subunit Rpb2</td>
<td>133.8</td>
<td>8.8×10E-101</td>
<td>29</td>
</tr>
<tr>
<td>RPB1_Human</td>
<td>Subunit Rpb1</td>
<td>217</td>
<td>2.6×10E-81</td>
<td>27</td>
</tr>
<tr>
<td>RPB3_Human</td>
<td>Subunit Rpb3</td>
<td>31.4</td>
<td>1.1×10E-20</td>
<td>6</td>
</tr>
<tr>
<td>RPB5_Human</td>
<td>Subunit Rpb5</td>
<td>24.5</td>
<td>6.8×10E-07</td>
<td>2</td>
</tr>
<tr>
<td><strong>Ribosome biogenesis</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>RL7A_Human</td>
<td>60S ribosomal protein L7a</td>
<td>29.9</td>
<td>4.6×10E-07</td>
<td>3</td>
</tr>
<tr>
<td>RL21_Human</td>
<td>60S ribosomal protein L21a</td>
<td>18.5</td>
<td>2.1×10E-11</td>
<td>3</td>
</tr>
<tr>
<td>RS2_Human</td>
<td>40S ribosomal protein RPS2</td>
<td>31.3</td>
<td>6.7×10E-05</td>
<td>3</td>
</tr>
<tr>
<td>RS3A_Human</td>
<td>40S ribosomal protein RPS3a</td>
<td>29.9</td>
<td>4.4×10E-04</td>
<td>3</td>
</tr>
<tr>
<td>RL23_Human</td>
<td>60S ribosomal protein RPL23</td>
<td>14.8</td>
<td>3.9×10E-04</td>
<td>2</td>
</tr>
<tr>
<td>RL23A_Human</td>
<td>60S ribosomal protein RPL23a</td>
<td>17.6</td>
<td>3.6×10E-04</td>
<td>1</td>
</tr>
<tr>
<td>RL24_Human</td>
<td>60S ribosomal protein L24</td>
<td>17.7</td>
<td>8.8×10E-01</td>
<td>1</td>
</tr>
<tr>
<td>RS26L_Human</td>
<td>40S ribosomal protein S26-like 1</td>
<td>12.9</td>
<td>9.8×10E-03</td>
<td>1</td>
</tr>
<tr>
<td>RS27A_Human</td>
<td>Ubiquitin-40S ribosomal protein S27a</td>
<td>17.9</td>
<td>0.23</td>
<td>1</td>
</tr>
<tr>
<td><strong>Other cellular proteins</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ACTB_Human</td>
<td>Actin, cytoplasmic 1</td>
<td>41.9</td>
<td>4.6×10E-51</td>
<td>15</td>
</tr>
<tr>
<td>ACTN1_Human</td>
<td>Alpha-actinin-1</td>
<td>102.9</td>
<td>8.4×10E-10</td>
<td>3</td>
</tr>
<tr>
<td>EF1A1_Human</td>
<td>Elongation factor 1-alpha 1</td>
<td>50.1</td>
<td>4.8×10E-3</td>
<td>3</td>
</tr>
<tr>
<td>ANXA5_Human</td>
<td>Annexin A5</td>
<td>35.9</td>
<td>9.7×10E-3</td>
<td>3</td>
</tr>
<tr>
<td>H2A1B_Human</td>
<td>Histone H2A type 1-B/E</td>
<td>14.12</td>
<td>0.15</td>
<td>3</td>
</tr>
<tr>
<td>RECO4_Human</td>
<td>ATP-dependent DNA helicase Q4</td>
<td>132.9</td>
<td>0.25</td>
<td>1</td>
</tr>
</tbody>
</table>

a Expectation value for peptide match (i.e. the number of times expected to obtain an equal or higher score, purely by chance). A lower value indicates a higher likelihood of the interaction.

Table 3.1. pUL79 protein partners identified by mass spectrometry

103
<table>
<thead>
<tr>
<th>Antigen</th>
<th>Clone No</th>
<th>Source</th>
<th>Type</th>
<th>Amount&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Dynabead&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normal mouse IgG (negative control)</td>
<td>100005292</td>
<td>Invitrogen</td>
<td>Mouse polyclonal</td>
<td>2.5 µg</td>
<td>Protein A/G</td>
</tr>
<tr>
<td>Normal rabbit IgG (negative control)</td>
<td>100005291</td>
<td>Invitrogen</td>
<td>Rabbit polyclonal</td>
<td>1 µg</td>
<td>Protein A/G</td>
</tr>
<tr>
<td>Normal rat IgG (negative control)</td>
<td>NI04</td>
<td>Millipore</td>
<td>Rat polyclonal</td>
<td>5 µg</td>
<td>Protein G</td>
</tr>
<tr>
<td>RNAP II, CTD phosphorylated Ser2 (pSer2)</td>
<td>ab5095</td>
<td>Abcam</td>
<td>Rabbit polyclonal</td>
<td>1 µg</td>
<td>Protein A/G</td>
</tr>
<tr>
<td>RNAP II, CTD phosphorylated Ser5 (pSer5)</td>
<td>3E8</td>
<td>Millipore</td>
<td>Rat monoclonal</td>
<td>5 µg</td>
<td>Protein G</td>
</tr>
<tr>
<td>RNAP II, unphosphorylated CTD</td>
<td>8WG16</td>
<td>Abcam</td>
<td>Mouse monoclonal</td>
<td>2.5 µg</td>
<td>Protein A/G</td>
</tr>
<tr>
<td>RNAP II, total</td>
<td>N-20</td>
<td>Santa Cruz</td>
<td>Rabbit polyclonal</td>
<td>1 µg</td>
<td>Protein A/G</td>
</tr>
</tbody>
</table>

<sup>a</sup> Amount of antibody used in a 2.5×10<sup>5</sup>-cell sample.

<sup>b</sup> Type of Fc-binding proteins conjugated to Dynabeads.

**Table 3.2. Antibodies used in chromatin immunoprecipitation assays**
<table>
<thead>
<tr>
<th>Primer ID</th>
<th>Sequence (5' - 3')</th>
<th>Target</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>YD-Pri692</td>
<td>TTTTCTCAACCGAGGAACCTTTC</td>
<td>UL44, forward</td>
<td>(27)</td>
</tr>
<tr>
<td>YD-Pri693</td>
<td>CGCCTGTCCCCGACGTAAT</td>
<td>UL44, reverse</td>
<td>(27)</td>
</tr>
<tr>
<td>YD-Pri744</td>
<td>CTGTTGTGCTAGGCAAATTCGT</td>
<td>GAPDH, forward</td>
<td>(52)</td>
</tr>
<tr>
<td>YD-Pri745</td>
<td>ACCCACTCCTCCACCTTTGAC</td>
<td>GAPDH, reverse</td>
<td>(52)</td>
</tr>
<tr>
<td>YD-Pri2074</td>
<td>GGTTCACGTGCTGGATGTCG</td>
<td>UL32, forward</td>
<td>(52)</td>
</tr>
<tr>
<td>YD-Pri2075</td>
<td>CACCAAACCCTCGCTCCGATTAC</td>
<td>UL32, reverse</td>
<td>(52)</td>
</tr>
<tr>
<td>YD-Pri2076</td>
<td>GTGTCACCCTCCCGACTCG</td>
<td>UL99, forward</td>
<td>(52)</td>
</tr>
<tr>
<td>YD-Pri2077</td>
<td>TTCAACACGTCCACCCACC</td>
<td>UL99, reverse</td>
<td>(52)</td>
</tr>
<tr>
<td>YD-Pri2489</td>
<td>CTTACGGGACTTTCTCTACGA</td>
<td>MIE, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2490</td>
<td>CGATCTGACGGCTTACAA</td>
<td>MIE, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2491</td>
<td>CTAAGCTGGATGACCTA</td>
<td>MIE, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2492</td>
<td>GTGACCCAGAGAATCAG</td>
<td>MIE, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2493</td>
<td>CAACAAAGACAGCTCGTT</td>
<td>UL54, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2494</td>
<td>GTCCCTGCGACCCGACATT</td>
<td>UL54, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2495</td>
<td>GTGTCGACTGACTCGTTA</td>
<td>UL54, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2496</td>
<td>GACAGCAGTGAGTTTACA</td>
<td>UL54, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2511</td>
<td>GGGTGTGGCCTCGACGCTACAA</td>
<td>UL75, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2512</td>
<td>ATGCGCTAATCCCGTGCTGCC</td>
<td>UL75, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2521</td>
<td>GTGTCACATACAAAGAGGATCAG</td>
<td>UL75, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2522</td>
<td>AAACCAACTCCACAGACTCG</td>
<td>UL75, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2527</td>
<td>CCAGGATGACACAGATCATC</td>
<td>UL32, forward</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri2528</td>
<td>AGAAAACCGCGTATCCGCCT</td>
<td>UL32, reverse</td>
<td>(46)</td>
</tr>
<tr>
<td>YD-Pri3021</td>
<td>GTAACCGAGATTCCACATT</td>
<td>18s rRNA, forward</td>
<td></td>
</tr>
<tr>
<td>YD-Pri3022</td>
<td>CCCATCAATCAGTTGAGTGCG</td>
<td>18s rRNA, reverse</td>
<td></td>
</tr>
<tr>
<td>YD-Pri3023</td>
<td>CAAGGCAAAGCGAAATTGCTT</td>
<td>RPL30, forward</td>
<td>(34)</td>
</tr>
<tr>
<td>YD-Pri3024</td>
<td>GCCCGCTCGATGCTTCTTGATT</td>
<td>RPL30, reverse</td>
<td>(34)</td>
</tr>
<tr>
<td>YD-Pri3027</td>
<td>CTGGGCAGTGTTAGCGAGA</td>
<td>RPL30, forward</td>
<td>(34)</td>
</tr>
<tr>
<td>YD-Pri3028</td>
<td>GTGCCGTGGATGCTGATGACT</td>
<td>RPL30, reverse</td>
<td>(34)</td>
</tr>
<tr>
<td>Hs03929097_g1</td>
<td>Taqman probe for ChIP (Invitrogen)</td>
<td>GAPDH</td>
<td></td>
</tr>
</tbody>
</table>

* Promoter regions

a Used in RT-qPCR for nuclear run-on assay only

b Used in RT-qPCR for total transcript quantification

c Table 3.3. Primers and probes used in ChIP and RT-qPCR analysis

105
REFERENCES


48. **Omotu, S., and E. S. Mocarski.** 2013. Cytomegalovirus UL91 is essential for transcription of viral true late (gamma2) genes. Journal of virology **87**:8651-8664.


64. **Strang, B. L., S. Boulant, and D. M. Coen.** 2010. Nucleolin associates with the human cytomegalovirus DNA polymerase accessory subunit UL44 and is necessary for efficient viral replication. Journal of virology **84**:1771-1784.


Chapter 4:

Summary and Future Directions
Summary

As an important human pathogen, HCMV possesses many essential genes whose functions remain unknown. In this dissertation, I characterize the function and potential mechanism of a novel essential protein pUL79 during viral infection.

To characterize the function of pUL79 during viral infection, we applied the dDFKBP-mediated, Shield-1-regulated protein genetic approach(3) to investigate its role during HCMV replication(20). Without pUL79, HCMV failed to replicate. Systemic analysis of ADddUL79 infection in the absence of Shield-1 showed that UL79 was required at late stages of HCMV infection. Acting as a viral early-late protein, pUL79 localized into viral replication compartments where viral replication and late gene transcription took place. In the absence of pUL79, the immediate-early and early viral proteins as well as viral DNA accumulated normally. However, accumulations of viral late transcripts and proteins were dramatically reduced. We conclude that UL79 acts as a viral transactivator after DNA replication to promote the accumulation of late viral transcripts.

To further dissect the potential mechanisms of pUL79 to regulate viral late gene expression, we applied an unbiased proteomic analysis to identify its protein partners. In chapter 3, we showed that HCMV pUL79 interacts with human RNAP II as well as other viral factors involved in late gene expression during viral infection. Since HCMV genes are transcribed by RNAP II during viral infection, we focused on the nature of the pUL79-RNAP II interactions, finding that pUL79 does not alter the protein levels of RNAP II or its recruitment to viral promoters. However, during late times of infection, pUL79 helps RNAP II to efficiently elongate along the viral DNA template to transcribe HCMV genes. While viral genes were regulated in this manner, we found that host genes are not regulated by this pUL79-mediated mechanism. Therefore, our study identifier a previously uncharacterized
mechanism in which RNAP II activity is modulated by the viral factor pUL79, and potentially other viral factors as well, for coordinated viral transcription.

Altogether, our significant finding is that pUL79 plays a critical role to drive viral late transcription after viral DNA replication. Since pUL79 is functionally conserved among beta- and gamma- herpesviruses, this provides an opportunity to understand the previously uncharacterized regulation of viral late transcription of these two viral sub-families. Inhibition of pUL79 and homologs represents a novel anti-viral therapeutic target against both beta- and gamma- herpesviruses. Moreover, since the function of pUL79 is mediated by its interaction with RNAP II, our study also serves as a novel probe to elucidate the mechanistic actions of the RNAP II complex during transcription initiation, elongation, or potentially termination.
Future Directions

Although we have taken important steps to characterize the function of pUL79, the precise mechanism by which pUL79 regulates HCMV late gene transcription awaits further investigation. Below we discuss future experiments that may aid in this discovery.

How does pUL79 specifically regulate transcription of viral loci during late times of infection? In chapter 3, we show that pUL79-mediated transcriptional regulation is limited to viral but not host genes. The transcription rates of the host genes investigated, including GAPDH and RPL30, were not regulated by pUL79 (Figures 3.6F and 3.6G). Consistent with these results, RNAP II did not stall on the GAPDH locus in the absence of pUL79 (Figure 3.4). We suspect that the specific regulation of viral genes by pUL79 may be partially due to its localization. pUL79 is enriched in the replication compartments where late viral transcription occurs, which suggests that pUL79 primarily associates with RNAP II transcribing viral genes (20). A recent study showed that the viral late promoters encode a non-canonical TATA box sequence that is well-conserved among late promoters of beta- and gamma-herpesviruses (9). EBV BcRF1, homolog of HCMV pUL87, is a viral TATA-box binding protein which preferentially bound to this non-canonical TATA box over the canonical sequence, suggesting that the viral transcription machinery directs RNAP II to viral late promoters during late stages of viral infection (9). In HCMV, multiple characterized viral late promoters also encode the same non-canonical TATA sequences which may allow for the targeting of RNAP II to their loci (8, 11-14, 16, 23). However, in the absence of pUL79, we also observed an overall drop in transcription rates among all different kinetic classes of viral genes during late times of infection (Figure 3.6). For example, UL54 is an early-late gene with a classical TATA sequence (12, 13). However, the transcription rate of UL54 still decreased in the absence of pUL79 (Figure 3.6C). We hypothesize that, during late stage of
viral infection, BcRF1 and its homologs (ex. pUL87 in HCMV) are expressed and form complex with pUL79 homologs. This viral protein complex, located in the viral replication compartment, replaces host TBP and other RNAP II associated factors to direct RNAP II transcribing viral genes belong to different kinetics. However, the binding preference of BcRF1 promotes RNAP II to transcribe more viral early-late and late genes during late times of viral infection. Further characterization is needed to dissect how pUL79 associates with these viral factors to regulate the rate of viral transcription.

**How does pUL79 modulate RNAP II elongation?** We considered two potential roles that pUL79 might play. First, pUL79 might act as a modulator to increase promoter clearance, a step in which RNAP II transfers from the initiation state to the elongation state. During the transcription cycle, RNAP II is recruited to promoters by host TATA-box binding protein (TBP) and other general transcription factors (GTFs) to form the pre-initiation complex (PIC). The PIC helps position RNAP II over transcription start sites, denatures DNA, and positions DNA into the RNAP II active site for transcription(5). After transcription initiates, the RNAP II dissociates from PIC and recruits elongation factors for efficient transcription. The dissociation of RNAP II from PIC is mediated by TFIIH and other cellular kinases to facilitate exchange between initiation factors and elongation factors(1, 25). Inefficient dissociation reduces the rate of RNAP II elongation with subsequent failure to transcribe genes(1). Several viral proteins have been shown to be viral transcription initiation factors and to potentially form a unique viral PIC. For example, UL87 and its homologs in beta- and gamma- herpesviruses were found to be viral TBPs and regulate late transcript accumulation(2, 9). MHV68 ORF30 and ORF34, homologs of HCMV UL91 and UL95, were also shown to be required for RNAP II recruitment to viral late promoters(24). EBV Rta, a homolog of CMV UL112/113, was found to associate with viral late promoters during
late times of viral infection (10). All of these viral factors might participate in viral PIC assembly by mediating RNAP II promoter positioning (Figure 3.7A). Strikingly, pUL79 co-purified with the factors mentioned above (Table 3.1). In the absence pUL79, the recruitment of RNAP II to viral promoters was not influenced, indicating that pUL79 is not required for RNAP II initiation (Figure 3.4). However, the elongation rates of RNAP II at viral loci dropped, suggesting that pUL79 is essential for certain transcription steps downstream of initiation. Since the viral PIC complex might not be recognized by host dissociation factors, we hypothesize that pUL79 plays a role in the release of RNAP II from viral PIC prior to elongation. The simplified model is shown in Figure 3.7A. To test this hypothesis, further analysis is required to determine the composition of RNAP II/viral PIC as well as their distribution on the viral DNA.

Second, pUL79 could play a role in the epigenetic regulation to modulate viral transcription. Previous studies showed that HCMV undergoes dedicated histone modification during infection, including increases in histone 3 acetylation (H3Ac) and histone 3 lysine 9 (H3K9) methylation, that activate gene expression(17). After the onset of viral DNA replication, newly synthesized viral DNA is wrapped with histone 3 lysine 4 methylation (H3K4me2), a modification that favors active transcription, suggesting that epigenetic regulation is involved in viral late transcription(18). However, pUL79 was not required for methylating H3K4(18). Notably, we showed that the rate of RPL30 transcription is not regulated by pUL79 (Figure 3.6G). RPL30 is a host ribosomal protein whose epigenetic modification was similar to viral DNA during late times of viral infection(18). Therefore, if pUL79 is an epigenetic reader for late viral transcription, it might recognize an as of yet uncharacterized histone modifications unique to viral DNA. We hypothesize that pUL79 act as an epigenetic reader to recognize these unique modified histones and unwrap viral DNA packaged by histones to facilitate RNAP II elongation (Figure 3.7B).
How does pUL79 interacts with RNA polymerase II (RNAP II) complex? To determine how pUL79 interacts with RNA polymerase II, we created truncation mutants targeting N-terminus and C-terminus of pUL79 (Figure 4.1A), and tested the ability of mutant pUL79 proteins to interact with RNAP II in 293T cells (Figure 4.1B). Both mutants were expressed at 293T cells. However, N-terminal truncation mutant had lower expression level, suggesting that the N-terminus of pUL79 might be required for proper protein folding and stability. As expected, full-length pUL79 co-immunoprecipitated with RNAP II. In contrast, both N- and C- terminuses mutants were unable to do so. We concluded that both N- and C- terminuses of pUL79 are required for pUL79-RPB1 interaction. To further explore how pUL79 interacts with RNAP II complex, we decided to focus on RPB1, the largest subunit of RNA polymerase II complex, with the help of bioinformatics tools.

In chapter 3, we identified RPB1 as one of pUL79 interacting partners. RPB1 plays a major role in the RNA polymerase II transcription cycle. The C-terminal domain (CTD) of Rpb1 functions as a binding platform for the recruitment of transcription-associated factors. During the transcription cycle, the Rpb1 CTD undergoes dynamic phosphorylation that defines the ability of the RNA pol II transcriptional complex to recruit the appropriate set of mRNA-processing and histone-modifying factors (4, 6). The pUL79-RPB1 interaction suggests that pUL79-regulated RNAP II transcription elongation might be mediated by this interaction. To test this hypothesis, we performed bioinformatics analysis (MAFFT program) to identify the potential domain of pUL79 required for binding to Rpb1. We found that pUL79 potentially has conserved domains similar to a CTD-interacting domain (CID) present in proteins such as Pcf11 and SCAF4/8, which are essential and evolutionarily conserved factors for polyadenylation-dependent and -independent 3’-RNA processing, respectively (7, 15) (Figure 4.2A).
To confirm this finding, we first created pUL79 mutants that contained an internal deletion covering the CID-like domains (pUL79 del aa. 40-46, del. aa. 96-103, del. aa. 192-200, see Figure 4.1). We examined their ability to complement pUL79 null virus and interact with Rpb1. All three mutants failed to complement the growth of pUL79 null virus or rescue the defect in late protein accumulation (Figure 4.3). We also performed co-immunoprecipitation experiments to test the interaction between Rbp1 and these pUL79 mutants. The preliminary study also suggests that all three mutants fail to interact with Rpb1 during virus infection (Figure 4.2). However, we noticed that these pUL79 mutant proteins were expressed at low levels, suggesting that the internal-deletion mutants might be unstable and degraded. To overcome this, we created two different pUL79 mutant libraries. First, we created additional internal-deletion mutants before and after the CID-like domains. We expected that these mutants would serve as controls and help us understand whether the CID-like domains are critical for protein folding and stability. Second, we created site-directed mutagenesis mutants covering most of the three CID-like domains that might avoid protein stability issues and allows us pinpoint the critical residues of protein interaction. The preliminary results of complementation tests show several loci within or outside CID-like domains are important for viral growth (Figure 4.4B and 4.4C). However, we noticed that several pUL79 mutant proteins, even when expressed at similar levels compared to the wild-type pUL79, failed to translocate into nucleus (Figure 4.4D). Further analyses are required to define whether these regions are critical for proper protein folding to pass through nuclear pore. Meanwhile, further analyses, including co-immunoprecipitation and in vitro binding assays, will help to confirm whether and how pUL79 interacts with RNAP II complex through RPB1 CTD domain.
**Interactions between pUL79 and viral DNA replication factors.** During lytic infection, HCMV genes are expressed in a highly ordered temporal cascade. The transcript accumulation of early genes is independent of viral DNA synthesis; however, the continued accumulation of a subset of these genes (i.e., early-late) at late times is enhanced by the onset of viral DNA synthesis (22). Following viral DNA synthesis, late viral genes start to transcribe. In chapter 2, we showed that the pUL79-mediated viral transcription requires DNA replication (20). Over-expression of neither pUL79 alone nor the combination of all late gene regulators currently identified in cells altered the expression kinetics of viral genes, especially viral late genes (19, 20). In chapter 3, we found that pUL79 co-purified with viral DNA replication factors. It is possible that viral DNA replication directs viral late transcription through the direct interactions between replication factors and pUL79. Interestingly, a recent study of the HCMV latent transcriptome in CD14+ and CD34+ cells showed that pUL79, pUL87, pUL95 were consistently expressed in latently-infected cells (21). Further analysis will help to determine whether and how these beta-gamma genes (pUL79, pUL87, pUL92, or pUL95) act as sensors of viral DNA replication to direct late transcription during acute infection and viral reactivation during viral latency.
**FIGURES**

**A.**

![Diagram of pUL79 truncation and internal-deletion mutants analyzed in this study.](image)

**B.**

<table>
<thead>
<tr>
<th>Input</th>
<th>IP:RPB1</th>
</tr>
</thead>
<tbody>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Figure 4.1. Both N- and C- terminuses of pUL79 are required for pUL79-RNAP II interaction. (A) Diagram of pUL79 truncation and internal-deletion mutants analyzed in this study. (B) HA-tagged UL79 wild-type and truncation mutant proteins were expressed in 293T cells by transfection. Cells were collected at 72 hours and lysates were immunoprecipitated with RNAP II antibody (RPB1). Cell lysates (Input) and eluted proteins (IP:RPB1) were analyzed by immunoblotting.
A. 

Figure 4.2. pUL79 potentially contains RPB1 CTD-binding domains (A) Amino acid alignment of UL79 protein and the CTD-binding domains (CID) of two human transcription termination factors, SCAF4 and SCAF8, using the MAFFT program. Amino acid residues conserved in all three proteins are indicated in black, and residues conserved in any two proteins were indicated in gray. The conserved residues required for CTD-CID interaction are boxed in red. (B) MRC5 cells over-expressing pUL79 wild-type and internal deletion mutants were infected with ADddUL79 at a MOI of 3 in the absence of Shield-1. Cell lysates were collected at 72 hpi, immunoprecipitated with anti-HA antibody, and analyzed by immunoblotting.

![Amino acid alignment](attachment:image.png)

**Figure 4.2.** pUL79 potentially contains RPB1 CTD-binding domains (A) Amino acid alignment of UL79 protein and the CTD-binding domains (CID) of two human transcription termination factors, SCAF4 and SCAF8, using the MAFFT program. Amino acid residues conserved in all three proteins are indicated in black, and residues conserved in any two proteins were indicated in gray. The conserved residues required for CTD-CID interaction are boxed in red. (B) MRC5 cells over-expressing pUL79 wild-type and internal deletion mutants were infected with ADddUL79 at a MOI of 3 in the absence of Shield-1. Cell lysates were collected at 72 hpi, immunoprecipitated with anti-HA antibody, and analyzed by immunoblotting.
Figure 4.3. Deletion of pUL79’s CID-like domains reduces viral late protein accumulation and viral progeny production. (A) MRC5 cells over-expressing HA-tagged pUL79 wild-type or internal deletion mutants were infected with ADddUL79 at a MOI of 3 in the absence of Shield-1. Cell lysates were harvested at 72 hpi and the expression profile of viral proteins was analyzed by immunoblotting. (B) The supernatant was also harvested 7 days post infection and the presence of viral progeny was analyzed by TCID₅₀ assay. The detection limits of TCID₅₀ assay are indicated by dashed lines in the graphs.
A.

1st:

M A R D E E N P A V P R V T G K P S F T C A N H L I L Q I S E K M S R G Q P L S L R

del. aa. 40-46


del. aa. 40-46

P Y W P H L Y R B L R Q A F P G L D F E A A V F D E T R A A R L S Q R L C H P R L S G G L

aa. 101-102 RQ-aa

del. aa. 96-103


del. aa. 174-182

R H W C A E A Y E P L I R I I C Q M W Y F Y L I G T G K R T P D A F E I Q R S R H E

aa. 181-182 RH-EE-aa. 197-198 QK-II

del. aa. 174-182

del. aa. 192-200

del. aa. 201-200 MW-EE

aa. 183-184 WG-SS

aa. 185-186 AE-SS

aa. 189-190 EP-RR

aa. 193-194 RI-EE

aa. 195-196 IC-EE

T G I F T F I M E D Y R T F A G T L S R H P H R P H P Q Q Q Q Q Q Q H H P G P H P P L S H P

del. aa. 231-239

del. aa. 240-245

del. aa. 246-250

del. aa. 251-255

del. aa. 263-271

A S S C L S P E A V L A A R A L H M P T L A N D V

del. aa. 263-271

del. aa. 289-295

B.

TCID₅₀ per ml (log₁₀)
Figure 4.4. Identification of residues critical to pUL79-RPB1 mediated viral replication.
(A) Diagram of pUL79 internal-deletion and side-directed mutagenesis mutants analyzed in this study. (B&C) MRC5 cells over-expressing HA-tagged pUL79 internal deletion (B) or site-directed mutagenesis (C) mutants were infected with ADddUL79 at a MOI of 1 in the absence of Shield-1. The supernatant was also harvested 7 days post infection and the presence of viral progeny was analyzed by TCID$_{50}$ assay. The detection limits of TCID$_{50}$ assay are indicated by dashed lines in the graphs. (D) HA-tagged UL79 wild-type or truncation mutant proteins were expressed in 293T cells by transfection. Cell lysates were collected at 72 hours and cellular nuclear extracts were extracted for immunoblotting analysis.
REFERENCES


