Murine Cytomegalovirus Encodes Proteins that Regulate Viral Late Transcription

Travis Chapa
Washington University in St. Louis

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

Recommended Citation
Chapa, Travis, "Murine Cytomegalovirus Encodes Proteins that Regulate Viral Late Transcription" (2014). All Theses and Dissertations (ETDs). 1289.
https://openscholarship.wustl.edu/etd/1289

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.
Murine Cytomegalovirus Encodes Proteins that Regulate Viral Late Transcription

By

Travis James Chapa

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

St. Louis, Missouri
# TABLE OF CONTENTS

List of Figures ................................................................................................................................................ v  
List of Tables ................................................................................................................................................ vii  
List of Abbreviations ..................................................................................................................................... viii  
Acknowledgements ....................................................................................................................................... xi  
Abstract of the Dissertation .......................................................................................................................... xii  

**Chapter I: Introduction** .......................................................................................................................... 1  

Herpesviruses ............................................................................................................................................... 2  
Betaherpesviruses ........................................................................................................................................ 3  
Cytomegalovirus ........................................................................................................................................... 4  
Pathogenesis ................................................................................................................................................ 4  
Virus structure ............................................................................................................................................... 6  
Life cycle ..................................................................................................................................................... 10  
Early gene expression ................................................................................................................................. 16  
Late gene expression ................................................................................................................................... 19  
Development of herpesvirus BAC system ................................................................................................... 24  
Functional profiling in HCMV ....................................................................................................................... 25  
Aim and Scope of Thesis ............................................................................................................................ 25  
References .................................................................................................................................................. 30  
Tables.......................................................................................................................................................... 52  
Figures and Legends ................................................................................................................................ 54  

**Chapter II: Murine Cytomegalovirus protein pM79 Is a Key Regulator of Viral Late Transcription** 57  
Preface ........................................................................................................................................................ 58  
Summary ..................................................................................................................................................... 59  
Introduction ................................................................................................................................................ 60  
Results ......................................................................................................................................................... 63  

<table>
<thead>
<tr>
<th>Chapter III: Murine Cytomegalovirus Protein pM92 Is a Conserved Regulator of Viral Late Gene</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface ................................................................................... 101</td>
</tr>
<tr>
<td>Summary................................................................................... 102</td>
</tr>
<tr>
<td>Introduction.............................................................................. 103</td>
</tr>
<tr>
<td>Results .................................................................................. 105</td>
</tr>
<tr>
<td>Discussion .............................................................................. 111</td>
</tr>
<tr>
<td>Acknowledgements ................................................................... 114</td>
</tr>
<tr>
<td>Materials and Methods .......................................................... 115</td>
</tr>
<tr>
<td>References ............................................................................. 121</td>
</tr>
<tr>
<td>Tables ..................................................................................... 126</td>
</tr>
<tr>
<td>Figures and Legends .................................................................. 127</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Chapter IV: MCMV Late Transcription Regulator pM92 Interacts with pM79 and RNAP II During Infection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preface ................................................................................... 138</td>
</tr>
<tr>
<td>Introduction.............................................................................. 139</td>
</tr>
<tr>
<td>Results .................................................................................. 143</td>
</tr>
<tr>
<td>Future Directions .................................................................... 148</td>
</tr>
<tr>
<td>Discussion .............................................................................. 150</td>
</tr>
<tr>
<td>Materials and Methods .......................................................... 152</td>
</tr>
</tbody>
</table>
List of Figures

Chapter I: Introduction

Figure 1. Human cytomegalovirus genome schematic ................................................................. 54
Figure 2. A representative diagram depicting the strategy for creating recombinant MCMV BAC Clones ......................................................................................................................... 55
Figure 3. Life cycle of CMV in a host cell ...................................................................................... 56

Chapter II: Murine Cytomegalovirus protein pM79 Is a Key Regulator of Viral Late Transcription

Figure 1. Gene M79 is essential for MCMV growth in fibroblasts .................................................... 88
Figure 2. Expression of M79 gene products is markedly enhanced by viral DNA synthesis and protein pM79 ........................................................................................................................................ 90
Figure 3. pM79 localizes to replication compartments during infection ........................................ 91
Figure 4. pM79 is not required for viral DNA synthesis ................................................................. 92
Figure 5. pM79 is not required for the maturation of viral replication compartments .......... 94
Figure 6. pM79 is required for efficient expression of a representative viral late gene ................. 95
Figure 7. Tiled array analysis of genome-wide transcription during MCMV infection ................. 97
Figure 8. RT-qPCR analysis of representative PAA-sensitive transcripts in MCMV infection ....... 99

Chapter III: Murine Cytomegalovirus Protein pM92 Is a Conserved Regulator of Viral Late Gene Expression

Figure 1. pM92 is essential for MCMV replication in fibroblasts .................................................. 127
Figure 2. pM92 accumulates abundantly at late times of infection and localizes to viral nuclear replication compartments ................................................................. 129
Figure 3. pM92 is dispensable for viral DNA synthesis but is required for efficiently accumulation of M55 late gene products during infection .................................................... 131
Figure 4. pM92 is required for efficient accumulation of a panel of late transcripts .................. 133
Figure 5. CMV UL92/M92 proteins interact with UL79/M79 proteins during infection ............. 134
Figure 6. pM92 trans complements the growth of pUL92-deficient HCMV virus ..................... 135
Chapter IV: MCMV Late Transcription Regulator pM92 Interacts with pM79 and RNAP II During Infection

Figure 1. Proteomic analysis to identify binding partners of pM79 and pM92 ............................... 164
Figure 2. Immunoprecipitation analysis to confirm RNAP II interactions ....................................... 165
Figure 3. pM92 requires viral accessory proteins to facilitate interaction with pM79 ..................... 166
Figure 4. Amino acid alignment of the UL92 protein family ........................................................... 167

Chapter V: Conclusions and Future Directions

Figure 1. Amino acid alignment of the UL79 protein family ........................................................... 189
List of Tables

Chapter I: Introduction

Table 1. HCMV gene families .............................................................................................................. 52
Table 2. CMV mutagenesis and phenotype characterization .............................................................. 53

Chapter II: Murine Cytomegalovirus protein pM79 Is a Key Regulator of Viral Late Transcription

Table 1. Primers used in quantitative PCR analysis .......................................................................... 84
Table 2. M79-dependent expression of MCMV ORFs ....................................................................... 85

Chapter III: Murine Cytomegalovirus Protein pM92 Is a Conserved Regulator of Viral Late Gene Expression

Table 1. Primers used in PCR analysis ........................................................................................ 126

Chapter IV: MCMV Late Transcription Regulator pM92 Interacts with pM79 and RNAP II During Infection

Table 1. Mass spectrometry identities for sm79/flag immunoprecipitation ............................................ 160
Table 2. Mass spectrometry identities for sm/flag92 immunoprecipitation ............................................ 161
Table 3. Mass spectrometry identities for viral proteins ................................................................ 162
Table 4. Primers used in QuickChange Mutagenesis ................................................................... 163
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIDS</td>
<td>Acquired Immunodeficiency Syndrome</td>
</tr>
<tr>
<td>ARF</td>
<td>Anthony Roger Fehr</td>
</tr>
<tr>
<td>ARF1</td>
<td>Anthony R. French</td>
</tr>
<tr>
<td>ATP</td>
<td>Adenosine Triphosphate</td>
</tr>
<tr>
<td>BAC</td>
<td>Bacterial Artificial Chromosome</td>
</tr>
<tr>
<td>CA</td>
<td>Christopher Affolter</td>
</tr>
<tr>
<td>CCMV</td>
<td>Chimpanzee Cytomegalovirus</td>
</tr>
<tr>
<td>co-IP</td>
<td>Co-Immunoprecipitation</td>
</tr>
<tr>
<td>CPE</td>
<td>Cytopathic Effect</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>DY</td>
<td>Dong Yu</td>
</tr>
<tr>
<td>EBV</td>
<td>Epstein-Barr Virus</td>
</tr>
<tr>
<td>FLP</td>
<td>Flippase</td>
</tr>
<tr>
<td>FRT</td>
<td>Flippase Recombination Target</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescent Protein</td>
</tr>
<tr>
<td>HCMV</td>
<td>Human Cytomegalovirus</td>
</tr>
<tr>
<td>HDAC</td>
<td>Histone Deacetylase</td>
</tr>
<tr>
<td>HFF</td>
<td>Human Foreskin Fibroblasts</td>
</tr>
<tr>
<td>HHV-6</td>
<td>Human Herpes Virus 6</td>
</tr>
<tr>
<td>HHV-7</td>
<td>Human Herpes Virus 7</td>
</tr>
<tr>
<td>HIV</td>
<td>Human Immunodeficiency Virus</td>
</tr>
<tr>
<td>HSV-1</td>
<td>Herpes Simplex Virus 1</td>
</tr>
<tr>
<td>HSV-2</td>
<td>Herpes Simplex Virus 2</td>
</tr>
<tr>
<td>IE</td>
<td>Immediate Early</td>
</tr>
<tr>
<td>IRES</td>
<td>Internal Ribosome Entry Site</td>
</tr>
<tr>
<td>KSHV</td>
<td>Kaposi’s Sarcoma-associated Virus</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>LSJ</td>
<td>L. Steve Johnson</td>
</tr>
<tr>
<td>mC-BP</td>
<td>Minor Capsid Binding Protein</td>
</tr>
<tr>
<td>MCMV</td>
<td>Murine Cytomegalovirus</td>
</tr>
<tr>
<td>MCP</td>
<td>Major Capsid Protein</td>
</tr>
<tr>
<td>mCP</td>
<td>Minor Capsid Protein</td>
</tr>
<tr>
<td>MCV</td>
<td>Mark C. Valentine</td>
</tr>
<tr>
<td>MEF</td>
<td>Murine Embryonic Fibroblasts</td>
</tr>
<tr>
<td>MHV68</td>
<td>Murine Herpesvirus 68</td>
</tr>
<tr>
<td>MIEP</td>
<td>Major Immediate Early Promoter</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of Infection</td>
</tr>
<tr>
<td>ORF</td>
<td>Open Reading Frame</td>
</tr>
<tr>
<td>PAA</td>
<td>Phosphonoacetic Acid</td>
</tr>
<tr>
<td>PAGE</td>
<td>Polyacrylamide Gel Electrophoresis</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>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>PFGE</td>
<td>Pulse Field Gel Electrophoresis</td>
</tr>
<tr>
<td>PML</td>
<td>Promyelocytic Leukemia Protein</td>
</tr>
<tr>
<td>RCMV</td>
<td>Rat Cytomegalovirus</td>
</tr>
<tr>
<td>RNA</td>
<td>Ribonucleic Acid</td>
</tr>
<tr>
<td>RNAP II</td>
<td>RNA Polymerase II</td>
</tr>
<tr>
<td>RPA</td>
<td>Replication Protein A</td>
</tr>
<tr>
<td>RRE</td>
<td>RTA Responsive Element</td>
</tr>
<tr>
<td>RTA</td>
<td>Replication and Transcription Activator</td>
</tr>
<tr>
<td>RT-qPCR</td>
<td>Reverse Transcriptase-Quantitative PCR</td>
</tr>
<tr>
<td>SCP</td>
<td>Smallest Capsid Protein</td>
</tr>
<tr>
<td>SDS</td>
<td>Sodium Dodecyl Sulfate</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>-------------</td>
</tr>
<tr>
<td>TAF</td>
<td>TBP associated Factor</td>
</tr>
<tr>
<td>TBP</td>
<td>TATA Binding Protein</td>
</tr>
<tr>
<td>TFII</td>
<td>Transcription Factor II</td>
</tr>
<tr>
<td>TJC</td>
<td>Travis James Chapa</td>
</tr>
<tr>
<td>TORCH</td>
<td>Several Vertically Transmitted Infections Including Toxoplasmosis, Other Infections, Rubella, Cytomegalovirus, and Herpes Simplex Virus 2)</td>
</tr>
<tr>
<td>UL</td>
<td>Unique Long</td>
</tr>
<tr>
<td>US</td>
<td>Unique Short</td>
</tr>
<tr>
<td>UV</td>
<td>Ultra Violet Light</td>
</tr>
<tr>
<td>VHS</td>
<td>Virion Host Shut-off</td>
</tr>
<tr>
<td>VZV</td>
<td>Varicella Zoster virus</td>
</tr>
<tr>
<td>WMY</td>
<td>Wayne M. Yokoyama</td>
</tr>
<tr>
<td>YCP</td>
<td>Yi-Chieh Perng</td>
</tr>
</tbody>
</table>
Acknowledgements

First and foremost I would like to thank my thesis committee. The many thoughtful suggestions they have had for my project over the years have really shaped my research and challenged me to think about controls and validations at an advanced level. Their guidance in this project has really made my thesis what it is today. Particularly, I would like to thank former member Dr. Herbert “skip” Virgin for taking the time to help me develop as both a scientist and a professional. He took the time to talk to me about the importance of demeanor and communication in success. Without his guidance I could have sabotaged my own career. I would also like to pay particular thanks to Dr. Henry Huang who spent many hours outside of committee meetings thinking about my project. Henry was always available for a conversation, and always had invaluable advice. I was lucky to have my committee, who all positively contributed to my doctoral journey in innumerable ways. I am especially indebted to my former advisor, Dr. Dong Yu, for seeing the potential in a very unpolished young graduate student like myself and allowing me to join his research group. Dong set a high standard for quality, and devoted considerable time and energy to graduate student training. Dong taught me everything I know about writing, public speaking, and critical thinking. The successful completion of my thesis would not have been possible without his exceptional mentorship throughout the years. I am also grateful to Dr. Anthony French who did not hesitate to rescue my career in a time of turmoil. Without Anthony’s compassion and guidance, my doctoral journey would have been prolonged and made difficult; with the very real possibility of a negative outcome. Thus, I am especially thankful to Anthony and his research group for accommodating me for the passed year. I am also appreciative for the Washington University community for providing an excellent environment in which to do science and receive training. My thesis project would not have been possible without the expertise of collaborators within the department, or the core facilities supported by the university. Finally I would like to acknowledge the friends and family that have brought balance to my life outside of science. Most important on that list is my fiancé Stephanie Yurash. Stephanie has been my biggest supporter for the last seven years, and sometimes I think she values my dreams and goals more than her own. I am truly grateful to everyone who has had a positive impact on my life and career.
Human Cytomegalovirus (HCMV) is a Beta-herpesvirus that causes severe disease in immunocompromised individuals (including AIDS patients), and is the leading viral cause of congenital birth defects. Murine Cytomegalovirus (MCMV) is the primary surrogate model for HCMV, and resembles the human virus with respect to virion structure, genome organization, gene expression, tissue tropism, and clinical manifestations. In my graduate studies, I discovered the function of two novel MCMV proteins during \textit{in vitro} infection. I demonstrated that ORF M79 encoded protein pM79, and ORF M92 encoded protein pM92. Both pM79 and pM92 have homologs in HCMV, and I showed that they regulate late viral gene transcription. During infection, a mutant virus for either M79 or M92 accumulated representative viral immediate early gene products, early gene products, and viral DNA sufficiently but had severe reduction in the accumulation of late gene products, thus unable to produce infectious progeny. Analysis of the viral transcriptome via tiled array and quantitative PCR analysis revealed that many late transcripts sensitive to a viral DNA synthesis inhibitor (phosphonoacetic acid) were markedly reduced by pM79 or pM92 mutation. Co-immunoprecipitation and mass spectrometry analysis revealed the interactions between pM92 and pM79, and suggests that they are part of a larger virus transcription complex. A colleague demonstrated that this function was conserved in the HCMV homolog pUL79, and I showed that pM92 expression \textit{in trans} could complement the growth defect of pUL92 deficient HCMV. My work was the first evidence of this regulatory complex in MCMV. This complex represents a potential new
target for therapeutic intervention in CMV disease, and a gateway into studying a largely uncharted viral process that is critical to the viral life cycle.
Chapter I

Introduction

The final draft of this chapter was written and edited by TJC.
Herpesviruses

Herpesviruses are best known for causing lesions that form in the mouth and lips called "cold sores". Herpesviruses are also famous for causing chickenpox, and the sexually transmitted infection Genital Herpes. These three annoying, but relatively mild diseases are caused by herpes simplex virus type 1 (HSV-1), varicella-zoster virus (VZV), and herpes simplex virus type 2 (HSV-2), respectively (57, 174). The herpesvirus family consists of a group of over 100 viruses that are categorized by their large double-stranded DNA genomes (~130-250 kbp) and the common architecture of their infectious particles. Herpesviruses are further grouped into three subfamilies: alpha, beta, and gamma. These subfamilies are based on biological properties such as growth characteristics, cell tropism, and gene conservation. Though members of the herpesvirus family can infect a large variety of animal species including mammals, birds, reptiles, fish, and oysters, there are nine herpesviruses that infect humans (242). These nine human viruses span all three herpesvirus subfamilies. In the alpha subfamily the human viruses are the rapidly replicating, neurotropic HSV-1, HSV-2, and VZV. In the beta subfamily the human viruses are human cytomegalovirus (HCMV), human herpes virus 6a and b (HHV-6a and HHV-6b), and HHV-7, which are slow replicating viruses with a broad cell tropism. In the gamma subfamily the human viruses are lymphotropic and include Epstein-Barr virus (EBV) and Kaposi’s sarcoma-associated herpesvirus (KSHV), which are associated with Burkitt’s lymphoma and Kaposi sarcoma, respectively (50, 241).

One of the hallmark characteristics of herpesviruses is their ability to establish latent infections (13, 106, 150, 151, 216, 260, 303, 324). After an initial infection, viral DNA is harbored in a state of latency with minimal to no transcriptional activity. Each subfamily has a different location for latency; alphaherpesviruses are latent in trigeminal ganglia, betaherpesviruses are latent in progenitor cells of the bone marrow and monocytes of the blood, and gammaherpesviruses are latent in lymphocytes. Individuals infected with latent virus can remain without symptoms for months or years, or even their entire life. This is especially true for those persons that do not suffer from any long term immuno-suppression. However, reactivation of latent virus, where the latent viral DNA becomes transcriptionally active, can lead to recurrent disease (ex. recurring genital herpes or shingles). People who are immuno-compromised are at a higher risk for suffering reactivation and more severe diseases (ex. retinitis or sarcoma) (64, 256). The mechanisms by which herpesviruses establish latent infections and then reactivate are not fully understood,
but are under intense study (218).

The large number of herpesviruses and their complexity make it unwieldy to consider more than one virus in any detail in this work. Though there will be references to other herpesviruses throughout the remaining text, this work will mainly concentrate on a lineage of the betaherpesvirus subfamily, cytomegaloviruses.

**Betaherpesviruses**

Viruses in the betaherpesvirus subfamily are distinct from those in the alpha or gamma subfamilies because they replicate more slowly and are highly species specific for infection, not being able to cause infection beyond one specific host species. There are two major lineages in the betaherpesvirus subfamily, the cytomegaloviruses and the roseoloviruses. The best characterized members of these lineages are human cytomegalovirus (HCMV) and human herpes virus 6A (HHV-6A), respectively. Though the genomes of roseoloviruses are smaller (159-170 kbp) than those of cytomegaloviruses (248 kbp), the structure of the genomes and the functions of many of the encoded genes are conserved between the two lineages. There are 43 core genes that are thought to be inherited from the common herpesvirus ancestor that are encoded by most herpesviruses. HCMV and HHV-6 contain 40 and 41, respectively, of the 43 core genes. HCMV and HHV-6 lack homologues to two proteins that are well characterized in alphaherpesviruses, thymidine kinase and ribonucleotide reductase. However, there are four genes that are common to HCMV and HHV-6 that have homologues in the gammaherpesvirus subfamily but not the alphaherpesvirus subfamily (HCMV designation: UL49, UL79, UL87, and UL92) (16, 51, 52, 70, 135, 317, 321). This is not surprising as it has been shown that beta- and gammaherpesviruses are more similar to one another than either subfamily is similar to alphaherpesviruses, especially with regard to transcription (7, 52). In addition, there are 27 genes that are specific to the betaherpesvirus subfamily; 22 of which are conserved by sequence between HCMV and HHV-6. These include genes involved in tropism and inhibiting apoptosis, something that is important for viruses with long replication cycles (206).

The thesis which is the subject of this work focuses on the cytomegalovirus lineage of the betaherpesvirus subfamily. Thus, it is important to describe this genus in further detail.
Cytomegalovirus

Cytomegaloviruses are present in a wide range of mammalian species, including primates and rodents (242). Though chimpanzee cytomegalovirus has the largest genome of all sequenced herpesviruses, it is human cytomegalovirus (HCMV) that has the largest genome of human viruses. HCMV is the best studied cytomegalovirus, and the prototype of the betaherpesvirus subfamily for studies of the virus life cycle, especially gene expression and regulation. However, studies have been prolonged because lab adapted strains of HCMV (including AD169) have been found to contain mutations that have important functional consequences (6, 48, 53, 71, 80, 267). For example, the AD169 strain has mutations that limit the cell tropism of the virus. Wild-type HCMV replicates in macrophages, dendritic cells, colonic and retinal pigmented epithelial cells, endothelial cells, fibroblasts, smooth muscle cells, neuronal cells, glial cells, hepatocytes, and trophoblasts (89, 90, 115, 124, 163, 182, 232, 254, 261, 262, 264, 265). In contrast, the lab adapted strain AD169 preferentially infects and replicates in fibroblasts. Despite the differences between wild-type and lab adapted strains of HCMV, a lot of progress has been made in understanding the structure and viral life cycle of betaherpesviruses.

Pathogenesis

HCMV is a widespread pathogen that infects a majority of the world's population by early adulthood. In fact, by the age of 40, between 50% and 85% of adults are infected by HCMV as indicated by the presence of antibodies in much of the general population (82, 208, 257). Seroprevalence is age-dependent, as 58.9% of individuals aged 6 and older are seropositive for HCMV while 90.8% of individuals aged 80 and older are seropositive for HCMV (280). HCMV infection is typically asymptomatic in healthy individuals but can cause life-threatening disease for the immuno-compromised, such as HIV-infected persons, organ transplant recipients, or new born infants (36, 64, 82, 94, 96, 256). For immuno-compromised individuals, symptoms consist of spiking fever, leucopenia (decrease in white blood cells), malaise, hepatitis, pneumonia, gastrointestinal disease and/or retinitis (inflammation of the retina) (270). HCMV is also responsible for approximately 8% of infectious mononucleosis cases (105). HCMV represents the most significant viral cause of congenital birth defects in industrialized countries, and was estimated in the 1990’s to cost the United States health care system ~$2 billion annually (8, 17, 105).
The birth defects can range from deafness and mental retardation to fatal diseases that include encephalitis, pneumonia, hepatitis, and hydrops fetalis (a condition in the fetus that leads to heart failure) (147, 248). HCMV has been implicated in playing a role in inflammatory and proliferative diseases, including atherosclerosis and cancers (184, 205, 259, 270). The severity of medical problems associated with HCMV in these vulnerable populations underlies the necessity for developing safer and more effective antiviral treatments and vaccine strategies to control its infection.

Antivirals. There are no vaccines for any of the betaherpesviruses, and the available antiviral therapies have limited efficacy and high rates of adverse effects. The clinical use of drugs to treat viral diseases in humans is difficult due to the fact that viruses replicate within the host cells, using many elements of the host cell machinery. Therefore, there are fewer obvious targets for chemotherapy against viruses than for bacterial or parasitic pathogens. A better understanding of betaherpesvirus replication and pathogenesis is needed to delineate the functions of proteins involved in key viral processes. Ideally drugs would be designed to target very early steps in infection such as attachment and entry, or very late steps in infection like release. Inhibitors of early and late steps of infection do not have to enter the cell to exert their effect, and therefore may be less toxic. Unfortunately, no drugs currently exist for such steps of the herpesvirus life cycle.

The first antiviral drugs developed, and the most effective drugs for herpesviruses to date, are the nucleoside analogues acyclovir and ganciclovir (225, 302). Nucleoside analogues mimic the natural substrate of the viral DNA polymerase and act as competitive inhibitors to inhibit viral DNA synthesis. Acyclovir is primarily used in alphaherpesvirus infections, while ganciclovir has been shown to inhibit CMV replication, prevent CMV disease in AIDS patients and transplant recipients, and reduce the severity of some CMV syndromes, such as retinitis and gastrointestinal disease (225, 302). Ganciclovir treatment is made more effective when combined with immunoglobulin treatment, which reduces mortality for bone marrow transplant patients with HCMV caused pneumonia; but long-term survival of these patients is still very poor (56). Another drug approved for treating HCMV disease is foscarnet. Similar to ganciclovir, foscarnet inhibits the HCMV DNA polymerase (22, 301). However, foscarnet’s toxic effects are primarily renal and thought to be less severe than ganciclovir which can inhibit bone marrow function. The most
recently approved drug is the nucleotide analog cidofovir, which was approved for therapy of retinitis (22, 159). The side effects for cidofovir are nephrotoxic. Toxic effects aside, all herpesvirus drugs are proving ineffective due to the emergence of resistant virus (22, 95, 201).

Viral life cycle steps such as genome replication are great targets for antivirals because the process requires specific viral enzymes. Enzymes make good drug targets because they are usually at low concentrations inside cells, they can be well-understood mechanistically, and they normally interact with small-molecule substrates. Unfortunately, targeting one enzyme, like DNA polymerase, allows for escape mutations that confer resistance to antivirals (as mentioned above). Thus, new drug targets are always being sought after. One promising target is the cleavage of the viral genome and its packing into capsids, which is mediated by the viral terminase. This virus-specific enzymatic process appears to be especially vulnerable to antiviral inhibitors, and consequently, several substances have been described that interfere with the function of the terminase (43, 123, 195, 295). The drug AIC246 is in phase 2 clinical trials (102, 170).

An effective vaccine for HCMV has still not been found. As a member of the TORCH complex (several vertically transmitted infections including Toxoplasmosis, Other infections, Rubella, Cytomegalovirus, and Herpes Simplex virus 2), cytomegalovirus is a priority for vaccine development by the National Vaccine Program Office in the United States (17). The last vaccine candidate in 2009 had an efficacy of 50%, thus the protection was limited and many subjects contracted HCMV infection despite the vaccination (224). A better understanding of betaherpesvirus replication and pathogenesis is needed for developing novel strategies to prevent disease to this important pathogen.

**Virus structure**

Like all herpesviruses, HCMV has a highly ordered icosahedral-shape nucleocapsid of about 130 nm in diameter, which encases the viral DNA genome. The capsid is enclosed in a polymorphic lipid bilayer envelope containing multiple viral glycoproteins that are responsible for viral attachment and entry into host cells. Unlike most enveloped viruses, herpesviruses have a large space between the capsid and the envelope called the tegument. The tegument is a partially ordered proteinaceous layer that contains about 30 virus-coded proteins (138, 291). Tegument proteins are introduced into the cell upon viral entry and carry
out crucial functions during the early stages of HCMV infection. The following sections will describe each of
the viral structural elements in detail.

**Envelope.** The viral envelope is a lipid bilayer that contains both viral glycoproteins, and host membrane
proteins. Herpesvirus envelopment is believed to take place initially at the inner nuclear membrane, and
then further proceeds to acquire membranes from endosomes as well as the Golgi network (85, 100, 275).
Thus, the viral envelope contains lipid components that are associated with both the nuclear membrane and
the cytoplasmic membrane system. Whether these lipid components are important in maintaining the
integrity of the viral envelope has not been determined and their functional role in stabilizing virion structure
is unknown. A few examples of host proteins associated with the HCMV envelope include
β2-microglobulin, CD55 and CD59, and annexin II (107, 109, 194, 319). It’s thought that these molecules
may participate in the induction of host cellular responses and/or play a role in modulating virus attachment
(24, 59, 330).

In addition to host factors, viral envelopes contain virus-coded glycoproteins. Herpesviruses
encode a set of 20-80 glycoproteins, very few of which are conserved between all the Herpesviruses (203,
241, 242). HCMV encodes 57 putative glycoproteins, which is far more than other herpesviruses. The exact
organization of the glycoproteins in the envelope is not completely understood. It is known that some of the
glycoproteins are found to aggregate into three complexes: homodimers of glycoprotein B (gB; encoded by
UL55); heterodimer of glycoprotein M (gM; UL100) and glycoprotein N (gN; UL73); and a heterotrimer of
glycoprotein H (gH; UL75), glycoprotein L (gL; UL115), and glycoprotein O (gO; UL74) (121, 122, 180, 238).
gB is a type I integral membrane protein, which mediates a two step cell attachment starting with binding to
heparin sulfate proteoglycans, followed by a stabilizing interaction with another unknown receptor (32, 63).
Little is known about the function of gM (a type III membrane protein) or gN (a type I membrane protein), but
the gM/gN complex is the most abundant component of the viral envelope, and is essential for virus
replication (180, 181, 298). The gHgLgO heterotrimer is involved in mediating fusion of viral and host cell
membranes (121, 141, 274). The function of these glycoprotein complexes will be detailed in a later section
(the “Entry” subsection of the “Life Cycle” section).
**Capsid.** Cryo-microscopy and image reconstruction has been used to determine the structure of the HCMV capsid (65, 66, 327). The architecture of the herpesvirus capsid consists of a hexamer-pentamer cluster arranged in an icosadeltahedral lattice (44). The major structural subunit of the capsid is the caposomere, of which there are two types in the CMV capsid. The penton caposomeres form the 12 vertices of the icosadeltahedral, and the hexon caposomeres form the 150 subunits of the icosadeltahedral faces. The CMV capsid consists of four structural proteins: the major capsid protein (MCP; encoded by UL86), the minor capsid protein (mCP; UL85), the minor capsid protein-binding protein (mC-BP; UL46) and the smallest capsid protein (SCP; UL48A) (99-101, 127, 255). The penton and hexon caposomeres are composed of five or six copies of the MCP, respectively. The hexon caposomeres are further decorated with six copies of SCP (326). The caposomeres are linked together by mCP and mC-BP triplexes that stabilize the capsid structure (66, 215, 293).

**Tegument.** The tegument of herpesviruses occupies the space between the capsid and the envelope (138, 327). Since the capsid is ~125nm in diameter and the entire virion is ~220nm in diameter, the tegument represents a significant part of the virion space; containing approximately 40% of the herpesvirus virion protein mass (100). Many tegument proteins play roles in the early stages of virus infection, which is why the virus has evolved to package them within the virion for infection of a new host (208). Mass spectrometry analyses have been done to determine the protein content of the tegument. At least 30 virus-encoded proteins have been found in the HCMV tegument (100, 203). The five most abundant proteins are: the lower matrix protein pp65 (UL83); the basic phosphoprotein pp150 (encoded by UL32); the upper matrix protein pp71 (UL82); the membrane-associated myristylated protein pp28 (UL99); the high molecular weight tegument protein UL48, and its binding partner UL47 (19, 100, 203, 251).

pp65 is the most abundant tegument protein and accounts for more than 15% of the virion protein mass (100). pp65 has been reported to inhibit presentation of viral proteins by the major histocompatibility complex class I, and may inhibit the induction of host interferon responses (38). Furthermore, pp65 forms a large portion of the protein mass of the noninfectious B capsids and dense bodies (45% and 90%, respectively), which may be required to help the replicating virus evade immune surveillance (38, 100, 144). pp150 and UL48 are both essential for viral replication, and have been proposed to interact intimately with
nucleocapsids (83, 200). Blocking pp150 causes defects in tegumentation, and results in the accumulation of nucleocapsids around the centrosome of host cells (200). pp71 is also believed to be involved in direct interaction with the newly synthesized nucleocapsid, and is important for initiation of tegument assembly (292). Moreover, pp71 protein is a transcriptional activator that helps to induce the expression of the immediate-Early genes within the infected cells (172). pp28 has a wide array of functions, but is primarily responsible for the cytoplasmic envelopment of tegument proteins and capsids in HCMV during the assembly and egress process (138, 258). The function of UL47 is still unclear.

The requirement to incorporate specific proteins suggests that tegument morphogenesis may involve specific interactions between individual components, leading to the formation of a structure that is at least partially ordered. This hypothesis is supported by the fact that HCMV infections in cell culture give rise to the production of many particles composed of enveloped tegument lacking a capsid (dense bodies), suggesting that the tegument can spontaneously self-assemble (125). Cryo electron tomography has shown that the HCMV tegument can be divided into two sub-compartments: an inner and an outer tegument. The inner tegument consists of densely-packed proteins surrounding the capsid and the outer tegument contains components that are loosely packed between the inner tegument and the envelope (327). Moreover, immune-gold labeling experiments on thin sections of virus-infected cells can reproducibly locate tegument proteins to the outer or inner regions of the tegument (160, 161).

**Genome.** All herpesviruses contain a double-stranded linear DNA genome, which is between 130 and 250 kbp in length depending on the virus species. The size of the HCMV genome is 230kb, which is 51% longer than the HSV-1 genome, and as a result the HCMV capsid is 117% larger than HSV-1 (71, 193). Original estimates of the HCMV genome coding capacity was 165 open reading frames (ORFs). However, it has since been shown that HCMV may encode as many as 751 ORFs, making it one of the most complex pathogens to infect humans (285).

HCMV, and all members of the betaherpesvirus subfamily, have a genome consisting of two unique regions flanked by direct repeats. For cytomegaloviruses, the two unique regions are designated unique long (UL) and unique short (US) (Fig. 1) (208). There are two sets of repeats in the HCMV genome; the terminal direct repeats of approximately 300–600 bp, and internal repeats of the same length, which are a
duplication of the terminal repeats in an inverted orientation at the junction between the UL and Us segments (272, 273, 286). Due to these repeat sequences, recombination takes place during viral DNA replication to generate four isomers of the genome. Each isomer accounts for ¼ of the DNA packaged into virions, and each isomer has a different relative orientation of the UL and US segments. Interestingly, DNA of any one of the isomers is infectious, and limiting the ability to undergo recombination by removing the repeat sequences also results in infectious virus (112, 253). Gene duplication has been employed widely by large eukaryotic DNA viruses and their hosts as a means of generating diversity (236). There are 13 gene families in HCMV that are presumed to have arisen by gene duplication (Table 1) (70).

The genomes of other cytomegaloviruses show variations on the theme of interspersed unique and repeated DNA sequences. MCMV and RCMV have small terminal direct repeats of 30 and 504 bp, respectively, but no internal repeat sequences (187, 299). Higher primates such as CCMV, have two unique regions (UL and US) which are flanked by direct repeats (similar to HCMV) (71, 308). The biological reason for why cytomegaloviruses maintain their genome structure remains unknown.

Life cycle

All viruses must deliver their genomes to host cells to initiate infection. For enveloped viruses like HCMV, entry requires the use of virion envelope proteins to facilitate adherence to the cell surface and fusion between the virus envelope and the cellular membrane. For HCMV, fusion occurs both at the cell surface and in endosomes. Regardless of which entry location is used, fusion results in the deposition of virion components into the cytoplasm that facilitate transport of the capsid to the nucleus. The viral genome is then delivered to the nucleus through a nuclear pore in a process known as uncoating. The genome is then used for viral transcription, which ultimately produces the proteins responsible for inhibiting the immune response and high jacking the cellular machinery required for viral genome replication. At late times of infection, proteins involved in capsid assembly and subsequent packaging of the viral genome are expressed. Once assembled, the capsids travel from the nucleus to the Golgi—acquiring the viral tegument and envelope in the process—and then are released from the cell (208). This entire process will be described in detail in the following sub-sections (Fig. 2).
**Tropism.** To begin a discussion of the virus life cycle at the cellular level, one must first consider the basis of cellular tropism since receptors involved in entry are expressed on permissive cells. In the human host, HCMV causes systemic infection and exhibits a tropism for fibroblasts, endothelial cells, epithelial cells, monocytes/macrophages, smooth muscle cells, stromal cells, neuronal cells, neutrophils, and hepatocytes (213, 263). This exceptionally broad cellular tropism in the infected host is why HCMV disease manifests in a variety of organs and tissue types in the host. However, in vitro, HCMV has a restricted cell tropism. Though entry into many target cells is possible, productive HCMV infection in vitro is only supported by primary fibroblast, endothelial cells, and certain differentiated myeloid cells (61, 124, 220). The exact cellular receptors required for entry is unknown, and thus the reason for restricted tropism in vitro is still a mystery.

**Entry.** HCMV initially binds to heparin sulfate proteoglycans at the host cell surface via gB to initiate entry (61). HCMV engagement of heparin sulfate proteoglycans is thought to enhance binding of subsequent receptors on the cell surface that ultimately lead to fusion in a multi-step cascade (45, 60, 139). Studies for finding the exact HCMV receptor are complicated by the fact that HCMV has such a broad cell tropism and that no group of cellular receptors have been shown to be necessary on all infectable cell types. β-microglobulin, major histocompatibility complex class I molecules, annexin II, aminopeptidase N (CD13), and epidermal growth factor have been suggested to be virus receptors, but all have since been shown to be nonessential for virus infection and not present on all HCMV susceptible cells (20, 21, 40, 87, 107-109, 194, 220, 230, 233, 279, 290, 305, 319). Regardless of what receptor is used, receptor binding triggers fusion of the viral envelope with the cellular membrane. This requires the conserved glycoprotein complex gB/gH/gL. In this complex, gH appears to serve as the actual fusion glycoprotein (276, 297, 329).

Membrane fusion remains a poorly understood component of entry for any of the herpesviruses because of the absence of a direct fusion assay. Any role that the gB/gH/gL complex has in fusion is only inferred. To that end, entry of HCMV can be blocked by gB and gH dependent neutralizing antibodies at a post-attachment stage of entry, which could be at the level of fusion (27, 35, 142, 294, 296). Integrins have also been implicated in mediating HCMV fusion (88, 304). Integrins are a ubiquitous cell surface receptor that can cause cytoskeletal reorganization. Analysis of the effects of various integrin blocking antibodies
show that integrins function at a post-attachment stage of infection, and are involved in delivering tegument protein pp65 into infected cells (88). More work needs to be done to confirm that integrins are involved in fusion. Development of a reliable fusion assay is required to confirm what factors are truly involved in HCMV fusion.

After HCMV fusion with host cells the viral capsid is released into the cytoplasm. The tegument proteins help to transport the capsid to nuclear pores, presumably by dynein motors along the microtubular network, and then the DNA is released and enters the nucleus. Empty capsids remain at the cytoplasmic side of the nuclear pores for several hours until they disintegrate (227). Viral DNA in the nuclease is circularized, either by direct ligation of the ends or by recombination between the repeat sequences at the terminal ends of the genome (208). The circularized DNA is localized near nuclear structures known as nuclear domain 10 (ND10), where early transcription takes place (5, 93)

Transcription. Herpesvirus gene expression can be classified into at least four groups based on the time of their expression during the viral replication cycle. The Immediate-Early (IE) or alpha (α) genes are transcribed during the first several hours after infection; 0-2 hours in the case of HCMV. IE gene products are required to combat innate immune defenses, and to make the cell suitable for viral genome replication. Additionally, IE gene products function to set up a regulatory cascade that leads to the properly timed transactivation of Early or beta (β) genes. Early genes are expressed between 4 and 8 hours post infection (hpi) for HCMV. The Early gene products are mainly involved in viral genome replication. It’s only after genome replication is occurring that the expression of Late or gamma (γ) genes occurs. The majority of viral genes and all lytic genes are Late genes (~51% of the gene products in HCMV) (49, 51, 158, 185). There are two classes of Late genes, designated γ1 and γ2. Genes in the γ1 group begin to be transcribed at low levels before DNA replication, but their transcription is stimulated several fold after the onset of viral DNA replication. For this reason, the genes in the γ1 class are also called Delayed Early or Early-Late genes. Expression of γ2 genes begins only after DNA replication has initiated, and these genes are often called True Late genes. Gene expression will be discussed in extensive detail in a later section (“Gene Expression” section).
Genome replication. Genome replication of herpesviruses occurs in the nucleus of infected cells. The process takes place in virally induced structures called replication compartments, where viral proteins involved in DNA replication accumulate (314). These replication compartments form at the site of the cellular ND10 nuclear bodies where the initial circular viral genome localizes after release from the incoming capsid. Studies have shown that a cell can only support ~7 replication compartments during infection (146). Since each replication compartment is derived from a single genome, the virus bottleneck has been defined as seven entering particles which will have their genome expressed in an infected cell (145). At late times of infection viral DNA levels can equal cellular DNA content, which implies that each virus genome is replicated to high levels.

Herpesviruses initiate lytic DNA replication at defined sites on the viral genome called origin of DNA/lytic replication (ori-lyt). Some herpesviruses have a single ori-lyt, such as in the betaherpesviruses (including HCMV), while others have either two or three, as in the case of EBV or HSV-1, respectively (242). The reason for multiple origins of lytic replication in the biology of these viruses remains unclear. After initiation from the ori-lyt, herpesvirus DNA replication proceeds through either of two potential mechanisms: 1) genome circularization and theta form replication, which proceeds to a rolling circle (like bacteriophage lambda), or 2) producing complicated branch structures from linear genomes that are then resolved by homologous recombination and by viral cleavage machinery (25, 26, 132, 169, 219, 314). Either mechanism of synthesis results in genome concatamers that are cleaved and packaged into newly synthesized capsids.

HCMV encodes a virion-associated transcript that associates with the ori-lyt to form a three-stranded structure, which results in targeted unwinding to enable the assembly of a replication fork complex (235). HCMV encodes a core set of six conserved DNA synthesis enzymes that direct the synthesis of viral DNA during lytic infection: DNA polymerase (encoded by UL54), the processivity factor (UL44), a single stranded DNA binding protein (UL57), and a heterotrimeric helicase-primase complex (UL70, UL102, UL105). Following ori-lyt engagement (which is still not fully understood), an interaction involving UL57 leads to localized unwinding of the DNA and recruitment of other viral replication proteins. The helicase-primase complex has DNA dependent ATPase, helicase, and primase activities. Thus, this complex unwinds DNA and produces RNA oligonucleotides to aid polymerase elongation. The polymerase
enzyme consists of UL54 and processivity factor UL44. UL54 has the main enzymatic activity of polymerizing, but also contains a 3'-5' exonuclease activity for proofreading. The UL44 protein increases the processivity of the polymerase which is essential for DNA replication (26, 313, 314). Cellular enzymes such as topoisomerases are highly likely to be required for replication; however, a complete understanding of this process will require a defined cell-free assay.

Capsid assembly. The assembly pathway of the HCMV particle is highly similar across all herpesviruses species (100, 242). The assembly of the HCMV capsid occurs in the nucleus of infected cells starting with the assembly of the procapsid, which consists of the capsid shell and the internal scaffolding structure. Second, the procapsid is filled with viral DNA genome while the scaffolding proteins are simultaneously removed (227). This results in a major conformation change of the capsid shell, converting the procapsid into a mature nucleocapsid (214, 326, 328).

In order to package its genome, HCMV requires the terminase complex, which is comprised of viral proteins UL51, UL56 and UL89 (31). The terminase interacts with both the viral DNA and the portal protein (UL104) to initiate genome encapsidation (76-78, 149). The terminase recognizes specific sequences termed packaging signals on the viral genomes, docks at the portal vertex of the capsid, and by ATP hydrolysis provides the energy necessary for genome insertion, followed by cutting of the genome concatamer after exactly one genome length is packaged. However, the precise mechanism for how concatameric HCMV DNA is resolved into unit-length genomes that undergo packaging is not completely understood.

Maturation and envelopment. The HCMV envelope is acquired in the cytoplasm from Golgi membranes and secretory vesicles. However, the HCMV nucleocapsids is assembled in the nucleus, and must first escape this compartment before acquire its envelope. Since the nuclear pores are thought to be too small to allow escape of the HCMV capsid, it is generally believed that HCMV uses the nuclear membrane in an envelope/de-envelope mechanism to escape from the nucleus (126, 127, 227). In this model, HCMV buds through the inner nuclear membrane into the lumen between the inner and outer nuclear membranes, acquiring an envelope and a layer of tegument proteins. The nucleocapsid then loses the nuclear envelope
by fusion with the outer nuclear membrane, which releases the nucleocapsids into the cytoplasm with the acquired tegument proteins. These nucleocapsids subsequently reacquire an envelope by budding into Golgi membranes located in the viral assembly compartment (9). The viral assembly compartment results from extensive reorganization of the cellular secretory apparatus (67-69). As a result, the viral assembly compartment contains many cellular and viral components (including early endosomes, Golgi, viral structural proteins, and viral tegument proteins) that are necessary for virion maturation. The final step in egress is fusion of an exocytic vesicle with the plasma membrane, a process that is likely to follow cellular vesicle trafficking pathways (208).

**Latency.** A critical component for HCMV persistence in the non-immune compromised host is the ability of the virus to establish cellular sites of latency. Viral latency can be operationally defined as the maintenance of the viral genome in the absence of production of infectious virions but with the ability of the viral genome to reactivate under certain conditions. During latency, the HCMV genome is maintained within the cell with limited viral gene expression and little to no genome replication. The virus can be reactivated upon cellular stimulation, including radiation, UV, and other stress responses. Such reactivation from latency is thought to occur routinely in healthy virus carriers, but is limited by the host immune response. On the other hand, reactivation of HCMV in an immuno-compromised or immuno-suppressed host background is a well-established cause of morbidity and mortality (81, 244). Cytomegaloviruses latency occurs in CD34+ progenitor myeloid cells, but stay with the differentiated monocyte-macrophage lineage (152, 266). It is unclear why other lineages which arise from CD34+ progenitors, such as T and B cells, do not carry viral genomes or why even some cells of the myeloid lineage itself, such as polymorphonuclear cells, do not carry viral genomes (288, 289). Interestingly, the number of cells carrying viral genomes during latency is very small, estimated to be $10^4$–$10^5$ peripheral blood mononuclear cells or bone marrow cells (268). Within these carrier cells, the viral genomes are maintained as episomes (29). How the episome is maintained, and why lytic infection does not occur is not completely understood. What is known is that myeloid cell differentiation is crucial for virus reactivation and that reactivation is mediated by the expression of IE genes. Analysis of the major IE enhancer promoter (MIEP) suggests that DNA sequences within the MIEP not only interact with cellular transcription factors which activate the IE promoter but also
with cellular transcriptional repressors like chromatin (18, 98, 120, 178, 197, 247, 318). It is now well established that the regulation of promoter activity of many cellular genes involves regulation at the level of their chromatin structure, and that the chromatin profile during cell differentiation is such that there are more active promoters (84, 157, 179, 306). The same correlation between chromatin structure of the MIEP and virus reactivation has also been observed in models of experimental latency (240). Changes in chromatin structure of IE gene promoters likely control reactivation as seen with other herpesviruses (11, 15, 134, 156). Though it is clear that the absence of the IE protein is required for latency, however, both spliced and unspliced RNAs, termed cytomegalovirus latency-specific transcripts (CLTs) (153). Though some of them are increased or decreased during monocyte differentiation, when latent virus reactivates, no role for them has been confirmed in latency (103, 310). Understanding the relationships between HCMV and myeloid cells will be important for a full understanding of how this virus persists in the host and the complexities of its interaction with the host immune system. A thorough understanding of the molecular biology and immunology of HCMV persistence in vivo can only help to generate specific strategies that more effectively control virus reactivation.

**Early gene expression**

A viral process that was briefly described earlier is gene expression. Since this topic is the focus of this dissertation, much more elaboration is required. Below you will find details of Immediate Early and Early gene regulation, but the section will extensively focus on Late gene expression and regulation.

**Immediate Early genes.** A major function of the IE genes is to set up a regulatory cascade that leads to properly timed expression of other viral genes. The IE genes are the first viral genes transcribed after infection, and their transcription does not require de novo viral protein synthesis. These gene products optimize the cell for viral gene expression and replication. Downstream of the MIEP are two IE genes designated IE1 and IE2 in HCMV. In MCMV these genes are designated ie1 and ie3, respectively. Multiple proteins are encoded by these two IE genes through differential mRNA splicing throughout infection (217, 282-284). The resulting viral proteins are designated according to their apparent molecular weight; IE1 = IE72, IE2 = IE86. The minor products are designated IE38, IE55, and IE18, and have received less study.
and attention than IE72 and IE86. During the two hours after HCMV infection, the mRNA for the IE genes are expressed abundantly, but as the protein products are made, the IE72 and IE86 proteins negatively regulate the MIEP and decrease transcription of IE genes (14, 54, 162, 231, 278). After synthesis in the cytoplasm, the IE proteins are transported to the nucleus and targeted to ND10 nuclear bodies (mentioned earlier as the place where viral replication compartments eventually form). IE72 is involved in alleviating the repressive effects of ND10 on the viral genome presumably by binding to its associated proteins, such as PML, SP100, and hDaxx (3, 4, 166, 168, 315). Once the ND10 repressive effects are alleviated, the basal transcription initiation complex is activated and viral Early gene transcription can take place (128, 209, 287). It is unclear how the IE72 protein activates promoters, but an association of IE72 with TATA box-associated factors (TAFs) and transcription factors (Sp-1, E2F-1, CTF-1) has been proposed (176, 186). The mechanism by which the IE72 protein activates promoters may also be related to inhibition of HDAC-2 activity (222, 287). Interestingly, the IE72 gene is only required for efficient HCMV replication at low multiplicity of Infection (MOI). Deletion of the IE72 encoding gene, IE1, results in reduced virus replication due to insufficient levels of Early gene products, which results in low efficiency genome replication (97, 104, 204). However, at high MOI viral replication levels of IE1 mutant virus are similar to wild-type virus. This is likely due to virion-associated proteins present in infectious and non-infectious particles that compensate for the absence of functional IE72. The same phenotype is true for mutants of the homologue ie1 in MCMV infections.

On the other hand, the IE2 gene product, IE86, and its functional homologue in MCMV (ie3) are absolutely essential for virus replication (12, 183). Recombinant viruses with deletion of the HCMV IE2 gene or the MCMV ie3 gene are unable to activate Early viral gene expression, which abrogates viral DNA synthesis and Late gene expression. IE86 protein is considered a master regulator of productive virus infection. It regulates activation of transcription from viral and cellular promoters, negatively auto-regulates the MIEP, and induces cell cycle progression (118, 208, 309). Though IE86 localizes adjacent to ND10 where the viral DNA is located, IE86 does not interact with ND10 (128). Instead IE86 promotes gene expression by directly activating promoters. IE86 interacts with a wide variety of cellular transcription factors (TBP, TFIIB, TAF4, and histone acetyl-transferase) to link various transcription regulators to the viral transcription complex (42, 47, 93, 111, 176, 177). IE86 also regulates gene expression by interacting with
regulators of cell cycle progression. Replication of HCMV appears to be best in a cell that progresses to the G/S transition point, but is prevented from entering the S phase (46, 92, 110, 212, 271, 311, 312).

Though not well studied, minor Immediate Early proteins also have functions important to HCMV replication. Immediate Early proteins TRS1 and IRS1 were initially recognized as regulatory proteins working in conjunction with IE1 and IE2 in activation of delayed Early and Late gene expression (129, 281). The mechanism of their action is thought to be related to the interferon response (23, 55). HCMV genes UL36 and UL37 are also IE genes that encode cell death suppressors (188-190). These proteins inhibit apoptosis in infected cells, but are both dispensable for replication.

**Early genes.** Early gene products constitute a significant proportion of the essential genes for HCMV replication. Mutagenesis analysis has shown that 23 to 25 essential and augmenting genes expressed with Early or Early–Late kinetics (49). Most of the viral Early genes function either by participating in genome replication, or by creating an environment that is conducive for genome replication.

An experiment in which fragments of the HCMV genome were tested for their ability to support ori-lyt dependent DNA replication identified 11 loci required for viral genome replication (223). Of the genes encoded in these regions, six were the core set of conserved herpesvirus genes involved in DNA replication (UL44, UL54, UL57, UL70, UL102, and UL105, mentioned earlier in the “Genome replication” sub-section of the “Life Cycle” section). Each of these genes is expressed with Early kinetics (49, 269). These loci also encoded UL112–113 protein products which are important for the formation of viral replication compartments, and may play a role in the recruitment of additional factors to these sites (5, 131, 226). The IE86 proteins were also encoded by these loci, along with UL84. The product of the UL84 gene is essential for viral DNA synthesis and productive infection (83, 252, 322, 325). The UL84 protein interacts with IE86 in viral replication compartments. Though, the importance of this interaction is not known (171, 277). It is also thought that UL84 is involved in the formation of replication compartments (322). A protein that is important for viral DNA replication, but was not identified in the loci complementation experiment, is UL114. The UL114-encoded uracil-DNA glycosylase is not strictly required for growth in fibroblasts, but a mutant lacking this gene is delayed in the initiation of DNA replication (62, 234).
How Early gene expression makes host cells suitable for DNA replication is less clear. Early gene expression during infection is associated with the stimulation of host cell genes that encode proteins involved in host DNA synthesis and cell proliferation (ex. thymidine kinase and ornithine decarboxylase) (28, 39, 58, 86, 116). Along those lines, tumor suppressor protein p53 and host DNA synthesis proteins PCNA and RPA are sequestered to viral replication compartments (79, 92, 133). Early gene expression is also required to induce elevated levels of cyclin E and cyclin B and their associated kinase activities (133, 192, 246, 250). In contrast to the activation of cyclins E and B, the expression of cyclin A and its associated kinase activity is inhibited by Early proteins during infection (133). In this way the virus is able to both positively and negatively regulate host cell DNA synthesis and cell division (33, 34, 79, 133, 175). The challenge remains to determine which viral genes are involved and elucidate the mechanisms governing their activity. Given the large number of Early genes, most of which have not yet been studied, the task is not trivial.

Aside from their roles in DNA replication, some Early genes play roles at a later stage. Following synthesis, viral DNA is cleaved into genome-length segments and packaged into preformed capsids. Early genes UL89 and UL56 are involved in DNA cleavage (43, 155, 295). Four Early proteins, UL51, UL52, UL77, and UL104, are predicted to be involved in packaging cleaved DNA into progeny capsids. Recently, it has been shown that the TRS1 protein also may be involved in packaging at a step that occurs after the cleavage of the DNA (2).

Late gene expression

Our understanding of Late gene regulation in the betaherpesviruses lags far behind that of Early gene regulation. Late genes code for viral structural proteins, proteins involved in the assembly of capsids and packaging of viral DNA, and may other proteins whose functions are only beginning to be discovered. It is known that DNA replication is required for the transcription of Late genes. The γ1 genes can be transcribed at very low levels in the absence of viral DNA replication, but their expression is dramatically increased after viral DNA synthesis. In contrast, the transcription of γ2 genes is totally dependent on viral DNA synthesis, and occurs exclusively after genome replication has begun. The mechanism that links Late gene activation to DNA replication has not yet been determined. However, some of the theories include:
restriction on expression related to DNA structure; DNA modifications that promote expression; the availability of certain viral or cellular proteins that inhibit repressors or recruit transcription activators; or inefficient Late gene promoter that requires DNA abundance for efficient transcript accumulation. To that end, only a handful of HCMV Late gene promoters have been analyzed in detail. One of the primary reasons for this is that Late gene promoters are utilized with Early gene kinetics when put into an ectopic plasmid (74, 245). What DNA elements are required to recapitulate the appropriate kinetic class has been a mystery. However, it has been shown for gammaherpesviruses that Late gene regulation is only restored in ectopic plasmids if the plasmid contains the Ori-lyt element in cis (10, 73). Recently, a plasmid that contains the HCMV ori-lyt in cis with the Late promoter has been developed, and transient transfection of this plasmid does indeed express the reporter gene in the appropriate kinetic class (207). Thus, studies in the next few years are likely to provide great insight into the importance of the Late promoter sequence elements in regulation of Late gene expression.

Recent research has also identified a series of cellular and viral factors that are specifically involved in Late gene regulation during herpesvirus infection. Though the fields of alpha and gammaherpesviruses are more advanced than that of beta herpesvirus, research, including that done in this dissertation, has made progress in understanding the protein requirements for Late gene expression (16, 51, 52, 137, 317, 321). To give us a better understanding about how HCMV may be regulating Late gene expression, it is important to take into consideration what is already known.

**Promoter requirements.** Typically, Late promoters seem to require a TATA element but no further upstream sequences for transcriptional activation (75, 148, 196, 316). Similar findings have been observed for Late gene promoters of herpes simplex virus, where the TATA box and downstream elements are sufficient to direct Late gene activation (91, 119, 307). This is in contrast to the beta herpesviruses Early promoters, where specific upstream sequence elements capable of binding to cellular transcription factors are critical for promoter activation.

The most extensive studies on Late gene regulation have examined the promoter of the UL99 gene encoding the virion tegument protein pp28 (74, 143, 148). UL99 is expressed as a γ2 gene, and requires only sequences from -40 relative to the cap site to restrict expression to late times (148). Thus, it was
thought that Late gene promoters are relatively simple, consisting primarily of a TATA element and downstream sequences and lacking a requirement for multiple transcriptional regulatory elements. However, studies of other HCMV Late gene promoters suggest that regulation can be more complex (75, 196). Analysis of the UL75 promoter, which promotes expression of glycoprotein gH, sequences from −38 to +15 relative to the cap site were sufficient to activate this promoter (196). Closer examination of those sequences revealed that sequences downstream of the UL75 cap site appeared to function as a dominant regulatory element, because in the absence of this region, both activation and repressor elements were identified in the upstream sequences (196). The same type of dominant regulatory element was found downstream of the UL94 promoter cap site (316). In the absence of this downstream element, deletion of the upstream sequences enhanced promoter activity, suggesting the presence of an upstream negative regulatory element. Further analysis of the UL94 promoter identified two p53 binding sites that were important, but not sufficient for the repressive effects (316). These studies suggest that the regulation of Late gene expression may actually be more complex than previously estimated. However, it remains to be determined if these elements truly play a role in Late gene regulation in the context of a normal viral infection, or if these are artifacts resulting from the transient transfection systems used.

The regulation of Late gene promoters is even more interesting if you consider that a number of Late promoters are located in regions of IE and Early gene expression (164, 165, 237). In fact, some Late transcripts with alternate TATA elements and start sites have been identified within HCMV Early promoters (136, 164). Thus it might be expected that the association of repressors and activators may be critical for restricting viral Late gene expression within the promoter context. A system for studying Late promoters in context of their natural genomic location will be important for elucidating the mechanism behind their regulation. Intriguingly, a more global analysis of the putative promoter regions of HCMV Late genes revealed a palindromic GC-rich sequence (CCGCGGGCGCGG) in the promoters of 17% of viral Late genes (49). The significance of this sequence element in Late gene regulation remains to be determined.

**Regulatory cellular proteins.** Some efforts have been made to identify cellular regulatory proteins that are involved in the activation of HCMV Late promoters. Analysis of the UL75 promoter indicated an activation role for a cellular PEA3-related protein (196). Analysis of the UL94 promoter indicated a role for p53 in
transcriptional repression (316). Studies also suggest that the NF-κB transcription factor may also be involved in activating viral Late gene expression (72). As mentioned earlier, there are sequence elements that have been identified as having activating or repressive effects on Late gene promoters, but the specific proteins that are required for that domain are unknown (74, 75). In addition to cellular proteins, there is evidence for the role of viral factors in the activation of Late promoters.

**The role of IE genes.** It has been shown that the HCMV immediate–Early proteins can activate some Late promoters, although to minimal extents (74, 75, 243, 249, 281, 309, 320). Deletion of specific sequences of the IE2 gene results in reduced accumulation of the Late proteins pp65, pp28, and UL83, despite DNA replication being minimally affected (249, 309). Even more interesting is that certain sequence deletions in IE2 had a completely different effect. Certain deletions resulted in an IE2 protein that did not activate Early gene expression, but expressed certain Late genes at early times after infection (309). Sequence analysis identified a potential IE2-binding site within one of these Late gene promoters, which suggests that IE2 may be involved in repression of Late gene transcriptional (309). Thus, IE2 appears to play both a positive and negative regulator of Late gene expression during HCMV infection. The mystery still remains how it is that a gene expressed at Immediate Early times of infection can be involved in Late gene regulation without inducing Late gene expression at earlier times post infection. Though many possibilities exist, the front running theories are that the IE2 protein is modified at late times of infection (phosphorylation and/or sumoylation), or that Late promoters are epigenetically regulated in such a way that they are not accessible until late times of infection (113, 114, 117, 167).

**Late gene transactivators.** Recent studies have identified multiple HCMV genes that appear to specifically act as Late gene transactivators. The first genes of this nature were discovered in gammaherpesviruses. ORF18, ORF24, ORF30, ORF31, and ORF34 of gammaherpesvirus MHV68 encode proteins that are required for Late gene expression, but have no impact on DNA replication (16, 135, 317, 321). Homologues for these genes exist in HCMV and MCMV. To this end, UL79 (homologue of ORF18) and UL92 (homologue of ORF31) have been shown to be required for HCMV Late gene expression, but dispensable for viral DNA replication (16, 130, 135, 221, 229). However, the mechanism for
how these proteins regulate Late gene expression during HCMV replication is unknown. Identifying additional viral regulatory factors in Late gene expression will be critical to understand how these proteins function. The MCMV homologues M79 and M92, which are the focus of this thesis, remained uncharacterized.

**Direct role for DNA replication.** There has only been one study addressing the direct role of DNA replication in stimulating Late gene expression. A vector containing both the Late gene TRL7 and the ori-lyt in cis were transfected into mammalian cells and infected with HCMV (300). The ori-lyt prevented the plasmid from replicating outside of infection. What was observed is that the ability of the plasmid to replicate during HCMV infection increased the TRL7 promoter-directed transcript to levels higher than could be explained by plasmid replication alone. Furthermore, this increase in transcription was not dependent on methylation or specific upstream promoter elements. Therefore it is possible that factors required for DNA replication may directly be involved in the transcription of some Late genes as well, with replication providing a mechanism for bringing such factors into proximity of the promoter.

**Post-transcriptional regulation.** In addition to transcriptional regulation, viral genes are regulated at the post-transcriptional level as well. Analysis of the UL99 5' untranslated region revealed that these sequences influenced translational regulation, via a putative stem–loop structure (143). While the exact mechanism of this regulation has not been determined, it is thought that IE2 is involved (309). Post-transcriptional regulation for HCMV is in line with what has been observed for the other betaherpesviruses. Studies in HHV-6 and HHV-7 show that certain splice variants are only expressed at late times of infection (198, 202). Thus, post-transcriptional Late gene regulation is likely to be a conserved mechanism in the betaherpesviruses. Furthermore, post-transcriptional mechanisms also play an important role in the regulation of Late gene expression. During HSV1 infection it has been shown that reduction in gene expression can occur despite unaltered transcript levels, suggesting that post-transcriptional regulation may be conserved throughout the herpesvirus family (191, 228).
Development of herpesvirus BAC system

The large genome and complicated reverse genetics system for HCMV have made the creation of mutant viruses difficult. The only way to make mutant viruses was to use homologous recombination in mammalian cells, which was particularly difficult for HCMV due to the large genome size and slow-replication kinetics. Recently, the herpesvirus field has seen a revolution in the genetic system for large DNA viruses. A system was created whereby herpesvirus genomes are maintained as infectious bacterial artificial chromosomes (BACs) within Escherichia coli (30). This system allows researchers to do genetic manipulations of herpesviruses by using the well established bacterial genetic techniques. The creation of BACs required that virus genomes were inserted into F-plasmids, flanked by loxP sites. The F-plasmid maintained the genome in host cells, and allowed for reconstitution of infectious viral particles. Since the F-plasmid also encoded an intron-containing Cre recombinase, during viral reconstitution in mammalian cells, the F-plasmid was excised from the viral genome to prevent expression or packaging problems downstream. Due to the presence of an intron, the Cre is not expressed in E. coli cells, preventing any problems that Cre might cause during genetic modification of the BAC. This development has greatly increased the rate of making individual ORFs mutations in herpesvirus genomes. This method created a system where viral recombinants could be created in the absence of viral growth, thus one can create a viral mutant independently of its viral fitness. Furthermore, the viral genome could be characterized before reconstitution of viral progeny, reducing the likelihood of spurious point mutations or deletions occurring in the process of producing recombinant virus.

In HCMV, several strains have been cloned as BACs including the laboratory strains Towne and AD169, and several clinical isolates including TR, TB40, and Merlin. The ability to create recombinant clinical BACs will be critical to the future study of HCMV, as the laboratory strains contain large deletions and many point mutations and thus does not behave precisely like virus acquired directly from patients. It should also be noted that a BAC has been created for the MCMV smith strain. This will be relevant to the studies of this dissertation.
Functional profiling in HCMV

Not long after the advent of the BAC, several groups set out to create a complete library of ORF/gene mutations in HCMV and screen these mutants for their growth in fibroblasts. The functions of many of the 166 HCMV ORFs remain unknown, and thus these mutant libraries represented a significant step in the ability to understand the biology of this virus at the molecular level. Complete mutant libraries were created in both the Towne and AD169 strains (17, 143). These studies identified whether each gene of these HCMV strains was essential (no growth of mutant virus), nonessential (mutant virus grows like wild-type virus), or augmenting (mutant virus has >10 fold growth defect) for growth on fibroblasts. The conclusions of these two studies largely agreed with each other; and occasional differences have been attributed to the location of the mutation (i.e. UL35) (173). In libraries of HCMV BACs constructed to disrupt each unique ORF 41–45 of the ORFs examined appear essential for replication in fibroblasts, 117 are not required for viral replication in fibroblasts (83, 325). Of the non-essential genes, deletion in 88 of them had growth kinetics identical to wild type, and thus were truly non-essential, and 27 when deleted, give rise to a severe growth defect, and were designated augmenting. Interestingly, some of these dispensable ORFs (UL24, UL64, and US29) are required for viral growth in cell types other than fibroblasts, such as endothelial or epithelial cells (83, 325). In fact some of these nonessential genes have turned out to be tropism factors for these cell types. In addition, four of the mutants with non-essential genes deleted (UL10, UL16, US16, and US19) grow significantly better than the wild type in cell types other than fibroblasts. Furthermore, although many of the HCMV genes are dispensable for viral growth in cell culture, studies with MCMV or RhCMV suggest that many dispensable genes are important for modulating the virus–host interaction.

Aim and scope of thesis

Developing effective anti-HCMV therapeutics requires a better understanding of the molecular mechanisms by which the virus replicates and causes disease within the host. Late gene expression and regulation is a vital part of the HCMV life cycle. Proteins involved in this process would make for great drug targets. Due to the lack of a robust genetics system, comprehensive analysis of much of the HCMV genome has not been achieved. As mentioned above, the advent of the BAC system has allowed functional profiling of the entire HCMV genome. However, due to species restrictions, the attenuation of lab strains, and
difficulty of complementing mutant viruses, understanding the function of novel genes during HCMV pathogenesis has been difficult. Murine CMV (MCMV) is a useful surrogate model to bridge the knowledge gap between aspects of HCMV infection and the functions of undefined genes.

**MCMV as a surrogate model.** MCMV shares 45% sequence identity with HCMV, and is the most commonly used animal model for the study of CMV induced disease (140). MCMV allows researchers to investigate tissue tropism, virulence, latency, and reactivation of MCMV in mice with hopes of gaining insight into HCMV infection in humans. Remarkably, the effects of MCMV infection in mice resemble those of HCMV in humans with respect to pathogenesis during acute infection, persistent infection and reactivation from latency after immune-suppression. In addition, the major cell types and organs infected, the course of infection, and the type of pathology seen are identical for both viruses. There is the notable exception that MCMV does not cross the placenta to cause congenital defects in newborn pups (154). Nonetheless, the genetic and pathobiological similarities between these two viruses make the use of MCMV a sensible model system to study HCMV infection.

**Functional profiling.** Analysis of the complete nucleotide sequence of MCMV has revealed that the MCMV genome is collinear with HCMV over the central 180kb and that 78 open reading frames have significant sequence homology to those of HCMV (53, 71, 210, 211, 239). Characterization of these conserved genes should provide insight into the functions of their HCMV counterparts in viral infections in humans.

One of the most powerful approaches to identify the function of virus-encoded genes is to create and analyze viral mutants. The construction of herpesvirus mutants via site-directed homologous recombination and transposon-mediated insertional mutagenesis has been reported (41, 199). MCMV mutants can be generated from the BAC-based viral genome by both of the aforementioned strategies. The BAC-based mutagenesis approach provides a powerful and convenient strategy to generate viral mutants, facilitating studies of the functions of viral genes in tissue culture and in animals. Our experimental plan to identify MCMV protein function is based on the published sequence and ORF predictions of the Smith Strain of MCMV. There are 170 ORFs predicted for the MCMV genome assuming a minimum 100aa protein size and
<60% gene overlap (239). However, most of the MCMV genome coding capacity has not been confirmed (37).

We set out to mutate a subset of the 78 MCMV genes that have homologues in HCMV, and to characterize those mutants for their ability to replicate in tissue cultured fibroblasts (239). This would both expand our capacity to study the functions of the chosen homologous genes, and refine the use of MCMV as a model for studying HCMV biology and pathogenesis.

**Create insertion/deletion mutations in MCMV.** With genome wide mutagenesis outside the scope of my studies, we chose to focus on those genes in MCMV that have homologues in HCMV. A reasonable list of candidate genes was selected from the 78 conserved genes. For this study, we have decided not to include the conserved genes that either have been well studied in MCMV, such as M27, or have well-studied homologues in other human herpesviruses (ex. HCMV or HSV-1), such as core viral proteins like DNA polymerase or glycoprotein B (1). This process of elimination left us with 28 genes of interest to target (Table 2).

To inactivate a MCMV gene by insertion/deletion mutation, a kanamycin cassette flanked by FRT sites was recombined into the gene of interest using the MCMV BAC and linear homologous recombination. The insertion site was chosen such that it was as N-terminal as possible within the ORF but would not disrupt an overlapping ORF; often placed 200bp from the terminus of the neighboring ORF. The homologous recombination in the BAC was achieved by designing primers that had 20bp of homology to the Kan-containing vector on the 3’ end and 50bp homology to the MCMV gene target sequence on the 5’ end. After electroporation of the PCR product into BAC containing *E.coli* cells, Kan resistant bacteria were treated with arabinose. The arabinose treatment induced flippase expression from the bacteria which caused site specific recombination of the FRT sites. This removed the Kan-cassette and left behind an 88bp sequence including the FRT site. The remaining sequence caused a frame shift of the ORF and theoretical disruption of its expression downstream of the insertion (Fig. 3). The success of the insertion was confirmed by restriction digest, PCR analysis, and sequencing.
Analyze the growth of mutant MCMV. After the creation of the mutant MCMV BACs, it was necessary to determine the effect of the mutation on virus replication. Mutant BACs were transfected into mouse fibroblasts, and monitored via microscopy for seven days post transfection. Wild-type MCMV BAC and mutants that did not disrupt viral replication produced virus and lysed the cell monolayer by seven days post transfection. Mutations that were located in regions that were important for virus replication took longer than seven days to lyse the monolayer, or in some cases never lysed the monolayer at all. The mutants that did not lyse the monolayer even out to 15 days post transfection were categorized as “essential” genes. The genes that were able to lyse the monolayer by 15 days post transfection were further analyzed in a controlled growth curve analysis. Following reconstitution of the virus and titering of the stock, each virus was subjected to a multistep growth analysis on cultured fibroblasts to differentiate mutations as “non-essential” or “augmenting” genes. Each stock of virus was used to infect murine fibroblasts at a multiplicity of infection (MOI) of 0.01. Growth analysis was carried out over a course of 12 days post infection. Supernatants were collected every two days and tittered by plaque assay. Mutant viruses that grow to viral titers comparable to wild-type virus were classified as “non-essential” genes. Mutant viruses that produced progeny at a reduced level when compared to wild-type virus were categorized as “augmenting” genes.

Though most viruses were reconstituted to wild-type titer levels during the analysis, seven viruses emerged with a marked growth defect (Table 2). Mutant viruses with mutations in genes M49, M79, M92, M94 or M96 were able to express the SV40 driven GFP marker and cause cellular swelling as is characteristic of MCMV infection, but failed to spread to surrounding cells. Thus these genes were classified as “essential” for viral growth; which is consistent with the classification of their respective HCMV homologues. Mutants of M47 and M88 spread from cell to cell, but growth curve analysis showed that the virus titers were three and two logs lower than wild-type virus, respectively. Thus, M47 and M88 are augmenting genes, which is also consistent for functional profiling done in HCMV (325). The remaining 15 viruses, for which mutants were created, have demonstrated no defect during transfection or in subsequent infections into murine fibroblasts (data not shown). Six of the 28 mutant viruses that we set out to create proved to be difficult to obtain despite multiple attempts. As a result, those genes were excluded from the goal of the thesis.
**Complement mutant viruses.** In order to study the function of the essential viral proteins in MCMV infection it is necessary to produce a high titer stock of the mutant virus lacking the targeted gene. The complementation strategy that was used was to create cells over-expressing the viral gene *in trans* in order to complement the growth of a corresponding mutant MCMV BAC. This is a routine approach that is used in our laboratory with the pRetro-EBNA retroviral vector (323). A pRetro-EBNA retroviral vector containing our gene of interest was constructed and transfected into the Phoenix packaging cell line to produce retrovirus. Phoenix cells are derived from 293T cells but express the retroviral gag-pol and envelope proteins for generating amphotropic retrovirus. The resulting retrovirus was transduced into 10.1 fibroblasts to produce cells that express the gene of interest. The pRetro-EBNA vector contained a dsRED cassette, which allowed visual screening by fluorescent microscopy. Cells that expressed the gene of interest would show up as red under the fluorescent scope. Clonal cell lines could be created by limiting dilution and clonal expansion of dsRED expressing cells. The over-expressing cell lines were then used to reconstitute the corresponding mutant viruses from BAC-MCMV transfections.

Two of the three essential genes could be complemented by this system; M79 and M92. Defining the role of M79 and M92 will set the stage for using the MCMV model to elucidate the mechanism of action for this CMV gene family, and explore novel antiviral strategies targeting this viral factor. Determining the molecular mechanism of even one conserved CMV gene would greatly enhance our knowledge of CMV biology and may also provide fundamental discoveries in cell biology as many of these genes are likely to modulate critical virus-host cell interactions.
References


67. **Das, S., and P. E. Pellett.** 2007. Members of the HCMV US12 family of predicted heptaspanning membrane proteins have unique intracellular distributions, including association with the cytoplasmic virion assembly complex. Virology **361:**263-73.


91. **Flanagan, W. M., A. G. Papavassiliou, M. Rice, L. B. Hecht, S. Silverstein, and E. K. Wagner.** 1991. Analysis of the herpes simplex virus type 1 promoter controlling the expression of UL38, a


137. Kalamvoki, M., and B. Roizman. 2011. The histone acetyltransferase CLOCK is an essential component of the herpes simplex virus 1 transcriptome that includes TFIID, ICP4, ICP27, and


by herpes simplex virus ICP27 is posttranscriptional and does not require the endogenous promoter or polyadenylation site. J Virol 77:9872-84.


258. **Seo, J. Y., and W. J. Britt.** 2007. Cytoplasmic envelopment of human cytomegalovirus requires


### Table 1. HCMV gene families

<table>
<thead>
<tr>
<th>Family name</th>
<th>Genes in the family</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>RL11</td>
<td>RL5A, RL6, RL11, RL12, RL13, UL1, UL4, UL5, UL6, UL7, UL8, UL9, UL10, UL11</td>
<td>Membrane glycoproteins that share a common immunoglobulin domain</td>
</tr>
<tr>
<td>UL14</td>
<td>UL14, UL141</td>
<td>Membrane glycoprotein containing an immunoglobulin domain, putatively involved in NK cell evasion.</td>
</tr>
<tr>
<td>UL18</td>
<td>UL18, UL142</td>
<td>MHC-1-related membrane glycoproteins involved in immune evasion</td>
</tr>
<tr>
<td>UL25</td>
<td>UL25, UL35</td>
<td>Tegument proteins</td>
</tr>
<tr>
<td>GPCR</td>
<td>UL33, UL78, US27, US28</td>
<td>Chemokine receptors</td>
</tr>
<tr>
<td>DURP</td>
<td>UL31, UL72, UL82, UL83, UL84</td>
<td>Tegument proteins derived from dUTPase that have multiple roles in modulating the cellular response</td>
</tr>
<tr>
<td>UL120</td>
<td>UL120, UL121, UL119</td>
<td>Membrane glycoproteins</td>
</tr>
<tr>
<td>UL146</td>
<td>UL146, UL147</td>
<td>CXC chemokines</td>
</tr>
<tr>
<td>US1</td>
<td>US1, US31, US32</td>
<td>potentially regulating both innate and adaptive immunity</td>
</tr>
<tr>
<td>US2</td>
<td>US2, US3</td>
<td>Membrane glycoproteins involved in immune evasion</td>
</tr>
<tr>
<td>ORF</td>
<td>Mutant Created</td>
<td>HCMV Homolog</td>
</tr>
<tr>
<td>-------</td>
<td>----------------</td>
<td>--------------</td>
</tr>
<tr>
<td>M024</td>
<td>Yes</td>
<td>UL024</td>
</tr>
<tr>
<td>M025</td>
<td>Yes</td>
<td>UL025</td>
</tr>
<tr>
<td>M026</td>
<td>Yes</td>
<td>UL026</td>
</tr>
<tr>
<td>M028</td>
<td>Yes</td>
<td>UL028</td>
</tr>
<tr>
<td>M031</td>
<td>Yes</td>
<td>UL031</td>
</tr>
<tr>
<td>M035</td>
<td>Yes</td>
<td>UL035</td>
</tr>
<tr>
<td>M038</td>
<td>Yes</td>
<td>UL038</td>
</tr>
<tr>
<td>M071</td>
<td>Yes</td>
<td>UL071</td>
</tr>
<tr>
<td>M072</td>
<td>Yes</td>
<td>UL072</td>
</tr>
<tr>
<td>M076</td>
<td>Yes</td>
<td>UL076</td>
</tr>
<tr>
<td>M087</td>
<td>Yes</td>
<td>UL087</td>
</tr>
<tr>
<td>M091</td>
<td>Yes</td>
<td>UL091</td>
</tr>
<tr>
<td>M116</td>
<td>Yes</td>
<td>UL116</td>
</tr>
<tr>
<td>M118</td>
<td>Yes</td>
<td>UL118</td>
</tr>
<tr>
<td>M121</td>
<td>Yes</td>
<td>UL121-P</td>
</tr>
<tr>
<td>M049</td>
<td>Yes</td>
<td>UL049</td>
</tr>
<tr>
<td>M079</td>
<td>Yes</td>
<td>UL079</td>
</tr>
<tr>
<td>M092</td>
<td>Yes</td>
<td>UL092</td>
</tr>
<tr>
<td>M094</td>
<td>Yes</td>
<td>UL094</td>
</tr>
<tr>
<td>M096</td>
<td>Yes</td>
<td>UL096</td>
</tr>
<tr>
<td>M047</td>
<td>Yes</td>
<td>UL047</td>
</tr>
<tr>
<td>M088</td>
<td>Yes</td>
<td>UL088</td>
</tr>
<tr>
<td>M023</td>
<td>No</td>
<td>UL023-P</td>
</tr>
<tr>
<td>M034</td>
<td>No</td>
<td>UL034</td>
</tr>
<tr>
<td>M083</td>
<td>No</td>
<td>UL083</td>
</tr>
<tr>
<td>M95</td>
<td>No</td>
<td>UL95</td>
</tr>
<tr>
<td>M103</td>
<td>No</td>
<td>UL103</td>
</tr>
<tr>
<td>M114</td>
<td>No</td>
<td>UL114</td>
</tr>
</tbody>
</table>
Figure 1. Human cytomegalovirus genome schematic. Schematic map of the AD169 HCMV genome. The CMV genome is organized as two regions of unique sequences, unique long (UL) and unique short (US) (designated by the light shaded boxes), flanked by two sets of inverted repeats (TRL/IRL) (designated by black boxes) and (IRS/TRS) (designated by white boxes). Arrows indicate the relative orientations of the repeated and unique genome blocks.
Figure 2. A representative diagram depicting the strategy for creating recombinant MCMV BAC clones. An insertion mutation was created by two step recA-mediated homologous recombination. A site downstream of the translation initiation codon was chosen for insertion of a kanamycin cassette (indicated by red bar) flanked by two FRT sites. After the induction of flp recombinase, the BAC clone carried an 88-bp insert (indicated by black bar) downstream of the M92 translation initiation codon, resulting in a frame-shift mutation.
Figure 3. Life cycle of CMV in a host cell. CMV enters host cells (1) either through direct fusion or through the endocytic pathway. The virus attaches to the cell via interactions between viral glycoproteins and an unknown cell surface receptor(s), followed by the fusion of the envelope with the cellular membrane to release the capsid into the cytoplasm. The capsid is then translocated into the nucleus, where viral DNA is released. This initiates the expression of the IE genes (2). The IE genes make the environment suitable for viral DNA replication, and then transactivate the Early genes. Early genes facilitate viral DNA replication (3). Once viral DNA replication is occurring, the Late gene transcription program is initiated (4). Viral Late genes encode proteins involved in DNA packaging and capsid assembly. Once the nucleocapsid is assembled, it leaves the nucleus and acquires the virion envelope in the cytoplasm (5). Secondary envelopment occurs in the cytoplasm at the endoplasmic reticulum (ER)-Golgi intermediate compartment. This is followed by a complex two-stage final envelopment and egress process that leads to virion release by exocytosis at the plasma membrane.
Chapter II

Murine Cytomegalovirus Protein pM79 Is a Key Regulator of Viral Late Transcription
Preface

TJC designed and performed the majority of experiments, analyzed the data, generated the figures and wrote the first draft of the manuscript. CA assisted with the preparation of the RNA samples for the tiled array analysis, LSJ and MCV performed the computational analysis of the tiled array data, WMY provided key reagents and insights for the tiled array analysis, ARF and DY supervised the studies and greatly contributed to the design and analysis of the experiments, TJC, DY, and WMY greatly contributed to the writing and editing of the manuscript.

This chapter has been previously published:

Summary

Herpesvirus genes are temporally expressed during permissive infections, but how their expression is regulated at late times is poorly understood. Previous studies indicate that the human cytomegalovirus (CMV) gene, UL79, is required for late gene expression. However, the mechanism remains to be fully elucidated and UL79 homologues in other CMVs have not been studied. Here we characterized the role of the conserved murine CMV (MCMV) gene M79. We showed that M79 encoded a protein (pM79) which was expressed with early-late kinetics and localized to nuclear viral replication compartments. M79 transcription was significantly decreased in the absence of viral DNA synthesis but markedly stimulated by pM79. To investigate its role, we created the recombinant virus SMin79, in which pM79 expression was disrupted. While marker-rescued virus grew efficiently in fibroblasts, SMin79 failed to produce infectious progeny but was rescued by pM79 expression in trans. During SMin79 infection, representative viral immediate early and early gene products as well as viral DNA accumulated sufficiently. Formation of viral replication compartments also appeared normal. Pulsed field gel electrophoresis analysis indicated that the overall structure of replicating viral DNA was indistinguishable between wild-type and SMin79 infection. Viral tiled array and quantitative PCR analysis revealed that many late transcripts sensitive to a viral DNA synthesis inhibitor (phosphonoacetic acid) were markedly reduced by pM79 mutation. This study indicates that cytomegaloviruses use a conserved mechanism to promote transcription at late stages of infection, and that pM79 is a critical regulator for at least a subset of viral DNA synthesis-dependent transcripts.
Introduction

Cytomegalovirus (CMV) is the prototypical member of the β-herpesvirus subfamily. Human CMV (HCMV) is a ubiquitous human pathogen, which causes asymptomatic infection in healthy adults. However, in immuno-compromised hosts, such as neonates, transplant recipients, persons with advanced AIDS, and cancer patients, HCMV is a common cause of severe and even life-threatening disease (5, 10, 12, 38). The severity of medical problems associated with HCMV in these vulnerable populations underlies the necessity for developing safer and more effective antiviral treatments and vaccine strategies to control its infection. HCMV infection is limited to human hosts, as the members of the CMV family are species specific. Murine CMV (MCMV) infection provides a tractable small animal model to study CMV biology. MCMV shares conservation with HCMV in regard to its collinear genome, gene expression program, tissue tropism, and pathology (35, 37). Over 40% of MCMV genes have sequence or functional homologues in HCMV (34). This conservation provides us with excellent opportunities to explore MCMV as a tool to dissect the mechanistic basis of shared features of viral replication and pathogenesis. Furthermore, revealing the function of homologous viral genes in their respective hosts will allow antiviral therapeutics or vaccine candidates targeting these conserved genes to be tested in the mouse model.

Gene expression during herpesvirus lytic infection is highly coordinated and sequentially ordered such that viral genes are traditionally divided into three kinetic classes: immediate early (IE), early, and late. IE genes are transcribed following viral DNA translocation to the nucleus and require only incoming virion-associated proteins and cellular factors for their expression. Products of IE genes transactivate early genes and remodel the host cell to be permissive for virus replication. Early gene transcriptions are initiated prior to viral DNA synthesis but some persist at late times of infection, even after the onset of DNA synthesis. Early gene products are required for both viral DNA synthesis and formation of replication compartments, which are virus-induced subnuclear structures where viral DNA synthesis occurs. Transcription of late viral genes occurs after the onset of viral DNA synthesis, and peaks at late times of infection. Many late genes encode structural proteins required for virion assembly, maturation, and release. Largely consistent with this temporal regulation, transcription of many IE and early genes is resistant to viral DNA synthesis inhibitors, such as phosphonoacetic acid (PAA). Late transcripts, defined
by their abundant accumulation at late times of infection and dependency on viral DNA synthesis, are largely derived from late genes. In addition, some genes have both early and late properties, as their expression initiates prior to DNA synthesis but the transcripts continue to accumulate to high levels at late times in a DNA synthesis-dependent manner.

While IE and early transcription has been extensively studied in herpesviruses, little is known about how viral late transcription is regulated. Viral DNA synthesis is required \textit{in cis} for viral late promoter activity but the precise mechanism remains elusive (17, 26, 28). Herpes simplex virus 1 (HSV-1), the prototypical $\alpha$-herpesvirus, is perhaps the best studied example. Multiple HSV-1 proteins (ICP0, ICP4, ICP22, and ICP27) have been shown to regulate late expression (8, 32, 36, 41, 48, 49). For some late genes, the TATA box as well as DNA sequences downstream of the transcription start site are also main determinants of transcription (11, 16, 27). However, a majority of these HSV-1 genes lack homologues in $\beta$-herpesviruses (1), and evidence suggests that the requirement for these sequence elements is not universal (11, 14, 20, 27, 39).

Understanding how CMV regulates late gene expression is important to understand its biology and identify novel targets for antiviral therapeutics. The HCMV UL79 family is a viral gene family conserved between $\beta$- and $\gamma$-herpesviruses (2). We and others have recently shown that UL79 is required for HCMV late gene expression (15, 31). However, the MCMV homologue of UL79, M79, remains uncharacterized. Defining the role of M79 will set the stage for using the MCMV model to elucidate the mechanism of action for this CMV gene family, and explore novel antiviral strategies targeting this viral factor.

In this study, we characterized M79 during MCMV infection. We show that pM79, the protein product of M79, acts downstream of viral DNA synthesis to facilitate viral late transcription. Importantly, viral oligonucleotide tiled array analysis reveals at least two subsets of late transcripts. Both require viral DNA synthesis for their expression, but they have different degrees of dependence on pM79 for expression. As a result, abrogation of pM79 results in a complete failure in virus growth. These results, along with studies of HCMV UL79 and murine gammaherpesvirus 68 (MHV68) ORF18 (2, 31), suggest that divergent herpesviruses use similar mechanisms to promote late gene expression. Furthermore, our study provides evidence to support the model that CMV late transcription is tightly regulated beyond its
dependency on viral DNA synthesis, and that pM79 is a key regulator for at least a subset of MCMV late transcription, highlighting the complex regulatory mechanisms governing CMV late transcription.
Results

**M79 is essential for MCMV replication.** The homolog of M79 in HCMV, UL79, is essential for virus replication (15, 31), but the role of M79 has not been characterized. To investigate this, we created a mutant BAC clone of MCMV Smith strain, pSM\textit{in}79, by BAC recombineering. In this mutant clone, an 88-nt insertion was introduced at 403-nt downstream of the start codon of the M79 coding sequence, resulting in a frame-shift mutation (Fig. 1A). We anticipated that this site of insertion made it unlikely to interfere with expression of neighboring genes, particularly M80, an essential gene that overlaps with M79 at its N-terminus (3, 23, 47).

To reconstitute recombinant virus, both pSM\textit{in}79 and its wild-type parental clone, pSM\textit{gfp}, were electroporated into 10.1 mouse embryonic fibroblasts (MEF10.1). Cells transfected with pSM\textit{gfp} readily initiated virus production and spread. By 5 days post transfection, the monolayer of transfected cells demonstrated complete cytopathic effect (CPE) and full virus spread, indicated by virus-driven GFP. However, even though pSM\textit{in}79 transfection was efficient, evidenced by the presence of individual GFP-positive cells upon initial inspection at day 2, it repeatedly failed to produce any CPE or GFP spread even at two weeks post transfection (Fig. 1B). Transfection of multiple independently isolated pSM\textit{in}79 clones yielded the same result, suggesting that M79 is essential for MCMV replication.

To provide a means to propagate BAC-derived M79 mutant virus, we created multiple clonal MEF10.1 cell lines (10.1-M79\textit{flag}) that stably expressed C-terminally FLAG-tagged M79 by retroviral transduction. Transfection of pSM\textit{in}79 into 10.1-M79\textit{flag} cells supported efficient virus reconstitution, producing complete CPE and full spread of virus-driven GFP expression on the monolayer (Fig. 1B). Importantly, transfection of pSM\textit{in}79 in 10.1-M79\textit{flag} cells produced virus with titers similar to that of reconstituted wild-type virus, SM\textit{gfp} (Fig. 1B).

Finally, to provide definitive evidence for the essential role of M79 in MCMV infection, we performed growth curve analysis of SM\textit{in}79 in MEF10.1 cells. For this experiment, we also created SM\textit{rev}79, a marker-rescued virus of SM\textit{in}79. The SM\textit{in}79 virus failed to produce cell-free or cell-associated progeny virus in both multistep (data not shown) and single step growth analysis (Fig. 1C). On the other hand, SM\textit{rev}79 replicated indistinguishably from wild-type virus (Fig. 1D). Together, our results
indicate that the defect of SM\textit{in}79 is the direct result of M79 ablation, and that M79 is essential for MCMV replication at steps prior to virus release.

**Expression of M79 gene products is markedly enhanced by viral DNA synthesis and protein pM79.**

To better understand the function of the M79 gene, we first characterized its potential protein product (pM79). M79 is predicted to encode a protein of 258 amino acids (aa) with a molecular weight of 29 kDa. As no specific antibody to the M79 protein was available, we created a recombinant virus, SM79\textit{flag}, in which the M79 coding sequence was tagged with 3x FLAG at the C-terminus (Fig. 1A). SM79\textit{flag} was reconstituted efficiently from BAC transfection and grew indistinguishably from wild-type virus (Fig. 1D), indicating that the 3x FLAG tag did not interfere with M79 function. To determine expression of pM79 during infection, we infected MEF10.1 cells with SM79\textit{flag} and analyzed the accumulation of FLAG-tagged pM79 over the course of a single viral replication cycle by immunoblotting (Fig. 2A). FLAG-tagged pM79 migrated at the apparent molecular weight of 31 kDa, consistent with the predicted size, and was detected at 24-36 hours post infection (hpi) (Fig. 2A). We then profiled M79 transcription by reverse transcription-coupled quantitative PCR analysis (RT-qPCR) to more precisely characterize M79 expression (Fig. 2B). When viral DNA synthesis was inhibited by PAA, low levels of M79 transcription persisted and increased modestly from 10 hpi to 30 hpi, suggesting that a small amount of M79 transcripts were produced independent of viral DNA synthesis. Importantly however, the majority of M79 transcription was inhibited by PAA, and in the absence of PAA, M79 transcript levels accumulated to high abundance at late times (20-30 hpi). Together, our results indicate that low levels of M79 expression occurs independent of viral DNA synthesis but that the majority of its expression requires viral DNA synthesis, similar to the previously described viral gene expression pattern of early-late kinetics (28).

It was interesting to note that during SM\textit{in}79 infection, M79 transcript levels were drastically reduced relative to that of SM\textit{gfp} virus infection (Fig. 2B). As the small insertion mutation in SM\textit{in}79 was designed to abolish only the M79 protein but not its transcript, we interpreted this to suggest that pM79 enhances its own transcription, particularly at late times of infection, correlating with the previous report of its homolog UL79 during HCMV infection (31).
M79 protein localizes to viral nuclear replication compartments. We next examined the intracellular localization of pM79 during infection of SM79flag using a rabbit anti-FLAG antibody. At 24 hpi, FLAG staining was found exclusively in the nucleus, closely co-localized with the viral polymerase processivity factor pM44 in the nuclei of infected cells (Fig. 3A). As pM44 is a widely used marker of viral replication compartments, this result indicates that the nuclear pM79 localizes within replication compartments. The rabbit anti-FLAG antibody also produced a diffuse, weak cytoplasmic staining, which was likely nonspecific as it was also present in SMgfp infected control cells (Fig. 3A). To test this, a separate set of infected cells were stained with a mouse anti-FLAG antibody, even though this strategy precluded us from co-staining cells with the mouse anti-pM44 antibody to mark replication compartments (Fig. 3B). Nonetheless, in these cells there was only specific staining in the nuclei resembling replication compartments, and there was no cytoplasmic background staining observed. Collectively, we conclude that pM79 is a nuclear protein that localizes to replication compartments during MCMV infection.

M79 is not required for viral DNA synthesis or development of nuclear replication compartments. As pM79 localized to viral replication compartments during infection, we next determined if this protein was involved in viral DNA replication processes, particularly the ability of the virus to synthesize its genome and form nuclear replication compartments. To test if pM79 was required for viral DNA synthesis, we first examined the kinetics of viral DNA accumulation during SMin79 infection by quantitative PCR (qPCR) analysis. Viral DNA accumulation over the course of SMin79 infection was comparable to that in SMgfp infection, indicating that pM79 is not required for the virus to synthesize its DNA (Fig. 4A).

To probe if its overall structure was altered in the absence of pM79, viral DNA from infected cells was analyzed by a pulsed field gel electrophoresis (PFGE) (Fig. 4B). SMgfp infected cells produced both concatemeric replicating viral DNA, which was retained in the well, and cleaved 232-kb monomeric viral genome, which migrated into the gel (lane 2). In addition, a minor population of viral DNA with the apparent molecular weight greater than 232 kb also migrated into the gel, likely representing polymeric DNA intermediates. In contrast, SMin79 infected cells produced only concatemeric viral DNA and polymeric DNA intermediates without any monomeric viral genomes (lane 3). This suggests either a failure of the mutant virus to cleave replicating viral DNA into mature viral genomes, or altered viral DNA
structures that prevent monomeric viral genomes from migrating into the gel. To differentiate these possibilities, intracellular viral DNA was digested with Pac I, a restriction enzyme that cuts the MCMV genome four times. As expected, Pac I cut the 232-kb monomeric virion genome (lane 7) into linear fragments of 92.8 kb, 90.3 kb (which co-migrated together), 43.0 kb, 1.3 kb and 4.2 kb (the last two run off the gel) (lane 8). Furthermore, Pac I digested intracellular DNA from SMgfp infected cells produced an additional 135.8-kb linear fragment that resulted from joint ends within concatemeric viral DNA (lane 5). Pac I digestion of SMinM79 infected cells released the same fragments of 90.3-kb and 135.8-kb from concatemeric DNA, but did not produce the monomer-derived fragments (lane 6). Therefore, the overall structure of replicating viral DNA produced by pM79-deficient virus was not appreciably different from that of wild-type virus. Together, our data indicate that loss of M79 does not have a deleterious effect on viral genomic amplification.

We also tested if pM79 was involved in the development of replication compartments, virus-induced nuclear structures critical to successful viral DNA replication (30). We infected cells with wild-type or M79-deficient virus, and monitored the formation and maturation of replication compartments marked by pM44 staining. pM44 staining became evident at 12 hpi, forming multiple small nuclear foci (i.e. pre-replicative compartments), and these foci coalesced into larger and fewer mature replication compartments at 24 hpi (Fig. 5). Importantly, no appreciable difference was observed in this temporal progression of replication compartments between infections of SMgfp and SMin79 virus (Fig. 5). Therefore, pM79 does not appear to play a role in the development of replication compartments.

These data taken together suggest that the defect seen in SMin79 infection occurs downstream of viral DNA synthesis. Moreover, the failure of mutant virus to process replicating viral DNA into monomeric genomes suggests the defect occurs at or prior to genome cleavage and packaging, such as the step of late gene expression or capsid assembly.

**M79-deficient virus is defective in the efficient accumulation of representative viral Late gene products.** To continue to define the stage of the viral replication cycle where pM79 acts, we analyzed the accumulation of representative viral proteins from each kinetic class: immediate early protein IE1, early protein E1 (M112/113), and late protein gB (M55). M79-deficient virus demonstrated two defects in viral
protein accumulation (Fig. 6A). One was a modest delay in IE1 and E1 protein accumulation, whereas the other was the complete loss of gB. As this modest delay in IE1 and E1 had no deleterious consequence on viral DNA synthesis (Fig. 4), it was unlikely responsible for defects in events beyond viral DNA synthesis, such as late protein accumulation and ultimately virus growth. The delay in IE1 and E1 expression might be due to the characteristics of the mutant viral stocks, such as tegumentation or virion composition, which could be slightly different from wild type stocks because of the efficiency of complementation. To rule out any effect of IE1 and E1 expression, we infected cells with SM\textit{in}79 at a MOI of 5 and SM\textit{gfp} at a MOI of 1 to compensate for IE1 and E1 expression levels. Despite elevated IE1 and E1 accumulation in SM\textit{in}79 infection under this condition, gB remained absent even after 48 hpi (Fig. 6B). We interpreted this result to indicate that the major defect of SM\textit{in}79 was the inability to express viral gene products at late times of infection.

To test if this defect was at the transcriptional level, we analyzed the accumulation of E1 and M55 transcripts during wild-type and M79 mutant virus infection. E1 transcript accumulated at comparable levels during SM\textit{in}79 and SM\textit{gfp} infection, particularly at early times (10 hpi) (Fig. 6C). In contrast, accumulation of M55 transcript in SM\textit{in}79 infected cells was effectively reduced to levels comparable to those under PAA treatment. All transcripts detected were specific and were not the result of genomic DNA contamination as mock cells and reactions done in the absence of reverse transcriptase failed to produce any products (data not shown). This result supports the hypothesis that both pM79 and viral DNA synthesis are required for efficient viral late transcription.

\textbf{pM79 regulates the accumulation of a subset of viral late transcripts.} The failure of SM\textit{in}79 to express a representative late transcript could be due to a global down-regulation of MCMV late transcription, or it could be due to a down-regulation of only a subset of late transcripts. To differentiate these two possibilities, we profiled the entire MCMV transcriptome with or without pM79. We designed a high density oligonucleotide tiled array with probes to both the forward and reverse strand of the MCMV genome, allowing us to measure transcription across the entire viral genome. We first determined viral regions where transcription was dependent on viral DNA synthesis. To test this, MEF10.1 cells were infected with MCMV in the presence or absence of PAA. At 20 hpi, RNA was harvested, converted to
fluorescently labeled cDNA, and hybridized to the MCMV DNA array. The normalized mean intensity of probes across the genome was plotted in Fig. 7. The sensitivity of viral transcription to PAA was variable across the genome, ranging from no change to greater than 100-fold reduction. Although most viral regions exhibited some sensitivity to PAA, many regions that were highly regulated did not correspond to annotated ORFs. To minimize false positives, we considered a region of transcription to be dependent on viral DNA synthesis only if the intensity of the PAA-treated sample was at least 3-fold lower than that of the untreated sample in this analysis. By this criterion, among viral regions corresponding to the 172 annotated ORFs examined, efficient transcription of 115 ORFs was dependent on viral DNA synthesis (Table 2). Among them, 80 ORFs have been previously analyzed for their temporal expression, and 61 have been found to be expressed at elevated levels at late stages of MCMV infection (i.e. late genes) (19, 25). Therefore, transcription of a large set (76%) of annotated ORFs sensitive to PAA are also known to be highly expressed at late times, consistent with the notion that most viral DNA synthesis-dependent transcripts are derived from late genes.

We then examined the transcriptome profile in SM<sup>in79</sup> infection. Overall, the effect of pM79 mutation on viral transcription was less pronounced than that of PAA (Fig. 7). Compared to PAA treatment, fewer regions of transcription were reduced by greater than 3-fold in the absence of pM79 in array analysis. Interestingly, many un-annotated regions of RNA expression were affected by PAA but not by pM79 mutation. Transcription from regions corresponding to 43 ORFs was reduced by greater than 3-fold in the absence of pM79 by array analysis, and 41 of them (95%) were also reduced by PAA treatment (Table 2). Transcription from the regions corresponding to the remaining 74 PAA-sensitive ORFs was less affected by pM79 mutation based on the criterion used in our analysis. These results suggest that at least a subset of DNA synthesis-dependent viral transcripts also have a high dependence on pM79 for their expression.

This differentiated dependency of late transcripts on pM79 was validated by RT-qPCR analysis. Transcription of M74 and M116, which showed a greater dependency on pM79 in array analysis (Table 2), was reduced by 9.8 and 37.7 -fold by pM79 mutation in qPCR analysis (Fig. 8A). Conversely, transcription of M25 and M121 was reduced by only 5.5 and 4.5 -fold, respectively, in pM79 mutant virus
infection (Fig. 8B), thus consistent with the result of array analysis (Table 2) and showing relatively less dependency on pM79.

Together, our results show that M79 is critical for the accumulation of at least a subset of DNA synthesis-dependent viral late products (Fig. 8). These results, along with the observations of UL79 in HCMV and ORF18 in MHV68 (2, 15, 31), underscore a conserved function of the UL79 family of genes in β- and γ- herpesvirus replication cycles.
Discussion

MCMV is the commonly used model virus for HCMV, so revealing HCMV genes that are functionally conserved in MCMV will allow the use of the robust mouse genetic system to elucidate their role and test novel antivirals targeting these products. Previously, we have found that the HCMV gene UL79 regulates viral late gene expression (31). Here, we characterized the function of its MCMV sequence homologue, M79, by analyzing mutant MCMV virus in which pM79 expression was disrupted. pM79 accumulated with early-late kinetics in nuclear viral replication compartments, and if disrupted, abrogated the ability of MCMV to replicate. In particular, we showed that pM79 was critical for MCMV to promote expression of a set of late transcripts. Moreover, MCMV DNA was not only synthesized efficiently in the absence of pM79, PFGE analysis shows that the overall structure of replicating viral DNA was unimpaired. The development of viral replication compartments during mutant virus infection was indistinguishable from those during wild-type virus infection. This body of evidence further excludes the involvement of pM79 in viral DNA synthesis or other events preceding viral late transcription.

Furthermore, the failure of pM79-deficient virus to cleave viral concatameric replicating DNA is consistent with its defect in late viral transcription. For example, genes required for genomic cleavage (e.g. M56 and M104) or capsid assembly (e.g. M80, M85, and M86) are down-regulated during SM\textit{mM79} infection (Table 2). Our work, together with the reported role of its homologues, including pUL79 in HCMV and ORF18 in MHV68 (2, 15, 31), suggests a common mechanism governing late transcription among β and γ-herpesviruses. How viral late transcription is regulated remains largely unknown and viral/host factors involved are poorly defined. The pM79/pUL79 protein family represents an invaluable tool to gain insight into this key viral process.

Our oligonucleotide tiled array analysis has identified a set of annotated genes whose transcription is substantially reduced at late stages of virus infection when pM79 is abrogated (Fig. 7). Comparative analysis of viral transcriptomes among MCMV infections with PAA, without PAA, or in the absence of pM79 leads to two interesting observations. First, while viral DNA synthesis is required for transcription from genomic regions containing many previously reported late genes, it also facilitates the continued transcription from genomic regions containing several previously reported early genes at late times of infection. Second, viral transcripts that are dependent on viral DNA synthesis also have a
dependency on pM79 for their accumulation. There seems to be a striation in dependence, such that without pM79, some transcripts are markedly reduced, whereas others are reduced to a much less extent. Thus, pM79 is a key viral regulator of late transcription in MCMV infection. In future studies, similar transcriptome analysis should be applied to other viral genes known or predicted to be regulators of late gene expression. Examples are HCMV UL79 and MHV68 ORF18, as well as MCMV M87 and M95 (homologues of HCMV UL87 and UL95). Such analysis will reveal whether these viral regulators control an overlapping or distinct set of viral late gene expression.

How does pM79 regulate expression of late transcripts? A late transcript regulator could be involved in epigenetic regulation of replicating viral DNA. For instance, viral DNA-associated histones are modified when herpesviruses replicate their genomes, and this may render late promoters accessible for transcription (21, 22, 40). However, the pM79 coding sequence does not resemble those of histone modifying enzymes, such as histone acetyltransferase or histone deacetylase (data not shown). If pM79 had a role in epigenetic regulation of gene expression, it would likely act indirectly, for instance, by recruiting modification enzymes to histones associated with late viral promoters. Late regulators may also act as transcription factors. However, pM79 does not contain any identifiable putative DNA binding domains (data not shown). Therefore, in this capacity pM79 would have to act as a modulator of cellular and viral transcriptional regulators or RNA polymerase to facilitate viral late gene transcription. Finally, it is tempting to speculate that regulation of late gene expression may function as a switch to decide a lytic or latent viral infection. Cellular or viral repressors may associate with viral late promoters by default to keep them silent, but viral regulators such as pM79 could displace these repressors to favor a productive, lytic infection. Work is underway to identify viral and cellular binding partners of pM79 in order to reveal the mechanism of its activity.

The identification of pM79 as an essential protein for MCMV late transcription indicates that it is a functional homologue of HCMV pUL79. MCMV infection in mice is an important model for preclinical evaluation of antiviral compounds, most of which have been directed at highly conserved viral DNA replication proteins. The pM79/pUL79 protein family presents an attractive, alternative target for therapeutic intervention in CMV disease. The functional conservation between pM79 and pUL79 justifies MCMV as a credible model to test this novel antiviral strategy in vivo.
Acknowledgements

We thank Herbert Virgin and the members of his laboratory for helpful discussion and invaluable advice; Dr. Ulrich Koszinowski (Max von Pettenkofer-Institute, Ludwig Maximilians-University, Germany), Dr. Martin Messerle (Hannover Medical School, Hannover, Germany), and Dr. Wolfram Brune (Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Germany) for the MCMV BAC clone pSM3fr; Dr. Anthony Scalzo (University of Western Australia) for M44 and gB antibodies; Dr. Stipan Jonjic (University of Rijeka, Croatia) for IE1 and E1 antibodies; Jian Gao (Washington University) for assistance in array analysis, and members of the Yu lab for critical reading of the manuscript.

We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1RR024992 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH.

This study was supported by Public Health Service grants (RO1CA120768 and RO1AI51345). D.Y. holds an Investigators in the Pathogenesis of Infectious Disease award from the Burroughs Wellcome Fund and W.M.Y. is an investigator of the Howard Hughes Medical Institute.
Material and methods

Plasmids, antibodies, and chemicals. pYD-C245 and pYD-C571 were retroviral expression vectors derived from pRetro-EBNA (18). pYD-C245 expressed the red fluorescent protein (DsRed) (4) from an internal ribosome entry site (IRES). pYD-C571 was derived from pYD-C245. It carried the coding sequence of C-terminally 1x FLAG tagged M79 that was expressed together with DsRed as a bicistronic transcript. pYD-C191 carried a kanamycin selection cassette bracketed by two Flp recognition target (FRT) sites. pYD-C630 was derived from pGalK (44) and carried a FRT-bracketed GalK/kanamycin dual selection cassette (29). pYD-C746 was derived from pYD-C630, where a 3x FLAG sequence preceded the FRT-bracketed selection cassette.

The primary antibodies used in this study included: anti-actin (clone AC15; Abcam); anti-FLAG polyclonal rabbit antibody (F7425) and monoclonal mouse antibody (F1804) (Sigma); anti-MCMV IE1 (CROMA101) and E1 (CROMA103) (generous gifts from Dr. Stipan Jonjic, University of Rijeka, Croatia); anti-MCMV M44 (3B9.22A) and gB (2E8.21A) (generous gifts from Dr. Anthony Scalzo, University of Western Australia). The secondary antibody used for immunoblotting was horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG (Jackson Laboratory). The secondary antibodies used for immunofluorescence were Alexa Fluor 594-conjugated goat anti-mouse IgG and Alexa Fluor 488-conjugated goat anti-rabbit IgG (Invitrogen-Molecular Probes).

Other chemicals used in this study include phosphonoacetic acid (PAA) (284270-10G; Sigma Aldrich); L-(-)-Arabinose (A3256-25G; Sigma Aldrich); and TO-PRO3 iodide (T3605; Invitrogen).

Cells and viruses. Mouse embryonic fibroblast 10.1 cells (MEF10.1) (13) were propagated in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum, nonessential amino acids, and 1 mM sodium pyruvate. Cells were maintained at 37°C and 5% CO₂ in a humidified atmosphere. To create cell lines stably expressing FLAG-tagged M79 (10.1-M79flag), MEF10.1 cells were transduced three times with retrovirus reconstituted from pYD-C571 and allowed to recover for 48 hours. Clonal cells expressing DsRed were isolated by limiting dilution and expanded to produce stocks of cell lines. Individual clonal cell lines were tested by transfecting them with the recombinant MCMV BAC clone pSMin79 (see below) and determining the titer of reconstituted virus at 5 days post transfection. The cell line that yielded the
highest titer was used in this study. Recombinant MCMV viruses SM<sub>gfp</sub>, SM<sub>rev</sub>79, and SM79<sub>flag</sub> (see below) were reconstituted from electroporation of corresponding BAC clones into MEF10.1 cells. Recombinant virus SM<sub>in</sub>79 (see below) was reconstituted from electroporation of the BAC clone pSM<sub>in</sub>79 into 10.1-M79<sub>flag</sub> cells.

**BAC recombineering.** Recombinant MCMV BAC clones used in this study were derived from the self-excisable parental MCMV BAC clone, pSM3fr, which carried a full length genome of the MCMV Smith strain (43). All recombinant MCMV BAC clones in this study were created using the linear recombination-based BAC recombineering protocol that we have previously established (29). Recombination was carried out in *E. coli* strain SW105 that harbored an MCMV BAC clone and expressed an arabinose inducible *Flippase* gene for transient expression of Flp recombinase (44). We inserted the green fluorescent protein (GFP) expression cassette at the C-terminus of the IE2 loci within pSM3fr to produce the BAC clone pSM<sub>gfp</sub>. This clone was used to produce wild-type virus SM<sub>gfp</sub> as IE2 has been shown to be dispensable for MCMV infection *in vivo* and *in vitro* (6, 7, 24). We independently confirmed that the insertion of the GFP cassette at this locus had no deleterious consequence on virus growth in our infection system (data not shown). The BAC clone pSM<sub>in</sub>79 carried a frame-shift mutation in the viral gene M79 (Fig. 1A). To construct pSM<sub>in</sub>79, the FRT-bracketed GalK/kanamycin cassette was PCR amplified from pYD-C630 and recombined into pSM<sub>gfp</sub> at 403 nucleotide (nt) downstream of the start codon of the M79 coding sequence. Transformants were selected by kanamycin resistance. The selection cassette was subsequently removed by Flp-FRT recombination, leaving an 88-nt insert within the M79 coding sequence and creating a frame-shift mutation. The BAC clone pSM<sub>rev</sub>79 contained a C-terminally 3x FLAG-tagged M79 (Fig. 1A). To construct pSM<sub>rev</sub>79, a DNA fragment that contained the FRT-bracketed GalK/kanamycin selection cassette preceded by a 3x FLAG sequence was PCR amplified from pYD-C746 and recombined into the C-terminus of the M79 coding sequence. The selection cassette was subsequently removed by Flp-FRT recombination, resulting in the 3x FLAG fused in frame with the M79 coding sequence. The BAC clone pSM<sub>rev</sub>79 was derived from pSM<sub>in</sub>79 and contained the repaired M79 coding sequence. To create pSM<sub>rev</sub>79, a patched PCR fragment containing the wild-type M79 coding sequence followed by a FRT-
bracketed kanamycin selection cassette was recombined into pSMin79 to replace the M79 frame-shift mutation. The selection cassette was subsequently removed by Flp-FRT recombination, leaving an 81-bp sequence insert after the stop codon of the M79 coding sequence, which had no deleterious effect on virus replication (Fig. 1D). All the final BAC clones were validated by restriction digestion, PCR analysis, and direct sequencing as previously described (46).

To reconstitute recombinant viruses that did not require complementation to grow, confluent MEF10.1 cells were electroporated with 5 µg of MCMV BAC DNA and plated on a 10 cm plate. Culture medium was changed 24 hours post transfection, and virus was harvested by collecting cell free culture medium after the entire monolayer of cells was lysed. Alternatively, virus stocks were produced by collecting cell free supernatant from infected culture at a multiplicity of infection (MOI) of 0.001. Virus titers were determined in duplicate by a tissue culture infectious dose 50 (TCID$_{50}$) assay in MEF10.1 cells. To reconstitute, propagate, and titer SM$_{in}$79 virus, 10.1-M79flag cells were used as described above. In experiments where comparative analysis was performed between SM$_{in}$79 and other recombinant viruses, titers of all viruses were determined in 10.1-M79flag cells.

**Viral growth analysis.** MEF10.1 cells were seeded in 12-well plates overnight to produce a confluent monolayer. Cells were inoculated with recombinant MCMV viruses for 1 hour at a MOI of 2 for single step or 0.01 for multistep growth analysis. The inoculum was removed, the infected monolayer was rinsed with phosphate-buffered saline (PBS), and fresh medium was replenished. At various times post infection, cell-free virus was collected in duplicate by harvesting medium from infected cultures. Cell-associated virus was collected by rinsing infected cells once with PBS and scraping cells into fresh medium. Cells were lysed by one freeze-thaw cycle followed by sonication. Lysates were cleared of cell debris by low speed centrifugation and supernatants were saved as cell-associated virus. Virus titers were determined by TCID$_{50}$ assay.

**DNA and RNA analysis.** Intracellular DNA was measured by quantitative PCR (qPCR) as previously described (46). Briefly, MCMV-infected cells were collected in a lysis buffer (200 mM NaCl, 20 mM Tris [pH 8.0], 20 mM EDTA, 0.2 mg/ml proteinase K, 0.4% sodium dodecyl sulfate [SDS]), and lysed by
incubation at 55°C overnight. DNA was extracted with phenol-chloroform and treated with RNase A (100 µg/ml) at 37°C for 1 hour. Samples were extracted again with phenol-chloroform, precipitated with ethanol, and resuspended in nuclease-free water (Ambion). Viral DNA was quantified by qPCR using SYBR Advantage qPCR Premix (Clonetech) and a primer pair specific for the MCMV IE1 or M55 gene (Table 1). Cellular DNA was quantified using a primer pair specific for the mouse actin gene (Table 1) (45). A standard curve was generated using serially diluted pSMgf BAC DNA or DNA from infected cells, and used to calculate relative amounts of viral or cellular DNA in a sample. The amount of viral DNA was normalized by dividing IE1 or M55 equivalents over actin equivalents. The normalized amount of viral DNA in SMgf infected cells at 2 hour post infection (hpi) was set at 1.

Intracellular RNA was determined by reverse transcription-coupled qPCR (RT-qPCR) as previously described (46). Total RNA was extracted by Trizol reagent (Invitrogen) and treated with TURBO DNA-free reagents (Ambion) to remove contaminating DNA. First strand cDNA synthesis was performed with the High Capacity cDNA Reverse Transcription Kit using random hexamer primered total RNA (Applied Biosystems). Each sample also included a control without the addition of reverse transcriptase to determine the level of residual contaminating DNA. cDNA was quantified using SYBR Advantage qPCR Premix (Clonetech) and primer pairs specific for viral genes or the mouse actin gene (Table 1). A standard curve was generated for each gene using serially diluted cDNA from infected cells and used to calculate the relative amount of a transcript in each sample. The amounts of viral transcript were normalized by dividing viral transcript equivalents over actin equivalents. The normalized amount of transcript during SMgf infection at 10 hpi in the absence of PAA was set to 1.

Tiled array design, experimental procedure, and analysis. The array was designed using Agilent’s eArray package. The array consisted of 103,347 60-mer oligonucleotide probes for each strand of the MCMV genome, and each probe advanced 4-5nt. In addition, the array also contained probes complementary to Agilent’s RNA Spike-Ins, and 400 negative control probes against Arabidopsis with no homology to mouse or MCMV genomes.

To prepare viral RNA, total RNA was harvested from infected MEF10.1 cells at 20hpi using Trizol (Invitrogen) and mRNA was purified using RNAeasy columns (Qiagen) according to the manufacturer's
instructions. cDNA probes were synthesized, fluorescently labeled by random hexamer-primed polymerization using the SuperScript Plus Indirect cDNA Labeling module (Invitrogen), and hybridized to the array chip at Genome Technology Access Center of Washington University School of Medicine (GTAC). Agilent’s RNA Spike-In Kit was used to monitor the linearity, sensitivity, and accuracy of the array.

The total signal intensity for each probe was log2 transformed, and the mock signal was subtracted from the experimental signal after they were normalized to each other using the spike-in RNA signals. To enable comparisons among samples, raw data from each MCMV infected sample was normalized using spike-in controls. This normalization was further refined and validated using qRT-PCR data for several viral probes. Normalized intensities of experimental probes were mapped back to the MCMV genome, and mean fluorescence of each nucleotide was calculated from all overlapping probes. Changes in fluorescence intensity greater than 3-fold between compared samples were used for data interpretation. MCMV open reading frames (ORFs) were annotated based on the studies by Rawlinson and coworkers (34) and by Cheng and coworkers (9), and updated with details from additional publications whenever possible. Data were converted to gff3 file format and visualized using gBrowse (42). The data are in the process of being deposited into the Gene expression Omnibus (GEO) database.

**Pulsed-field gel electrophoresis (PFGE) and southern blot analysis.** PFGE was performed on a Bio-Rad CHEF Mapper XA Pulsed Field Electrophoresis System. To prepare DNA from infected cells, MEF10.1 cells were seeded onto a 60 mm dish at a density of 1.4 x 10^6 cells per dish and infected at a MOI of 2. At 36 hpi, cells were scraped off the dish, collected by centrifugation at 200 xg for 5 minutes, and resuspended in 180 μl of 55°C 1% low-melting-point agarose (NuSieve GTG Agarose, Lonza) in PBS. 90μl of cell suspension was casted into a disposable casting mold (Bio-Rad) and solidified at 4°C for 15 minutes. To prepare DNA from cell-free virions, liquid viral stock equivalent to 10^6 PFU was casted in one low-melting-point agarose block. Agarose blocks were transferred into lysis buffer (20 mM Tris.Cl pH8.0, 200 mM NaCl, 400 mM EDTA, 1% SDS, 1 mg/mL proteinase K) to lyse imbedded samples by incubation at 37°C overnight. Blocks were then rinsed five times with TE buffer (10mM Tris-Cl, pH8.0, 0.1mM EDTA) at 50°C for 15 minutes each and stored in TE Buffer at 4°C.
Restriction enzyme digestions were carried out by incubating one-half of a block (~45uL) in 200µl digestion buffer containing 50 units of Pac I at 4°C overnight, and then at 37°C for 6 hours. Digested blocks were loaded into wells of a 1% megabase agarose gel (Biorad Pulse Field Certified Agarose) in 0.5 x TBE (0.045 M Tris pH 8.0, 0.045 M boric acid, 0.001 M EDTA pH 8.0) and sealed with 1% low melting-point agarose in 0.5 x TBE. PFGE was performed at 6 V/cm at 14°C for 16 hours, with a linear pulse increase from 0.26 to 20.01 seconds over the course of the run.

Resolved viral fragments were analyzed by Southern blot analysis as previously described (33). Briefly, the gel was transferred to a Nytran SuPerCharge Membrane as described in the Turboblotter transfer system protocol (VWR), and the membrane was air dried and cross-linked by UV at 125mJ. The 32P-labeled probe was generated by random hexamer-priming of pSMgfp BAC DNA using the Prime-It II Labeling Kit (Stratagene), purified with ProbeQuant G-50 micro columns (Amersham), and denatured in 10mM EDTA at 90°C for 10 minutes before use. The membrane was pre-hybridized with 10 ml ULTRAhyb solution (Ambion) at 42°C for 1 hour, and hybridized with 10⁷ cpm of denatured probes at 42°C for 3 hours. The membrane was washed once with buffer I (2x SCC and 0.1% SDS) for 10 minutes followed by two washes with buffer II (0.1x SSC and 0.1% SDS) for 15 minutes before exposure on a Kodak film.

**Protein analysis.** Protein accumulation was analyzed by immunoblot. Cells were washed and lysates were collected in sodium dodecyl sulfate (SDS)-containing sample buffer. Proteins were resolved by SDS-containing polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto a polyvinylidene difluoride (PVDF) membrane. Proteins of interest were detected by hybridizing the membrane with specific primary antibodies followed by HRP-coupled secondary antibodies, and visualized by using SuperSignal West Pico enhanced chemiluminescent (ECL) substrate (Thermo Scientific).

Intracellular localization of proteins of interest was analyzed by immunofluorescence assay. Cells were seeded onto cover-slips 24 hours prior to infection. At various times, cells were washed with PBS, fixed and permeabilized with methanol (-20°C) for 10 minutes, and blocked with 5% FBS in PBS at room temperature for 1 hour. Cells were incubated with primary antibodies for 30 minutes at room temperature and subsequently labeled with secondary antibodies coupled to Alexa Fluor 488 or Alexa Fluor 594.
(Invitrogen-Molecular Probes). Cells were counterstained with TO-PRO3 and mounted on slides with
Prolong Gold antifade reagent (Invitrogen-Molecular Probes). Confocal microscopic images were
captured by a Zeiss LSM510 Meta confocal laser scanning microscope.
References


## TABLE 1. Primers used in quantitative PCR analysis

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMV IE1 forward</td>
<td>5'-CAGGGTGGGATCATGAGCCT-3'</td>
</tr>
<tr>
<td>MCMV IE1 reverse</td>
<td>5'-AGCCGCATCGAAAGACAACG-3'</td>
</tr>
<tr>
<td>MCMV M25 forward</td>
<td>5'-AAGACATGTACGCGACGGA-3'</td>
</tr>
<tr>
<td>MCMV M25 reverse</td>
<td>5'-CTATTGCCCATCATCGCCCG-3'</td>
</tr>
<tr>
<td>MCMV gB (M55) forward</td>
<td>5'-GCAGATGTCGAGTGCTGTCAAG-3'</td>
</tr>
<tr>
<td>MCMV gB (M55) reverse</td>
<td>5'-GCACAGCGCGGCTCGAATAAC-3'</td>
</tr>
<tr>
<td>MCMV M74 forward</td>
<td>5'-AGGAGGCTGTGAACCTTGAAA-3'</td>
</tr>
<tr>
<td>MCMV M74 reverse</td>
<td>5'-CTCATCAGCGTCTCAGAAAAG-3'</td>
</tr>
<tr>
<td>MCMV M79 forward</td>
<td>5'-CTACCTGAGCGCTGGAGAAAG-3'</td>
</tr>
<tr>
<td>MCMV M79 reverse</td>
<td>5'-TAGTCCTGGATACAGAGGAAAAG-3'</td>
</tr>
<tr>
<td>MCMV M112/113 (E1) forward</td>
<td>5'-GAATCCGAGGAGGAAGACGAT-3'</td>
</tr>
<tr>
<td>MCMV M112/113 (E1) reverse</td>
<td>5'-GGTGAACGGTTTGCTCAGATCTC-3'</td>
</tr>
<tr>
<td>MCMV M116 forward</td>
<td>5'-TCCTTGGGTGTGATGCGGT-3'</td>
</tr>
<tr>
<td>MCMV M116 reverse</td>
<td>5'-GCATCCCGTACCTGACCA-3'</td>
</tr>
<tr>
<td>MCMV M121 forward</td>
<td>5'-CCGTTCGGCTTTGTAACACTG-3'</td>
</tr>
<tr>
<td>MCMV M121 reverse</td>
<td>5'-GCTTCTCGAGGCAGCAGCA-3'</td>
</tr>
<tr>
<td>Mouse actin forward</td>
<td>5'-GCTGTATTCCCCTTCTCAGTG-3'</td>
</tr>
<tr>
<td>Mouse actin reverse</td>
<td>5'-CAAGGTGCTCTCTATTGTA-3'</td>
</tr>
<tr>
<td>ORF</td>
<td>Strand</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>m011</td>
<td>Y</td>
</tr>
<tr>
<td>m012</td>
<td>Y</td>
</tr>
<tr>
<td>m014</td>
<td>Y</td>
</tr>
<tr>
<td>M032</td>
<td>C</td>
</tr>
<tr>
<td>M035</td>
<td>Y</td>
</tr>
<tr>
<td>m039</td>
<td>C</td>
</tr>
<tr>
<td>m040</td>
<td>C</td>
</tr>
<tr>
<td>M046</td>
<td>C</td>
</tr>
<tr>
<td>M053</td>
<td>Y</td>
</tr>
<tr>
<td>M055</td>
<td>C</td>
</tr>
<tr>
<td>M056</td>
<td>C</td>
</tr>
<tr>
<td>M072</td>
<td>C</td>
</tr>
<tr>
<td>M073</td>
<td>Y</td>
</tr>
<tr>
<td>M074</td>
<td>C</td>
</tr>
<tr>
<td>M075</td>
<td>C</td>
</tr>
<tr>
<td>M076</td>
<td>Y</td>
</tr>
<tr>
<td>M077</td>
<td>Y</td>
</tr>
<tr>
<td>M080</td>
<td>Y</td>
</tr>
<tr>
<td>M084</td>
<td>C</td>
</tr>
<tr>
<td>M085</td>
<td>C</td>
</tr>
<tr>
<td>M086</td>
<td>C</td>
</tr>
<tr>
<td>M096</td>
<td>Y</td>
</tr>
<tr>
<td>M100</td>
<td>C</td>
</tr>
<tr>
<td>M104</td>
<td>C</td>
</tr>
<tr>
<td>M114</td>
<td>C</td>
</tr>
<tr>
<td>M115</td>
<td>C</td>
</tr>
<tr>
<td>M116</td>
<td>C</td>
</tr>
<tr>
<td>M118</td>
<td>C</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>m119.1</td>
<td>C</td>
</tr>
<tr>
<td>m131</td>
<td>C</td>
</tr>
<tr>
<td>m155</td>
<td>C</td>
</tr>
<tr>
<td>m156</td>
<td>C</td>
</tr>
<tr>
<td>m157</td>
<td>C</td>
</tr>
<tr>
<td>m158</td>
<td>C</td>
</tr>
<tr>
<td>m159</td>
<td>C</td>
</tr>
<tr>
<td>m160</td>
<td>C</td>
</tr>
<tr>
<td>m161</td>
<td>C</td>
</tr>
<tr>
<td>m162</td>
<td>C</td>
</tr>
<tr>
<td>m163</td>
<td>C</td>
</tr>
<tr>
<td>m165</td>
<td>C</td>
</tr>
<tr>
<td>m168</td>
<td>Y</td>
</tr>
<tr>
<td>m164</td>
<td>C</td>
</tr>
<tr>
<td>m166</td>
<td>C</td>
</tr>
<tr>
<td>m001</td>
<td>C</td>
</tr>
<tr>
<td>m002</td>
<td>Y</td>
</tr>
<tr>
<td>m003</td>
<td>Y</td>
</tr>
<tr>
<td>m007</td>
<td>Y</td>
</tr>
<tr>
<td>m015</td>
<td>Y</td>
</tr>
<tr>
<td>m016</td>
<td>Y</td>
</tr>
<tr>
<td>m018</td>
<td>C</td>
</tr>
<tr>
<td>m019</td>
<td>Y</td>
</tr>
<tr>
<td>m021</td>
<td>Y</td>
</tr>
<tr>
<td>m022</td>
<td>Y</td>
</tr>
<tr>
<td>M023</td>
<td>C</td>
</tr>
<tr>
<td>m023.1</td>
<td>Y</td>
</tr>
<tr>
<td>M024</td>
<td>C</td>
</tr>
<tr>
<td>M025</td>
<td>Y</td>
</tr>
</tbody>
</table>
a "C", complementary strand.

b "Y", PAA treatment reduces the mean intensity of a transcriptional region by at least 3-fold in array analysis.

c "Y", M79 mutation reduces the mean intensity of a transcriptional region by at least 3-fold in array analysis.

d Transcription of an ORF is considered to be late kinetics ("L") when it is elevated by at least 50% from 12 to 24 hpi as described by Marcinowski and coworkers (25), or absent in 6.5 hpi but becomes detectable at 24 hpi as described by Lacaze and coworkers (19). "ND", kinetics class has not been previously described.
Figure 1. Gene M79 is essential for MCMV growth in fibroblasts. (A) Diagram depicting recombinant MCMV BAC clones used in this study. The BAC clone pSM79flag carried a 3xFLAG tag that was fused in frame at the C-terminus of the M79 coding sequence (indicated by shaded region). The BAC clone pSMin79 carried an 88-bp insert (indicated by black region) at 403-nt downstream of the start codon of the M79 coding sequence, resulting in a frame-shift mutation. See Materials and Methods for details. (B)
Growth of SM\textit{in}79 virus on MEF10.1 cells expressing FLAG-tagged M79 (10.1-M79\textit{flag}). Left panel; virus-driven GFP expression in normal MEF10.1 cells or 10.1-M79\textit{flag} cells 7 days post transfection with pSM\textit{in}79. Right panel; titers of wild-type virus (SM\textit{gfp}) and M79 mutant virus (SM\textit{in}79) produced at 72 hour post infection (hpi) in 10.1-M79\textit{flag} cells that were infected at a multiplicity of infection (MOI) of 2. Shown is a representative from at least two reproducible, independent experiments. (C-D) Growth kinetic analysis of M79 recombinant viruses used in this study. Normal MEF10.1 cells were infected at a MOI of 2, cell free and cell associated viruses were collected at indicated times, and viral titers were determined by tissue culture infectious dose 50 (TCID\textsubscript{50}) assay in 10.1-M79\textit{flag} cells. The detection limit of TCID\textsubscript{50} assay is indicated by a dashed line. Shown is a representative from two reproducible, independent experiments.
Figure 2. Expression of M79 gene products is markedly enhanced by viral DNA synthesis and protein pM79. (A) Accumulation of the M79 protein product during MCMV infection. MEF10.1 cells were infected with SM79flag virus at a MOI of 2, total cell lysates were collected at indicated times and analyzed by immunoblotting. The M79 protein was detected with anti-FLAG antibody. Actin was used as a loading control, and viral proteins IE1, E1 and gB were used as representative immediate early, early, and late proteins, respectively. Shown is a representative from three reproducible, independent experiments. (B) Accumulation of the M79 transcript during MCMV infection. MEF10.1 cells were infected with SMgfp in the presence or absence of viral DNA synthesis inhibitor phosphonoacetic acid (PAA) (200 µg/ml) or with SMin79 at a MOI of 2. Total RNA was isolated at indicated times, and the amount of M79 transcript was measured by reverse transcription-coupled quantitative PCR (RT-qPCR) analysis with the primers listed in Table 1. The values were normalized to that of actin, and the normalized amount of M79 transcript during SMgfp infection at 10 hpi in the absence of PAA was set to 1. Shown is a representative from three reproducible, independent experiments.
Figure 3. pM79 localizes to replication compartments during infection. MEF10.1 cells were mock infected or infected with SMgfp or SM79flag at a MOI of 2. At 24 hpi, cells were fixed with methanol (which quenches GFP fluorescence), and stained with either rabbit polyclonal (A) or mouse monoclonal (B) anti-FLAG antibody for detection of the tagged M79 protein (green). In panel A, cells were co-stained with antibody to pM44 to mark replication compartments (red). The last row of images in panel A represent the magnified view of an infected nucleus where the M79 protein (pM79) localized. Cells were counterstained with TO-PRO3 to visualize the nuclei (blue). Scale bars are equivalent to 20 µm. Shown is a representative from four reproducible, independent experiments.
Figure 4. pM79 is not required for viral DNA synthesis. (A) Accumulation of viral DNA during pM79-deficient virus infection. MEF10.1 cells were infected with SMgfpr or SMm79 at a MOI of 2 and total DNA was harvested from infected cells at indicated times. Viral DNA accumulation was analyzed by qPCR using primers specific for viral genes IE1 (left panel) or M55 (right panel) (see Table 1 for primer
sequences), and the values were normalized to that of actin. The normalized values of viral DNA during SMgfp infection at 2 hpi were set to 1. Shown is a representative from five reproducible, independent experiments. (B) Pulsed field gel electrophoresis (PFGE) analysis of intracellular viral DNA during M79-deficient virus infection. MEF10.1 cells were infected with SMgfp or SMin79 at a MOI of 2, and collected at 36 hpi. Cell-free virions or infected cells were suspended in a low melting agarose block, lysed, digested with Pac I, and subjected to pulsed field gel electrophoresis. The gel was then transferred to a membrane, hybridized with a $^{32}$P-labeled probe specific to the entire SMgfp BAC sequence. The positions of wells, high-molecular-weight viral DNA (HMW DNA), 232-kb monomer viral DNA, and prominent digested viral DNA fragments are indicated. Shown is a representative from two reproducible, independent experiments. The top panel shows the schematic diagram of monomer and concatamer viral DNA with Pac I sites indicated.
Figure 5. **pM79 is not required for the maturation of viral replication compartments.** MEF10.1 cells were infected with SMgfp or SMin79 virus at a MOI of 2. At 12 and 24 hpi, cells were fixed with methanol and stained with antibody to pM44 to mark replication compartments (red). Cells were also counterstained with TO-PRO3 to visualize the nuclei (blue). Scale bars are equivalent to 20 µm. Shown is a representative from three reproducible, independent experiments.
Figure 6. pM79 is required for efficient expression of a representative viral late gene. (A) Accumulation of representative viral proteins during pM79-deficient virus infection. MEF10.1 cells were infected with SMgf or SMm79 at a MOI of 2, and the accumulation of viral proteins IE1, E1, and gB at indicated times was analyzed by immunoblotting. Shown is a representative from five reproducible, independent experiments. (B) Failure of gB accumulation during pM79-deficient virus infection cannot be rescued by elevating expression of viral early genes. Cells were infected with SMgf or SMm79 at a MOI of 1 or 5, respectively, so SMm79 infected cells expressed E1 protein at levels comparable to that in SMgf infected cells. Cell lysates were collected at indicated times and analyzed by immunoblotting. Shown is a representative from three reproducible, independent experiments. (C) Accumulation of viral transcripts during pM79-deficient virus infection. Cells were infected with SMgf in the presence or
absence of 200 μg/ml PAA, or with SMi79 as described in (A). Total RNA was isolated at indicated times, and amounts of viral E1 and M55 transcript were measured by RT-qPCR with the primers listed in Table 1. Shown is a representative from at least three reproducible, independent experiments. The values were normalized to that of actin, and normalized values of viral transcript during SMgfp infection at 10 hpi in the absence of PAA were set to 1.
Figure 7. Tiled array analysis of genome-wide transcription during MCMV infection. MEF10.1 cells were infected with SMgfp in the presence or absence of PAA, or with SMin79 at a MOI of 2. Total RNA was isolated at 20 hpi, reverse transcribed, and labeled cDNAs were hybridized to an oligonucleotide tiled array of the MCMV genome. The mean fluorescence of probes overlapping each nucleotide position is plotted on a log2 scale underneath the annotated viral genomic sequence. Blue or red arrows represent annotated open reading frames on the positive or negative strand of the viral genome, respectively. The colored lines represent the transcriptional signals from SMgfp infection (black), SMgfp infection with PAA (red), and SMin79 infection (blue). These probe intensities were compared on a nucleotide by nucleotide basis between SMgfp infections with or without PAA treatment, or between infections of SMgfp and SMin79. Regions in which the fluorescent intensity was reduced by greater than 3-fold by PAA treatment or mutation of M79 are plotted on a linear scale below the transcriptional intensity plots, and labeled as "PAA" or "in79", respectively; scales for these plots reflect fold-reduction. (A) Genomic sequence 1-80 kb. (B) Genomic sequence 80-160 kb. (C) Genomic sequence 160-230 kb. Shown is a representative from two reproducible, independent experiments.
Figure 8. RT-qPCR analysis of representative PAA-sensitive transcripts in MCMV infection.

MEF10.1 cells were infected with SMgfp in the presence or absence of 200 μg/ml PAA, or with SMin79 at a MOI of 2. Total RNA was isolated at indicated times, and amounts of indicated late transcripts were measured by RT-qPCR with the primers listed in Table 1. The values were normalized to that of actin, and normalized values of viral transcript during SMgfp infection at 10 hpi in the absence of PAA were set to 1.
Chapter III

Murine Cytomegalovirus Protein pM92 is a Conserved Regulator of Viral Late Gene Expression
Preface

TJC designed and performed the majority of experiments, analyzed the data, generated the figures and wrote the first draft of the manuscript. YCP developed the UL92 and M92 MRC5 complementing cells and created the UL92 mutant HCMV, ARF\(^1\) provided reagents, advice, and support for completion of the study, DY supervised the studies and greatly contributed to the design and analysis of the experiments, TJC and DY greatly contributed to the writing and editing of the manuscript.

This chapter has been previously published:

Summary

In this study, we report that murine cytomegalovirus (MCMV) protein pM92 regulates viral late gene expression during virus infection. Previously, we have shown that MCMV protein pM79 and its human cytomegalovirus (HCMV) homologue pUL79 are required for late viral gene transcription. Identification of additional factors involved is critical to dissecting the mechanism of this regulation. Here we showed that pM92 accumulated abundantly at late times of infection in a DNA synthesis-dependent manner and localized to nuclear viral replication compartments. To investigate the role of pM92, we constructed a recombinant virus SMint92, in which pM92 expression was disrupted by an insertional/frame-shift mutation. During infection, SMint92 accumulated representative viral immediate early gene products, early gene products, and viral DNA sufficiently but had severe reduction in the accumulation of late gene products, thus unable to produce infectious progeny. Co-immunoprecipitation and mass spectrometry analysis revealed the interactions between pM92 and pM79 as well as between their HCMV homologues pUL92 and pUL79. Importantly, we showed that the growth defect of pUL92-deficient HCMV could be rescued in trans by pM92. This study indicates that pM92 is an additional viral regulator of late gene expression, that these regulators (represented by pM92 and pM79) may need to complex with each other for their activity, and that pM92 and pUL92 share a conserved function in CMV infection. pM92 represents a potential new target for therapeutic intervention in CMV disease, and a gateway into studying a largely uncharted viral process that is critical to the viral life cycle.
Introduction

Human cytomegalovirus (HCMV), the prototypical member of the betaherpesvirus subfamily, is a ubiquitous pathogen limited to the human host (37). After the resolution of acute infection, HCMV establishes a persistent, life-long infection characterized by alternate stages of virus production and latency (37). In immuno-competent hosts, the infection is typically asymptomatic. However, in immuno-compromised hosts, lytic infection, during both primary infection and reactivation from latency, can cause significant morbidity and mortality (37, 48). HCMV is the leading viral cause of birth defects, such as deafness and mental retardation, in perinatally infected infants (1, 20). It is a major cause of retinitis and blindness in AIDS patients (48). It is a common source of infectious complications in transplant recipients and cancer patients (50). Emerging evidence also provides possible association of HCMV infection with cardiovascular disease and proliferative diseases such as cancer (47). Currently there is no vaccine to this virus, and antiviral therapies are limited by poor toxicity scores, low availability, and emergence of resistant viruses (6). Understanding the role of viral genes in lytic infection is paramount and will yield novel targets for antiviral therapies.

Murine CMV (MCMV) is the homologue of HCMV and model of choice to study CMV biology and pathogenesis. It shares conserved features with HCMV with regard to virion structure, genome organization, gene expression, tissue tropism, and clinical manifestations (18, 42-44). Many genes of MCMV are conserved in HCMV and its ability to infect mice provides a tractable small animal model to investigate virus infection in vivo. The use of MCMV to explore conserved viral genes will shed light on the roles of their counterparts in replication and pathogenesis of HCMV.

The lytic replication cycles of herpesviruses are characterized by highly ordered cascades of gene expression, which can be sequentially divided into immediate early (IE), early (E), and late (L) phases (37). Expression of IE genes only requires cellular factors and viral proteins associated with incoming virions. IE proteins transactivate expression of early genes that are required for viral DNA synthesis. Many of the early proteins localize to viral nuclear replication compartments, where viral DNA synthesis, late gene transcription, and viral genome encapsidation take place (37). Following viral DNA synthesis, late genes, many of which encode structural proteins, are expressed to allow virion assembly, maturation, and egress. It is also worth noting that some genes have both early and late properties; their
transcriptions start prior to viral DNA synthesis, but the accumulation of their transcripts is enhanced considerably by DNA synthesis.

Though the regulation of IE and E gene expression has been studied extensively, less is known regarding the regulation of late gene expression during CMV infection. Previously we have shown that MCMV protein pM79 is dispensable for viral DNA synthesis but is a key regulator of late gene transcription during MCMV infection (10). pM79 is conserved in both beta- and gamma-herpesviruses, and its homologues in HCMV (pUL79) and MHV68 (ORF18) have been shown to play similar roles during virus infection (3, 27, 41). Identifying additional viral regulatory factors in late gene expression will be critical to understand this process. In this study, we report that pM92 is another key regulator of MCMV late gene transcription. Like pM79, pM92 is also conserved in both beta- and gamma-herpesviruses. Its homologues include pUL92 of HCMV, ORF31 of MHV68, and U63 of HHV-6. However, the role of pM92 during MCMV infection has not been defined, even though genome-wide mutagenesis analyses have previously shown that both MHV68 ORF31 and HCMV pUL92 are essential for lytic virus replication. Here we create a pM92-deficient MCMV mutant virus, and show that in the absence of pM92, MCMV is capable of synthesizing its DNA at wild-type levels but unable to efficiently produce late gene products. The M92 gene products abundantly accumulate at 20 hours post infection (hpi) and localize to nuclear replication compartments. We also provide evidence that pM92 interacts with pM79 during virus infection, and that their HCMV homologues pUL92 and pUL79 interact as well. These results support the role of pM92 as a key regulator of viral late transcription, and suggest that pM92 and pM79 are part of multi-component regulatory complex controlling late transcription. Finally, we demonstrate the functional conservation between MCMV pM92 and HCMV pUL92 by rescuing the growth defect of pUL92-deficient HCMV virus with pM92 expression in trans. pM79 and pM92 offer attractive targets for novel antivirals, and MCMV provides a powerful system to dissect the regulatory mechanism of CMV late gene transcription as well as to test antivirals targeting steps other than viral DNA synthesis.
Results

**pM92 is essential for MCMV replication in fibroblasts.** The M92 open reading frame (ORF) is predicted to encode a gene product of 231 amino acids (aa), and is a sequence homologue to the HCMV ORF UL92 (202aa). UL92 is essential for virus replication during HCMV infection (61), but the importance of M92 in MCMV infection has not yet been established. To investigate this, a frame-shift mutation was introduced at nt 358 of the predicted M92 ORF by BAC recombineering (10, 40), producing mutant BAC pSM{$\text{in}$_92} (Fig. 1A). This insertion is not expected to interfere with the expression of neighboring genes, particularly the 5’-terminally overlapping M93. Transfection of wild-type pSM{$\text{gfp}$_BAC} in MEF10.1 cells produced virus (termed SM{$\text{gfp}$_BAC}), resulting in complete CPE of the monolayer, and full spread of the virus-driven GFP expression at 5 days post transfection; whereas pM92-deficient virus (termed SM{$\text{in}$_92}) failed to show any sign of CPE even at two weeks post transfection (Fig. 1B). However, pSM{$\text{in}$_92} was rapidly reconstituted to wild-type levels from pSM{$\text{in}$_92} upon transfection into MEF 10.1 cell that stably expressed N-terminally 1×FLAG-tagged pM92 (10.1-$\text{flag}$_M92). Thus, the defect of SM{$\text{in}$_92} is the direct result of pM92 ablation.

To more precisely define the growth defect of SM{$\text{in}$_92}, we performed growth curve analyses to quantify the defect of the recombinant virus and validate the essentiality of pM92 (Fig. 1C). SM{$\text{in}$_92} failed to produce detectable levels of cell-free or cell-associated progeny through the entire course of analysis, indicating that pM92 is essential for MCMV replication at steps prior to virus release.

**pM92 is a 25kDa protein that accumulated at high levels at late times of infection.** A thorough search of available nucleotide and amino acid sequence databases failed to identify any significant homology of pM92 to proteins with known function. To acquire basic information and gain insights into the role of pM92, we first characterized potential protein and transcript products from this gene. As no antibody was available for detecting the M92 protein product, we created a recombinant MCMV BAC, pSM{$\text{flag}$_92}, in which the M92 coding sequence was tagged with 3×FLAG at the N-terminus (Fig. 1A). Transfection of pSM{$\text{flag}$_92} BAC in MEF 10.1 cells rapidly reconstituted recombinant virus, termed SM{$\text{flag}$_92}. Both single-step and multi-step growth curve analyses indicated that SM{$\text{flag}$_92} grew similarly to SM{$\text{gfp}$_BAC} (Fig. 1D), suggesting that the 3×FLAG tag did not interfere with the function of M92 or
neighboring gene M91 (Fig. 1A). 3×FLAG-tagged pM92 (pflagM92) migrated at the expected size of 25 kDa, became detectable at 24 hpi, and accumulated at more abundant levels at 48 hpi during SMflag92 infection (Fig. 2A). To profile its transcription, we determined the accumulation of M92 transcript by reverse transcription-coupled quantitative PCR analysis (RT-qPCR). In agreement with its protein accumulation profile, M92 transcript levels were low at 10 hpi but increased 8-fold at 20 hpi (Fig. 2B). Importantly, M92 transcription was dramatically reduced when viral DNA synthesis was inhibited by phosphonoacetic acid (PAA) at 20 hpi (Fig. 2B). Therefore, M92 gene products accumulate abundantly in a viral DNA synthesis-dependent manner at late times of infection.

**pM92 localizes to viral nuclear replication compartments during infection.** To further characterize pM92, we next examined the intracellular localization of 3×FLAG-tagged pM92 during infection of SMflag92. Infected cells were fixed and permeabilized with methanol, which also quenched GFP fluorescence, thus allowing visualization of pM92 localization using Alexa Fluor 488 conjugated mouse anti-flag antibody by indirect immunofluorescence. pM92 localized to the nucleus of infected cells, and in particular, it localized to subnuclear structures resembling those of nuclear replication compartments (Fig. 2C). This led us to hypothesize that pM92 localized to viral replication compartments during infection. To test this, we compared intracellular localization of 3x FLAG-tagged pM92 to that of viral protein pM44 (i.e., the viral polymerase processivity factor and commonly used marker for replication compartments) during infection of SMflag92. As the anti-pM44 antibody is of mouse origin, we used a rabbit anti-FLAG antibody to co-stain pM92 in this experiment. At 24hpi, FLAG staining strongly co-localized with pM44 in the nucleus of SMflag92-infected cells (Fig. 2D), indicating that the majority of pM92 localizes within replication compartments. The rabbit anti-FLAG antibody has been previously shown to have high background staining (10). In this experiment, it also produced a diffuse, weak cytoplasmic staining, which likely represented nonspecific background as it was also present in SMgfp infected control cells. Collectively, we conclude that pM92 is a nuclear protein that localizes to replication compartments during MCMV infection.
PM92 is dispensable for viral DNA synthesis but required for efficient Late gene expression. To define the function of PM92, we first determined where it acted in the viral lifecycle, hypothesizing that PM92 might be required for viral DNA synthesis as it localized to replication compartments (Fig. 2). However, quantitative PCR analysis showed that viral DNA accumulation over the course of SM\textit{in}92 infection was comparable to that during SM\textit{gfp} infection (Fig. 3C). This result indicates that PM92 is not required for viral DNA synthesis.

To test if PM92 is required for late gene expression, a viral event immediately downstream of viral DNA synthesis, we next examined the accumulation of immediate early protein IE1, early protein E1 (M112/113), and late protein gB (M55) during SM\textit{in}92 infection by immunoblot analysis. Compared to wild-type control SM\textit{gfp}, SM\textit{in}92 appeared to have two defects during infection. The first was a modest decrease in the accumulation of E1 protein at early times of infection. The second, and more striking defect, was that accumulation of late protein gB was reduced to undetectable levels in SM\textit{in}92 infection (Fig. 3A). Early genes primarily function prior to viral DNA synthesis. Since no measurable defect was observed in DNA synthesis during SM\textit{in}92 infection (Fig. 3C), the modest decrease in E1 accumulation was unlikely the main cause for the growth defect of SM\textit{in}92. It is unclear why SM\textit{in}92 has this minor defect in early gene expression. It may be a result of sub-optimal complementation by the 10.1-\textit{flag}M92 cells used to produce the mutant virus. It is also reminiscent of a similar observation reported for pM79-deficient virus (10). We therefore hypothesized that the inability of SM\textit{in}92 to replicate was likely due to the failure to efficiently produce late proteins during infection.

To determine if the defect was at the transcriptional level, we measured transcript accumulation of representative immediate early (IE1), early (E1), and late (M55) genes in the presence or absence of PM92 during virus infection by RT-qPCR. The expression kinetics of these genes was validated by treatment with the DNA synthesis inhibitor phosphonoacetic acid (PAA). As expected, both IE and E1 gene expression was resistant to PAA, whereas late gene M55 expression was markedly sensitive to PAA (Fig. 3B). Importantly, in the absence of PM92, IE1 and E1 transcripts accumulated at wild-type levels, but the accumulation of late transcript M55 was significantly reduced (Fig. 3B). Therefore, PM92 is required for efficient transcript accumulation of late gene M55 during MCMV infection.
This result led us to hypothesize that pM92 is a new member of the viral late transcription regulators, which includes recently reported pM79 (10). To test if pM92 had a global regulatory role in viral late transcription, we examined transcript accumulation of multiple early genes (M34, M37, M45, M102) and late genes (M46, M74, M85, M96, M116) during SM\textit{in}92 infection. Transcription of early genes was independent of viral DNA synthesis, and consequently only modestly affected by PAA treatment (< 2-fold), consistent with previous reports (Fig. 4A) (10). Importantly, the effect of pM92 mutation on early gene transcription was as modest as PAA treatment, indicating that pM92 is not required for efficient early gene expression. In stark contrast, late gene transcription was dependent on viral DNA synthesis, and therefore significantly sensitive to PAA (Fig. 4B). Importantly, transcription of these genes was also significantly reduced during SM\textit{in}92 infection. It was also noted that transcription of individual genes showed various dependency on pM92 relative to that on viral DNA synthesis. Reduction in M116 and M46 transcriptions in the absence of pM92 was comparable to that with PAA treatment, whereas the reduction in M96 and M74 transcription in the absence of pM92 was less pronounced. This is reminiscent of the previous report that different late gene transcriptions have different dependency on the viral late transcription regulator pM79 (10). Our results indicate that pM92 plays an important role in regulating late gene transcription during MCMV infection.

**CMV UL92/M92 proteins interact with UL79/M79 proteins during infection.** We have previously found that MCMV protein pM79 regulates viral late transcription (10). As we showed here that pM92 also played a critical role in late gene expression, we hypothesized that these two proteins might interact and form a functional complex to exert this regulatory activity. To test this, we created a retroviral vector expressing the C-terminally HA-tagged M79 ORF (pM79ha) or MCMV M38 ORF as a control (pM38ha), and transfected it into MEF10.1 cells to generate expression cells. Transfected cells were subsequently infected with SM\textit{flag}92 and cell lysates were collected at 48 hpi. FLAG-tagged pM92 complexes were immunoprecipitated using the mouse anti-FLAG antibody and analyzed by immunoblotting (Fig. 5). pM79ha co-immunoprecipitated with p\textit{flag}M92, indicated by its relative abundance in the eluted sample as compared to that in the flow-through sample. The interaction was specific, as the control pM38ha, a
protein not thought to be involved in late gene regulation, was only detected in the flow-through wash but not in the eluted sample. These results suggest that pM79 and pM92 interact during MCMV infection.

As MCMV M79 and HCMV UL79 play a similar role in late gene expression during infection (10, 41), we wanted to determine whether pUL79 also interacted with pUL92, the HCMV homologue of pM92, during infection. To test this, we infected human foreskin fibroblasts (HFFs) cells with recombinant HCMV expressing C-terminally 3×FLAG-tagged pUL79 (AD\textsubscript{flag} UL79), or wild-type HCMV expressing GFP (AD\textsubscript{gfp}), collected lysates at 72 hpi, and isolated pUL79-containing complex by co-immunoprecipitation using anti-FLAG antibody. Immunoprecipitants were resolved by SDS-PAGE gel followed by silverstain analysis. Protein bands present in AD\textsubscript{flag} UL79-infected samples but absent in AD\textsubscript{gfp} control samples were extracted, and their identity was determined by mass spectrometry analysis. pUL92 was among the pUL79-associated viral proteins identified by this analysis.

Taken together, our results suggest that two viral regulators of late gene expression, pM79 and pM92, interact during MCMV infection, and this interaction is conserved between MCMV and HCMV.

**pM92 trans complements the growth of pUL92-deficient HCMV virus.** MCMV pM92 and HCMV pUL92 share 50% identity and 71% similarity, and notably pM92 has an additional 30aa at the N-terminus (Fig. 6A). As pM92 and pUL92 share significant sequence homology and a similar interaction partner (i.e., pM79 and pUL79, respectively), we hypothesized that pM92 and pUL92 were functional homologues. To test this, we first constructed a pUL92-deficient HCMV recombinant BAC clone, pAD\textsubscript{in}UL92, by FLP/FRT-mediated BAC recombineering (Fig. 6B). pAD\textsubscript{in}UL92 carried an 88-nt insertion at nt 124 of the UL92 ORF to replace a 282-nt segment of the coding sequence (Fig. 6A). The location of the mutation was expected not to interfere with expression of neighboring genes, namely the overlapping 3’-terminus of UL91 or 5’-terminus of UL93. Transfection of pAD\textsubscript{in}UL92 BAC into MRC5 cells failed to produce any infectious virus even after 4 weeks of incubation, whereas cells transfected with the wild-type BAC pAD\textsubscript{gfp} readily developed complete CPE and full spread of virus-driven GFP expression. This was in accordance with previous reports that UL92 is essential for HCMV viral replication in fibroblasts (61). To reconstitute pUL92-deficient virus, we constructed a lentivirus carrying the C-terminally 3×FLAG tagged UL92 ORF, and subsequently generated pUL92 expressing cells by lentiviral transduction (MRC5-UL92\textsubscript{flag}).
Transfection of pAdinUL92 into MRC5-UL92flag cells could now reconstitute infectious progeny virus, ADinUL92, with wild-type titers. Therefore, pUL92 is essential for HCMV replication, and the growth defect of the HCMV recombinant virus ADinUL92 was due to the disruption of pUL92 expression.

To determine if pM92 is the functional homologue of pUL92, we tested if pM92 expression could trans complement the growth of pUL92-deficient virus. We created a lentivirus containing the C-terminally 3×FLAG-tagged M92 ORF, and subsequently generated pM92-expressing MRC5 cells by lentiviral transduction (MRC5-M92flag). We then infected MRC5 cells expressing pUL92 (MRC5-UL92flag), pM92 (MRC5-M92flag), or empty vector (MRC5-vector) with ADinUL92 at an MOI of 0.01, and determined the titer of the cell free virus produced at 14 dpi. Both infected MRC5-UL92flag and MRC5-M92flag showed spread of the pUL92-deficient virus at 14 dpi whereas MRC5-vector showed little sign of CPE (Fig. 6C). Analysis of the final titers of infected culture supernatants indicated that pM92 could complement ADinUL92 to titers similar to that by pUL92 (Fig. 6D). PCR analysis of genomic DNA from MRC5-UL92flag and pM92 MRC5-M92flag confirmed that there was no cross-contamination of these two cell types (Fig. 6E). Therefore, pM92 and pUL92 have a conserved function during CMV infection.
Discussion

The expression of late genes is an essential step for CMV to complete its lytic infection cycle. Key viral factors required in this process could be attractive targets for antiviral strategies to prevent CMV infection and disease. In this study we have determined the role of MCMV protein pM92 during virus infection. pM92 accumulated at late times, and localized to nuclear replication compartments during infection (Fig. 2). When pM92 was abolished, the accumulation of early gene products and viral DNA was minimally affected, but the accumulation of late gene products was markedly reduced (Figs. 3-4). As a result, the mutant virus failed to complete the infection cycle to produce progeny virus (Fig. 1). Therefore, pM92 is a novel regulator of viral late gene expression and thus plays an essential role in the MCMV lytic infection cycle.

Our study provides additional evidence that the regulatory mechanism of viral late gene expression is conserved between MCMV and HCMV. We have previously shown that MCMV protein pM79 and its HCMV homologue pUL79 regulate viral late gene expression (10, 41). In this study, we demonstrated that pM92 interacted with pM79 during MCMV infection; likewise pUL92 could interact with pUL79 during HCMV infection (Fig. 5). This suggests that during betaherpesvirus infection, a complex containing similar components of virus-encoded factors forms to promote late gene expression. How this complex functions and what additional protein components are in this complex remain important questions. Furthermore, we demonstrated that viral protein pUL92, the HCMV homologue of pM92, was also essential for virus infection (Fig. 6). Importantly, pM92 could trans complement the growth of pUL92-deficient HCMV recombinant virus (Fig. 6). These experiments do not specify whether the compensation occurs at the transcriptional or translational level during HCMV infection, and further work is required to define the exact mechanism at play. Regardless, this work suggests a conserved function for pM92 homologues among betaherpesviruses.

Our study also provides additional evidence that viral DNA synthesis is necessary but not sufficient to drive late viral gene expression during herpesvirus infection. Inhibition of the viral polymerase by PAA abolishes the accumulation of late transcripts (23, 25, 26, 51). This dependence on viral DNA synthesis has been linked to the origin of lytic replication (oriLyt), as the oriLyt sequence is required in cis for proper expression of late transcripts in many herpes viruses (2, 11, 14, 29, 55). However, our previous
data and the data presented here demonstrate that viral gene expression at late times of infection depends not only on viral DNA in cis, but also on viral factors such as pM92 and pM79 in trans (10). In the absence of these viral factors, DNA synthesis kinetics was indistinguishable from wild-type virus despite a defect in late transcript accumulation. Therefore, pM92 does not function as a viral DNA synthesis protein; rather it specifically acts on gene expression at late times of infection.

What is the mechanism of pM92 activity? It has been established that herpesviral genomes associate with histones during infection and require epigenetic regulation for gene expression (7, 12, 21, 30, 34, 38, 39). One possible mechanism is that the pM92/pM79 complex may activate late gene transcription by remodeling the chromatin structure of the viral genome. This could be accomplished by recruiting chromatin-remodeling complexes to the late gene loci, and rendering their promoters accessible for transcription. Such activity has been observed for HSV-1 late gene trans factors (13, 24). Alternatively, pM92/pM79 could play a more direct role in transcription. Herpesvirus genes are transcribed by the cellular RNA polymerase II (RNAPII) (5, 33, 53, 58), a 12-subunit multi-protein enzyme that requires a host of accessory scaffold and regulatory proteins for its activity. The pM92/pM79 complex could play an essential role in the recruitment or assembly of these components on viral late gene promoters. Such a mechanism has been suggested for regulation of late gene expression in both alpha and gammaherpesviruses (30, 59). Finally, the activity of pM92/pM79 complex could be required at the stage of post transcriptional modification. It is clear that accumulation of late transcripts was defective during mutant virus infection, but it remains to be determined if this defect results from a failure in transcription or an alteration in mRNA stability. HSV-1 endoribonuclease VHS-RNase is tightly regulated by at least four other viral proteins in order to prevent it from degrading viral mRNAs (32, 45, 46, 52). Precedent for viral regulation of viral mRNA accumulation also exists in HCMV, as viral protein IE2 can inhibit its own transcription by binding to the MIEP promoter (35, 49). Thus we cannot rule out the possibility that during pM92-deficient mutant virus infection, a defect in RNA trafficking, stability, or processing results in higher RNA turnover rates.

Our efforts are currently underway for a better mechanistic understanding of the role of pM92 in late gene regulation. Identification of cellular and viral factors that interact with pM92 is anticipated to provide important insights into its function. Though pM79 is one interaction partner of pM92, it is almost
certain that many additional partners exist. Furthermore, genetic and protein analysis to identify functional
domains and structural elements of pM92 will be invaluable to understand the mechanistic basis for its
activity and to determine additional functions that pM92 may have. Finally, it is tempting to speculate that
late gene expression regulators such as pM92 and pM79 could play a role in the establishment of latency.
As both proteins are essential for the lytic viral life cycle, regulation of their activity and/or expression may
be a deciding factor for viral latency and reactivation.

In summary, we have identified pM92 as a novel late gene regulator in MCMV lytic infection,
shown its interaction with another late gene regulator pM79, and demonstrated its conserved function
with its HCMV homolog, pUL92. pM92 represents a potential new target for therapeutic intervention in
CMV disease, and a gateway into studying a largely uncharted viral process that is critical to the viral life
cycle.
Acknowledgements

We thank Herbert Virgin and the members of his laboratory for helpful discussion and invaluable advice; Dr. Ulrich Koszinowski (Max von Pettenkofer-Institute, Ludwig Maximilians-University, Germany), Dr. Martin Messerle (Hannover Medical School, Hannover, Germany), and Dr. Wolfram Brune (Heinrich Pette Institute, Leibniz Institute for Experimental Virology, Germany) for the MCMV BAC clone pSM3fr; Dr. Anthony Scalzo (University of Western Australia) for M44 and gB antibodies; Dr. Stipan Jonjic (University of Rijeka, Croatia) for IE1 and E1 antibodies; and members of the Yu lab for critical reading of the manuscript.

This study was supported by Public Health Service grants RO1CA120768. D.Y. holds an Investigators in the Pathogenesis of Infectious Disease award from the Burroughs Wellcome Fund.
Material and methods

Plasmids, antibodies, and chemicals. pYD-C433, pYD-C569, pYD-C245, and pYD-C618 were retroviral vectors derived from pRetro-EBNA (31). pYD-C433 and pYD-C569 contained the C-terminally HA tagged M38 and M79 coding sequences, respectively. pYD-C245 expressed the red fluorescent protein (DsRed) (4) from an internal ribosome entry site (IRES). pYD-C618 was derived from pYD-C245, and carried the N-terminally 1×FLAG tagged M92 coding sequence that was expressed together with DsRed as a bicistronic transcript. pYD-C755 (gift from Roger Everett, University of Glasgow Center for Viral Research), pYD-C678, and pYD-C780 were pLKO.1-based lentiviral expression vectors that carried a puromycin resistance marker (15, 17). Both pYD-C780 and pYD-C678 were derived from pYD-C755, and they carried the C-terminally 3×FLAG tagged M92 and UL92 coding sequences, respectively. pYD-C191 carried a kanamycin selection cassette bracketed by two Flp recognition target (FRT) sites. pYD-C630 was derived from pGalK (57) and carried a FRT-bracketed GalK/kanamycin dual selection cassette (40). pYD-C746 was derived from pYD-C630, where a 3×FLAG sequence preceded the FRT-bracketed selection cassette.

The primary antibodies included: anti-actin (clone AC15, Abcam); anti-FLAG polyclonal rabbit antibody (F7425) and monoclonal mouse antibody (F1804) (Sigma); rat anti-HA (11867423001, Roche); anti-MCMV IE1 (CROMA101) and E1 (CROMA103) (generous gifts from Dr. Stipan Jonjic, University of Rijeka, Croatia); anti-MCMV M44 (3B9.22A) and gB (2E8.21A) (generous gifts from Dr. Anthony Scalzo, University of Western Australia). The secondary antibodies used for immunoblotting were horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG, goat anti-rabbit IgG, and goat anti-rat IgG (Jackson Laboratory). The secondary antibodies used for immunofluorescence were Alexa Fluor 594-conjugated goat anti-mouse IgG and Alexa Fluor 488-conjugated goat anti-rabbit IgG (Invitrogen-Molecular Probes).

Other chemicals used in this study include phosphonoacetic acid (PAA) (284270-10G; Sigma Aldrich); L- (+)-Arabinose (A3256-25G; Sigma Aldrich); TO-PRO3 iodide (T3605; Invitrogen); Dyanbeads (Novex, Life technologies); Benzonase (Novagen, Fisher Scientific).

Cells and viruses. Mouse embryonic fibroblast 10.1 cells (MEF10.1) (22), human embryonic lung fibroblasts (MRC5) (19, 28), and human foreskin fibroblasts (HFF) were propagated in Dulbecco modified
Eagle medium supplemented with 10% fetal bovine serum, nonessential amino acids, 1 mM sodium pyruvate and 100 U/mL penicillin-streptomycin. Cells were maintained at 37°C and 5% CO₂ in a humidified atmosphere. To create cell lines stably expressing N-terminally 1×FLAG tagged pM92 (10.1-flagM92), MEF10.1 cells were transduced three times with pYD-C618-derived retrovirus and allowed to recover for 48 hours. DsRed-positive cells were cloned by limiting dilution. Clonal cell lines were tested for their ability to produce virus upon transfection with the mutant MCMV BAC pSMin92 (see below). The cell line yielding the highest titer at 5 days post transfection was used in this study. To create cells expressing C-terminally 3×FLAG-tagged pM92 (MRC5-M92flag) or pUL92 (MRC5-UL92flag), MRC5 cells were transduced with lentivirus reconstituted from pYD-C780 or C678, respectively, and allowed to recover for 48 hours (16). To create cells containing the vector control (MRC5-vector), MRC5 cells were transduced with pYD-C755-derived vector lentivirus. Transduced cells were then selected with 1 µg/mL puromycin, and then maintained with 0.5 µg/mL puromycin.

To reconstitute recombinant MCMV or HCMV viruses, confluent MEF10.1 or MRC5 cells were electroporated with corresponding MCMV or HCMV BAC DNA (see below), respectively. Recombinant MCMV SMgfp and SMflag92 were reconstituted in MEF10.1 cells. SMin92 was reconstituted in 10.1-flagM92 cells. To reconstitute recombinant HCMV, BAC-HCMV DNA, pp71-expression plasmid, and G403-expression plasmid were co-transfected into MRC5-UL92flag cells by electroporation (62). Cells were plated on a 10 cm plate, medium was changed 24 hours post transfection, and virus was harvested by collecting cell-free culture medium after the entire monolayer of cells was lysed. Alternatively, virus stocks were produced by collecting cell-free supernatant from infected culture at a multiplicity of infection (MOI) of 0.01-0.001. Virus titers were determined in duplicate by a tissue culture infectious dose 50 (TCID₅₀) assay in the appropriate cell type. In experiments where comparative analysis was performed between SMin92 or ADinUL92 with other recombinant viruses, titers of all viruses were determined in 10.1-flagM92 or MRC5-UL92flag cells, respectively.

**BAC recombineering.** Recombinant BAC clones in this study were created using the linear recombination-based BAC recombineering protocol that we have previously established (40). Recombination was carried out in *E. coli* strain SW105 that harbored either the MCMV or HCMV BAC
clone, and expressed an arabinose inducible Flippase gene for transient expression of Flp recombinase (57).

Recombinant MCMV BAC clones were derived from the parental clone pSM3fr that carried a full length genome of the MCMV Smith strain (56). pSMgf, used as the wild-type clone in this study, contained the green fluorescent protein (GFP) expression cassette at the C-terminus of the IE2 loci, which has been shown to be dispensable for MCMV infection in vivo and in vitro (8, 9, 36). The clone pSMin92 carried a frame-shift mutation in the MCMV gene M92 (Fig. 1A). To construct pSMin92, the FRT-bracketed GalK/kanamycin cassette was PCR amplified from pYD-C630 and recombined into pSMgf at nucleotide (nt) 358 of the M92 coding sequence. The selection cassette was then removed by arabinose induction of Flp recombinase and subsequent Flp-FRT recombination (40), leaving an 88-nt insert within M92 to create a frame-shift mutation. The clone pSMflag92 contained an N-terminally 3×FLAG-tagged M92 coding sequence (Fig. 1A). To construct pSMflag92, a fragment containing the FRT-bracketed GalK/kanamycin selection cassette preceded by a 3×FLAG sequence was PCR amplified from pYD-C746 and recombined into the N-terminus of the M92 coding sequence of pSMgf. The selection cassette was subsequently removed by Flp-FRT recombination, resulting in the 3×FLAG fused in frame with the M92 coding sequence. Recombinant HCMV BAC clones were derived from the parental clone pAD/Cre that carried the full length genome of HCMV strain AD169 (62). pADgf, used as the wild-type clone in this study, had a GFP gene in place of the viral US4-US6 region (54, 62). The clone pADinUL92 was created in a similar manner to that of pSMin92, except that the insertion replaced nt 125-406 of the UL92 coding sequence carried in pADgf. All of the BAC clones were validated by restriction digestion, PCR analysis, and direct sequencing as previously described (60).

**Viral growth analysis**. MEF10.1 or MRC5 cells were seeded in 12-well plates overnight to produce a confluent monolayer. Cells were inoculated with recombinant viruses for 1 hour at an MOI of 2 for single step or 0.01 for multistep growth analysis. The inoculum was removed, infected monolayer was rinsed with phosphate-buffered saline (PBS), and fresh medium was replenished. At various times post infection, cell-free virus was collected in duplicate by harvesting medium from infected cultures. Cell-associated virus was collected by rinsing infected cells once with PBS and scraping cells into fresh medium. Cells
were lysed by one freeze-thaw cycle followed by sonication. Lysates were cleared of cell debris by low speed centrifugation and supernatants were saved as cell-associated virus. Virus titers were determined by TCID<sub>50</sub> assay.

**DNA and RNA analysis.** Intracellular DNA was measured by quantitative PCR (qPCR) as previously described (10). Briefly, cells were collected in a lysis buffer (200 mM NaCl, 20 mM Tris [pH 8.0], 20 mM EDTA, 0.2 mg/ml proteinase K, 0.4% sodium dodecyl sulfate [SDS]), and lysed by incubation at 55°C overnight. DNA was extracted with phenol-chloroform and treated with RNase A (100 µg/ml) at 37°C for 1 hour. Samples were extracted again with phenol-chloroform, precipitated with ethanol, and resuspended in nuclease-free water (Ambion). Viral DNA was quantified by qPCR using SYBR Advantage qPCR Premix (Clonetech) and a primer pair specific for the MCMV IE1 gene (10). Cellular DNA was quantified using a primer pair specific for the mouse actin gene (10). A standard curve was generated using serially diluted pSM<sub>gfp</sub> BAC DNA or cellular DNA, and was used to calculate relative amounts of viral or cellular DNA in a sample. The amount of viral DNA was normalized by dividing IE1 equivalents over actin gene equivalents. The normalized amount of viral DNA in SM<sub>gfp</sub> infected cells at 10 hours post infection (hpi) was set at 1.

Intracellular RNA was determined by reverse transcription-coupled qPCR (RT-qPCR) as previously described (10). Total RNA was extracted by Trizol reagent (Invitrogen) and treated with TURBO DNA-free reagents (Ambion) to remove contaminating DNA. First strand cDNA synthesis was performed with the High Capacity cDNA Reverse Transcription Kit using random hexamer primered total RNA (Applied Biosystems). Each sample also included a control without the addition of reverse transcriptase to determine the level of residual contaminating DNA. cDNA was quantified using SYBR Advantage qPCR Premix (Clonetech) and primer pairs specific for viral genes or cellular β-actin (Table 1). A standard curve was generated for each gene using serially diluted cDNA from infected cells, and was used to calculate the relative amount of a transcript in each sample. The amounts of viral transcript were normalized by dividing viral transcript equivalents over actin equivalents. The normalized amount of transcript during SM<sub>gfp</sub> infection at 10 hpi in the absence of PAA was set to 1.
Protein analysis. Protein accumulation was analyzed by immunoblotting. Cells were washed and lysates were collected in sodium dodecyl sulfate (SDS)-containing sample buffer. Proteins were resolved by SDS-containing polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto a polyvinylidene difluoride (PVDF) membrane. Proteins of interest were detected by hybridizing the membrane with specific primary antibodies followed by HRP-coupled secondary antibodies, and visualized by using SuperSignal West Pico enhanced chemiluminescent (ECL) substrate (Thermo Scientific).

Intracellular localization of proteins of interest was analyzed by immunofluorescence assay. Cells were seeded onto cover-slips 24 hours prior to infection. At 24 hours post infection, cells were washed with PBS, fixed and permeabilized with methanol (-20°C) for 10 minutes, and blocked with 5% FBS in PBS at room temperature for 1 hour. Cells were incubated with primary antibodies for 30 minutes at room temperature and subsequently labeled with secondary antibodies coupled to Alexa Fluor 488 or Alexa Fluor 594 (Invitrogen-Molecular Probes). Cells were counterstained with TO-PRO3 and mounted on slides with Prolong Gold antifade reagent (Invitrogen-Molecular Probes). Confocal microscopic images were captured by a Zeiss LSM510 Meta confocal laser scanning microscope.

Protein interactions were analyzed by co-immunoprecipitation. For MCMV, MEF 10.1 cells transiently expressing C-terminally HA tagged M38 or M79 were infected with SMflag92 at an MOI of 2. Cells were collected at 48hpi and lysed by incubation with extraction buffer (50 mM Tris pH 8.0, 300 mM NaCl, 0.5% NP-40) for 15 minutes. In the meantime, 1 µg FLAG-antibody was conjugated to 25µL Dynabeads (Novex, Life technologies) by incubation in conjugation buffer (0.02% Tween 20 in PBS, pH 7.2) for 20 minutes at room temperature. Lysates were cleared by centrifugation, pellet of cellular debris was saved, and supernatant was incubated with FLAG antibody-conjugated Dynabeads in the presence of endonuclease Benzonase (800 U/mL) (that digested DNA and prevented DNA-mediated, nonspecific interactions among DNA-binding proteins). After an overnight incubation, beads were washed four times with extraction buffer, and supernatant was saved as flow-through. Washed beads were mixed with NuPAGE LDS sample buffer (Invitrogen) and boiled to elute FLAG-associated proteins. Pellets, flow-through, and eluted samples were analyzed by a SDS-PAGE gel followed by immunoblotting. For HCMV, HFF cells were infected with ADgfp or ADflagUL79 (41) at an MOI of 3. Infected cells were collected and lysed at 72 hpi, and immunoprecipitation was performed as described above. After elution, the samples
were resolved on a SDS-PAGE gel, and protein bands were visualized by silverstain using the ProteoSilver Plus Silver Stain Kit (Sigma).
References


<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>MCMV IE1 forward</td>
<td>5’-CAGGGTGAGATCATGAAACCC-3’</td>
</tr>
<tr>
<td>MCMV IE1 reverse</td>
<td>5’-AGGCGCATCGAAAGACAACG-3’</td>
</tr>
<tr>
<td>MCMV M34 forward</td>
<td>5’-TACTCTGATCGCAACAGCA-3’</td>
</tr>
<tr>
<td>MCMV M34 reverse</td>
<td>5’-GGCTTGGTTGCTCTGTCGG-3’</td>
</tr>
<tr>
<td>MCMV M37 forward</td>
<td>5’-ATGACGGGCGGTCTCCTCAT-3’</td>
</tr>
<tr>
<td>MCMV M37 reverse</td>
<td>5’-ACGTGTACGTCTCCGTCGG-3’</td>
</tr>
<tr>
<td>MCMV M45 forward</td>
<td>5’-GGAACTCCTGGTTCATGG-3’</td>
</tr>
<tr>
<td>MCMV M45 reverse</td>
<td>5’-CTGAAACCATAGCAGCAG-3’</td>
</tr>
<tr>
<td>MCMV M46 forward</td>
<td>5’-CTGAACCTATAGCAGCAG-3’</td>
</tr>
<tr>
<td>MCMV M46 reverse</td>
<td>5’-ACTTATAAGGCGATGG-3’</td>
</tr>
<tr>
<td>MCMV M55 (gB) forward</td>
<td>5’-GCGATGTCAGCTGGTCTAAG-3’</td>
</tr>
<tr>
<td>MCMV M55 (gB) reverse</td>
<td>5’-CGACCAGCAGGTTCGGATA-3’</td>
</tr>
<tr>
<td>MCMV M74 forward</td>
<td>5’-AGGAGGCTGACTGAAA-3’</td>
</tr>
<tr>
<td>MCMV M74 reverse</td>
<td>5’-CTCATCAGCGGTATCGAG-3’</td>
</tr>
<tr>
<td>MCMV M85 forward</td>
<td>5’-TTTCATGAGAAGATGTCG-3’</td>
</tr>
<tr>
<td>MCMV M85 reverse</td>
<td>5’-CTCAGATCTCCTCAGAT-3’</td>
</tr>
<tr>
<td>MCMV M92 forward</td>
<td>5’-AAACCCACCGAGAATGCG-3’</td>
</tr>
<tr>
<td>MCMV M92 reverse</td>
<td>5’-ACGAACAGGTGACCT-3’</td>
</tr>
<tr>
<td>MCMV M96 forward</td>
<td>5’-TCGAGGCGGTTCCTGAT-3’</td>
</tr>
<tr>
<td>MCMV M96 reverse</td>
<td>5’-GCATTCTCAGGATAC-3’</td>
</tr>
<tr>
<td>MCMV M102 forward</td>
<td>5’-AGACCAGTACGGCGATCC-3’</td>
</tr>
<tr>
<td>MCMV M102 reverse</td>
<td>5’-AGCTTCTTCTAGGGTTCG-3’</td>
</tr>
<tr>
<td>MCMV M112/113 (E1) forward</td>
<td>5’-GAATCCGAGGAGAGACG-3’</td>
</tr>
<tr>
<td>MCMV M112/113 (E1) reverse</td>
<td>5’-GTTGAACGTTTGCTGAC-3’</td>
</tr>
<tr>
<td>MCMV M116 forward</td>
<td>5’-TCCTTGGTGGATGCGGCT-3’</td>
</tr>
<tr>
<td>MCMV M116 reverse</td>
<td>5’-GCATCCCATACCTGACCA-3’</td>
</tr>
<tr>
<td>Mouse actin forward</td>
<td>5’-GCTGTATTCCCTCCTCAGTG-3’</td>
</tr>
<tr>
<td>Mouse actin reverse</td>
<td>5’-CAGGGTGTCCTAGGTTCG-3’</td>
</tr>
<tr>
<td>M92 forward $^a$</td>
<td>5’-ATGGTCCACAGGGCGGAGA-3’</td>
</tr>
<tr>
<td>M92 reverse $^a$</td>
<td>5’-CTAGCGGTCTGCAAAACG-3’</td>
</tr>
<tr>
<td>UL92 forward $^a$</td>
<td>5’-ATGTTGCGACGCGCTCGGGC-3’</td>
</tr>
<tr>
<td>UL92 reverse $^a$</td>
<td>5’-AACGCGCATCCGAATACAG-3’</td>
</tr>
</tbody>
</table>

$^a$ primers used in PCR analysis of cellular DNA from pM92- or pUL92-expressing MRC5 cells.
Figure 1. pM92 is essential for MCMV replication in fibroblasts. (A) Diagram depicting the construction of MCMV BACs used in this study by BAC recombineering. To create pSMin92, an 88-nt insert was introduced (indicated by the black bar) at nt 358 of the M92 coding sequence, resulting in a frame-shift mutation. pSMflag92 was created by fusing a 3×FLAG tag in frame at the N-terminus of the M92 coding sequence (indicated by the shaded bar). (B) Growth of SMin92 virus in MEF 10.1 cells expressing FLAG-tagged pM92 (10.1-flagM92). Left panels are fluorescent images of virus-driven GFP expression in MEF 10.1 cells or 10.1-flagM92 cells, both of which were transfected with pSMin92. Images were taken at 7 days post transfection. Right panels are titers of cell free virus at 72 hours post infection (hpi) from 10.1-flagM92 cells infected with either SMgfp or SMin92 at a multiplicity of infection (MOI) of 2. (C-D) Growth kinetic analysis of MCMV recombinant viruses used in this study. MEF 10.1 cells were
infected with indicated viruses at an MOI of 2 (for single-step growth analysis) or 0.001 (for multi-step growth analysis). Cell free and cell associated viruses were collected at indicated times and titers were determined by tissue culture infectious dose 50 (TCID$_{50}$) assay in 10.1-\textit{flag}M92 cells. The detection limit of the TCID$_{50}$ is indicated with a dashed line.
Figure 2. pM92 accumulates abundantly at late times of infection and localizes to viral nuclear replication compartments. (A) Accumulation of the M92 protein product during MCMV infection. MEF 10.1 cells were infected with SMflag92 virus at an MOI of 2. Total cell lysates were harvested at indicated times post infection and analyzed by immunoblotting. The M92 protein was detected with the mouse anti-FLAG antibody, and actin was included as a loading control. (B) Accumulation of the M92 transcript during MCMV infection. MEF 10.1 cells were infected with SMgfp at an MOI of 2 in the presence or absence of viral DNA synthesis inhibitor phosphonoacetic acid (PAA) (200 µg/ml). Total RNA was isolated at indicated times post infection and the M92 transcript was measured by reverse transcription-coupled quantitative PCR (RT-qPCR) analysis with the primers listed in Table 1. The values were normalized to β-actin, and the normalized amount of M92 transcript at 10 hpi in the absence of PAA was
set to the value of 1. (C-D) The M92 protein localizes to nuclear replication compartments during infection. MEF 10.1 cells were mock infected or infected with SMgf or SMflag92 at an MOI of 2. At 24 hpi, cells were permeabilized and fixed with methanol, and stained with either rabbit polyclonal (C) or mouse monoclonal (D) anti-FLAG antibody to detect the FLAG-tagged M92 protein (green). In (C), cells were also stained with mouse antibody to pM44, which served as a marker for replication compartments. Cells were counterstained with TO-PRO3 to visualize the nuclei. Scale bars are equivalent to 20 µm.
Figure 3. pM92 is dispensable for viral DNA synthesis but is required for efficiently accumulation of M55 late gene products during infection. (A) Accumulation of representative viral proteins during pM92-deficient virus infection. MEF 10.1 cells were infected with SMgfp or SMin92 at an MOI of 2. Cell lysates were harvested at indicated times and analyzed by immunoblotting. (B) Accumulation of viral transcripts during pM92-deficient virus infection. MEF 10.1 cells were infected with SMgfp in the presence or absence of viral DNA synthesis inhibitor PAA (200 µg/mL), or with SMin92, at an MOI of 2. At 10 and 20 hpi, total RNA was harvested and specific transcripts were quantified by RT-qPCR analysis using the primers listed in Table 1. The values were normalized to β-actin, and the normalized amount of transcript during SMgfp infection at 10 hpi in the absence of PAA was set to the value of 1. (C) Accumulation of viral DNA during pM92-deficient virus infection. MEF 10.1 cells were infected as described in (A), total DNA
was isolated from infected cells at the indicated times, and viral DNA synthesis was analyzed by quantitative PCR (qPCR). The values were normalized to β-actin and the quantity of DNA during SMgfp infection at 10 hpi was set to the value of 1.
Figure 4. pM92 is required for efficient accumulation of a panel of late transcripts. MEF 10.1 cells were infected with SMgfpl in the presence or absence of viral DNA synthesis inhibitor PAA, or with SMin92, at an MOI of 2. At indicated times, total RNA was harvested and then representative early transcripts (A) and late transcripts (B) were quantified by RT-qPCR analysis using the primers listed in Table 1. The values were normalized to β-actin, and the normalized amount of transcript during SMgfpl infection at 10 hpi in the absence of PAA was set to the value of 1.
Figure 5. CMV UL92/M92 proteins interact with UL79/M79 proteins during infection. MEF10.1 cells expressing either HA-tagged M79 (pM79ha) or HA-tagged M38 (pM38ha) were infected with SMflag92 at an MOI of 2 and cell lysates were collected at 48 hpi. Cell lysates were separated into insoluble cell debris (pellet, “P”) and supernatant by centrifugation. Supernatant was then immunoprecipitated with the mouse anti-FLAG antibody, washed, and eluted. Cell debris (“P”), flow-through fraction collected by wash (“FT”), and eluted fraction (“E”) were analyzed by immunoblotting using anti-FLAG and anti-HA antibody. Molecular weight markers (in kDa) are also shown.
Figure 6. pM92 trans complements the growth of pUL92-deficient HCMV virus. (A) Coding sequence alignment of MCMV M92 with its homologues HCMV UL92 and MHV-68 ORF31. (B) Diagram depicting the construction of pUL92-deficient recombinant HCMV BAC, pADinUL92, by BAC recombineering. pADinUL92 carried an 88-nt insertion (indicated by black bar) at nt 124 of the UL92 ORF to replace a
282-nt segment of the coding sequence. (C-D) Growth of ADinUL92 virus in MRC5 cells expressing C-terminally 3×FLAG-tagged pUL92 or pM92. MRC5 cells expressing tagged pUL92 (MRC5-UL92flag), tagged pM92 (MRC5-M92flag), or the empty vector (MRC5-vector) were infected with ADinUL92 at an MOI of 0.01. At 14 dpi, cells were examined under a fluorescent microscope for virus-driven GFP expression (C), and titers of cell free virus were determined by TCID50 assay in MRC5-UL92flag cells (D). The detection limit of the TCID50 is indicated with a dashed line. (E) Cellular DNA of MRC5-UL92flag and MRC5-M92flag cells was isolated, and the presence of UL92 and M92 DNA in these cells was determined by PCR analysis using the primers listed in Table 1.
Chapter IV

MCMV Late Transcription Regulator pM92 Interacts with pM79 and RNAP II During Infection
Preface

TJC designed and performed the majority of experiments, analyzed the data, and generated the figures presented in this chapter. ARF provided reagents, advice, and support for completion of the study, DY supervised the studies and greatly contributed to the design and analysis of the experiments. The final draft of this chapter was written and edited by TJC.
Introduction

Human cytomegalovirus (HCMV), the prototypical member of the beta herpesvirus subfamily, is a ubiquitous, species specific pathogen that infects 90% of the general population (40). Due to its broad cell tropism, virus can be found in most organs and cell types after infection. In immuno-competent hosts, the infection is typically asymptomatic. However, in immuno-compromised hosts, HCMV infection causes significant morbidity and mortality (40, 56). HCMV is the leading viral cause of birth defects in perinatally infected infants, and is the major cause of retinitis and blindness in AIDS patients (1, 21, 56). It is also a prominent cause of disease in patients under immuno-suppressants, such as bone marrow and solid-organ transplant recipients, or cancer patients receiving chemotherapy (58). HCMV infection has also been associated with cardiovascular disease and proliferative diseases such as cancer (54). After the resolution of acute infection, HCMV establishes persistent life-long infections characterized by alternate stages of virus productivity and latency (40). Lytic infection, both during primary infection and after reactivation from latency, can cause severe disease. There is no vaccine to HCMV, and antiviral therapies have poor toxicity scores, low availability, and are quickly becoming outdated by resistant viruses (6). For these reasons, it is important to understand the role of viral genes involved in lytic infection, as they may yield new targets for anti-viral therapies.

Due to species restriction and a slow replication cycle, studies with HCMV can be difficult. Murine CMV (MCMV) is a model to study HCMV biology due to its conservation with regard to virion structure, genome organization, gene expression, and tissue tropism. MCMV infection recapitulates many clinical and pathohistological features of HCMV infection (18, 49, 50, 53). In addition MCMV contains many functionally conserved genes and provides a tractable small-animal model for experimental MCMV infection in vivo. The study of conserved lytic genes in MCMV will significantly contribute to the understanding of their roles in viral replication and in the pathogenesis of HCMV.

Viral gene expression during lytic infection is temporally regulated and sequentially divided into immediate early, early, and late phases (40). The Immediate early (IE) genes are expressed first requiring only the incoming virion associated proteins and/or cellular transcription factors. The IE genes transactivate the early genes (E) and modulate the cellular microenvironment to permit viral genome replication. These proteins localize to a subnuclear domain, termed the viral replication compartment,
where viral DNA synthesis, late gene transcription, and viral genome encapsidation take place (40). Following viral DNA replication, late genes (L) coding for structural proteins are expressed. On a finer scale, an intermediate category of L genes is defined and classified as early-late (also called delayed-early or leaky-late), depending on their requirement for viral DNA replication. Early-late genes are expressed preceding viral DNA replication, and their transcription rate is enhanced considerably during DNA synthesis. True-late genes absolutely require that viral DNA replication occurs before expression (40).

Though the regulation of IE and E gene expression has been studied extensively, very little is known regarding the regulation of late gene expression during MCMV infection. Previous studies have shown that the activation of both MCMV and HCMV late gene promoters is dependent on the origin of viral DNA synthesis (ori-lyt) in cis (2, 14, 42). This further supports the notion that late gene transcription is tightly coupled to viral DNA synthesis. However, additional levels of regulation for viral late gene expression remain poorly defined.

During cytomegalovirus infection, host cell RNA polymerase II (RNAP II) and associated basal transcription machinery direct the transcription of CMV genes. RNAP II is a 12-subunit DNA-dependent RNA polymerase that is responsible for transcribing nuclear genes encoding messenger RNAs and several small nuclear RNAs (45). RNAP II and its general initiation factors assemble on promoter DNA, creating a large multi-protein DNA complex that supports accurate transcription initiation. Transcriptional activators and coactivators, regulate the rate of RNA synthesis from each gene in response to various developmental and environmental signals. Host cell transcription factors are influenced by virus-encoded transactivators that modulate expression of both viral and host genes during infection (51, 59, 68). It has been shown in HSV-1 that ICP4 can recruit RNAP II via accessory proteins to viral promoters and aid in the process of transcription (7, 31, 55, 61). It has also been shown that proteins ICP22 and UL13 are required for changing the phosphorylation status of RNAP II, which presumably has an impact on the transcription of viral genes (19, 36). ICP22 and ICP4 also play roles in regulating the transcription accessory proteins during transcription, as ICP4 and ICP22 are involved in the recruitment of TAF and acetylated histones to viral promoters (24, 31, 52, 55). Likewise, it has been shown that ORF30 and ORF34 are involved in recruiting RNAP II to the promoter of gammaherpesviruses genes during infection.
This result indicates a critical role of ORF30 and ORF34 in stimulating the assembly of the transcription complex on the late gene promoters. However, the role of RNAP II in gammaherpesviruses transcription has not been further explored, and ORF30 and ORF34 do not have homologues in HSV-1. Analysis of the RNAP II complex during CMV infection has not been done. Thus, the mechanism by which RNAP II is recruited to viral promoters during CMV infection is still unknown.

Previously, it has been demonstrated that HCMV encodes five essential proteins, UL79, UL87, UL91, UL92, and UL95, which were shown to be key regulators of late gene expression during infection (28, 47, 48). In the absence of these genes, transcription of known late genes was abrogated during HCMV infection; however, DNA replication was unaffected (48). The homologues of UL79 and UL92 in MCMV, M79 and M92 respectively, have also been shown to be required for late gene expression (9, 11). It was shown that M92 can complement HCMV deficient in UL92, implying that the mechanism of function is similar between the human and mouse virus. Homologs of UL79, UL87, UL91, UL92, and UL95 are conserved in beta and gamma herpesviruses, including HHV-6, EBV, KSHV, and MHV68. In MHV68 these homologues, ORF18, ORF24, ORF30, ORF31, and ORF34, respectively, have been found to regulate late gene transcript accumulation during infection (3, 30, 65, 66). These proteins do not have homologues in HSV-1, and the mechanism by which these genes function has not been determined (41, 46). Understanding how these late transactivators function during MCMV infection will be important to understand how late genes are regulated during HCMV infection and will give insight into how gammaherpesviruses regulate late gene expression.

In this study, we will further investigate the molecular bases for gene functions by proteomic analysis to identify binding partners of pM79 and pM92. We use an unbiased approach for this objective by utilizing the co-immunoprecipitation (co-IP)-coupled mass spectroscopy strategy. This analysis revealed that both pM79 and pM92 interacted with a panel of viral and host proteins, including host proteins involved in transcription and putative viral late transactivators. The mass spectroscopy analysis was confirmed by directed co-IP in which pM79 and pM92 were tested for their interaction with RNAP II during infection. It was found that RNAP II copurified with both pM79 and pM92. However, the interaction between pM92 and RNAP II appeared to be at least partially facilitated by pM79, as pM79 was required for this interaction. To get a better understanding of the mechanism by which pM92 functions we propose
to try to identify the functional domains of pM92 by creating a series of pM92 mutants and screening them for their ability to maintain interaction with binding partners RNAP II and pM79. Though these studies are not complete, we can at least conclude that during late times of infection pM79 and pM92 form a complex with RNAP II. Analyzing the interactome of these late gene transactivators and identifying the domains necessary for late gene expression will provide important insights into how late gene regulation occurs during betaherpesvirus infection.
Results

Recover proteins bound to late transcription regulators by mass spectrometry analysis. In an initial effort to identify proteins associated with pM79 and pM92 during MCMV infection, we first created a recombinant MCMV in which the M79 or M92 coding sequence was tagged with the 3×FLAG sequence. The M79 ORF was tagged on the C-terminus (SM79flag) and the M92 ORF was tagged on the N-terminus (SMflag92) so that protein complexes containing either recombinant protein could be isolated by co-immunoprecipitation (co-IP) with an anti-FLAG antibody. It has been shown previously that both single-step and multi-step growth curve analyses in MEF 10.1 fibroblasts indicated that SM79flag and SMflag92 grew similarly to SMgfp (wild-type), suggesting that the 3×FLAG tag did not interfere with the function of the tagged gene or neighboring genes (9, 11). We then infected MEF 10.1 cells with SM79flag or SMflag92 at an MOI 2 and collected lysates at 48hpi for co-IP with the anti-FLAG antibody. The lysates were incubated with an extraction buffer designed to isolate whole cell proteins, the lysates were clarified by centrifugation, and the supernatant was incubated with magnetic beads conjugated to anti-FLAG antibody. The same procedure was performed for cells infected with SMgfp as a negative control. The endonuclease Benzonase was also added to each sample in an effort to reduce non-specific binding of proteins that are only indirectly associated with pM79 and pM92 via DNA interactions (57). Immunoprecipitated proteins were resolved by SDS-PAGE and visualized by silver-staining (Fig. 1). Silver-stain analysis revealed a number of protein bands that were unique to the pM79 and pM92 containing samples that could not be seen in the SMgfp control lane (Fig. 1). We extracted the protein bands that were the most easily visualized, and determined their identities by mass spectrometry. As a negative control, we also extracted gel bands from the SMgfp sample with migrating positions corresponding to those of the SM79flag and SMflag92 specific protein bands. Table 1 and Table 2 lists the full set of cellular proteins that were identified by this approach and unique to SM79flag and SMflag92 co-IP, respectively. Table 3 lists the full set of viral proteins that were identified by this approach and unique to SM79flag and SMflag92 Co-IP samples.

It can be seen that pM79 and pM92 share many interaction partners during infection. This is not surprising since viruses deficient for either of these proteins have similar phenotypes. The list of cellular proteins can be categorized into several functional groups. The prominent group of cellular proteins that
copurify with the viral transactivators are proteins involved in transcription. That includes RNA helicases, histones, and most notably, three out of the twelve RNA polymerase II (RNAP II) subunits were identified (Table 1 and 2). The largest subunit, Rpb1, has been shown to be very important for regulation of viral transcription in other herpesviruses, primarily through regulation of the C-terminal domain, which can be phosphorylated to regulate transcription activity and interaction with various transcription accessory proteins. In Table 3 it can be seen that several viral proteins that are conserved in beta and gammaherpesviruses were found to complex with pM79 and pM92. In HCMV pUL87 and pUL95 together with pUL79 or pUL92, all of which are homologous to the MCMV proteins of similar name, have been shown to be essential for viral late gene expression (15, 28, 47, 67). The functions of pM87 and pM95 are not known in MCMV, but it is likely that they interact with pM79 and pM92 to facilitate late gene expression. Interestingly, proteins involved in viral DNA synthesis, including the pM112/113 and the viral polymerase pM54, were found to interact with pM79 or pM92 (41). This suggests that these late gene transactivators may partially act by regulating viral DNA replication. Finally, several cellular proteins involved in post-transcription mRNA activities, including proteins involved in splicing and translation were also copurified with pM79 and pM92. The list of proteins that copurified with pM79 and pM92 suggests that these two viral proteins are directly involved in late gene transcription during infection.

**pM79 and pM92 interact with RNAP II during virus infection.** We next validated the mass-spec analysis to confirm the interaction of pM79 and pM92 with RNAP II. MEF 10.1 cells were infected with either SM79flag, SMflag92, or SMgfp (negative control), cell lysates were collected at 48hpi, and lysates were subjected to co-IP using anti-FLAG antibody. The IP elutions were resolved by SDS-PAGE gel and immunoblotted with both FLAG and RNAP II antibody to the Rpb1 subunit. The results indicate that RNAP II copurifies with both pM79 and pM92 (Fig. 2 and data not shown). This interaction was specific for the FLAG tagged protein as RNAP II did not co-immunoprecipitate in the SMgfp control sample. In addition, the use of benzonase in this assay rules out the possibility that the observed interaction is a byproduct of DNA/RNA binding by any of these proteins. Taken together, these results indicated that pM79 and pM92 interact with RNAP II during infection, and that this association is not mediated by nucleic acids.
We then determined whether the viral proteins could interact with RNAP II independent of additional viral factors. To achieve this, we transfected MEF10.1 cells with a plasmid expressing either HA tagged pM79 (pM79ha) or FLAG tagged pM92 (pflagM92). Lysates were collected at 48hpi, and subjected to co-IP using anti-HA or anti-FLAG antibody. The IP elutions were resolved by SDS-PAGE gel and immunoblotted with both FLAG and RNAP II antibody to the Rpb1 subunit. Cells transfected with unrelated viral protein M38 with an HA tag (pM38ha or M44 with a FLAG tag (pM44flag) were used as negative controls. RNAP II copurified with pM79ha in the absence of viral factors, an interaction that was specific due to a failure for RNAP II to copurify with pM38ha (data not shown). Interestingly, RNAP II did not copurify with pflagM92. This implies that pM79 and pM92 are functioning in late gene regulation at different capacities.

**pM92 requires viral accessory proteins to facilitate interaction with pM79.** Though many viral proteins were common partners for pM92 and pM79, these late gene regulators were not identified as interacting partners for each other. This could be due to an extraction issue (both proteins run near the IgG band) or / and a sensitivity issue (proteins could be present in low abundance). Thus, an experiment was set up in which a vector containing M79 C-terminally tagged with HA (pM79ha) was transfected into MEF 10.1 cells. These cells were subsequently infected with SMflag92 at an MOI 2. Cell lysates were collected at 48hpi and subjected to co-IP using an anti-FLAG antibody. The IP eluted proteins were resolved by SDS-PAGE gel and immunoblotted with both anti-FLAG and anti-HA antibodies. Cells transfected with unrelated viral protein pM38 with an HA tag (pM38ha) were used as negative controls. The blot revealed that pM79ha copurified with SMflag92 (Fig. 3A). This interaction was specific due to a failure of pM38ha to copurify with SMflag92. This data suggests that pM79 and pM92 are found in a complex during virus infection.

Next we wanted to determine if pM92 and pM79 could interact independent of accessory viral proteins. We transfected MEF 10.1 cells with a vector containing pM79ha and a vector containing pflagM92. These cells were allowed to incubate for 36 hours before cell lysates were collected and subjected to co-IP using either anti-HA or anti-FLAG antibody. The IP elutions were resolved by SDS-PAGE gel and immunoblotted with both anti-FLAG and anti-HA. Cells transfected with pM79ha and
pM44flag or pflagM92 and pM38ha were used as negative controls. The blot revealed that pM79 and pM92 were unable to complex in the absence of viral infection (Fig. 3B and data not shown). Thus, viral accessory proteins are required for complex formation between pM92 and pM79.

**Creation of M92 mutants.** We propose to follow up the pM92 interaction studies by better characterizing the interaction between pM92, pM79, and RNAP II. We intend to do this by creating a series of M92 mutant proteins. Mutagenesis will be done at highly conserved residues of the M92 ORF to determine which regions are important for function, interaction with pM79, and interaction with RNAP II. This type of analysis will help to map the functional domains of pM92 by potentially identifying regions that are differentially important for binding pM79, and for maintaining its function as a transcriptional regulator. With luck, we will be able to identify mutants that maintain binding to pM79, but are unable to complement SM*in92* growth, giving us insight into the downstream functions of pM92 (ex. domains involved in the recruitment of other proteins).

In an effort to minimize the mutagenesis efforts, we constructed two retro viral vectors; each containing a truncated version of the M92 ORF. One vector contained 110aa of the N-terminal portion of M92, and the second vector contained 130aa of the C-terminal portion of M92. Both fragments of M92 were tagged with the FLAG epitope for downstream immunoassays. Each vector was transfected individually into fibroblasts, along with a vector containing HA tagged M79 (pM79ha). A vector containing the full length M92 protein was also transfected into fibroblasts with the M79 containing vector as a positive control. The cell lysates were then collected and subjected to co-IP using an anti-FLAG antibody. We then performed a western blot on the collected samples, blotting with an antibody to the HA tag. The blot revealed that neither pM92 fragment could pull-down pM79ha (data not shown). Western blot analysis of the pM92 fragments revealed that the N-terminal fragment was not expressed in transfected cells; likely due to rapid degradation. However, the C-terminal fragment was expressed at abundant levels (data not shown). Thus, it appears that the C-terminal fragment is not sufficient for direct interaction with pM79.

Since our initial analysis implies that the N- and C-terminal domains of pM92 are insufficient to complement pM92 function, a more specific mutagenesis strategy will be required to parse out the
domains of pM92 required for its function. We analyze the M92 ORF and identified all of the residues that are conserved between representative homologues. The alignment of the M92 homologues from betaherpesviruses RhCMV, MCMV, HCMV, RCMV, HHV-6, and gammaherpesviruses MHV68, EBV, and KSHV revealed that there are 15 residues that are identical between all homologues (Fig. 4). QuickChange mutagenesis was designed to replace each of the first nine conserved residues with an alanine. The primers designed for this purpose are listed in Table 4. There are five cysteine residues among these mutants which are known to be important for protein structure; so it is very likely that we will see a phenotype with one of these mutants that will inform us of the function of the N-terminus of pM92. For the remaining six conserved residues, a single gene was synthesized that contained all six mutants. The gene synthesis was done by the company Genewiz. This will tell us whether any of the conserved residues are important in the C-terminus without having to design a QuickChange mutagenesis strategy for each residue. It is possible that a single residue mutation will not disrupt the function of pM92 enough to see a phenotype, which is why we choose to group the mutations in the C-terminus.
Future directions

At the time of writing this dissertation, the planned experiments for this study are not completed. Though they will be completed at a later time, for the purposes of this dissertation, a description of the planned experiments will have to suffice.

Impact of M92 mutants on protein-protein interactions. The mutants that were created for pM92 will be tested for their ability to complement the growth of SM\textit{in}92 mutant virus. The mutant M92 ORFs have an N-terminal 3×FLAG tag in a retro-viral vector (pRetroflagM92). Retrovirus will be made for each of the mutant containing vectors, and then 10.1 MEF cells will be transduced with one of these retroviruses. The expressing cells will then be infected with SM\textit{in}92, and tested for their ability to complement the mutant virus growth. Any M92 mutant expressing cells that are able to complement the growth of SM\textit{in}92 will be deemed uninteresting for this study. Though those M92 mutant proteins may have other important phenotypes under different conditions, we are only interested in M92 mutants that recapitulate the essential phenotype of SM\textit{in}92. Following the initial screen, we will test for the ability of the candidate M92 mutant ORFs to produce protein. This is an important check because we are only interested in mutant proteins that are stably expressed and folded during infection. Since we are mutating many cystiene residues, it is likely that we will find some pM92 mutants that are misfolded and degraded. Such proteins will not be informative about pM92 functional domains.

Once we have a candidate set of pM92 mutants that fail to complement SM\textit{in}92 but are stably expressed, then we are ready to test the mutant proteins for their ability to facilitate late gene transcription. This function will be tested by infecting mutant expressing cells with SM\textit{in}92, isolating RNA by Trizol extraction, and performing RT-qPCR on representative transcripts from each kinetic class of expression. We expect that many of the M92 mutant proteins will fail to compensate the late transcription defect seen in SM\textit{in}92. If it is observed that transcript accumulation is unaffected with some of these pM92 mutants, then we’ll know that pM92 has functions beyond transcription which are important for the lytic replication of MCMV. However, following up on such mutant proteins is outside of the scope of the proposed work.
We will focus on those pM92 mutants that fail to complement the transcription defect of SMn92, and analyze them for their functional interactions. We know that pM92 can interact with RNAP II during infection (Fig 2). We will test whether the selected pM92 mutants can still interact with RNAP II. Mutants that are able to maintain the interaction with RNAPII, but unable to restore late gene transcript accumulation, will inform us that the interaction with RNAP II is not sufficient for the function that pM92 plays in late gene transcription. This may imply that pM92 recruits additional factors to the transcription complex, or that pM92 has an enzymatic activity that is required for transcription initiation but dispensable for protein binding.

Mutants that fail to bind RNAP II will be tested for their ability to bind to pM79. We saw that pM79 facilitates the interaction between pM92 and RNAP II (data not shown). Thus, it would be interesting to determine if there are pM92 mutants that are able to bind pM79 despite a failure to associate with RNAP II. This would suggest that the pM92-pM79 complex actually has additional functions outside of RNAP II activity. Since we know that pM79 can directly interact with RNAP II, the ability to maintain association with pM92 in the absence of RNAP II implies that the two proteins form multiple complexes together which are likely involved in multiple processes important for lytic replication. This is supported by the fact that both pM79 and pM92 have multiple overlapping binding partners not known to be involved in transcription (Table 1 and 2).

In this proposed study, we will elucidate the many functional domains of the pM92 protein. All of this information will no doubt increase the resolution of our understanding about late gene expression in CMV, and about how these late transactivators function during infection.
Discussion

The expression of late genes is a critical step in the lytic infection of cytomegalovirus. Key proteins involved in this process could be attractive targets for specific antiviral strategies. In this study, we discovered a novel regulatory mechanism of viral transcription mediated by MCMV protein pM79 and pM92. We identified cellular RNA polymerase II (RNAP II) as a key factor that interacted with these two viral late transactivators. Co-immunoprecipitation (co-IP)-coupled mass spectroscopy studies demonstrated that pM92 and pM79 interact with a wide array of cellular proteins involved in transcription. Of these, RNAP II was confirmed by subsequent co-IP analysis. The interaction between pM79 and RNAP II is a direct interaction, but pM92 requires viral accessory proteins to mediate complex formation with RNAP II. We also showed that pM92 and pM79 interact during viral infection, but not when expressed together in trans without viral infection. This suggests that pM92, pM79, and RNAP II form a viral transcription complex that requires additional viral proteins to facilitate the interaction. The fact that neither a M92 deficient virus nor a M79 deficient virus completes the viral replication cycle implies that both members of the complex are required to facilitate transcription. This suggests that even though pM79 directly binds to RNAP II, this interaction is not sufficient to facilitate transcription. How this complex functions and what is the protein composition of the complex remain interesting questions.

We also set out to identify the pM92 residues required for its function as a viral transcription factor. Though these studies were not completed at the time of this dissertation, the experiments have been planned and will be completed in the near future.

How is late gene transcription regulated? It is known that herpes virus genomes associate with histones during infection, and require epigenetic regulation for gene expression (8, 12, 22, 31, 37, 43, 44). Thus, it’s possible that pM92 and pM79 activate late gene transcription by remodeling the chromatin structure of the viral genome. In this way, it’s easy to imagine that pM92 and pM79 could play a role in recruiting chromatin-remodeling complexes to the late gene promoters making them more accessible for transcription. Such activity has already been observed for HSV-1 late gene trans factors (13, 26). Another distinct possibility is that pM92 and pM79 are involved directly in transcription. Herpesvirus transcription is driven by RNAP II, and RNAP II requires a host of accessory scaffold proteins and enzymes for its function (5, 35, 60, 62). It’s easy to imagine that pM92 and pM79 could be involved in the
recruitment or assembly of these accessory components on late gene promoters. This behavior has been observed in HSV-1, and has been suggested as a mechanism for gammaherpesvirus late gene regulation (31, 66). It has been shown that late promoters of beta and gammaherpesviruses contain a non-canonical TATA box sequence (23, 27, 64). In HCMV, several characterized viral late promoters contain the TATT non-canonical TATA sequence, which is presumably recognized by the UL87 TATA box binding protein (20, 29, 32, 34, 39, 63). Studies of the UL87 homologue in EBV, ORF24, suggest that this family of proteins preferentially bind to non-canonical TATA sequences. Therefore, pM79 and pM92 may function to guide transcription machinery to assemble around the UL87 binding site.

There is clearly much more to learn about pM92 and its role in late gene regulation. A few areas of future research can be suggested. First, following up on some of the other cellular and viral factors that interact with pM92 could provide significant insights into the protein's function. In HSV-1 it has been shown that Dead-box helicases play an important role in late gene regulation during infection (38). Exactly what role dead-box helicases play during herpesvirus infection is still under investigation. Second, further analysis of the pM92 mutant proteins will be important to identify the comprehensive list of pM92 functions during infection. We only identified the domains important for binding to RNAP II and pM79, it is likely that other domains of pM92 have additional functions. Lastly, it would be interesting to see how pM79 and pM92 participate in viral DNA synthesis. As both transactivators interact with viral DNA synthesis proteins by mass-spec analysis (Table 3), it is possible that pM92 and pM79 regulate gene expression by modulating viral DNA replication.

In summary, we have used a systematic proteomic approach to identify RNAP II as a binding partner for both pM79 and pM92, and are on our way to determining what residues of pM92 are important for both of these interactions. Further study of pM92 and its relationship to pM79 and RNAP II is required to elucidate the molecular mechanism governing late gene regulation, potentially presenting new targets for therapeuic intervention in CMV related disease. More importantly, this protein will serve as a gateway into studying a viral process that is little defined and critical to the viral life cycle of both beta and gammaherpesviruses.
Material and methods

Plasmids, antibodies, and chemicals. pYD-C433, pYD-C569, pYD-C245, and pYD-C618 were retroviral vectors derived from pRetro-EBNA (33). pYD-C433 and pYD-C569 contained the C-terminally HA tagged M38 and M79 coding sequences, respectively. pYD-C245 expressed the red fluorescent protein (DsRed) (4) from an internal ribosome entry site (IRES). pYD-C618 was derived from pYD-C245, and carried the N-terminally 1×FLAG tagged M92 coding sequence that was expressed together with DsRed as a bicistronic transcript. pYD-C755 (gift from Roger Everett, University of Glasgow Center for Viral Research), pYD-C678, and pYD-C780 were pLKO.1-based lentiviral expression vectors that carried a puromycin resistance marker (16, 17). pYD-C780 was derived from pYD-C755, and carries the C-terminally 3×FLAG tagged M92 coding sequence.

The primary antibodies included: anti-FLAG polyclonal rabbit antibody (F7425) and monoclonal mouse antibody (F1804) (Sigma); rat anti-HA (11867423001, Roche); and rabbit anti-RNAP II (sc-899X, Santa Cruz). The secondary antibodies used for immunoblotting were horseradish peroxidase (HRP)-conjugated goat anti-mouse IgG, goat anti-rabbit IgG, and goat anti-rat IgG (Jackson Laboratory).

Other chemicals used in this study include phosphonoacetic acid (PAA) (284270-10G; Sigma Aldrich); Dyanbeads (Novex, Life technologies); Benzonase (Novagen, Fisher Scientific).

Cells and viruses. Mouse embryonic fibroblast 10.1 cells (MEF10.1) (25) were propagated in Dulbecco modified Eagle medium supplemented with 10% fetal bovine serum, nonessential amino acids, 1 mM sodium pyruvate and 100 U/mL penicillin-streptomycin. Cells were maintained at 37°C and 5% CO_2 in a humidified atmosphere. To create cell lines expressing N-terminally 3×FLAG tagged pM92, MEF10.1 cells were transduced three times with pYD-C618-derived retrovirus and allowed to recover for 48 hours. Transduced cells were then selected with 1 µg/mL puromycin, and then maintained with 0.5 µg/mL puromycin. Expressing cells were tested for their ability to produce virus upon transfection with the mutant MCMV SmIn92 (see below).

Virus stocks were produced by collecting cell-free supernatant from infected culture at a multiplicity of infection (MOI) of 0.01-0.001. Virus titers were determined in duplicate by a tissue culture
infectious dose 50 (TCID$_{50}$) assay in the appropriate cell type. MCMV SMgfp and SMflag92 were reconstituted in MEF10.1 cells. SMrin92 was reconstituted in 10.1-flagM92 cells.

**Viral growth analysis.** MEF10.1 cells expressing various vectors containing M92 mutant ORFs were seeded in 12-well plates overnight to produce a confluent monolayer. Cells were inoculated with recombinant viruses for 1 hour at an MOI of 2. The inoculum was removed, infected monolayer was rinsed with phosphate-buffered saline (PBS), and fresh medium was replenished. At various times post infection, cell-free virus was collected in duplicate by harvesting medium from infected cultures. Cell-associated virus was collected by rinsing infected cells once with PBS and scraping cells into fresh medium. Cells were lysed by one freeze-thaw cycle followed by sonication. Lysates were cleared of cell debris by low speed centrifugation and supernatants were saved as cell-associated virus. Virus titers were determined by TCID$_{50}$ assay.

**RNA analysis.** Intracellular RNA was determined by reverse transcription-coupled qPCR (RT-qPCR) as previously described (10). Total RNA was extracted by Trizol reagent (Invitrogen) and treated with TURBO DNA-free reagents (Ambion) to remove contaminating DNA. First strand cDNA synthesis was performed with the High Capacity cDNA Reverse Transcription Kit using random hexamer primed total RNA (Applied Biosystems). Each sample also included a control without the addition of reverse transcriptase to determine the level of residual contaminating DNA. cDNA was quantified using SYBR Advantage qPCR Premix (Clonetech) and primer pairs specific for viral genes or cellular β-actin (Table 1). A standard curve was generated for each gene using serially diluted cDNA from infected cells, and was used to calculate the relative amount of a transcript in each sample. The amounts of viral transcript were normalized by dividing viral transcript equivalents over actin equivalents. The normalized amount of transcript during SMgfp infection at 10 hpi in the absence of PAA was set to 1.

**Protein analysis.** Protein accumulation was analyzed by immunoblotting. Cells were washed and lysates were collected in sodium dodecyl sulfate (SDS)-containing sample buffer. Proteins were resolved by SDS-containing polyacrylamide gel electrophoresis (SDS-PAGE) and transferred onto a polyvinylidene
difluoride (PVDF) membrane. Proteins of interest were detected by hybridizing the membrane with specific primary antibodies followed by HRP-coupled secondary antibodies, and visualized by using SuperSignal West Pico enhanced chemiluminescent (ECL) substrate (Thermo Scientific).

Protein interactions were analyzed by co-immunoprecipitation. For MCMV, MEF 10.1 cells transiently expressing C-terminally HA tagged M38 or M79 were infected with SM/flag92 at an MOI of 2. Cells were collected at 48hpi and lysed by incubation with extraction buffer (50 mM Tris pH 8.0, 300 mM NaCl, 0.5% NP-40) for 15 minutes. In the meantime, 1 µg FLAG-antibody was conjugated to 25µL Dynabeads (Novex, Life technologies) by incubation in conjugation buffer (0.02% Tween 20 in PBS, pH 7.2) for 20 minutes at room temperature. Lysates were cleared by centrifugation, pellet of cellular debris was saved, and supernatant was incubated with FLAG antibody-conjugated Dynabeads in the presence of endonuclease Benzonase (800 U/mL) (that digested DNA and prevented DNA-mediated, nonspecific interactions among DNA-binding proteins). After an overnight incubation, beads were washed four times with extraction buffer, and supernatant was saved as flow-through. Washed beads were mixed with NuPAGE LDS sample buffer (Invitrogen) and boiled to elute FLAG-associated proteins. Pellets, flow-through, and eluted samples were analyzed by a SDS-PAGE gel followed by immunoblotting.

For silver stain analysis, after the elution, the samples were resolved on a SDS-PAGE gel, and protein bands were visualized by silverstain using the ProteoSilver Plus Silver Stain Kit (Sigma).
References


<table>
<thead>
<tr>
<th>Score</th>
<th>Expectation</th>
<th>Protein Name</th>
<th>Function</th>
<th>MW</th>
<th>%Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1829</td>
<td>0</td>
<td>DNA-directed RNA polymerase II subunit RPB1</td>
<td>Subunit of RNA polII</td>
<td>217039</td>
<td>32.3</td>
</tr>
<tr>
<td>1522</td>
<td>0</td>
<td>DNA-directed RNA polymerase II subunit RPB2</td>
<td>Subunit of RNA polII</td>
<td>133825</td>
<td>40.2</td>
</tr>
<tr>
<td>374</td>
<td>1.90E-33</td>
<td>DNA-directed RNA polymerase II subunit RPB3</td>
<td>Subunit of RNA polII</td>
<td>31424</td>
<td>47.3</td>
</tr>
<tr>
<td>248</td>
<td>7.90E-21</td>
<td>Isoform 2 of Heterogeneous nuclear ribonucleoprotein M</td>
<td>splicing</td>
<td>73692</td>
<td>13.6</td>
</tr>
<tr>
<td>165</td>
<td>1.70E-12</td>
<td>Histone H1.2</td>
<td>Histone</td>
<td>21254</td>
<td>19.3</td>
</tr>
<tr>
<td>145</td>
<td>1.50E-10</td>
<td>60S ribosomal protein L4</td>
<td>ribosomal protein</td>
<td>47124</td>
<td>15.8</td>
</tr>
<tr>
<td>140</td>
<td>5.60E-10</td>
<td>Uncharacterized protein</td>
<td>Other</td>
<td>30030</td>
<td>23.3</td>
</tr>
<tr>
<td>120</td>
<td>4.60E-08</td>
<td>ATP-dependent RNA helicase DDX3X</td>
<td>RNA helicase</td>
<td>73056</td>
<td>8.2</td>
</tr>
<tr>
<td>118</td>
<td>7.70E-08</td>
<td>60S ribosomal protein L6</td>
<td>ribosomal protein</td>
<td>33489</td>
<td>17.9</td>
</tr>
<tr>
<td>103</td>
<td>0.0000027</td>
<td>A-kinase anchor protein 8</td>
<td>RNA helicase</td>
<td>76246</td>
<td>3.6</td>
</tr>
<tr>
<td>72</td>
<td>0.0034</td>
<td>Histone H1.5</td>
<td>Histone</td>
<td>22562</td>
<td>14.3</td>
</tr>
<tr>
<td>66</td>
<td>0.012</td>
<td>60S ribosomal protein L8</td>
<td>ribosomal protein</td>
<td>28007</td>
<td>10.5</td>
</tr>
<tr>
<td>65</td>
<td>0.017</td>
<td>Protein 2210010C04Rik</td>
<td>Other</td>
<td>26405</td>
<td>8.1</td>
</tr>
<tr>
<td>64</td>
<td>0.021</td>
<td>ATPase family AAA domain-containing protein 1</td>
<td>Other</td>
<td>40718</td>
<td>4.4</td>
</tr>
<tr>
<td>59</td>
<td>0.066</td>
<td>ATP-dependent RNA helicase A</td>
<td>RNA helicase</td>
<td>149596</td>
<td>2.9</td>
</tr>
<tr>
<td>41</td>
<td>4</td>
<td>Cardiotrophin-like cytokine factor 1</td>
<td>Other</td>
<td>25245</td>
<td>3.6</td>
</tr>
<tr>
<td>38</td>
<td>8.8</td>
<td>Isoform 2 of Serine/arginine-rich splicing factor 7</td>
<td>splicing</td>
<td>27361</td>
<td>16.8</td>
</tr>
<tr>
<td>37</td>
<td>9.5</td>
<td>Isoform 2 of Heterogeneous nuclear ribonucleoprotein F</td>
<td>splicing</td>
<td>43658</td>
<td>6.8</td>
</tr>
<tr>
<td>34</td>
<td>21</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Transcription?</td>
<td>36519</td>
<td>20.1</td>
</tr>
<tr>
<td>33</td>
<td>27</td>
<td>Protein Gm8225</td>
<td>ribosomal protein</td>
<td>31486</td>
<td>7.9</td>
</tr>
<tr>
<td>32</td>
<td>31</td>
<td>Neurofilament medium polypeptide</td>
<td>Other</td>
<td>95859</td>
<td>0.8</td>
</tr>
<tr>
<td>31</td>
<td>38</td>
<td>Nocturnin</td>
<td>Other</td>
<td>48270</td>
<td>1.9</td>
</tr>
<tr>
<td>30</td>
<td>53</td>
<td>Dual specificity protein phosphatase 10</td>
<td>Other</td>
<td>52498</td>
<td>2.1</td>
</tr>
<tr>
<td>30</td>
<td>55</td>
<td>Isoform 2 of Protein prune homolog 2</td>
<td>Other</td>
<td>37105</td>
<td>5.2</td>
</tr>
<tr>
<td>27</td>
<td>110</td>
<td>Isoform 2 of Coiled-coil domain-containing protein 132</td>
<td>Other</td>
<td>107106</td>
<td>3.6</td>
</tr>
<tr>
<td>26</td>
<td>130</td>
<td>cAMP-specific 3',5'-cyclic phosphodiesterase 4D (Fragment)</td>
<td>Other</td>
<td>85498</td>
<td>5.8</td>
</tr>
<tr>
<td>Score</td>
<td>Expectation</td>
<td>Protein Name</td>
<td>Function</td>
<td>MW</td>
<td>%Coverage</td>
</tr>
<tr>
<td>-------</td>
<td>-------------</td>
<td>--------------</td>
<td>----------</td>
<td>------</td>
<td>-----------</td>
</tr>
<tr>
<td>2109</td>
<td>0</td>
<td>DNA-directed RNA polymerase II subunit RPB1</td>
<td>Subunit of RNA polII</td>
<td>217039</td>
<td>35.9</td>
</tr>
<tr>
<td>1382</td>
<td>0</td>
<td>DNA-directed RNA polymerase II subunit RPB2</td>
<td>Subunit of RNA polII</td>
<td>133825</td>
<td>37.1</td>
</tr>
<tr>
<td>625</td>
<td>1.50E-58</td>
<td>Heat shock cognate 71 kDa protein</td>
<td>splicing</td>
<td>70827</td>
<td>32.8</td>
</tr>
<tr>
<td>260</td>
<td>5.50E-22</td>
<td>DNA-directed RNA polymerase II subunit RPB3</td>
<td>Subunit of RNA polII</td>
<td>31424</td>
<td>36.4</td>
</tr>
<tr>
<td>180</td>
<td>4.70E-14</td>
<td>60S ribosomal protein L6</td>
<td>ribosomal protein</td>
<td>33489</td>
<td>21.3</td>
</tr>
<tr>
<td>175</td>
<td>1.70E-13</td>
<td>Voltage-dependent anion-selective channel protein 2 (Fragment)</td>
<td>Mitochondria</td>
<td>30427</td>
<td>16.6</td>
</tr>
<tr>
<td>166</td>
<td>1.30E-12</td>
<td>Histone H1.3</td>
<td>Histone</td>
<td>22086</td>
<td>17.6</td>
</tr>
<tr>
<td>118</td>
<td>7.50E-08</td>
<td>Uncharacterized protein</td>
<td>Other</td>
<td>30030</td>
<td>18.4</td>
</tr>
<tr>
<td>88</td>
<td>0.000073</td>
<td>Voltage-dependent anion-selective channel protein 3</td>
<td>Mitochondria</td>
<td>30832</td>
<td>8.1</td>
</tr>
<tr>
<td>85</td>
<td>0.00016</td>
<td>ADP/ATP translocase 2</td>
<td>Mitochondria</td>
<td>32910</td>
<td>19.8</td>
</tr>
<tr>
<td>81</td>
<td>0.0004</td>
<td>Heterogeneous nuclear ribonucleoprotein M</td>
<td>ribosomal protein</td>
<td>72022</td>
<td>8.5</td>
</tr>
<tr>
<td>78</td>
<td>0.00077</td>
<td>DEAD (Asp-Glu-Ala-Asp) box polypeptide 5</td>
<td>RNA helicase</td>
<td>69223</td>
<td>5.7</td>
</tr>
<tr>
<td>74</td>
<td>0.0021</td>
<td>Histone H1.5</td>
<td>Histone</td>
<td>22562</td>
<td>8.5</td>
</tr>
<tr>
<td>69</td>
<td>0.0061</td>
<td>60S ribosomal protein L4</td>
<td>ribosomal protein</td>
<td>47124</td>
<td>8.1</td>
</tr>
<tr>
<td>66</td>
<td>0.014</td>
<td>Glyceraldehyde-3-phosphate dehydrogenase</td>
<td>Transcription?</td>
<td>36519</td>
<td>4.4</td>
</tr>
<tr>
<td>66</td>
<td>0.011</td>
<td>Isoform Mt-VDAC1 of Voltage-dependent anion-selective channel protein 1</td>
<td>Mitochondria</td>
<td>30737</td>
<td>11</td>
</tr>
<tr>
<td>65</td>
<td>0.015</td>
<td>Myeloid leukemia factor 2</td>
<td>Other</td>
<td>28037</td>
<td>11.3</td>
</tr>
<tr>
<td>61</td>
<td>0.036</td>
<td>ATPase family AAA domain-containing protein 1</td>
<td>Other</td>
<td>40718</td>
<td>4.4</td>
</tr>
<tr>
<td>60</td>
<td>0.057</td>
<td>60S ribosomal protein L8</td>
<td>ribosomal protein</td>
<td>28007</td>
<td>27.6</td>
</tr>
<tr>
<td>53</td>
<td>0.26</td>
<td>Isoform 2 of UAP56-interacting factor</td>
<td>RNA export</td>
<td>32271</td>
<td>4.6</td>
</tr>
<tr>
<td>51</td>
<td>0.41</td>
<td>Putative ATP-dependent RNA helicase P10</td>
<td>RNA helicase</td>
<td>73095</td>
<td>3.5</td>
</tr>
<tr>
<td>48</td>
<td>0.84</td>
<td>A-kinase anchor protein 8</td>
<td>RNA helicase</td>
<td>76246</td>
<td>8.4</td>
</tr>
<tr>
<td>48</td>
<td>0.84</td>
<td>Prohibitin-2</td>
<td>Transcription/ Mitochondria?</td>
<td>33276</td>
<td>7.4</td>
</tr>
<tr>
<td>47</td>
<td>1.1</td>
<td>Protein 2210010C04Rik</td>
<td>Other</td>
<td>26405</td>
<td>8.1</td>
</tr>
<tr>
<td>46</td>
<td>1.4</td>
<td>Mitochondrial carrier homolog 2</td>
<td>Mitochondria</td>
<td>32323</td>
<td>4.8</td>
</tr>
<tr>
<td>43</td>
<td>2.3</td>
<td>Uncharacterized protein</td>
<td>Other</td>
<td>29152</td>
<td>12.1</td>
</tr>
<tr>
<td>33</td>
<td>28</td>
<td>Isoform 2 of RNA-binding protein 39</td>
<td>splicing</td>
<td>58647</td>
<td>13.7</td>
</tr>
<tr>
<td>28</td>
<td>73</td>
<td>Trifunctional enzyme subunit alpha, mitochondrial</td>
<td>Mitochondria</td>
<td>82617</td>
<td>2.8</td>
</tr>
<tr>
<td>27</td>
<td>98</td>
<td>Isoform 2 of Serine/arginine-rich splicing factor 7</td>
<td>splicing</td>
<td>27361</td>
<td>7.1</td>
</tr>
<tr>
<td>25</td>
<td>160</td>
<td>60S ribosomal protein L7 (Fragment)</td>
<td>ribosomal protein</td>
<td>32507</td>
<td>17.6</td>
</tr>
<tr>
<td>22</td>
<td>280</td>
<td>Myosin light chain 3</td>
<td>Other</td>
<td>22407</td>
<td>9.3</td>
</tr>
<tr>
<td>15</td>
<td>1600</td>
<td>Isoform 1 of Paired amphipathic helix protein Sin3b</td>
<td>transcription</td>
<td>109325</td>
<td>3</td>
</tr>
</tbody>
</table>
**Table 3. Mass Spectrometry Identities for Viral Proteins**

<table>
<thead>
<tr>
<th>Score</th>
<th>Expectation</th>
<th>Protein Name</th>
<th>Function</th>
<th>MW</th>
<th>%Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>3112</td>
<td>0</td>
<td>M87</td>
<td>Potential Transcription</td>
<td>102346</td>
<td>49.9</td>
</tr>
<tr>
<td>2354</td>
<td>0</td>
<td>M95</td>
<td>Potential Transcription</td>
<td>45744</td>
<td>54.7</td>
</tr>
<tr>
<td>1008</td>
<td>2.70E-99</td>
<td>M79</td>
<td>Transcription</td>
<td>29354</td>
<td>66.3</td>
</tr>
<tr>
<td>539</td>
<td>2.00E-52</td>
<td>m139</td>
<td>unknown</td>
<td>71713</td>
<td>19.3</td>
</tr>
<tr>
<td>439</td>
<td>2.40E-42</td>
<td>M112andM113</td>
<td>Transcription/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DNA replication</td>
<td></td>
<td></td>
</tr>
<tr>
<td>340</td>
<td>5.20E-29</td>
<td>IE3</td>
<td>viral transactivator</td>
<td>68061</td>
<td>23.9</td>
</tr>
<tr>
<td>156</td>
<td>4.90E-14</td>
<td>M69</td>
<td>RNA export</td>
<td>92977</td>
<td>9.3</td>
</tr>
<tr>
<td>114</td>
<td>7.10E-10</td>
<td>M25</td>
<td>Unknown</td>
<td>103152</td>
<td>7.4</td>
</tr>
<tr>
<td>86</td>
<td>4.10E-07</td>
<td>M86</td>
<td>Major capsid protein</td>
<td>151329</td>
<td>5.9</td>
</tr>
<tr>
<td>76</td>
<td>0.0000039</td>
<td>m138</td>
<td>Unknown</td>
<td>63024</td>
<td>3.2</td>
</tr>
<tr>
<td>67</td>
<td>0.00033</td>
<td>M31</td>
<td>Unknown</td>
<td>56622</td>
<td>6.2</td>
</tr>
<tr>
<td>56</td>
<td>0.00045</td>
<td>M50</td>
<td>Unknown</td>
<td>34691</td>
<td>8.9</td>
</tr>
<tr>
<td>29</td>
<td>0.21</td>
<td>M49</td>
<td>Unknown</td>
<td>60761</td>
<td>2.8</td>
</tr>
<tr>
<td>23</td>
<td>0.95</td>
<td>M54</td>
<td>DNA polymerase</td>
<td>123759</td>
<td>1.1</td>
</tr>
<tr>
<td>21</td>
<td>1.3</td>
<td>m25.2</td>
<td>Unknown</td>
<td>35360</td>
<td>3.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>M95</td>
<td>Potential Transcription</td>
<td>45744</td>
<td>49.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M112andM113</td>
<td>Transcription/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>alternative splice site</td>
<td>DNA replication</td>
<td>34568</td>
<td>53.9</td>
</tr>
<tr>
<td></td>
<td></td>
<td>M112andM113</td>
<td>Transcription/</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>alternative splice site</td>
<td>DNA replication</td>
<td>65988</td>
<td>21.1</td>
</tr>
<tr>
<td>274</td>
<td>6.50E-26</td>
<td>m138</td>
<td>Unknown</td>
<td>71713</td>
<td>13.7</td>
</tr>
<tr>
<td>209</td>
<td>2.10E-19</td>
<td>M87</td>
<td>Potential Transcription</td>
<td>102346</td>
<td>11.9</td>
</tr>
<tr>
<td>184</td>
<td>7.80E-17</td>
<td>M92</td>
<td>Viral transactivator</td>
<td>25430</td>
<td>23</td>
</tr>
<tr>
<td>138</td>
<td>2.60E-12</td>
<td>M49</td>
<td>Transcription</td>
<td>60761</td>
<td>4.9</td>
</tr>
<tr>
<td>106</td>
<td>4.80E-09</td>
<td>M25</td>
<td>Unknown</td>
<td>103152</td>
<td>8.3</td>
</tr>
<tr>
<td>72</td>
<td>0.000011</td>
<td>M31</td>
<td>Unknown</td>
<td>56622</td>
<td>3.5</td>
</tr>
</tbody>
</table>
**TABLE 4. Primers used in QuickChange Mutagenesis**

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Primer Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA 58 Cysteine forward</td>
<td>5'-GGAATATGTACGTGGCCGTGCGCTGTCACCG-3'</td>
</tr>
<tr>
<td>AA 58 Cysteine reverse</td>
<td>5'-CGGTGACAGCGACAGGCCACGATCATATTTCC-3'</td>
</tr>
<tr>
<td>AA 61 Cysteine forward</td>
<td>5'-CGTGTCGCCGCGCTCACCCGACACTCACCTTCGAC-3'</td>
</tr>
<tr>
<td>AA 61 Cysteine reverse</td>
<td>5'-CGCAGAGGGTGAGGCGTGACGCGACACACAGCGC-3'</td>
</tr>
<tr>
<td>AA 65 Histidine forward</td>
<td>5'-GCTGTCACCGCAGCGCTACCGTCTCGCGATCTACGC-3'</td>
</tr>
<tr>
<td>AA 65 Histidine reverse</td>
<td>5'-CGTAGATCGCAGGAGCGTGCTGACGACG-3'</td>
</tr>
<tr>
<td>AA 67 Cysteine forward</td>
<td>5'-GCTGTCACCGCAGCGCTACCGTCTCGCGATCTACGC-3'</td>
</tr>
<tr>
<td>AA 67 Cysteine reverse</td>
<td>5'-CGTGCCGTCGATCGGAGGTCGCTGACGACG-3'</td>
</tr>
<tr>
<td>AA 73 Cysteine forward</td>
<td>5'-CGATCTACGCGAAGCAGCGTCTGCGGTCGACGACG-3'</td>
</tr>
<tr>
<td>AA 73 Cysteine reverse</td>
<td>5'-CGTGACCGATCGGAGGTCGCTGACGACG-3'</td>
</tr>
<tr>
<td>AA 84 Cysteine forward</td>
<td>5'-CCACACGCAAGGACGCTCCGTCGACGACGACG-3'</td>
</tr>
<tr>
<td>AA 84 Cysteine reverse</td>
<td>5'-GGTGAGGCGCTCGGATCGGACGACGACGACG-3'</td>
</tr>
<tr>
<td>AA 87 / 88 Threonine / Glycine forward</td>
<td>5'-GCTCCGTCGACGCTACACCCGACGACGACGACG-3'</td>
</tr>
<tr>
<td>AA 87 / 88 Threonine / Glycine reverse</td>
<td>5'-GGAGCCGATCGGAGGCGCCCGCCGACGACGACGACG-3'</td>
</tr>
<tr>
<td>AA 160 Phenylalanine forward</td>
<td>5'-GGAGAATGCGATCTTTACGCGCAACCGCGTCTGACGACG-3'</td>
</tr>
<tr>
<td>AA 160 Phenylalanine reverse</td>
<td>5'-GCTGCTTGAACGCGGCTGCGTTGCGTAAAATAGATCGCATTCC-3'</td>
</tr>
</tbody>
</table>

Red text indicates Alanine residues designed to replace the original sequence
Figure 1. Proteomic analysis to identify binding partners of pM79 and pM92. Identification of pM79 and pM92 interacting partners. MEF 10.1 cells were infected with SMwt, SM79flag, or SMflag92 at MOI 2, collected at 48 hpi, and were immunoprecipitated in the presence of benzonase with FLAG antibody. Eluted proteins were run on an SDS-containing polyacrylamide gel and silver stained. The bands indicated with an arrow were extracted for mass spectrometry. Potential alignment of proteins identified by mass spec with candidate bands. The silver stained gel has been labeled with potential alignment of proteins identified from the mass-spec results.
Figure 2. Immunoprecipitation analysis to confirm RNAP II interactions. (A) FLAG-tagged pM79 interacts with RNAP II during MCMV infection. MEF 10.1 cells were infected with SMwt or SM79flag at MOI 2, collected at 48 hpi, and lysates were subjected to co-immunoprecipitation with FLAG mouse monoclonal antibody. Cell pellets, lysates, and eluted proteins were analyzed by immunoblotting with indicated antibodies. Blots were cropped to save space but were from the same lane and exposed film. (B) FLAG-tagged pM92 interacts with HA-tagged pM79 in MCMV infection. MEF 10.1 cells were transfected with a vector expressing HA-tagged pM79. 24 hours after transfection, cells were infected with SMflag92 virus at MOI 2. Lysates were collected at 48hpi and subjected to co-immunoprecipitation with FLAG antibody. Cell pellets (P), lysates (FT), and eluted proteins (E) were analyzed by immunoblotting with indicated antibodies.
**Figure 3.** pM92 requires viral accessory proteins to facilitate interaction with pM79. (A) pM92 interacts with pM79 during infection. MEF10.1 cells expressing either HA-tagged M79 (pM79ha) or HA-tagged M38 (pM38ha) were infected with SMflag92 at an MOI of 2 and cell lysates were collected at 48 hpi and subjected to co-immunoprecipitation with anti-FLAG antibody. Cell pellets, lysates, and eluted proteins were analyzed by immunoblotting with anti-FLAG and anti-HA antibody. Molecular weight markers (in kDa) are also shown. (B) pM92 does not interact with pM79 independent of viral infection. Lysates from MEF 10.1 cells expressing HA-tagged M79 and FLAG-tagged M92 were collected and analyzed as in (A).
Figure 4. **Amino acid alignment of the UL92 protein family.** Amino acid alignment of UL92 proteins from mouse, rat, human, and rhesus, CMVs, with HHV6, MHV68, KSHV, and EBV. Amino acids highlighted in red are conserved across all species.
Chapter V

Conclusions and Future Directions

The final draft of this chapter was written and edited by TJC.
Conclusions

**Functional profiling.** Despite the ready availability of the MCMV Smith strain as an infectious BAC clone and the wide use of MCMV as an animal model of CMV infection, the 78 open reading frames that are conserved between HCMV and MCMV have never undergone comparative functional analysis. After excluding the genes that are well studied in MCMV, or genes that have well studied homologues, we were left with 28 genes of interest to target. Mutagenesis of the BAC-MCMV was achieved by PCR-based linear recombination followed by FRT/FLP recombination. Twenty-two of the twenty-eight target genes have been successfully mutated and confirmed via PCR, restriction digest, and sequencing. The 22 target genes that have been successfully mutated have been reconstituted in fibroblasts and screened for a phenotype via multi-step growth curve analysis. The screen identified seven genes that were either essential for viral replication or augmenting for viral growth. Identification of essential and augmenting genes that are conserved across cytomegaloviruses will help herpes virologists focus their effort on proteins that may be involved in important conserved viral processes.

**Late gene transactivators.** MCMV is the commonly used model virus for HCMV, so revealing HCMV genes that are functionally conserved in MCMV will allow the use of the robust mouse genetic system to elucidate their role and test novel antivirals targeting these gene products. We have found that the products of MCMV genes M79 and M92 regulate viral Late gene expression. In particular, we showed that the encoded proteins, pM79 and pM92 respectively, were critical for MCMV to promote accumulation of a set of Late gene transcripts (10, 13). Moreover, MCMV DNA was not only synthesized efficiently in the absence of these Late gene transactivators, but in the case of pM79 deficient virus Pulse Field Gel Electrophoresis analysis showed that the overall structure of replicating viral DNA was unimpaired and that the development of viral replication compartments during mutant virus infection was indistinguishable from those during wild-type virus infection. This body of evidence further excludes the involvement of pM92 and pM79 in viral DNA synthesis or other events preceding viral transcription at late times of infection. This phenotype has also been observed for the HCMV homologues UL79 and UL92 (66). We also determined that pM79 accumulates with Early-Late kinetics in nuclear viral replication compartments during infection, which suggests that it is expressed slightly earlier than the other lytic genes. It is likely
that this Early-Late expression is required to initiate Late gene transcription. How pM79 expression is regulated remains a mystery. pM92 is expressed at late times of infection, and is likely regulated by pM79 as other Late genes are during infection. Furthermore, we demonstrated that viral protein pUL92, the HCMV homologue of pM92, was also essential for virus infection. Another group further showed that pUL92 was important for viral transcript accumulation at late times of infection, a phenotype that mimics what was seen for pM92 deficient MCMV (65). Importantly, pM92 could trans-complement the growth of pUL92 deficient HCMV recombinant virus. These experiments do not specify whether the compensation occurs at the transcriptional or translational level during HCMV infection, and further work is required to define the exact mechanism at play. Regardless, this work suggests a conserved function for pM92 homologues among betaherpesviruses.

Our work, together with the reported role of its homologues, including pUL79 and pUL92 in HCMV and ORF18 and ORF31 in MHV68 (3, 41, 43, 65, 66), suggests a common mechanism governing Late transcription among beta and gammaherpesviruses. How viral Late transcription is regulated remains largely unknown and viral/host factors involved are poorly defined. These late transcription transactivator proteins represent an invaluable tool to gain insight into this key viral process.

**Viral DNA synthesis.** Our study also provides additional evidence that viral DNA synthesis is necessary but not sufficient to drive viral Late gene expression during herpesvirus infection. Inhibition of the viral polymerase by PAA abolishes the accumulation of Late gene transcripts (33, 34, 39, 79). This dependence on viral DNA synthesis has been linked to the origin of lytic replication (ori-lyt), as the ori-lyt sequence is required in cis for proper expression of Late transcripts in many herpes viruses (2, 14, 19, 45, 82). However, our previous data and the data presented here demonstrate that viral gene expression at late times of infection depends not only on viral DNA replication in cis, but also on viral factors such as pM92 and pM79 in trans (10, 11, 13). In the absence of these viral factors, DNA synthesis kinetics were indistinguishable from wild-type virus despite a defect in Late gene transcript accumulation. Therefore, pM79 and pM92 do not function as viral DNA synthesis proteins; rather they specifically act on gene expression at late times of infection.
Several theories have been proposed to explain why Late genes are dependent on DNA replication. It has been shown in alpha and gamma herpesviruses that Late gene promoters can be unusual in structure, sometimes not even containing a canonical TATA element (26, 40, 86, 89). These unusual structures make it hard for the cellular transcription complex to assemble, since the viral promoter region lacks some of the recognition sequences for transcription accessory proteins. Thus, a popular hypothesis is that the transcription complex is recruited by the viral DNA replication complex to Late gene promoters, bypassing the requirement for conventional activators. Such a mechanism would be very effective to link viral DNA replication to the activation of Late gene transcription. This activity has been seen in HSV-1, where ICP8, a component of the viral DNA replication complex, interacts with the RNAP II complex to assist in recruitment to the replicating viral DNA, thereby stimulating Late gene expression (15, 64, 91). This mechanism could be conserved in beta herpesviruses, especially since we were able to copurify DNA replication components with the Late gene transactivators, pM79 and pM92.

**Protein interactions.** In this study, we demonstrated that pM92 interacted with pM79 during MCMV infection; likewise pUL92 could interact with pUL79 during HCMV infection (13). This suggests that during cytomegalovirus infection, a complex containing conserved virus-encoded factors forms to promote Late gene expression. How this complex functions and what additional protein components are in this complex remain important questions. Cytomegalovirus genes are transcribed by the cellular RNA polymerase II (RNAP II), a 12-subunit multi-protein enzyme that requires a host of accessory scaffold and regulatory proteins for its activity (4, 52, 80, 85). Our study also identified RNAP II as an interaction partner for both pM92 and pM79 during infection. This implies that pM92 and pM79 play a direct role in transcription. The pM92/pM79 complex could play an essential role in the recruitment or assembly of cellular transcription components on viral Late gene promoters. Such a mechanism has been suggested for regulation of Late gene expression in both alpha and gamma herpesviruses (47, 88). However in what capacity the interaction of pM79 and pM92 with RNAP II aids in transcription is unknown. Further studies will need to be done to determine how these proteins modulate transcription. One possibility is that pM79 and/or pM92 modulate phosphorylation of the RNAP II C-terminal domain (CTD), which impacts initiation, elongation, and termination of transcription. Secondly, pM79 and/or pM92 could recruit RNAP II to the
promoters of Late genes, potentially by recruiting TAF or TFIIID accessory factors. Third, pM79 and pM92 could modulate the architecture of the promoter by regulating methylation or acetylation of DNA or histones in the area. Confirmation of any of these strategies will greatly advance our understanding of herpesvirus Late gene transcription and regulation.

**Latency.** Finally, it is tempting to speculate that Late gene expression regulators such as pM92 and pM79 could play a role in the establishment of latency. As both proteins are essential for the lytic viral life cycle, regulation of their activity and/or expression may be a deciding factor for viral latency and reactivation.

**Implications.** The work described here has many implications for herpesviruses, virology research, and medicine. With the dearth of drugs available to treat HCMV, identifying new drug targets is important. Despite being discovered 40 years ago, many of the fundamental pathways of the HCMV life cycle are undefined. Here we identify major players in one of the most important steps for lytic replication in CMV. The development of small molecule inhibitors of pM92, pM79, or their homologues may result in drugs that effectively stop CMV at the lytic stages of infection, and can be combined with the acyclovir cocktail to prevent the evolution of escape viruses. It is also important to note that this research effectively adds an additional level of regulation to Late gene expression that succeeds viral DNA replication. We now know that DNA synthesis is required but not sufficient for Late gene expression in cytomegalovirus replication. Though the pathway has not been mapped out completely, this thesis comprises substantial work toward elucidating the details of this process. Understanding how DNA replication and transcription are related will give us a better understanding of how these crucial lytic steps can be inhibited during infection. Understanding how the virus hijacks host machinery is also of great interest to the field of virology. Though decent progress has been made for identifying important protein interactions for alphaherpesviruses, little to no progress has been made for beta or gammaherpesviruses, both of which are responsible for severe disease.
Late gene transcription in other herpesviruses. Though we have identified a few important elements for regulating Late gene transcription during cytomegalovirus infection, there are still many elements of Late gene expression that need to be defined. In the studies moving forward, it will be important to consider what is known about Late gene regulation in other herpesviruses. In this way, future experiments can be directed at testing for strategies that are likely to be conserved in this family of viruses.

HSV-1 is the most extensively studied herpesvirus, and as a result there is much known about the regulation of Late gene transcription. It is known that ICP4, ICP22, ICP27, ICP8, and ICP0 are important for Late gene expression. ICP4 is the major activator of Early and Late genes and interacts with the viral transcription complex (9, 59). ICP4 has a DNA binding domain that interacts with Early and Late promoters. However, these specific sites aren’t required for activation of transcription (18, 21). The sequence that is required for Late promoter activity is the INR which is downstream of the TATA box (27, 29, 30, 37). In fact, the INR can initiate transcription in the absence of a TATA box (51, 73). The INR helps ICP4 stabilize RNAP II accessory protein, TFIID complex, which contains TBP (the TATA-binding protein), that is required for transcription. Specifically, ICP4 interacts with the TAF250 subunit of the TFIID complex (9). Interestingly, it was shown that RNAP II accessory protein TFIIA is required to stabilize TBP on Early promoters, but was not required for Late promoters. At Late times of infection TFIIA is down regulated, which may help control the switch between Early gene transcription and Late gene transcription.

ICP4 localizes to the nucleus in viral replication compartments where it can be found in a viral transcription complex, which includes RNAP II and ICP22 (16, 17, 53). ICP22 is involved in HSV-1 Late gene transcription. ICP22 binds to pTEFb and causes alterations to RNAP II phosphorylation, which presumably regulates viral transcription (24, 28). It is unclear how ICP22 activity is achieved, but it may have something to do with the binding partner pTEFb, which has the ability to phosphorylate RNAP II to modulate transcription. ICP22 also interacts with circadian histone acetyltransferase CLOCK, which has been shown to be important for viral gene expression (22, 32, 46, 47, 50). CLOCK interacts with ICP0, which has been shown to displace HDACs from the REST/CoREST complex to enable Early and Late gene transcription.

Unfortunately, the regulatory proteins in HSV-1 do not have homologues in the beta or
gammaherpesviruses. So though the information from HSV-1 might shed light on some potential themes to test, it is more likely that gammaherpesviruses will give us better insight into what HCMV might be doing. As mentioned earlier, the homologues ORF18, ORF24, ORF30, ORF31, and ORF34 have a similar phenotype as the HCMV homologues M79, M87, M91, M92, and M95 respectively (3, 41, 43, 65, 66, 87, 88). In fact, ORF30 and ORF34 have been implicated in recruiting RNAP II to the promoters of MHV68 Late genes. Further studies will be needed to determine exactly how these proteins regulate RNAP II, and what the impact is on late gene transcription.

Understanding how the expression of the late gene regulators themselves are regulated will also be important to understanding how late gene transcription is regulated. In MHV68, some information has been discovered about the regulation of ORF18 during infection. All gammaherpesviruses encode a gene called replication and transcription activator (RTA). RTA controls the switch from latency to lytic infection. It has been shown that RTA can bind directly to responsive elements (RREs) or can act in conjunction with other binding partners (5, 6, 20, 74-76). It was found that ORF18 has a RRE element in its promoter, which allows it to be activated by RTA (36). Since the ORF18 promoter is in the ori-lyt of MHV68, it’s possible that DNA replication frees up the RRE so that RTA can bind and transactivate ORF18 expression, which starts the Late gene cascade. At this point a homologue to RTA does not exist in betaherpesviruses, but it is possible that one of the late gene transactivators plays a similar role. pM79/pUL79 is a good candidate for this role, as it is expressed at Early-Late times of infection.

In addition to the Late gene transactivators, MHV68 also regulates its Late gene transcription by mRNA degradation, which is mediated by SOX endonuclease (1). SOX was shown to degrade viral mRNA from all kinetic classes of expression, but a mutant in SOX had its primary defect due to missing components of the tegument. The altered tegument composition resulted in the inability of progeny virus to undergo lytic infection. Instead, all progeny virus defaulted to latent infection. Since tegumentation is a process that occurs late in infection, it is assumed that degradation of Late gene transcripts is important to maintain the appropriate balance of lytic gene expression during virus infection. The exact mechanism in effect here has yet to be determined.
Future directions

The work that has been done in this thesis has only touched the surface of understanding Late gene expression and the regulation of CMV transcription. There are many more studies to be done, and many directions with which to start.

Analyze various levels of Late gene expression. Our oligonucleotide tiled array analysis identified a set of annotated genes whose transcription is substantially reduced at late stages of virus infection when pM79 is abrogated. Comparative analysis of viral transcriptomes among MCMV infections with PAA, without PAA, or in the absence of pM79 leads to two interesting observations. First, while viral DNA synthesis is required for transcription from genomic regions containing many previously reported Late genes, it also facilitates the continued transcription from genomic regions containing several previously reported Early genes at Late times of infection. Second, viral transcripts that are dependent on viral DNA synthesis also have a dependency on pM79 for their accumulation (10). There seems to be a hierarchy in dependence, such that without pM79, some transcripts are markedly reduced, whereas others are reduced to a much less extent. Thus, pM79 is a key viral regulator of Late gene transcription in MCMV infection, but it is not the only regulator. In future studies, similar transcriptome analysis should be applied to other viral genes known or predicted to be regulators of Late gene expression. Examples are HCMV UL79 and MHV-68 ORF18, as well as MCMV M87 and M95 (homologues of HCMV UL87 and UL95) (3, 62, 65, 66). Such analysis will reveal whether these viral regulators control the expression of an overlapping or distinct set of viral Late genes.

I think RNAseq analysis will yield a more high-throughput view of the cytomegalovirus transcriptome. Comparing the abundance of transcripts between pM79 or pM92 deficient virus and wild-type virus will let us know all of the genes affected by these Late gene transactivators. RNAseq is more beneficial than an oligonucleotide array because the method does not rely on annotated open reading frames for analysis. This is especially important in light of the recent discovery that HCMV produces over 700 transcripts during infection, most of which are not annotated (78). Thus, this experiment has the potential to identify novel transcripts that are involved in MCMV lytic replication. It is important to know exactly what genes are affected by pM79 and pM92 expression, and by what magnitude these genes are
regulated. Such findings may reveal additional levels of lytic gene regulation, and may lead us to clues of additional viral regulators involved in both Late gene transcription and viral DNA synthesis regulation.

**Functional domain study of pM79 to define functionally important protein regions.** How does pM79 regulate expression of Late gene transcripts? A Late gene transactivator could be involved in epigenetic regulation of replicating viral DNA. For instance, viral DNA-associated histones are modified when herpesviruses replicate their genomes, and this may render Late gene promoters accessible for transcription (54, 55, 72). However, the pM79 coding sequence does not resemble those of histone modifying enzymes, such as histone acetyltransferase or histone deacetylase (data not shown). If pM79 had a role in epigenetic regulation of gene expression, it would likely act indirectly, for instance, by recruiting modification enzymes to histones associated with Late gene promoters. Late gene regulators may also act as transcription factors. However, pM79 does not contain any identifiable putative DNA binding domains (data not shown). Therefore, in this capacity pM79 would have to act as a modulator of cellular and viral transcriptional regulators or RNAP II to facilitate viral Late gene transcription. To this end, we have shown that pM79 can interact with RNAP II during infection (Chapter IV). Furthermore, our research group has unpublished data that the HCMV homologue, pUL79, interacts with RNAP II during infection. During HCMV infection, UL79 can modulate the phosphorylation state of RNAP II to control transcript elongation (Perng et al, unpublished). Determining what domains of pM79 are important for its function during infection will lead us to answer some of the posed questions above, and further understand the mechanism of its activity.

A mutagenesis strategy will be designed that targets the highly conserved residues of pM79 (Fig. 1). Much like the Quickchange mutagenesis of pM92 done in Chapter IV, conserved residues in pM79 will be converted to alanine. These pM79 mutants will be tested for their ability to complement the growth of SM\textit{in79} mutant virus. Any mutants that are able to complement the growth of SM\textit{in79} will be deemed uninteresting. Though those mutant may have other important phenotypes under different conditions, we are only interested in mutants that recapitulate the essential phenotype of a M79 deficient virus. pM79 mutants will also be tested for their ability to facilitate Late gene transcription. This function will be tested by performing RT-qPCR on transcripts representative of genes from each kinetic class. pM79 mutants
that fail to complement the transcription defect of SMin79 will be analyzed for functional interactions. We would do co-IP-coupled mass spectrometry analysis to determine the difference in interaction partners between all selected pM79 mutants and wild-type pM79. The results of this experiment would identify cellular and viral factors that interact with pM79 in a domain specific way. This method could provide significant insights into the function of pM79. Any and all results will no doubt increase the resolution of our understanding about late gene expression in CMV, and about how pM79 functions during infection.

**Continue the functional domain study for pM92 to define protein regions important for function.** Though some functions of pM92 have been elucidated, it is likely that this protein is multifunctional. It would be interesting to do a different screen with the pM92 mutants to determine if the domains that were dispensable for RNAP II and pM79 interaction are important for another interaction or process. Furthermore, it is important to consider putting the pM92 mutants that had no phenotype during the *in vitro* screen into mice. Mouse experiments may reveal additional functions of this Late gene transactivator that are not observed *in vitro*.

**Determine promoter sequence elements for Late genes that are important for expression.** We know that pM79 and pM92 interact with RNAP II, and we know that RNAP II is recruited to Late gene promoters at late times of infection. However, why Late gene promoters are not engaged until late times of infection by the transcription machinery is a mystery. Initiation of transcription requires the assembly of TFIIA, B, D, E, F, H and RNAP II onto the core promoter elements (61). TFIID consists of the cellular TATA-binding protein (TBP) and TBP-associated factors (TAFs). TAFs interact with initiator elements (INRs) to aid transcription (48, 49). TFIID is stabilized by TFIIA and TFIIIB. It is known that kinetic classes are regulated in part by the promoter architecture (84, 90). In HSV-1 there is a decrease in promoter complexity from IE to Early to Late genes (83, 85). IE gene promoters have viral and cellular cis acting sequences in the promoter; Early gene promoters only have cellular elements (sp1 and CTF); Late gene promoters lack any influential upstream cis-acting elements (23, 30, 35, 38, 44). For True Late gene promoters, activators have not been described, but the main elements are the sequences downstream from the TATA box that are important for Late gene regulation (27, 29, 30, 37, 38, 77). It has actually been shown
in HSV-1 that both Early and Late gene promoters are engaged by transcription factors at early times of infection, yet they still have different kinetics for expression. Two theories are that there is relocalization of key proteins to the viral replication compartments at late times of infection, or changes in the DNA damage response at late times of infection could stimulate Late gene promoter transcription (16, 42, 56).

Similar themes have been seen in the gammaherpesviruses. Studies in EBV and MHV68 have shown that unlike viral Early gene promoters, the promoters of Late genes map to the regions that lack regulatory sequences upstream of the TATA box (23, 30, 35, 38, 44). However, the role of the TATA box, and the sequence that defines the functional TATA box in gammaherpesviruses Late genes has only recently been uncovered. It has been shown in EBV that the region comprising the Late gene promoter includes a short core sequence required for expression that includes an unconventional TATT sequence, rather than a canonical TATA box (2, 71). In KSHV, a region of 12 bp that also contained a TATT sequence was found to be sufficient to support activation of a Late gene promoters (81).

The most striking characteristic of this atypical TATT sequence is a T at the fourth position, which is only found in 10% of TATA box containing promoters of eukaryotic genes (8). The functional significance of this fourth T has been examined for other herpesvirus Late gene promoters by creating a T-to-A point mutation. It was shown that such mutation in the two Late gene promoters of EBV result in a loss of promoter function (2, 71). However, these results are not universal for Late gene promoters, because a study has been done to show that mutagenesis of the ORF52 Late promoter indicates that both TATATA and TATTA/TA support the Late gene promoter activity, but TATAAA does not (86). This implicates that nucleotides neighboring the TATA/TATT element are also important for Late gene regulation.

Minimal analysis has been done to analyze the TATA elements of Late genes in the cytomegalovirus genome. The UL44 Late gene promoter of HCMV has a noncanonical TATT sequence, and a T-to-A point mutation greatly reduces the level of the UL44 Late gene transcript (40). Attempts have also been made to map the promoter regions of Late genes in both HCMV and MCMV using bioinformatic programs. It was observed that the sequences of TATTA/TA and TATA/TTA can been found in the promoter regions of many Late genes, such as UL14, UL18, UL31, UL32, and UL48 of HCMV and M25, M31, M34, M69, M97, M100, and M102 of MCMV (86). This analysis supports the theory that the
canonical TATA box is not likely to mediate transcription of herpesvirus Late gene promoters. The inability of a canonical TATA box consensus sequence to function as a core element for the Late gene promoter is an important piece of information in understanding the regulation of Late gene expression.

However, formal analysis of Late gene promoters is scarce in betaherpesvirus literature. Even taking the above information into account, many questions remain. Is the required core sequence of CMV similar to the 12bp KSHV core sequence (81)? How does such a short core sequence govern the expression of Late genes during virus infection? How does the TATT sequence affect recruitment of transcription initiation factors? To address these questions, it will be important to analyze the role of individual nucleotides of Late gene promoters in controlling Late gene expression. To do this we would utilize a luciferase reporter vector containing the CMV ori-lyt, which is essential for the activation of Late promoters (63). We would decipher the cis elements required for activation of Late gene transcription by putting Late gene promoter regions upstream of the luciferase in the reporter vector. Mutagenesis of this promoter region would give insight into what sequences are important and sufficient for Late gene expression and what effect the mutations have on transcription factor binding.

A study was done in MHV68 that analyzed the effects of promoter engagement by ORF24, the homologue of UL87 in HCMV. ORF24 and UL87 are thought to be virally encoded TATA binding proteins (62, 86, 89). The study showed that the binding of ORF24 to Late gene promoters was consistently reduced upon the mutation of the TATT sequence. However, the mutation still allowed at least some transcription from the Late gene promoter, which suggests that additional regulatory factors play a role in Late gene expression. Thus, future studies in HCMV will likely reveal the importance of protein binding to Late gene promoter sequences, and define how the actual sequence of the Late promoter regulates its kinetic class.

**Confirm the other interaction partners important for pM79/pM92 function.** Though RNAP II is critical for viral gene transcription, there are many other binding partners that copurified with pM79 and pM92 during infection. One such protein is a dead-box helicase. Though the function of this family of helicases is not completely defined in virus infection, a recent study in HSV-1 showed that a dead-box helicase could bind to actively replicating viral DNA and RNAP II during infection (60). The dead-box helicase and
its interaction with DNA and RNAP II were required for efficient expression of both Early and Late viral genes. Thus, it would be interesting to see if dead-box helicases have a similar role in CMV infection and, if so, to define what exactly that role is.

Furthermore, it would be interesting to define the role of the other mystery proteins that copurified with pM79 or pM92. The identification of any of these proteins as having a role in Late gene transcription will be informative for the field.

Determine the impact of mutations in pM79 and pM92 on the viral replication cycle in mice. MCMV infection in mice is an important model for preclinical evaluation of antiviral compounds, most of which have been directed at highly conserved viral DNA replication proteins. The Late transactivator proteins present an attractive, alternative target for therapeutic intervention in CMV disease. The functional conservation between pM79/pM92 and pUL79/pUL92 justifies MCMV as a credible model to test the effect of targeted antivirals \textit{in vivo}. However, none of the Late gene transactivator mutants have been tested for their growth in mice. In HSV-1 a study was done where the ICP22 mutant was put into mice. Interestingly, the mutant virus was able to grow to high titers in the mouse model. This was not surprising as the ICP22 mutant has an essential phenotype that is cell type specific (reviewed in (70)). Since most of the studies for beta and gammaherpesvirus transactivators have been done \textit{in vitro} and in the same cell lines, it will be important to test what impact pM79 and pM92 mutants have on \textit{in vivo} infection.

If these genes are essential for \textit{in vivo} infection, then it is very possible that drugs could be designed to target the transactivators during infection. It is also possible that transactivator mutants could serve as vaccine candidates. However, this would depend on just how many lytic genes are down regulated in the absence of a single Late gene transactivator. The immune response is primarily designed to recognize lytic antigens; so a mutant virus that produces no Late genes would be a poor vaccine candidate.

Determine what role pM79 may have in regulating latency. Finally, it is tempting to speculate that Late gene expression regulators such as pM79 could play a role in the establishment of latency. As both proteins are essential for the lytic viral life cycle, regulation of their activity and/or expression may be a
deciding factor for viral latency and reactivation. This concept is supported by the fact that RTA seems to regulate the expression of ORF18 in MHV68 virus infection (36). Since RTA is the lytic switch for gammaherpesviruses, it is possible that a similar mechanism is found in betaherpesviruses.

No molecular definition of CMV latency exists, but latency has been defined operationally as the inability to detect infectious virus despite the presence of viral DNA. Both HCMV and MCMV infect a variety of cell types in their respective host, but latency seems to be confined to cells of the myeloid lineage (7, 25, 31, 67, 68). Similar to HCMV, chromatinization and recruitment of cellular repressors to the viral DNA and to the major immediate early (MIE) gene locus are critically involved in the in vivo establishment of MCMV latency (57, 58). The latent viral genome is associated with repressive chromatin in immature myeloid cells, whereas virus reactivation is accompanied by chromatin remodeling and initiation of transcription at the MIE locus during cell-differentiation (69). The IE genes regulated by the MIE promoter (MIEP) act as essential transactivators of Early and Late genes (62). Therefore, MIEP transcriptional activity is generally considered an important checkpoint in CMV latency and reactivation.

To test the involvement of pM79 in the reactivation of HCMV from latency, we would analyze the chromatinization of the M79 promoter region, and assess how the state of chromatin repression is altered after the induction of reactivation. This analysis would be followed by monitoring the cells for expression of pM79. Since DNA replication is required for late gene expression during lytic infection, it would be interesting to see if the requirement is altered during reactivation from latency. The fact that MCMV BAC DNA can be reconstituted in the absence of tegument proteins makes it unclear what the requirement for a Late gene transactivator would be during reactivation. However, it has not been determined exactly what role pM79 may have in viral DNA synthesis, especially in light of the interaction data in Chapter IV.
Concluding Remarks

We have identified pM79 and pM92 as novel Late gene regulators in MCMV lytic infection, shown their interaction with not only each other but with RNAP II, and demonstrated their conserved functions with the HCMV homologs pUL79 and pUL92. We have published two papers that detail our research of M79 and M92 (12, 13). pM79 and pM92 represent potential new targets for therapeutic intervention in CMV disease, and a gateway into studying a largely uncharted viral process that is critical to the viral life cycle. Our current efforts aim at a better mechanistic understanding of the role of pM92 in Late gene regulation. With the mass spectrometry results (Chapter IV), we have identified cellular and viral factors that interact with pM92. Though pM79 and RNAP II are proven partners of pM92, it is almost certain that all additional partners will provide important insight into the function of pM92. Furthermore, genetic and protein analysis to identify functional domains and structural elements of pM92 will be invaluable to understand the mechanistic basis for its activity and to determine additional functions that pM92 may have. We plan to publish an additional paper to document the mass spectrometry interaction data for pM79 and pM92, and also to outline the functional domains of pM92. The work of this thesis has greatly contributed to understanding a very important process in betaherpesvirus biology, and has given insights that will help our understanding of novel host gene functions, and how nuclear replicating viruses regulate transcription of their genome.
References


47. **Kalamvoki, M., and B. Roizman.** 2011. The histone acetyltransferase CLOCK is an essential component of the herpes simplex virus 1 transcriptome that includes TFIID, ICP4, ICP27, and ICP22. Journal of virology **85:**9472-7.


57. **Liu, X. F., S. Yan, M. Abecassis, and M. Hummel.** Biphasic recruitment of transcriptional repressors to the murine cytomegalovirus major immediate-early promoter during the course of infection in vivo. J Virol **84:**3631-43.


60. **Mallon, S., B. T. Wakim, and B. Roizman.** Use of biotinylated plasmid DNA as a surrogate for HSV DNA to identify proteins that repress or activate viral gene expression. Proc Natl Acad Sci U S A **109:**E3549-57.


Figure 1. Amino acid alignment of the UL79 protein family. Amino acid alignment of UL79 proteins from mouse, rat, human, and rhesus, CMVs, with HHV6, MHV68, KSHV, and EBV. Amino acids highlighted in red are conserved across all species.