ANALYSIS OF THE HUMAN CYTOMEGALOVIRUS TRANSCRIPTOME AND IDENTIFICATION AND CHARACTERIZATION OF A HCMV GENE INVOLVED IN DISRUPTION OF INTERFERON SIGNALING

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Bindu Raghavan, M.Sc

* * * * *

The Ohio State University

2008

Dissertation Committee:

Joanne Trgovcich, Ph.D., Adviser
Daniel D. Sedmak, M.D.
W. James Waldman, Ph.D
Marshall Williams, Ph.D

Approved by

______________________________

______________________________

Adviser

Molecular Cellular and Developmental Biology Graduate Program
ABSTRACT

Human cytomegalovirus (HCMV) is a ubiquitous pathogen infecting 60-100% of the world population. Infection of an immunocompetent host by HCMV is largely asymptomatic. In populations that are immunosuppressed including AIDS patients, transplant recipients, and neonates who have immature immune systems, this virus causes severe disease and morbidity. In fact, HCMV is the leading infectious cause of congenital defects resulting in deafness, blindness, cerebral palsy and mental retardation in children. After primary infection of the host via contact with contaminated body fluids the virus is able to persist in the host and establish latency. Although the host immune responses can control the spread of the virus, the virus persists in the host for his or her lifetime. The virus reactivates from latency multiple times and is shed in the body fluids of the host. The ability of HCMV to establish such lifelong infections is attributed in part to its ability to modulate various aspects of host immune responses.

Central to the antiviral immune response of the host is interferon (IFN) mediated activation of cellular gene products that directly or indirectly interfere with various aspects of the viral life cycle. It is known that IFN mediated signaling is attenuated in HCMV infected cells. We set out to identify the HCMV gene products that are involved in modulating IFN signaling and elucidate their mode of action.
For this purpose we constructed a cDNA library of the HCMV laboratory strain AD169. Characterization of this cDNA library revealed an abundance of non-protein coding transcripts and transcripts in antisense orientation to known or predicted HCMV genes in the HCMV transcriptome. We screened this cDNA library for genes that disrupt IFN signaling, and the *UL123* gene was the first candidate gene identified. *UL123* encodes the immediate early transcriptional transactivator IE1. Our results indicate that the IE1 protein can attenuate both type I and type II IFN signaling. Further investigation of the type II IFN signaling pathway in IE1 overexpressing cells revealed that IE1 does not inhibit the proximal events of IFN signaling. However we did discover a defect in binding of STAT1 to target DNA elements in response to type II IFN stimulation in IE1 overexpressing cells. Using a series of truncation mutants we were able to map this function of IE1 to the carboxy terminal domain of the IE1 protein. Our data indicate that a novel mechanism is employed by IE1 to interfere with the association of STAT1 with its target DNA.

The research findings presented here reveal novel aspects of HCMV biology including the discovery of the increased complexity of the HCMV transcriptome and a potentially novel mechanism of viral attenuation of host IFN mediated responses. This enhanced understanding of HCMV biology may have implications in the development of antiviral drugs and vaccines.
DEDICATION

Dedicated to Anil, my husband, without whose support this would not be possible
ACKNOWLEDGMENTS

First and foremost I would like to thank my advisor Dr. Joanne Trgovcich, who gave me the opportunity to work in her lab. Joanne I really appreciated your open door policy and your willingness to discuss my work at any time. I am quite sure I do not know any other person who is so optimistic and enthusiastic about science. I would also like to acknowledge the contribution of my advisory committee members Dr. Daniel Sedmak, Dr. James Waldman and Dr. Marshall Williams for their helpful suggestions and feedback over the course of my research. I thank American Society for Microbiology for copyright permission for chapter two. I must also thank members of the Trgovcich lab for all their help throughout my research. My special thanks to Guojuan Zhang for being such a wonderful co-worker and all the work she put into the construction of the cDNA library. I thank Mark Kotur, Jackie Cheatham and Mary Sivulich for all their help. I am very grateful to Dr. Susheela Tridandapani for allowing me to use her nucleoporator several times over the course of the last two years. I thank my husband Anil, for his support especially during the difficult times over the last five years. Finally I would like to thank my parents Meethal Raghavan and Padmaja Raghavan for letting me pursue my dreams.
VITA

November 17, 1973.................................................................Born- Kannur, India

1995..............B.Sc Agriculture, University of Agricultural Sciences, Bangalore

1998..........M.Sc Biotechnology, University of Agricultural Sciences, Bangalore

1999-2003...........................................Scientist at Monsanto Research Center, Bangalore

2003- till present..............................Graduate Teaching and Research Associate

at The Ohio State University

Fellowships:

The Ohio State University Alumni Grant for Graduate Research and Scholarship

(Spring 2007)

PUBLICATIONS


**FIELDS OF STUDY**

Major Field: Molecular Cellular and Developmental Biology
### TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Chapter</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abstract</td>
<td>ii</td>
</tr>
<tr>
<td>Dedication</td>
<td>iv</td>
</tr>
<tr>
<td>Acknowledgments</td>
<td>v</td>
</tr>
<tr>
<td>Vita</td>
<td>vi</td>
</tr>
<tr>
<td>List of Tables</td>
<td>x</td>
</tr>
<tr>
<td>List of Figures</td>
<td>xi</td>
</tr>
<tr>
<td>Chapters:</td>
<td></td>
</tr>
<tr>
<td>1. Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Human Cytomegalovirus – an Introduction</td>
<td>1</td>
</tr>
<tr>
<td>Virus Structure</td>
<td>4</td>
</tr>
<tr>
<td>Life Cycle of HCMV</td>
<td>7</td>
</tr>
<tr>
<td>HCMV Pathogenesis</td>
<td>11</td>
</tr>
<tr>
<td>Immune modulation by HCMV</td>
<td>15</td>
</tr>
<tr>
<td>Role of Interferons in antiviral defense</td>
<td>20</td>
</tr>
<tr>
<td>Viral Evasion of Interferon action</td>
<td>30</td>
</tr>
<tr>
<td>Prelude to thesis</td>
<td>33</td>
</tr>
<tr>
<td>2. Antisense Transcription in the Human Cytomegalovirus Transcriptome</td>
<td></td>
</tr>
<tr>
<td>Introduction</td>
<td>34</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>Table</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Affect of HCMV on Antigen Presentation to immune cells</td>
</tr>
<tr>
<td>1.2</td>
<td>Affect of HCMV on immune cell function</td>
</tr>
<tr>
<td>1.3</td>
<td>Role of Interferons in intrinsic antiviral activity</td>
</tr>
<tr>
<td>1.4</td>
<td>Role of Interferons in adaptive immunity</td>
</tr>
<tr>
<td>2.1</td>
<td>Primers for RT-PCR and 5’ and 3’ RACE</td>
</tr>
<tr>
<td>2.2</td>
<td>Orientation of cDNA Clones</td>
</tr>
<tr>
<td>2.3</td>
<td>Genes for which \textit{cis}-Natural S-AS Pairs were Identified and Their Properties</td>
</tr>
<tr>
<td>2.4</td>
<td>Genes with Predicted S-AS Pairs</td>
</tr>
<tr>
<td>2.5</td>
<td>Orphaned AS Transcripts</td>
</tr>
<tr>
<td>2.6</td>
<td>Functional Classes of Genes with Verified or Predicted \textit{cis}-Natural Sense-Antisense Pairs</td>
</tr>
</tbody>
</table>
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Diagrammatic representation of an HCMV virion</td>
<td>6</td>
</tr>
<tr>
<td>1.2</td>
<td>The Type I IFN signaling pathway</td>
<td>24</td>
</tr>
<tr>
<td>1.3</td>
<td>The Type II IFN signaling pathway</td>
<td>25</td>
</tr>
<tr>
<td>2.1</td>
<td>Verification of virally-derived AS transcripts</td>
<td>67</td>
</tr>
<tr>
<td>2.2</td>
<td>Verification of antisense transcripts in endothelial cells infected with a clinical strain of HCMV</td>
<td>69</td>
</tr>
<tr>
<td>2.3</td>
<td>Northern blot analysis of S and AS transcripts</td>
<td>74</td>
</tr>
<tr>
<td>2.4</td>
<td>Northern blot analysis to examine relative abundance of transcripts</td>
<td>77</td>
</tr>
<tr>
<td>3.1</td>
<td>Reduced CD2 cell surface levels in cells expressing IE1</td>
<td>99</td>
</tr>
<tr>
<td>3.2</td>
<td>Diminished interferon induced gene expression in human fibroblasts expressing IE1</td>
<td>100</td>
</tr>
<tr>
<td>3.3</td>
<td>Schematic of the IE1 protein and truncation mutants generated in this study</td>
<td>102</td>
</tr>
<tr>
<td>3.4</td>
<td>The carboxy-terminal region of IE1 is required for disruption of IFNγ signaling</td>
<td>103</td>
</tr>
<tr>
<td>3.5</td>
<td>Disruption of ND10 structures in cells expressing IE1</td>
<td>105</td>
</tr>
</tbody>
</table>
3.6 IE1 does not affect the state or abundance the initial signaling molecules of IFNγ pathway

3.7 Nuclear translocation of STAT1 is not obstructed by IE1

3.8 Reduced binding of STAT1 to GAS elements in the presence of IE1

3.9 Coimmunoprecipitation studies fail to reveal a stable interaction between IE1 and STAT1

4.1 Reduced CD2 cell surface levels in cells expressing ssUS3

4.2 Stat1 phosphorylation at Immediate Early and Late times of infection

4.2 Stat1 phosphorylation at Early times of infection
CHAPTER 1

INTRODUCTION

Human Cytomegalovirus- an introduction

Human Cytomegalovirus (HCMV) belongs to the herpes virus family (herpesviridae). It is the prototypical virus of the subfamily betaherpesvirinae. It is estimated that 60%-100% of the human population is infected with cytomegalovirus (223). However in the immunocompetent host this is not a cause for concern as the virus does not cause serious disease. The virus remains in a latent or low-level persistent state after primary infection of the host. It reactivates from time to time during the lifetime of the host usually in inflammed tissues. It is now speculated that HCMV could be a cofactor in the development of inflammatory diseases like atherosclerosis, transplant vascular sclerosis, rheumatoid arthritis, etc. It is however, not clear whether HCMV actually plays a causative role in these diseases (222). HCMV has also been implicated in immune senescence that may lead to severe health problems in the elderly (5, 121). However it is in the immunocompromised host that HCMV becomes a major pathogen. In immunocompromised populations like AIDS patients, transplant
recipients who undergo immunosuppressive therapy and neonates who have immature immune systems, HCMV causes serious morbidity and mortality. In fact it is the leading infectious cause of congenital defects (29). Presently there are four drugs that are approved for CMV treatment, which can control virus replication but cannot completely eradicate the virus. The nucleoside analog ganciclovir is the most commonly used drug to treat CMV associated disease (56). In addition to this the nucleotide analog cidofovir (58) and the pyrophosphate analog foscarnet (194) are currently marketed as systemic anti-CMV drugs. There is also an antisense RNA targeting the viral immediate early genes called fomivirsen, which is administered by intraocular injection to treat CMV retinitis (91). All these drugs with the exception of fomivirsen target the viral DNA polymerase and prolonged use is associated with the development of drug resistant strains of the virus. These drugs are also associated with high toxicities, poor efficacy and oral bioavailability (19). Furthermore no effective vaccines are currently available to protect from HCMV infections. Therefore, there is an urgent need to find new therapies for HCMV.

In order to develop better antiviral drugs and possibly vaccines against HCMV it is essential to understand the biology of the virus. This includes knowledge of the full coding potential of the large HCMV genome of more than 230Kb. Analysis of the full coding potential of the HCMV genome may identify novel targets for antiviral drug development. It is estimated that the wild-type HCMV genome codes for approximately 165 protein coding genes. In addition to this there are
several noncoding transcripts and antisense transcripts encoded by the HCMV genome (282). The function of a majority of these gene products is not known. Understanding the function of more HCMV genes could provide more insight into the biology of the virus. Of the predicted 165 protein coding genes of HCMV only 45-57 are required for viral replication (193, 223). So what is the function of the majority of the viral genes? In order to answer this question, two important aspects of HCMV biology should be considered. One is that HCMV induces both humoral and cell mediated immune responses in the host (77, 193). The other is that inspite of these immune responses the virus is able to persist for the lifetime of the host. Taken together these facts imply that a majority of the viral genes are probably devoted to modulation of host immune system, to allow the virus to maintain a persistent infection in the host. An understanding of these genes and the mechanisms by which they modulate the host immunological responses is key to winning the war against this supreme pathogen. This chapter will cover aspects of HCMV biology including virus structure, life cycle of HCMV, pathogenesis caused by HCMV, immune evasion strategies employed by HCMV, the role of interferons in controlling viral infections and, viral evasion of interferon action.
**Virus Structure**

Human cytomegalovirus has a linear dsDNA genome. The viral genome consists of the Unique Long (UL) region and the Unique Short (US) region. At the ends of the US and UL regions there are the inverted repeat sequences called terminal repeats (TRS and TRL) and at the junction of the UL and US region there are the inverted repeats called internal repeats (IRS and IRL). In addition to these there are direct repeats at both ends of the genome and also at the US and UL junction between the IRS and IRL sequences. The viral genome is carried inside a proteinaceous capsid that has icosahedral geometry. The major capsid protein is pUL86 and unlike the capsid proteins of other herpes viruses it is not very immunogenic. The UL46 and UL85 gene products constitute the minor capsid proteins. pUL46 is located internally in the capsid and may function to anchor the viral genome. In addition to these there is a small capsid protein (SCP) derived from the UL48/49 genomic region that is essential viral growth. The nucleocapsid is encased in an envelope composed of lipids derived from the inner nuclear envelope or the membranes of endosomes or golgi. The viral envelope is rich in glycoproteins encoded by the virus. There are six major envelope glycoproteins in HCMV including gB, gM, gN, gO, gH, gL, and the most abundant of these glycoproteins is gB. gB is required for interaction with host cell receptors and neutralizing antibodies to gB can inhibit viral entry. gM and gN form a heterodimer that is implicated to have a role in tethering to the host cell.
cell surface via interactions with heparin sulphate proteoglycans on the host cell (117). gH, gL and gO form a heterotrimeric complex which, along with gB is required for fusion with the host cell membrane (53). Another glycoprotein which is a product of the UL4 gene is a minor envelope constituent (160). The region between the envelope and the nucleocapsid is called the tegument. The tegument is composed predominantly of two viral proteins pp65 and pp150 (109, 112). pp71 is a minor tegument protein in terms of abundance compared to pp65 and pp150, but it plays an important role as the transactivator of Immediate Early (IE) genes (144). In addition to this many other proteins and RNAs derived from both the virus and the host cell get packaged into the tegument (243). A schematic representation of a HCMV virus particle is shown in Figure 1.1.
Figure 1.1  Diagrammatic representation of an HCMV virion
Life cycle of HCMV

The initial step in infection of the host cell by HCMV involves tethering to the host cell surface. This tethering ensues through interactions between the viral envelope glycoproteins gB or gM/gN heterodimer and heparin sulphate proteoglycans on the host cell membrane (54). Following this the virion stably attaches to the host cell surface and this interaction is proposed to take place via the interaction of gB with the cellular receptor. It was proposed that the Epidermal Growth Factor Receptor (EGFR) is the cellular receptor for HCMV as virus binding to the host cell caused the same signaling cascade as treatment of cells with EGF (264). In addition to this, β1 and β3 integrins in the host cell membrane act as co-receptors for HCMV (68). It should be noted here that EGFR is expressed on most cells that HCMV infects, however dendritic cells, neutrophils and monocytes / macrophages do not encode the EGFR, but the virus is able to gain entry into these cells. More importantly monocytes are considered to be one of the sites of HCMV latency. Therefore HCMV must be using some other receptor for entry into these cells. Recently it has been reported that the EGFR receptor is not required for virus entry but rather the integrins are primary cellular receptors for HCMV (111). From these reports we can conclude that integrins play an important role in attachment of HCMV to the host cells and the role of the EGFR receptor is still under contention. However these reports do not rule out the possibility of the existence of other receptors for HCMV. Once the virus particle attaches to the host cell membrane there is fusion of the viral
envelope with the plasma membrane of the cell releasing the tegument and the
viral nucleocapsid into the cytoplasm of the cell. The nucleocapsid associates with
the cellular microtubule network to facilitate movement to the nucleus, where it
docks at the nuclear pore to release the viral genome into the nucleus (175). This
initiates a temporally regulated cascade of viral gene expression. First, the
tegmentum proteins activate the transcription of the viral IE genes. Following
transcription and translation the IE proteins translocate to the nucleus and activate
the transcription of the early (E) genes. The early gene products then initiate viral
DNA replication, which acts as trigger for the transcription of the late (L) genes.
The late gene products are mainly the viral structural proteins that after being
translated in the cytoplasm, translocate to the nucleus and form the viral capsid.
These capsids package the newly synthesized viral DNA forming new
nucleocapsids. After this point the nucleocapsids acquire a tegument and envelope
in order to form the mature virion. This process is not entirely understood. It has
been suggested that an envelope is acquired while the nucleocapsids bud through
the inner nuclear envelope and this is lost while exiting through the perinuclear
space (180). In the cytoplasm the virus acquires its tegument consisting mainly of
viral matrix proteins and viral RNAs. Following this the virus acquires the final
envelope either by budding into early endosomes (249) or the trans golgi network
(104, 212). Finally the virus containing vesicles fuse with the plasma membrane
of the cell releasing mature virus particles to the outside. The life cycle discussed
above is the typical lytic life cycle. The virus may spread from cell to cell in this
manner or through the blood to reach sites of latency. During latency the virus persists in the host cells without production of any infectious virus. The regulated cascade of gene expression seen during a lytic infection is absent in latency. Nevertheless there is some viral gene expression during latency. CMV latency specific transcripts (CLT) have been observed in experimentally infected granulocyte monocyte progenitor cells (GMP). These are spliced and unspliced transcripts derived form both strands of the HCMV major IE gene region (122). The viral IL-10 homolog has also been seen in this experimental model of latency (113). However the role of these genes in natural latency is not known. In naturally latent monocytes the UL138 transcript has been found and loss of the UL138 renders the virus unable to establish latency (84). Another transcript that has been found in naturally latent monocytes is UL82AS. This transcript is antisense to both UL81 and UL82. Since UL82 codes for the viral transcriptional transactivator pp71, it is proposed that the UL82AS may act to prevent expression of pp71 so as to prevent transactivation of the viral immediate early genes (12). Apart from this not much is known about latency specific transcripts of HCMV. There is now a consensus that CD34+ progenitor cells and monocytes harbor latent virus. Differentiation of monocytes or CD34+ progenitor cells into macrophages or DCs results in reactivation of the virus. However a productive reactivation is seen only in DCs (198). It is not clear what the specific signals are for reactivation of the virus. An impediment to understanding the reactivation process is that in an infected individual there are sites of true latency
and yet other sites of low level lytic infection. Whereas in true latency no infectious virions are produced, in a low level lytic infection, infectious virions are produced but at subclinical levels. Another confounding factor is that the full spectrum of sites of HCMV latency is not known.
HCMV Pathogenesis

Primary infection with HCMV is normally asymptomatic in immunocompetent individuals. Sometimes it may result in mononucleosis like symptoms including fever, myalgia and cervical adenopathy similar to symptoms of Epstein Barr Virus (EBV) mononucleosis. HCMV mononucleosis may account for 8% of all mononucleosis cases (29). Epithelial cells are the targets of primary infection by HCMV. This facilitates the spread of progeny virus to underlying tissue which may account for the prevalence of HCMV in fibroblasts in tissue samples from infected patients (217, 218). Replication of virus in epithelial cells also allows for release of progeny virus in bodily secretions. The virus is spread through close personal contact with contaminated body fluids including saliva, breast milk, semen, blood products etc. It can also be transmitted by solid organ and bone marrow transplants (25, 220). The virus is usually acquired in early childhood through breast milk or via close personal contact in day care centers. An immunocompetent host can control the virus but cannot completely eradicate it. This is in part due to the ability of the virus to establish latent infections. From the site of initial infection usually epithelial cells and fibroblasts the virus travels by cell to cell transfer or through blood and reaches sites of latency. HCMV is known to establish latency in cells of the myeloid lineage, although all the cell types capable of harboring latent virus have not been defined. Presently cells of the myeloid lineage including CD34+ progenitor cells and monocytes are
considered the major reservoir of latent HCMV (215, 216). The virus reactivates from these latent reservoirs of the immunocompetent host, from time to time and is shed in body fluids.

There is some evidence that suggests that in the immunocompetent host, HCMV may be associated with inflammatory diseases like atherosclerosis, restenosis and transplant vascular sclerosis (86, 124, 288). HCMV has been found in the bowels of patients with inflammatory bowel disease (191). There is some evidence for the association of HCMV with other autoimmune diseases like rheumatoid arthritis, sjogrens syndrome and systemic lupus erythematosus (169, 210). HCMV is also implicated in certain cancers including colon cancer, malignant glioma, prostate cancer (51, 96, 205). However it is not clear whether HCMV plays a causative role in the development of these diseases or the inflammation associated with these diseases promotes HCMV reactivation and replication. It has been reported that HCMV reactivation from latency is associated with elevated serum levels of IFNγ, TNFα (Tumor necrosis factor α) and IL-2(Interleukin 2) as is seen during immune response involving activated T cells(224, 225). Studies have revealed that 50% of the CD8+ T cell repertoire of elderly people is devoted towards HCMV (162). One of the implications of this finding is that the virus reactivates multiple times during the lifetime of the host requiring the host to mount an immune response to keep it in check. The virus on the other hand has evolved various strategies to modulate host immune responses and cellular processes including the cell cycle, cell migration, morphogenesis, and angiogenesis.
and apoptosis. The ability of the virus to influence these cellular processes may contribute to its role as a cofactor in inflammatory diseases and cancer (15, 222, 260).

While HCMV disease in the immunocompetent host may be largely asymptomatic and the association with inflammatory diseases may not presently be concrete, HCMV disease in the immunocompromised host results in serious morbidity. Under conditions of immune suppression HCMV becomes a major pathogen. Depending on the nature of the immunosuppression, HCMV disease presents with different symptoms. In AIDS patients HCMV reactivation causes retinitis and gastrointestinal disease. The advent of highly aggressive antiretroviral therapy (HAART) for AIDS, has reduced the incidence of HCMV disease in this patient population (172). However in AIDS patients with CMV retinitis, the elevated CD4+ T cell counts following HAART therapy results in immune recovery vitritis. Immune recovery vitritis involves inflammation of the vitreous that may result in loss of visual acuity (116). In case of immunosuppressed transplant recipients HCMV disease presents with spiking fever, leucopenia, malaise, hepatitis and pneumonia. In fact there is 20-40% occurrence of HCMV disease in solid organ transplant recipients (204). HCMV associated disease initially starts with the transplanted organ and later becomes systemic leading to hepatitis, pneumonitis and enteritis (245). HCMV pneumonitis is the major complication resulting from hematopoietic stem cell transplants. This is reported to result in mortality rates as high as 60-80% (73,
Perhaps the most severe disease manifestation caused by HCMV is through in utero infection of the fetus, due to primary infection of the mother during gestation. Neonates infected in utero with HCMV present with birth defects including blindness, deafness, cerebral palsy and mental retardation (7, 29). HCMV is the leading infectious cause of congenital defects in the developed world including the US. In fact the health care cost associated with congenital HCMV was estimated to be 1.8 billion dollars annually in the US in the 1990s (70). In most parts of the world HCMV is acquired in the first few years of life and an intact immune system can control the spread of the virus. Due to better health care, hygiene and prevalence of formula feeding of infants, the age of primary infection with HCMV is delayed in developed countries. Consequently women of child bearing age may be exposed to a primary infection with HCMV during pregnancy, resulting in increased incidence of congenital HCMV in the developed parts of the world (7). Therefore congenital HCMV poses an increased health risk and this underscores the need for development of effective prophylactic measures against this virus.
Immune modulation by HCMV

In order for any virus to persist for the lifetime of the host it must be able to:
infect cells without causing cytopathology, maintain the viral genome in the host
cell and, avoid elimination by the host immune system. HCMV fulfills all these
three criteria. It has been successful in establishing a persistent infection in the
human host by evolving numerous strategies to modulate host immune responses
(99, 223).

Upon invasion by an intracellular pathogen an infected cell signals to cells of the
immune system by presenting pathogen derived antigens in the context of Major
Histocompatibility Complex (MHC) class I. In addition to this, professional
antigen presenting cells including dendritic cells, macrophages and B cells present
pathogen derived antigens to immune cells in the context of MHC class II.

HCMV has various strategies to block processing and presentation of antigens in
the context of MHC class I, and MHC class II. Table 1.1 details the different
tactics used by HCMV to downregulate antigen presentation, the viral genes
involved in these, and their modes of action. The affect of HCMV on antigen
processing and, the function of antigen presenting cells like DCs, are also
described in Table 1.1. The immune cells, namely T cells, B cells and NK cells,
recognize infected cells and are activated by signals from the infected cells.

HCMV has mechanisms to inhibit activation of the immune cells. Furthermore,
the immune cells kill infected cells by inducing apoptosis of infected cells upon
recognition. Here too the virus has mechanisms to inhibit apoptosis. Antigen
presenting cells and immune cells recruit more immune cells to the site of infection by producing chemoattractant cytokines. HCMV has devices to modulate cytokine and chemokine signaling either by encoding viral cytokine/chemokine homologs or decoy receptors for several cytokines and chemokines. Table 1.2 details the different strategies used by HCMV to inhibit immune cell function, the viral genes involved in these, and their modes of action. It is clear from this table that the virus has a plethora of tactics to modulate both the innate and the adaptive immune responses.

In addition to this the virus also has means to disarm the first line of defense against viruses namely IFN. IFNs are cytokines that stimulate pathways leading to the induction of proteins that confer an antiviral state upon the cell. They also regulate the function of cells involved in the adaptive immune response. Chapter 3 will discuss this in more detail and present a detailed account of the HCMV UL123 gene product that disrupts type I and type II IFN signaling, by a novel mechanism.
<table>
<thead>
<tr>
<th>Immune Evasion Strategy</th>
<th>Viral Gene</th>
<th>Mechanism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>MHC class I presentation</td>
<td>US2</td>
<td>Export of MHC class I heavy chains from the ER via the Sec61 translocon</td>
<td>(271)</td>
</tr>
<tr>
<td></td>
<td>US3</td>
<td>Retains MHC class I molecules in the ER and tapasin dependent peptide loading on MHC class I</td>
<td>(2, 115, 179)</td>
</tr>
<tr>
<td></td>
<td>US6</td>
<td>Inhibits TAP mediated peptide entry into the ER</td>
<td>(3, 100, 135)</td>
</tr>
<tr>
<td></td>
<td>US11</td>
<td>Shuttles MHC class I heavy chains from the ER to the cytosol</td>
<td>(270)</td>
</tr>
<tr>
<td></td>
<td>UL82 (pp71)</td>
<td>Retains MHC class I in the ER or cis golgi</td>
<td>(250)</td>
</tr>
<tr>
<td>MHC class II presentation</td>
<td>US2</td>
<td>Degrades MHC class II alpha chain and HLA DM</td>
<td>(248)</td>
</tr>
<tr>
<td></td>
<td>US3</td>
<td>Binds MHC class II alpha and beta chains and prevents association with Invariant chain</td>
<td>(98)</td>
</tr>
<tr>
<td></td>
<td>UL83 (pp65)</td>
<td>Directs HLA DR to LAMP-1 and rab7 positive vacuoles where HLA DR is degraded</td>
<td>(174)</td>
</tr>
<tr>
<td>Antigen processing</td>
<td>UL83 (pp65)</td>
<td>Inhibits HCMV IE1 antigen presentation to CD8+ T cells</td>
<td>(79)</td>
</tr>
<tr>
<td>Dendritic cell function</td>
<td>Unknown</td>
<td>Inhibits migration of immature DCs to infection sites by downregulating CCR1 and CCR5 on the surface of DCs.</td>
<td>(256)</td>
</tr>
</tbody>
</table>

Table 1.1: Affect of HCMV on Antigen Presentation to immune cells
<table>
<thead>
<tr>
<th>Immune Evasion Strategy</th>
<th>Viral Gene</th>
<th>Mechanism</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Affect on NK cell function</td>
<td>UL16</td>
<td>Reduced cell surface expression of the ligand for NK cell activating receptor NKG2D Resistance to perforin, granzymeB and porcine NK lysin</td>
<td>(173, 203, 253, 266)</td>
</tr>
<tr>
<td></td>
<td>UL40</td>
<td>Upregulation of HLA E and HLA G which binds killer inhibitory receptor</td>
<td>(177, 182, 246)</td>
</tr>
<tr>
<td></td>
<td>UL18</td>
<td>MHC class I homolog, inhibits NK cell function</td>
<td>(201)</td>
</tr>
<tr>
<td></td>
<td>UL141</td>
<td>Blocks cell surface expression of CD155 ligand for NK cell activating receptor</td>
<td>(247)</td>
</tr>
<tr>
<td>Affect on B cell function/ Humoral immunity</td>
<td>TRL11/ IRL11 UL119/118</td>
<td>Viral Fc receptor binds human IgG</td>
<td>(8, 141)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Increased expression of CD35, CD46, CD55 resulting in inhibition of complement mediated lysis</td>
<td>(227, 229, 230)</td>
</tr>
<tr>
<td>Affect on T cell function</td>
<td>CMV IL-10 homolog</td>
<td>Inhibits mitogen stimulation of PBMC, production of proinflammatory cytokines and activation of T cells</td>
<td>(228)</td>
</tr>
<tr>
<td></td>
<td>Unknown</td>
<td>Induces TNFα mediated release of ProstaglandinE2 which inhibits T cell proliferation</td>
<td>(170)</td>
</tr>
<tr>
<td>Affect on apoptosis</td>
<td>UL122, UL123</td>
<td>Inhibits apoptosis by unknown mechanisms</td>
<td>(14, 289)</td>
</tr>
<tr>
<td></td>
<td>UL36</td>
<td>Binds prodomain of caspase 8, to interfere with apoptotic processes</td>
<td>(219)</td>
</tr>
<tr>
<td></td>
<td>UL37</td>
<td>Inhibits apoptosis mediated by Fas/FasL interaction</td>
<td>(82, 195)</td>
</tr>
</tbody>
</table>

**Table 1.2: Affect of HCMV on immune cell function**

continued
Table 1.2 (continued)

<table>
<thead>
<tr>
<th>Affect on cytokines and their receptors</th>
<th>US28, US27, UL33</th>
<th>Chemokine receptor homolog which binds $\beta$ chemokines</th>
<th>(44, 167)</th>
</tr>
</thead>
<tbody>
<tr>
<td>UL78</td>
<td></td>
<td>Potential GPCR homolog</td>
<td>(43)</td>
</tr>
<tr>
<td>UL146, UL147</td>
<td></td>
<td>CXC chemokine homolog, attracts neutrophils to site of infection to facilitate viral spread</td>
<td>(183)</td>
</tr>
<tr>
<td>UL144</td>
<td></td>
<td>TNF receptor homolog</td>
<td>(14)</td>
</tr>
</tbody>
</table>
Role of Interferons in antiviral defense

In 1957 Isaacs and Lindenmann discovered that virus infected chick embryos produce a secreted factor that could confer a virus resistant state upon cells (110). This factor was named interferon and research over the past 50 years has revealed the components of the IFN signaling system and the mechanisms by which IFN induced proteins act as antiviral agents, in addition to other biological affects of IFNs. Today IFNs are classified into three categories namely: Type I IFN (including α, β, ω, ε, κ), Type II IFN (IFNγ) and Type III IFN (IFNλ) (23).

There are 13 different IFN αs and only one type of IFN β, IFNω, IFNε and IFN κ in humans. IFN β is referred to as the fibroblast IFN and IFN α is referred to as the leukocyte IFN based on the cell types that are the major source of these IFNs. However, Type I IFNs can be produced by all cell types with the exception of IFNε and IFN κ, which are tissue specific (95, 97, 128). The genes encoding all the Type I interferons are located on chromosome 9 and are devoid of introns (178, 251). All Type I IFNs belong to the helical cytokine family of proteins. Elucidation of the three dimensional structure of the Type I IFN proteins revealed that they contain five alpha helices which are held together by two disulphide bonds (159). Type III interferons are similar to Type I IFNs and induce a similar signaling cascade. They differ from Type I IFNs in that they bind to a different cell surface receptor composed of IL10R2 (Interleukin 10 Receptor 2) and IL28RA (Interleukin 28 Receptor Alpha) (Figure1.2). Type III interferons include IL29 (IFNλ1), IL28A (IFNλ2) and IL28B (IFNλ3). They are encoded by three
genes clustered on chromosome 19. Unlike the Type I IFN genes Type III IFN genes do contain introns. The IFN\(\lambda\) proteins are monomeric proteins with a size of 20Kda (23). There is only one type of Type II IFN, namely, IFN\(\gamma\). IFN\(\gamma\) is referred to as the immune interferon as it is produced by immune cells, including NK cells and activated T cells. A single gene located on chromosome 12 encodes IFN\(\gamma\). The IFN\(\gamma\) gene has three introns and a repetitive DNA element is present in the first intron (87, 291). The IFN\(\gamma\) protein is a single glycosylated protein of 140 amino acids.

Type I IFNs bind to a cell surface receptor, which is expressed on the surface of all cells and is composed of two proteins IFNAR1 (IFN Alpha Receptor 1) and IFNAR2 (IFN Alpha Receptor 2). The cytoplasmic tail of the receptors is associated with protein tyrosine kinases, JAK1 (Janus Kinase 1) and TYK2 (Tyrosine Kinase 2). Binding of Type I IFN to its cognate receptor leads to phosphorylation of JAK1 and TYK2 and tyrosine residues on the cytoplasmic tails of the receptors. This leads to docking of STAT1 (Signal Transducer and Activator of Transcription 1) and STAT2 (Signal Transducer and Activator of Transcription 2) to the receptor cytoplasmic tails via SH2 (Src Homology 2) domains. This results in tyrosine phosphorylation of the STATs. The phosphorylated tyrosine of one STAT1 interacts with the SH2 domain of the other leading to the formation of STAT1-STAT2 heterodimers. The STAT1-STAT2 heterodimers then form a complex with IRF9 called ISGF3 (Interferon Stimulated Gene Factor 3) that translocates to the nucleus and binds ISRE
(Interferon Stimulated Response Element) in the promoters of type I IFN induced genes (186, 208). Type III IFNs induce the same signaling cascade as the Type I IFNs although they use different receptors. The Type III IFN receptor IL10R2 is expressed on all cells like the type I IFN receptors, whereas the IL28RA is not ubiquitously expressed. Type II IFNs bind to a different receptor composed of IFNγR1 (Interferon Gamma Receptor 1) and IFNγR2 (Interferon Gamma Receptor 2). A functional IFNγ receptor consists of 2 chains IFNγR1 and 2 chains of IFNγR2 (Figure 1.3). The cytoplasmic domain of these receptors is associated with the tyrosine kinase JAK1 and JAK2 (Janus Kinase 2). IFNγ signaling involves binding of two molecules of IFNγ to the receptor resulting in the phosphorylation of the cytoplasmic tails of the receptors, the associated tyrosine kinases JAK1 and JAK2 and STAT1. IFNγ signaling induces formation of STAT1 homodimers that translocate to the nucleus and which bind GAS (Gamma Activated sequence) elements in the promoters of type II IFN inducible genes (Figure1.3) (23, 208, 233).

This classical JAK-STAT signaling pathway described for IFN mediated signaling assumes that the STAT proteins are confined to the cytoplasm in unstimulated cells and rapidly accumulate in the nucleus upon cytokine stimulation. However there is mounting evidence for the constant nucleocytoplasmic shuttling of STATs in unstimulated cells. STAT2 has been found to be associated with IRF9 in unstimulated cells and a constitutive NLS in IRF9 allows for nuclear import of STAT2 (11). Upon cytokine stimulation
STAT2 is phosphorylated and heterodimerizes with STAT1 bringing about a conformational change which exposes a new NLS (Nuclear Localization Signal) leading to rapid nuclear accumulation and binding to ISREs. STAT1 also shuttles between the nucleus and cytoplasm in unstimulated cells. In fact, there is evidence that STAT1 forms antiparallel dimers in unstimulated cells that can enter the nucleus but they have low affinity for GAS elements and therefore do not play a physiologically relevant role in transcriptional activation. Upon IFN stimulation STAT1 homodimers are formed in the parallel orientation that rapidly enter the nucleus and bind GAS elements with high affinity (267, 286). Therefore it can be said that the STATs are in constant movement and IFN stimulation causes a shift in the equilibrium of STATs toward the nucleus leading to enhanced nuclear accumulation and binding to ISRE and GAS elements in the promoters of IFN stimulated genes.
Figure 1.2 The Type I IFN Signaling Pathway
Figure 1.3 The Type II IFN Signaling Pathway
Interferons can stimulate the expression of over 300 cellular genes (59). The genes whose expression is induced by IFNs are known as ISGs (Interferon Stimulated Genes). The proteins encoded by ISGs have antiviral and immunomodulatory activity. A list of IFN induced genes with a direct antiviral activity is provided in Table 1.3, along with their mode of action. The affect of these gene products may be interference with the viral life cycle by degrading viral nucleic acids, inhibiting viral protein synthesis, interference with viral trafficking or inducing apoptosis leading to death of virally infected cells. 2’, 5’ Oligoadenylate synthetase (OAS) is an ISG that is constitutively expressed at low levels. Upon stimulation with Type I IFN there is elevated synthesis of OAS. The dsRNA of invading viruses activates tetramerization of OAS monomers. This active form of OAS then catalyzes formation of 2’-5’ phosphodiester bonds between ATP (adenosine triphosphate) molecules resulting in the formation of oligoadenylates (196). Oligoadenylates then bind inactive Ribonuclease L (RNaseL) monomers and induce the formation of active dimers. Activated RNaseL dimers can cleave the RNA of the invading virus (50). Similar to OAS, there is a low level basal expression of Protein Kinase R (PKR) in unstimulated cells. Type I IFN binds to the ISRE in the promoter of the PKR gene and upregulates its transcription. However the protein remains as an inactive monomer until activated by the RNA of invading viruses (78, 166). This results in the formation of the active dimer, that phosphorylates Eukaryotic translation Initiation Factor 2α (EIF2α) leading to a block in protein synthesis (60). Unlike
OAS and PKR, Myxovirus resistance (Mx) proteins are not constitutively expressed, but are rapidly induced by Type I IFN through the ISRE in the promoters of their genes. They were first identified as proteins that conferred resistance to influenza A virus in mice (105). The Mx proteins associate with intracellular membranes and bind viral nucleocapsids and interfere with intracellular trafficking (1). Mx proteins also bind influenza virus polymerase and inhibit their function (251). In addition to their direct antiviral role, IFNs can control viral pathogenesis by enhancing adaptive immune responses, enhancing antigen presentation to cells of the immune system and recruiting immune cells to the sites of infection. In particular, type II IFN is important in stimulating the cytotoxic activity of macrophages and T-cells. Type II IFN also activates innate immunity by stimulating NK cells. Table 1.4 summarizes the role of IFNs in adaptive immunity. It can be concluded from the data in Table 1.3 and 1.4 that IFNs are potent immunomodulatory and antiviral agents.
<table>
<thead>
<tr>
<th>Interferon Type</th>
<th>Target</th>
<th>Mode of action</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>I and II</td>
<td>OAS, RNAseL</td>
<td>Degrades viral and cellular RNA</td>
<td>(61, 69, 214)</td>
</tr>
<tr>
<td>I</td>
<td>PKR</td>
<td>Phosphorylates eIF2α and inhibits protein synthesis</td>
<td>(72, 155, 273)</td>
</tr>
<tr>
<td>I</td>
<td>Mx1</td>
<td>Wraps around viral nucleocapsids and interferes with intracellular trafficking</td>
<td>(94, 105, 142, 231)</td>
</tr>
<tr>
<td>I</td>
<td>p56</td>
<td>Binds to EIF3 and blocks initiation of protein synthesis</td>
<td>(207, 262)</td>
</tr>
<tr>
<td>II</td>
<td>GBP1</td>
<td>GTPase with antiviral activity</td>
<td>(6)</td>
</tr>
<tr>
<td>I and II</td>
<td>PML</td>
<td>Intrinsic antiviral activity by unknown mechanism</td>
<td>(199)</td>
</tr>
<tr>
<td>I and II</td>
<td>ISG15</td>
<td>Protein modification by ISGylation</td>
<td>(136, 176)</td>
</tr>
<tr>
<td>I and II</td>
<td>Phospholipid Scramblase</td>
<td>Phospholipid migration and DNA binding enhancing expression of ISGs</td>
<td>(13, 62)</td>
</tr>
<tr>
<td>I and II</td>
<td>ISG12</td>
<td>Antiviral action by unknown mechanism</td>
<td>(126, 148)</td>
</tr>
<tr>
<td>I and II</td>
<td>ISG20</td>
<td>3’ exonuclease of RNA and DNA, degrades viral nucleic acids</td>
<td>(65)</td>
</tr>
<tr>
<td>I and II</td>
<td>ADAR1</td>
<td>Adenosine deaminase of dsRNA causes viral RNA editing and altered translation</td>
<td>(145)</td>
</tr>
<tr>
<td>I and II</td>
<td>Cig5( viperin)</td>
<td>Antiviral activity against HCMV by unknown mechanism</td>
<td>(49)</td>
</tr>
<tr>
<td>II</td>
<td>iNOS</td>
<td>Involved in the antimicrobial activity of macrophages</td>
<td>(153)</td>
</tr>
<tr>
<td>II</td>
<td>Nup98, Nup96</td>
<td>RNA and protein transporters associated with the nuclear pore</td>
<td>(64, 67)</td>
</tr>
</tbody>
</table>

Table 1.3: Role of Interferons in intrinsic antiviral activity
<table>
<thead>
<tr>
<th>Interferon Type</th>
<th>Target</th>
<th>Mode of action</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>I and II</td>
<td>MHC class I</td>
<td>Enhanced presentation of viral antigens to immune cells (CD8+ Tcells)</td>
<td>(20)</td>
</tr>
<tr>
<td>II</td>
<td>MHC class II</td>
<td>Enhanced presentation of viral antigens to immune cells (CD4+ Tcells)</td>
<td>(52)</td>
</tr>
<tr>
<td>II</td>
<td>CIITA</td>
<td>Transcriptional activation of MHC class II</td>
<td>(38, 234)</td>
</tr>
<tr>
<td>I and II</td>
<td>IRF1</td>
<td>Transcriptional activation of MHC class II and other ISGs</td>
<td>(240)</td>
</tr>
<tr>
<td>II</td>
<td>LMP</td>
<td>Component of the immunoproteasome responsible for generating pathogen derived peptides</td>
<td>(71)</td>
</tr>
<tr>
<td>II</td>
<td>TAP</td>
<td>Transporter pathogen derived antigens into ER for loading on to MHC class I</td>
<td>(279)</td>
</tr>
<tr>
<td>II</td>
<td>ICAM1</td>
<td>Promotes migration of leukocytes to site of infection</td>
<td>(106)</td>
</tr>
</tbody>
</table>

**Table1.4: Role of Interferons in adaptive immunity**
Viral Evasion of Interferon Action

Coexistence of the viral pathogen with its host requires the maintenance of a delicate equilibrium between host antiviral responses and viral countermeasures. In the absence of antiviral responses the virus could destroy the host, which is actually detrimental to the virus as it is dependent on the host for its own survival. On the other hand the virus has to develop countermeasures against antiviral agents including interferons in order to be able to replicate in the host and produce progeny virus. Most viruses have developed mechanisms to block IFN synthesis, interfere with IFN mediated signaling and interfere with the function of ISGs. For example, Hepatitis B virus (HBV) blocks IFNβ synthesis via its ORF-C gene product (269). Influenza A virus NS1 protein inhibits IFN induction (239). One of the herpes viruses Human Herpes Virus -8 (HHV-8) codes for a viral IRF1 protein that inhibits IFN induction by interfering with cellular IRF1 (292). Many viruses target the JAK-STAT signaling pathway involved in IFN mediated signaling to attenuate the antiviral activity of IFNs. The adenovirus E1A protein reduces STAT1 and IRF9 levels in the cell and inhibits ISGF3 formation (137, 138). Proteins encoded by many DNA viruses including Human Papilloma Virus (HPV) E7 protein, HBV terminal protein, polyoma virus large T protein interfere with the IFN induced signaling pathway. Herpes viruses also encode proteins that adversely affect IFN signaling, including EBV EBNA1 and EBNA2 and proteins of HHV-8 and HCMV (18, 83). The EBV immediate early protein BZLF1 has been shown to interfere with STAT1 phosphorylation and nuclear translocation.
Similarly, the Herpes Simplex Virus (HSV) ICP27 protein inhibits STAT1 phosphorylation and nuclear translocation (114). The HSV virion host shutoff protein encoded by the \textit{UL41} gene is implicated to play a role in the decreasing the levels of the IFN signaling mediators JAK1 and STAT2 in infected cells (42). The US3 kinase protein of HSV brings about posttranslational modification of the IFNGRα which adversely affects the induction of IFNγ induced genes (140).

In the case of HCMV infected cells it has been reported that there is a defect in the IFNγ stimulated induction of MHC II. Furthermore there is a blockade in the IFNγ mediated activation of the transcription factors required for MHC II induction namely Class II Transcriptional Activator (CIITA) and IRF1. These studies revealed that this blockade was brought about by the HCMV mediated degradation of JAK1 during early to late times of infection (156). Furthermore HCMV is also able to interfere with the induction of various ISGs by Type I IFN. The inhibition of ISG induction in response to IFN α was found to be due the degradation of IRF9 protein at late times of infection and JAK1 degradation starting earlier (158). In murine cytomegalovirus (MCMV) infected cells, STAT2 is targeted for proteosomal degradation by the viral pM27 protein. Surprisingly this causes a defect in both Type I and Type II IFN signaling revealing a novel role for STAT2 in the Type II IFN signaling pathway (290). Studies have also detected reduced IFNγ stimulated CIITA induction starting at immediate early times of infection with HCMV. This defect is not associated with decreased levels of JAK1 or any other signaling molecule in IFNγ pathway. It is suggested
that this defect in IFNγ signaling occurs downstream of nuclear translocation of STAT1 homodimers (132). It has also been reported that there is reduced binding of IFNγ- induced STAT1 homodimers to target GAS elements starting at immediate early times of infection (157). In MCMV infected macrophages there is a defect in IFNγ induced activation of ISGs. This blockade is found to occur at the level of transcriptional induction of the ISG and is suggested to adversely affect IFN- induced promoter assembly and chromatin remodeling (187). Recent reports have implicated a role for the HCMV immediate early transactivator protein IE1, in attenuating Type I IFN signaling by sequestering the ISGF3 complex away from the ISRE elements in the promoters of ISGs (181). These reports indicate that HCMV targets multiple levels of the Type I and Type II IFN signaling pathway to attenuate this important antiviral defense mechanism of the host cell. However, the viral gene products that are responsible for most of these phenomena are not known. Identification of these gene products and elucidation of the precise molecular mechanisms they employ to modulate IFN signaling is expected to improve our understanding of the biology of the virus. Furthermore these studies may suggest novel targets for antiviral drug development.
Prelude to thesis

Defects in type I and type II IFN induced signaling have previously been reported in HCMV infected cells (158). This has been suggested to be an outcome of the degradation of JAK1 that starts at early times of infection and continues till late times of infection (156). However there is also a defect in STAT1 homodimer binding to target promoter elements and induction of CIITA starting at IE times of infection (132, 157). This indicates that there may be more than one mechanism employed by HCMV to block IFN mediated signaling. We set out to identify HCMV genes involved in disruption of IFN signaling. Our aim was to construct a cDNA library of the HCMV laboratory strain AD169 and use it to screen for genes involved in interfering with IFN induced signaling. After constructing and characterizing this library we discovered that about 50% of our library was composed of non-coding genes and up to 56% of the library represented genes in antisense orientation to known or predicted HCMV genes. These novel findings are detailed in chapter 2. Upon screening the library for genes involved in disruption of IFN signaling the first gene we identified was UL123 that encodes the viral immediate early transactivator IE1. IE1 has been implicated to play a role in disruption of Type I IFN signaling, but our results show that IE1 also can affect Type II IFN signaling. We examined the role of IE1 in attenuating Type II IFN- induced signaling and these studies, which are described in chapter 3, indicate that HCMV IE1 employs a novel mechanism to affect IFN\(\gamma\) mediated activation of target genes.
CHAPTER 2

Reproduced with permission from American Society for Microbiology

ANTISENSE TRANSCRIPTION IN THE HUMAN CYTOMEGALOVIRUS TRANSCRIPTOME

Introduction

HCMV is classified within the betaherpesvirinae subfamily and most humans are infected with HCMV during their lifetime. Like other herpesviruses, HCMV cannot be completely eliminated by the immune system, and remains either as a low level persistent infection or in a quiescent latent state for the lifetime of the infected person. Although infections in immunocompetent adults are usually benign, considerable morbidity is seen in congenitally-infected infants and in persons with compromised immunity (reviewed in (180). HCMV infections are especially important in transplant patients, who may suffer graft loss, vascular disease, and other manifestations including hepatitis and pneumonitis (reviewed in (90, 180). An increasing number of studies have also raised the possibility that HCMV infections may be linked to some chronic diseases, including atherosclerosis, autoimmune diseases, and cancer (reviewed in (222).

Unfortunately, no effective vaccine is currently available to prevent or ameliorate
HCMV infections.

Progress in developing effective antiviral drugs and vaccine candidates will likely rely upon detailed knowledge of viral gene products and how they function in pathogenesis processes. For this reason, there is an intense interest in defining the gene products of HCMV as well as other herpes viruses. The HCMV genome of 230kb is among the largest of the herpes viruses. The genome is comprised of two unique regions known as the unique long (UL) and unique short (US) regions that are bounded by terminal (TRL or TRS) and internal (IRL and IRS) repeat regions. Although the entire sequence of the laboratory adapted AD169 strain of HCMV was first available in 1989 (43), the precise number and nature of viral genes and gene products is still in question. After sequencing the HCMV genome, Chee and colleagues predicted approximately 200 open reading frames (ORFs) capable of coding for proteins (10, 43). It was subsequently discovered that the AD169 laboratory strain harbored a deletion of approximately 15kb relative to clinical isolates. This region was predicted to encode for 19 additional ORFs, suggesting that the HCMV genome encoded up to 220 genes (36). More recently, comparison of the HCMV genome with the chimpanzee cytomegalovirus genome led to a revised estimate for the protein coding genes of AD169 to 145 (57). Likewise, application of an in-silico approach based on the Bio-Dictionary gene-finder algorithm to define the coding potential of HCMV supported elimination of 37 previously annotated ORFs (164). However, comparison of the AD169 genomic sequence to that of clinical isolates, as well as use of proteomic experimental
approaches, predict that additional unannotated ORFs exist (165, 257). Most studies of herpes genomes have focused on protein coding potentials of these viruses. However, rapid advances in understanding the role of noncoding gene products and antisense transcripts are dramatically changing the paradigms applied to gene definitions, gene regulation, and gene functions (reviewed in (130, 150, 163, 281). In particular, the application of bioinformatics approaches to analyze expressed sequences databases have revealed that antisense transcription is widespread in human and other genomes (reviewed in (130, 163).

Sense-antisense (S-AS) transcript pairs have been described in genomes from archaebacteria to humans (154, 237, 283), and a subset of S-AS pairs are recognized to be conserved across species (283). While some pioneering studies have estimated that between 1 and 15% of human or mouse genes were influenced by S-AS pairs (66, 120, 134, 213, 278), more recent estimates of up to 20-26% of human genes (46, 283) and 72% of mouse genes (118) suggest that AS-mediated gene regulation may be much more common than previously appreciated. Natural antisense transcripts (NATs) are classified as cis or trans in nature. Cis-NATs are transcribed from opposite strands of the same genomic locus and are predicted to have longer and more perfect complementary sequences to sense transcripts compared to trans-NATs derived from separate loci. Regulatory functions of NATs are predicted to derive from double-stranded RNA-dependent and –independent mechanisms including, RNA editing, RNA interference (RNAi), chromatin remodeling, transcriptional interference, and
masking of RNA elements involved in splicing, localization, transport, and translation of RNAs (reviewed in (130, 163).

In this study, we examined the transcriptional products of HCMV during lytic infection of fibroblasts. Remarkably, of the 604 HCMV cDNA clones analyzed in this study, at least 45% were derived from genomic regions predicted to be noncoding. Of similar interest was our finding that 55% of the cDNA clones in this study were completely or partially antisense to known or predicted HCMV genes. Moreover, cis-NAT pairs were identified or predicted for 56 of the 191 genes currently annotated at the Los Alamos National Laboratory Sexually Transmitted Diseases Sequences Database (STD database, now at the Oral Pathogens Sequences Database). We conclude that genomic maps based on open reading frame analyses and other in-silico analyses may drastically underestimate the true complexity of viral gene products. In addition, the abundance of antisense transcription in the HCMV viral genome raises the distinct possibility that antisense-dependent gene regulatory mechanisms influence both viral gene expression and gene organization. These noncoding and antisense transcripts may offer new insights into HCMV pathogenesis and may serve as novel targets for developing intervention strategies and treatments for HCMV-related diseases.
Materials and Methods

Cells and virus. Human foreskin fibroblasts (HFFs), telomerase immortalized HFF cells [HFF-TEL, a kind gift of Tom Shenk (27)], MRC-5 human fibroblasts, and human umbilical vein endothelial cells (HUVECs) were used in this study. HFF and HFF-TEL cells were maintained in Dulbecco’s modified Eagle’s medium (DMEM) supplemented with 10% fetal bovine serum (FBS), while MRC-5 cells were maintained in modified Eagle’s medium supplemented with 10% fetal calf serum, 1.7mM sodium bicarbonate, 1.4mM sodium chloride, essential and nonessential amino acids, vitamins, and sodium pyruvate at manufacturer-recommended concentrations (Sigma-Aldrich, Invitrogen). All three cell lines were used between passages 2-10. HUVECs were isolated from vessels as previously described (209) and propagated in endothelial cell growth medium (ECGM) consisting of M-199 (GibcoBRL) supplemented with 20% FBS (Hyclone), 22.5 μg/ml bovine brain extract (BioWhitaker, Inc., Walkersville MD), 12 U/ml sodium heparin (Sigma, St. Louis MO), and 20 mM HEPES buffer. All growth surfaces for HUVECs were pretreated with human fibronectin (25 μg/ml, Upstate Biotechnology). Cells were passed weekly by brief trypsin digestion at a ratio of 1:4, and used in experiments at passages 5-7.

HCMV strain AD169 was obtained from ATCC, and propagated and titrated on MRC-5 cells by plaque assay (268). CMV strain VHL/E, originally isolated from duodenal biopsy material from a bone marrow transplant recipient (261), was propagated in HUVEC as detailed elsewhere (259) to preserve it’s natural
endothelial cytopathogenicity.

**Extraction of HCMV genomic DNA.** HCMV genomic DNA was extracted as described previously (244). Briefly, confluent HFF-TEL cells in two 175 cm² flasks were exposed to 1 PFU per cell of AD169 or Towne strains. The cells were harvested at 72 hr post-infection and collected by centrifugation. The cell pellets were re-suspended in 5 mL of 150 mM NaCl, 10 mM Tris (pH 7.4), and 1.5 mM MgCl₂. After incubation on ice NP-40 was added to a final concentration of 0.1%. The lysate was centrifuged at 3,700 rpm for 20 min using a Beckman GS-6R centrifuge. The supernatant was collected and brought to a final concentration of 0.2% SDS, 0.5 mM EDTA, and 50 mM β-mercaptoethanol. After incubation on ice and extraction with phenol/chloroform, the genomic DNA was precipitated with ethanol, re-suspended in 1 ml of TE buffer, and treated with RNAseH (Sigma-Aldrich). The genomic DNA was further purified by centrifugation in a linear 5-20% W/V potassium acetate gradient at 40,000 rpm for 3.5 hr at 20°C in a Beckman L7 Ultracentrifuge SW60 rotor. Following centrifugation, the DNA was collected, precipitated with ethanol, and re-suspended in 50 µl distilled water. The purified genomic DNA was digested with Mse I, followed by phenol/chloroform extraction and ethanol precipitation. The digested genomic DNA was finally re-suspended in 50 µl sterile water.

**Construction of HCMV AD169 cDNA libraries.** RNA was extracted from infected HFF cells cultured in 175cm² flasks under conditions that selected for immediate early (IE), early (E) and late (L) viral transcripts. For all conditions,
cells were exposed to 20 PFU per cell of the AD169 strain of HCMV. To select for IE transcripts, cells were treated with 100 µg/mL of cyclohexamide (Sigma-Aldrich) for 1 hr prior to infection and throughout the 24 hr infection period, when cells were harvested for RNA isolation. To select for E transcripts, 100 µM ganciclovir (Roche Pharma) was added to the media after the infection period, and cells were harvested 72 hr later. To select for late viral transcripts, untreated infected cells were harvested 72 hr after infection. Prior to isolation of total RNA, cells from a small section of the flasks were scraped and collected. Cells were disrupted in SDS-PAGE sample buffer (25 mM Tris-Cl, pH 6.8, 2.5% β-mercaptoethanol, 5% glycerol and 0.5% SDS), boiled, and sonicated. Proteins were separated by denaturing gel electrophoresis and transferred to nitrocellulose (Amersham Bioscience). Efficacy of drug treatments was verified by immunoblot analyses for immediate early UL122/123 products, IE1/2 (Rumbaugh Goodwin #1203), the early UL55 product, glycoprotein B (gB) (Rumbaugh Goodwin #1201) and the late UL99 product, pp28 (Rumbaugh Goodwin #1207) (Data not shown).

Unless otherwise stated, all extractions of total cellular RNA were performed using the TRIZOL Reagent (Invitrogen) following the instructions of the manufacturer. Polyadenylated mRNA was isolated using a Oligotex kit (Qiagen) according to the manufacturer’s instruction. cDNA libraries were constructed using two cloning vectors, pAcCMV (287), derived from the pAcSG2 Baculovirus transfer vector (BD Biosciences), and pcDNA3.1(+) (Invitrogen).
These vectors were modified by introducing recognition sites for two restriction enzymes, PacI and PmeI, that are absent in the AD169 genome. Specifically, sequences recognized by the PacI and PmeI enzymes were inserted between the StuI and KpnI sites of vector pAcCMV using the following oligonucleotides: 5’ pCCTGTTTAAACCTAGGCGGCCGCTTAATTAAGGTAC and 5’ pCTTAATTAAGCGGCCGCTTAATTAAACAGG. The PmeI site at position 1007 in pcDNA3.1 (+) was replaced with sequences specifying a PacI cutting site using a site-directed mutagenesis kit (Stratagene) and the following oligonucleotides: 5’ CTAGAGGGCCCGTTTAATTAAGCTGATCA GCCCTGACTG and 5’ CAGTCGAGGCTGATCAGCTTAATTAAACGG GCCCTCTTAG.

cDNA libraries were constructed by following the instruction manual of SuperScript Plasmid System with Gateway Technology for cDNA Synthesis and Cloning (Invitrogen) with some minor modifications. Briefly, a poly (T)-tailed PacI primer-adapter was used for first strand cDNA synthesis (5’CGGGCCGCTTAATTAACC(T)\textsuperscript{15}). After 2\textsuperscript{nd} strand synthesis, an EcoRI-PmeI adapter was added to the 5’ end of the cDNA that was generated from the following oligonucleotides: 5’ AATTCAGGCGCTGTGTTAAACCG and pCGTTTAACAGGCCTG. cDNA fragments were used in ligation reactions with modified pAcCMV or pcDNA3.1(+) vectors previously digested with EcoRI and PacI restriction enzymes. Recombinant plasmids harboring cDNA sequences were transformed into XL1-Blue Supercompetent \textit{E. coli} cells (Stratagene).
cDNA library screening by colony hybridization. Random transformed bacterial colonies were picked individually and transferred onto agarose grid plates (approximately 100 clones on each plate). Colonies on grid plates were transferred to Hybond-N+ nylon membranes (82 mm in diameter, Amersham Bioscience) and processed for hybridization as described by Hirsch(101). Mse I-digested HCMV genomic DNA was labeled using a DIG High Prime DNA Labeling Detection Starter Kit II (Roche Applied Science). Probes were incubated with the membranes according to the manufacturer’s instruction.

DNA sequencing and sequence analysis. Bacterial colonies harboring HCMV-derived cDNA sequences were inoculated into 4mL LB broth supplemented with 50 µg/ml ampicillin. Plasmid DNA was purified from overnight culture using QIAprep Spin Miniprep Kits (Qiagen). cDNA inserts were sequenced from the 5’ end using T7 primer for pcDNA3.1(+) and a pAcCMV-specific primer (5’GGAGACGCCATCCACGCTGTTTTGACC), respectively, at the OSU Plant-Microbe Genomics Facility. In total 870 clones were submitted for sequence analysis from the 5’ ends. Sequences were compared to the AD169 genome [NC_001347] using mega BLAST (285). Matched AD169 gene sequences were downloaded and aligned to corresponding cDNA clone sequences manually using BioEdit software (http://www.mbio.ncsu.edu/BioEdit/bioedit.html) (93). Selected clones were subjected to a second round of sequence analysis using primers specific for the 3’ ends of the cDNA inserts for the pcDNA3.1(+) (5’GCACCTTCCAGGGTCAAGGAAG) and pAcCMV vectors
(5′GAGGTGCGTCTGGTGCAAAC). These primers failed to generate sequence from a subset of clones, presumably because of difficulties in reading through poly (A) tracts. Therefore, in some cases, 3′ ends were sequenced using standard poly (T) primers. Selected cDNA inserts were also sequenced using specifically designed internal primers. AD169 genomic positions [NC_001347] corresponding to the sequences of the cDNA clones were determined using the SPIDEY program for cDNA to genomic alignments (http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/).

**RT-PCR for detection of antisense transcripts.** Confluent MRC-5 cells in 6-well plastic tissue culture plates were exposed to 2 PFU per cell of the AD169 strain of HCMV. Cells were harvested at 24, 48, and 72 hr after infection. Also, confluent HUVEC monolayers in 6-well plastic tissue culture plates were inoculated with sonicated cell lysates containing the VHL/E of HCMV (1 PFU/cell). To enhance infection efficiency of endothelial cells, plates were centrifuged at 300 × g for 30 minutes at room temperature. Cells were then incubated an additional 30 minutes before removal of inoculum, followed by 2 washes with PBS, and addition of fresh medium. Cells were harvested at 24, 48, 72, 96, and 120 h after infection. Total RNA was isolated and treated with DNase I (Invitrogen). The RT-PCR analysis was performed using a ThermoScript kit (Invitrogen) following manufacturer’s instructions. cDNA synthesis was performed in the first step using total RNA and gene-specific primers at 55°C for 35 min (Table 1, reverse primers). In the second step, PCR was performed using primers specific for the
gene of interest. A complete list of PCR primers is listed in Table 1. Reactions were carried out at 94°C for 2min; 30 cycles of 94°C for 30 sec, 59°C for 45sec, 72°C 50 sec; and final extension at 72°C for 10min. RT-PCR products were analyzed by agarose gel (1% w/v) electrophoresis. No reverse transcriptase and no template controls were run in parallel.

**Northern blots.** Confluent MRC-5 cells in 6-well plastic tissue culture plates were exposed to 2 PFU per cell of the AD169 strain of HCMV. Cells were harvested at 24, 48, and 72 hr after infection. Total RNA was isolated and subjected to denaturing agarose gel (1% w/v) electrophoresis in the presence of formaldehyde. Nucleic acids were transferred to positively charged nylon membranes (Millipore) using a Posi-Blot 30-30 Pressure Blotter (Strategene). All probes were labeled using a DIG Northern Starter kit (Roche Applied Science) according to the manufacturer’s instructions.

Plasmids pE10422 and pE104114 harboring cDNAs with gene sequences in sense and antisense orientation were used to generate probes complementary to antisense and sense transcripts, respectively, of UL36 (see Supplemental Table 1 for description of cDNA clones). Plasmids pE10422 and pE104114 were linearized with BssHII and NdeI, respectively. RNA probes were generated using T7 polymerase. The AS probe corresponds to nucleotides 49791-49577 and 49473-48961, while the S probe to nucleotides 49135-50065, respectively, of the AD169 genome. Plasmids pL5312 and pL8212 harboring cDNAs with gene sequences in sense and antisense
orientation were used to generate probes complementary to antisense and sense transcripts, respectively, of UL24. Plasmids pL5312 and pL8212 were linearized with SalI and NcoI, respectively. RNA probes were generated using T7 polymerase. The AS probe corresponds to nucleotides 29806-29136, while the S probe for corresponds to nucleotides 28949-29362, respectively, of the AD169 genome.

Plasmid pL222 carrying AS-UL102 sequence was digested with EcoRI and XbaI, (corresponding to nucleotides 150184-149378 of the AD169) and this fragment was inserted into pBluescript II KS+ (Stratagene). The plasmid was linearized with EcoRI to generate a probe complementary to S transcripts of UL102 using T7 promoter or linearized with XbaI to generate a probe complementary to the AS transcripts using T3 promoter.

Plasmid pE1033 carrying AS-RL5 and S-RL4 sequences was digested with BamHI and SalI (corresponding to nucleotides 4555-3958 or 185843-186440 of AD169) and inserted into pBluescript II KS+. This plasmid was linearized with BamHI to generate a probe complementary to AS-RL4 transcripts using the T3 promoter. Because S-specific probe generated from this entire fragment exhibited nonspecific binding, this plasmid was linearized with AvaII to generate a probe complementary to S transcripts of β2.7 using T7 promoter (corresponding to nucleotides 4555-4399 or 185843-185999 of AD169).

Plasmid pE10335 carrying AS-UL61 and AS-UL62 sequences was digested with EcoRI and XhoI (corresponding to nucleotides 94467-94735 of the AD169) and
this fragment was inserted into pBluescript II KS+. This plasmid was linearized with EcoRI to generate a probe complementary to AS-UL61 and -UL62 transcripts using T3 promoter. Because S-specific probe generated from the T7 promoter exhibited nonspecific binding, a second sense-specific probe was generated by linearizing plasmid pE10335 with MluI (corresponding to nucleotides 94467-95174 of AD169) and using the T7 promoter.

RACE. Race was performed to determine 5’ and 3’ ends of the antisense clones of UL36 (early), and the 5’ ends of antisense clones for UL24 (early) and UL102 (late). Total RNA was isolated from MRC-5 cells in 6-well plastic tissue culture plates exposed to 2 PFU per cell of the AD169 strain of HCMV at 48 (for UL36 RACE) and 72 h (for UL24 and UL102 RACE) after infection. RNA was treated with DNase I (Roche Applied Science). 5’ or 3’ cDNA ends were amplified using the 5’/3’ RACE kit (Roche Applied Science) following the manufacturer’s instructions. The products of the RACE reactions were inserted into a TOPO TA vector (Invitrogen) and sequenced at the OSU Plant-Microbe Genomics Facility. Primers used for RACE experiments are listed in Table 2.1.
<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5’-3’)</th>
<th>Genomic position* or (reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AS-UL36-37-F</td>
<td>GGTGCGCCGATTCTGCTCCAGA</td>
<td>49677-49698</td>
</tr>
<tr>
<td>AS-UL36-37-R</td>
<td>CCGTCTGCTTGTGACGAATGC</td>
<td>50566-50543</td>
</tr>
<tr>
<td>AS-US30-31-F</td>
<td>GCTCCAAGAGCGACATGAGATCG</td>
<td>223730-223708</td>
</tr>
<tr>
<td>AS-US30-31-R</td>
<td>GGGAGCGCAATCAGTACCGTATTG</td>
<td>223019-223042</td>
</tr>
<tr>
<td>AS-UL61-62-F</td>
<td>CGAGGTGGGTGGTGAGCCGAA</td>
<td>94987-95007</td>
</tr>
<tr>
<td>AS-UL61-62-R</td>
<td>CGCTCTAGGCTCTCGACGCAAT</td>
<td>95612-95590</td>
</tr>
<tr>
<td>AS-UL67-F</td>
<td>CGTACACGGCTATAGCGCAGA</td>
<td>97937-97958</td>
</tr>
<tr>
<td>AS-UL67-R</td>
<td>GGTGCGTGAGGCGATGATGATG</td>
<td>98720-98698</td>
</tr>
<tr>
<td>AS-IGS-RL2-3-F</td>
<td>CCGACACTTCTCTGCTCTCA</td>
<td>2939-2919-or 187459-187479</td>
</tr>
<tr>
<td>AS-IGS-RL2-3-R</td>
<td>GGAGTGGAAGGGGTGAAGCAA</td>
<td>2140-2161 or 188258-188237</td>
</tr>
<tr>
<td>AS-TRL8-9-F</td>
<td>CCGTTTCTCCTCAGCTGCGT</td>
<td>7791-7771-or 182607-182627</td>
</tr>
<tr>
<td>AS-TRL8-9-R</td>
<td>CCACAGAGGGAGACGGGACGAAG</td>
<td>7159-7181 or 183239-183217</td>
</tr>
<tr>
<td>AS-UL27-28-F</td>
<td>GCTGCTGGGTTGCGGATGGA</td>
<td>34614-34633</td>
</tr>
<tr>
<td>AS-UL27-28-R</td>
<td>GATGAGATCGTAGAGGGCGCG</td>
<td>35289-35267</td>
</tr>
<tr>
<td>AS-UL70-F</td>
<td>GCTGACGTTGGTTCGTTAGGTG</td>
<td>101764-101785</td>
</tr>
<tr>
<td>AS-UL70-R</td>
<td>CGAGCGGCACACAGAGGAGA</td>
<td>102523-102506</td>
</tr>
<tr>
<td>AS-UL87-88-F</td>
<td>GCTCGCCGACCCGTACCGTCT</td>
<td>132216-132196</td>
</tr>
<tr>
<td>AS-UL87-88-R</td>
<td>CAGGTGCAAGAAGGCGGCTGCA</td>
<td>131608-131630</td>
</tr>
<tr>
<td>AS-UL102-F</td>
<td>CGTGCACCTGTTAGCGTTGAGG</td>
<td>150083-150061</td>
</tr>
<tr>
<td>AS-UL102-R</td>
<td>GGAATGGTCATAACGGCGGTAG</td>
<td>149395-149418</td>
</tr>
<tr>
<td>AS-TRL3-F</td>
<td>CCAAGCCAGATTCTACGTCATGAG</td>
<td>3567-3543 or 186831-186855</td>
</tr>
<tr>
<td>AS-TRL3-R</td>
<td>GCTTTGAGAGCGAAGAAGTGCGA</td>
<td>2915-2938 or 187483-187460</td>
</tr>
<tr>
<td>AS-UL23-F</td>
<td>CGTTTTCTGGGGAGCGTGCTCA</td>
<td>27954-27974</td>
</tr>
<tr>
<td>AS-UL23-R</td>
<td>GAAGGACGGGAGTGACACGCA</td>
<td>28566-28546</td>
</tr>
<tr>
<td>AS-UL24-F</td>
<td>CGTTTGCGGGCGCTGAAAAGTC</td>
<td>28985-29005</td>
</tr>
<tr>
<td>AS-UL24-R</td>
<td>CGTTTTGCGGGCGGACAGCAA</td>
<td>29601-29581</td>
</tr>
<tr>
<td>AS-UL25-F</td>
<td>CGAACGCGCTCTGGTTGTTGACC</td>
<td>30979-30958</td>
</tr>
<tr>
<td>AS-UL25-R</td>
<td>GGGGCGGAGATAAACAGACAGAG</td>
<td>30467-30489</td>
</tr>
<tr>
<td>AS-UL47-F</td>
<td>CGTACGCGCATGACGCGTCAGCAG</td>
<td>61984-61965</td>
</tr>
<tr>
<td>AS-UL47-R</td>
<td>CGGAGCGCGTTAAATACCGTGAC</td>
<td>61388-61410</td>
</tr>
<tr>
<td>AS-UL88-F</td>
<td>GCTTTCGCCGCAACATTGCCCAAG</td>
<td>133032-133011</td>
</tr>
<tr>
<td>AS-UL88-R</td>
<td>GGTGACGCTGGCAAGCTGCTTGG</td>
<td>132425-132446</td>
</tr>
<tr>
<td>gB E1</td>
<td>TCCAAACCCACACTACCAGCTT</td>
<td>(45)</td>
</tr>
<tr>
<td>gB E2</td>
<td>CGGAAACGAATGGGTCAGTTTG</td>
<td>(45)</td>
</tr>
</tbody>
</table>

*nucleotide positions correspond to the AD169 genome NC_001347

Table 2.1 Primers for RT-PCR and 5’ and 3’ RACE

continued
Table 2.1 (continued)

<table>
<thead>
<tr>
<th>Primer Set</th>
<th>Primer Sequence</th>
<th>Length</th>
</tr>
</thead>
<tbody>
<tr>
<td>GAPDH-F</td>
<td>CCATGGGGAAGGTGAAGGTCGGA GC</td>
<td></td>
</tr>
<tr>
<td>GAPDH-R</td>
<td>GGTTGGTGACGGCATTTGCTGATG (5)</td>
<td></td>
</tr>
<tr>
<td>pL8212-5RACEsp1</td>
<td>GCTCTAGACCTTATCCCCATCGTC GTG</td>
<td>29138-29119</td>
</tr>
<tr>
<td>pL8212-5RACEsp2</td>
<td>GATCTAGACGGAGCCATACCCGG GTG</td>
<td>29081-29061</td>
</tr>
<tr>
<td>pL5312-5RACEsp1</td>
<td>GCTCTAGATGAGGCACACACAGCG GTG</td>
<td>29555-29577</td>
</tr>
<tr>
<td>pL5312-5RACEsp2</td>
<td>CGTTCACGAAGCGACCCACGAAG (5)</td>
<td>29606-29628</td>
</tr>
<tr>
<td>PL222-5RACEsp1</td>
<td>GATCTAGACTTCGTCTTTCGAGTGG CAC G</td>
<td>149904-149924</td>
</tr>
<tr>
<td>PL222-5RACEsp2</td>
<td>CGTCTAGAACGCTTACGGTCACG (5)</td>
<td>150066-150083</td>
</tr>
<tr>
<td>pE104114-5RACEsp1</td>
<td>GATCTAGAGGTCCGGGCTTTATG CCG</td>
<td>49426-49407</td>
</tr>
<tr>
<td>pE104114-5RACEsp2</td>
<td>GCTCTAGAAGGACTTTCTGCGGAA CCG</td>
<td>49316-49297</td>
</tr>
<tr>
<td>pE104114-3RACEsp1</td>
<td>CGGAATTCGGTGGGATGAAACCC AC</td>
<td>50904-50924</td>
</tr>
</tbody>
</table>
Results

Construction and Analysis of HCMV cDNA libraries. Three cDNA libraries were generated from primary human fibroblasts infected with HCMV. We chose to use AD169 HCMV for these studies because it is the most extensively characterized with respect to analyses of viral gene products. HCMV, like other hepesviruses, is known to express genes in a temporal and regulated cascade. The most clearly recognized temporal classes are IE, E and L. In order to make predictions regarding temporal relationships of viral transcripts, RNA was isolated under conditions that selected for expression of IE, E, or L genes as described in Materials and Methods. We employed positive selection using viral genomic DNA as a probe to isolate cDNA clones harboring viral gene sequences. A total of 870 bacterial colonies that were positive by this screening method were individually expanded. Plasmid DNA isolated from each culture was subjected to DNA sequence analysis, and usable sequence data was obtained for 618 clones. After excluding 14 clones (2.3%) that were recombinants of viral gene sequences or viral and cellular gene sequences, 604 clones were subjected to analysis (109 from the IE library, 180 from the E library and 315 from the L library). A complete list of the cDNA clones and their positions relative to genomic DNA is listed in Supplemental Table1(http://jvi.asm.org/cgi/content/full/81/20/11267). We assigned these 604 clones to 184 transcript groups based on the similarity of cDNA sequences. Many indicators suggest that these libraries accurately reflect temporal regulation,
splicing, and abundance of viral gene products. cDNA clones isolated includes sequences overlapping 125 of the 191 unique genes currently annotated for the AD169 strain of HCMV in the STD database, including full length sequences for 92 annotated genes. Examples of correct temporal expression and splicing of viral genes in these libraries was supported by clones representing the UL123 and UL4 genes. The major transcriptional activator (IE72) encoded by the UL123 gene is known to be expressed during IE times (235). At least fifteen cDNA clones isolated from the IE library were full length and fully spliced transcripts capable of coding for the IE72 protein. Also, transcripts with short and long 5’ untranslated regions (UTR) for the UL4 gene have been characterized to accumulate at early and late times after infection, respectively (4, 33, 39). We identified 14 transcripts harboring longer 5’ UTR regions, all of which were derived from the L library. We also identified 15 clones that had shorter 5’ UTRs, the majority of which (9) were found in the E library. In addition, studies from other laboratories suggest that the most abundant transcripts found in infected cells overlap the TRL4/IRL4 genes (TRL/IRL region genes are henceforth referred to RL) (88) and this was reflected in our libraries. Indeed, 169 clones in our E and L libraries harbored RL4 gene sequences, 141 of which were classified within the same transcript group (See Supplemental Table 1, group 148). Taken together, these and other indicators suggest that our libraries reasonably reflect the range and abundance of HCMV transcriptional products.

A more detailed comparison of the temporal profiles of transcripts isolated in this
study relative to microarray and northern studies from other laboratories is shown in Supplemental Table 2 (http://jvi.asm.org/cgi/content/full/81/20/11267). This table also lists the clones characterized in this study organized by gene name and provides references for other transcript mapping studies performed for HCMV. The primary conclusion from this analysis is that the tentative temporal class assignments made in this study are largely congruent with temporal class assignments made by northern and microarray analyses. The major difference is that we found many more genes represented in the transcripts of the IE class relative to other methods. Specifically, we identified 45 genes with at least one clone isolated in the IE library. For many of these genes, the majority of clones were isolated at E or L times with only one or a few clones isolated at IE times. There are two non-mutually exclusive explanations to account for these unexpected transcripts found in the IE library: (1) they represent tegument-associated viral transcripts delivered by the virion rather than newly synthesized viral transcripts; and (2) these IE transcripts reflect leaky control of E and L gene expression in the IE period in HCMV-infected fibroblasts. Although latter possibility could be due to incomplete blockade of protein synthesis by cycloheximide treatment, the efficacy of drug treatments used to generate the cDNA libraries was verified by measuring viral protein accumulation (data not shown). It is worthy of note that while the cDNA library approach employed in this study is subject to different biases relative to other methods (see discussion and reference 37), it is free of bias introduced by employing gene-specific probes.
Thus, it is possible that a cDNA library approach offers a specific advantage relative northern and microarray approaches in capturing minor transcripts available at various temporal phases of infection.

Another unexpected feature of our libraries was the prevalence of transcripts overlapping genes predicted to be noncoding genes. While we expected transcripts overlapping the RL4 gene to be abundant (48), we also isolated numerous cDNA clones from other repeat region genes. We found that 230 of the 604 clones analyzed mapped to the RL2-RL9 region. We also found 29 clones derived from the UL61 to UL68 gene region, and 13 clones derived from the UL106 to UL111 gene region, also recently revised as likely to be noncoding regions (57). Altogether, we obtained 274 clones (45% of the total) overlapping annotated genes predicted to be noncoding.

**Antisense Transcription in the HCMV Transcriptome.** One of the most striking features of our transcriptome study was the prevalence of transcripts in antisense orientation to known or predicted genes. For this analysis, we compared our cDNA sequences to the genomic map of the AD169 strain [NC_001347] of HCMV and the STD database. This map includes up to date revisions in annotation proposed by Davison and colleagues, including annotation of those genes revised as noncoding (57).

Of the 604 sequences we analyzed, 257 represented one or more genes strictly in the S orientation (Table 2.2). Remarkably 347 sequences were partially or completely AS to genes annotated on the STD database map, representing 57% of
the cDNA clones isolated in our libraries. Because experimental evidence verifying the existence of gene products derived from a number of viral genes is lacking, we considered the possibility that the only products derived from a subset of these genes are those we identified in our library, and that our calculation for the number of AS transcripts could be overestimated. When we excluded those genes for which we could find no evidence in our libraries or in the literature for a product derived from the S orientation of the gene (orphaned AS transcripts, Table 2.5), we estimated that 271 clones (45%) represented transcriptional products strictly in one orientation with respect to gene sequences (designated the S orientation), and 333 clones (55%) were completely or partially in AS orientation. The AS sequences fell into two categories: a minority (54 clones) overlapped one or more genes strictly in AS orientation whereas 279 clones overlapped more than one gene with sequences both in S and AS orientations. ORF analysis of AS clones predict, that these are predominantly noncoding (Supplemental Table 1). When these clones are included in the calculations for coding and noncoding transcripts, we estimate that up to 49.5% of the clones isolated in this study are noncoding in nature.
<table>
<thead>
<tr>
<th>Classification</th>
<th>Relative to STD Database Gene Map</th>
<th>Revised Based on Experimental Evidence&lt;sup&gt;1&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Number of cDNA clones</td>
<td>Percentage&lt;sup&gt;2&lt;/sup&gt;</td>
</tr>
<tr>
<td>cDNA clones in only in S orientation</td>
<td>257</td>
<td>43</td>
</tr>
<tr>
<td>cDNA clones completely or partially in AS orientation</td>
<td>347</td>
<td>57</td>
</tr>
<tr>
<td>cDNA clones only in AS orientation&lt;sup&gt;3&lt;/sup&gt;</td>
<td>51</td>
<td>8</td>
</tr>
<tr>
<td>cDNA clones with both AS and S orientations&lt;sup&gt;4&lt;/sup&gt;</td>
<td>296</td>
<td>49</td>
</tr>
</tbody>
</table>

<sup>1</sup> Revised analysis based on excluding those AS clones for which we found no experimental evidence for a sense transcript.

<sup>2</sup> Relative to total number of clones in the cDNA library.

<sup>3</sup> Clones spanning one or more genes only in AS orientation

<sup>4</sup> Orientation of gene regions for clones spanning two or more genes

**Table 2.2 Orientation of cDNA Clones**
A complete list of genes for which we identified S-AS pairs is shown in Table 2.3. We obtained direct evidence for the existence of S-AS transcript pairs derived from 38 known or predicted viral genes. We also predict that S-AS pairs exist for 18 additional annotated genes listed in Table 2.4. In these cases, we identified AS but not S transcripts overlapping the indicated gene sequences. Considering that evidence for S products derived from these genes were published by other laboratories, we predict S-AS pairs also exist for these genes. As mentioned earlier, we also identified AS transcripts derived from gene sequences of 9 genes but could find no evidence of a corresponding S product in virus-infected cells or in the literature. We designated this group as orphaned AS transcripts (Table 2.5). Together our analysis indicates that cis-natural S-AS pairs are generated during infection for at least 56 of the 191 known or predicted unique genes annotated at the STD database for the AD169 strain of HCMV.

The abundance of the S and AS transcripts forming pairs fell into two groups. For most genes, we isolated between 1-6 clones representing S-oriented transcripts, and 1-6 clones representing AS-oriented transcripts. There were two dramatic exceptions to this finding: the gene region from UL61 to UL67, and the gene region from RL2 to RL5. cDNA clones representing transcripts from these two gene regions were among the most abundant in our libraries, and this was reflected in the number of S-AS pairs. For example, we found 7 clones overlapping RL3 gene sequences in S orientation and 189 clones overlapping RL3 gene sequences in the AS orientation. One notable feature of both classes of
genes was dominance of either S or AS transcripts of a pair. We observed that clones in one orientation outnumbered those in opposite orientations between 1.5 and 41 fold. In fact, an inverse relationship (at least 2:1) in abundances of S and AS transcripts was observed for 30 of the 38 S-AS pairs identified in this study. Based upon the library in which the clones were isolated, we made predictions regarding the temporal association of S and AS transcripts derived from the same gene. Accordingly, seven S-AS pairs were discordant relative to the library in which they were identified, whereas most (28) were concordant inasmuch that they were isolated from the same library. Keeping in mind that the E library is expected to contain both IE and E transcripts, and the L library could contain transcripts expressed at IE, E or L temporal classes, together these findings suggest that the majority of S-AS pairs are concordantly and inversely expressed during infection.

We also classified S-AS pairs with respect to the nature of the complementary overlapping sequences (Table 2.3). We used combinations of classification schemes proposed by others (118, 283) divided into one of four categories: full overlap, intronic, convergent, and divergent. Full overlap was defined as one gene sequence being completely contained within the gene sequence of the other member of the pair. Intronic was defined as one gene sequence starting within the intron of the other member of the pair, and ending beyond the start of its pair. Divergent was defined as S-AS pairs exhibiting overlap in their 5’ regions in a head-to-head manner. Finally, convergent was defined as S-AS pairs exhibiting
overlap in their 3’ regions in a tail-to-tail manner. Each potential pair was assigned to only one category and the order of stringency was full overlap, intronic, divergent and convergent. Using this scheme, we identified S-AS pairs that fell into each of these groups. Although the least abundant class, intronic pairs were observed for S-AS pairs overlapping 7 genes. Convergent overlap was common, with one or more S-AS pairs for 14 genes falling into this class. Pairs exhibiting full overlap were also abundant. We found one or more S-AS pairs overlapping 19 genes in this class. However, it should be noted that in the absence of 3’ end sequence for all of the cDNA clones, it is possible that the number of genes with S-AS pairs exhibiting full overlap is currently overestimated. Finally, we found that 29 of 38 genes included one or more S-AS pairs that were divergent in their overlap. Also, of those genes with multiple S-AS pairs, the divergent class was typically most abundant. Thus, while we observe a diversity classes of S-AS pairs derived from HCMV genes, pairs with divergent or head-to-head overlap were most common.
**Bold** indicates sequence analysis confirmed a genuine poly(A) tail of the antisense member of the pair

**Italicized** “full overlap” designations are tentative because the 3’ ends were not sequenced

<table>
<thead>
<tr>
<th>Pair</th>
<th>Gene</th>
<th># S clones</th>
<th>Library</th>
<th># AS clones</th>
<th>Library</th>
<th>Predicted temporal association</th>
<th>Predicted classification of AS pairs**</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>UL13</td>
<td>6</td>
<td>IE, L</td>
<td>1</td>
<td>E</td>
<td>Discordant</td>
<td>Divergent, Full Overlap</td>
</tr>
<tr>
<td>3</td>
<td>UL21</td>
<td>3</td>
<td>IE</td>
<td>2</td>
<td>IE</td>
<td>Concordant</td>
<td>Divergent</td>
</tr>
<tr>
<td>4</td>
<td>UL22</td>
<td>2</td>
<td>IE</td>
<td>3</td>
<td>IE</td>
<td>Concordant</td>
<td>Intronic</td>
</tr>
<tr>
<td>5</td>
<td>UL23</td>
<td>3</td>
<td>IE, L</td>
<td>1</td>
<td>L</td>
<td>Discordant</td>
<td>Divergent</td>
</tr>
<tr>
<td>6</td>
<td>UL24</td>
<td>2</td>
<td>L</td>
<td>3</td>
<td>L</td>
<td>Concordant</td>
<td>Divergent, Convergent</td>
</tr>
<tr>
<td>7</td>
<td>UL25</td>
<td>5</td>
<td>L</td>
<td>1</td>
<td>IE</td>
<td>Discordant</td>
<td>Divergent, Intronic</td>
</tr>
<tr>
<td>8</td>
<td>UL29</td>
<td>3</td>
<td>IE</td>
<td>1</td>
<td>L</td>
<td>Discordant</td>
<td>Divergent</td>
</tr>
<tr>
<td>9</td>
<td>UL36</td>
<td>3</td>
<td>IE, E</td>
<td>1</td>
<td>E</td>
<td>Concordant</td>
<td>Divergent, Full Overlap</td>
</tr>
<tr>
<td>10</td>
<td>UL37</td>
<td>5</td>
<td>IE, L</td>
<td>1</td>
<td>E</td>
<td>Discordant</td>
<td>Convergent</td>
</tr>
<tr>
<td>11</td>
<td>UL52</td>
<td>6</td>
<td>E, L</td>
<td>1</td>
<td>L</td>
<td>Concordant</td>
<td>Divergent</td>
</tr>
<tr>
<td>12</td>
<td>UL61</td>
<td>2</td>
<td>L</td>
<td>10</td>
<td>E, L</td>
<td>Concordant</td>
<td>Divergent</td>
</tr>
<tr>
<td>13</td>
<td>UL62</td>
<td>3</td>
<td>L</td>
<td>12</td>
<td>E, L</td>
<td>Concordant</td>
<td>Divergent, Full Overlap</td>
</tr>
</tbody>
</table>

Table 2.3. Genes for which *cis*-Natural S-AS Pairs were Identified and Their Properties  

continued
|   |   |   |   |   |   |   |   |
|---|---|---|---|---|---|---|
| 14 | UL63 | 19 | E, L | 3 | L | Concordant | Convergent, Divergent, Full Overlap |
| 15 | UL64 | 4 | L | 21 | E, L | Concordant | Convergent, Divergent, Full Overlap |
| 16 | UL65 | 20 | E,L | 3 | L | Concordant | Convergent, Divergent, Full Overlap |
| 17 | UL66 | 1 | L | 18 | E,L | Concordant | Divergent, Full Overlap |
| 18 | UL67 | 1 | L | 17 | E,L | Concordant | Divergent, Full Overlap |
| 19 | UL70 | 1 | L | 4 | IE, L | Concordant | Divergent, **Full Overlap** |
| 20 | UL72 | 1 | L | 3 | E,L | Concordant | Divergent, **Full Overlap** |
| 21 | UL73 | 3 | E,L | 1 | L | Concordant | Full Overlap |
| 22 | UL74 | 1 | L | 3 | E,L | Concordant | **Full Overlap**, Intronic |
| 23 | UL88 | 1 | L | 2 | L | Concordant | Divergent |
| 24 | UL89 | 1 | L | 2 | L | Concordant | Divergent, Convergent |
| 25 | UL92 | 1 | L | 1 | L | Concordant | Divergent, Intronic |

continued
<p>| | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>26</td>
<td><strong>UL102</strong></td>
<td>2</td>
<td>L</td>
<td>2</td>
<td>L</td>
<td>Concordant</td>
<td>Convergent, Divergent,</td>
<td></td>
</tr>
<tr>
<td>27</td>
<td>UL108</td>
<td>2</td>
<td>IE, L</td>
<td>1</td>
<td>IE</td>
<td>Concordant</td>
<td>Convergent</td>
<td></td>
</tr>
<tr>
<td>28</td>
<td>UL111</td>
<td>6</td>
<td>IE, L</td>
<td>2</td>
<td>IE, L</td>
<td>Concordant</td>
<td>Convergent, Full Overlap</td>
<td></td>
</tr>
<tr>
<td>29</td>
<td><strong>UL111A</strong></td>
<td>1</td>
<td>L</td>
<td>2</td>
<td>L</td>
<td>Concordant</td>
<td>Full Overlap</td>
<td></td>
</tr>
<tr>
<td>30</td>
<td>UL112</td>
<td>5</td>
<td>IE, L</td>
<td>1</td>
<td>L</td>
<td>Concordant</td>
<td>Divergent</td>
<td></td>
</tr>
<tr>
<td>31</td>
<td>UL115</td>
<td>4</td>
<td>IE, L</td>
<td>2</td>
<td>L</td>
<td>Concordant</td>
<td>Divergent, Intronic, <em>Full Overlap</em></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>UL116</td>
<td>2</td>
<td>IE</td>
<td>2</td>
<td>L</td>
<td>Discordant</td>
<td>Divergent, Intronic, <em>Full Overlap</em></td>
<td></td>
</tr>
<tr>
<td>33</td>
<td><strong>RL2</strong></td>
<td>6</td>
<td>E, L</td>
<td>15</td>
<td>E, L</td>
<td>Concordant</td>
<td>Convergent, Divergent, Full Overlap</td>
<td></td>
</tr>
<tr>
<td>34</td>
<td>RL3</td>
<td>7</td>
<td>E, L</td>
<td>189</td>
<td>E, L</td>
<td>Concordant</td>
<td>Convergent, Divergent, Full Overlap</td>
<td></td>
</tr>
<tr>
<td>35</td>
<td>RL4</td>
<td>167</td>
<td>E, L</td>
<td>4</td>
<td>E, L</td>
<td>Concordant</td>
<td>Convergent</td>
<td></td>
</tr>
<tr>
<td>36</td>
<td><strong>TRS1</strong></td>
<td>1</td>
<td>IE, E</td>
<td>1</td>
<td>L</td>
<td>Concordant</td>
<td><em>Full Overlap</em></td>
<td></td>
</tr>
<tr>
<td>37</td>
<td>US32</td>
<td>1</td>
<td>L</td>
<td>1</td>
<td>L</td>
<td>Concordant</td>
<td>Divergent</td>
<td></td>
</tr>
<tr>
<td>38</td>
<td>US33</td>
<td>1</td>
<td>L</td>
<td>4</td>
<td>IE, E, L</td>
<td>Concordant</td>
<td>Convergent, Divergent, Intronic</td>
<td></td>
</tr>
<tr>
<td>Predicted pair</td>
<td>Gene*</td>
<td># AS clones</td>
<td>Library</td>
<td>Reference for S gene product</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---------------</td>
<td>-------</td>
<td>-------------</td>
<td>---------</td>
<td>-------------------------------</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>UL2</td>
<td>3</td>
<td>E, L</td>
<td>(37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>UL3</td>
<td>32</td>
<td>IE, E, L</td>
<td>(37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>UL11</td>
<td>1</td>
<td>E</td>
<td>(102)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>UL27</td>
<td>1</td>
<td>L</td>
<td>(190)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>UL30</td>
<td>1</td>
<td>L</td>
<td>(85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>UL39</td>
<td>1</td>
<td>L</td>
<td>(85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>UL47</td>
<td>1</td>
<td>IE</td>
<td>(108)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>UL48</td>
<td>4</td>
<td>E, L</td>
<td>(9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>UL68</td>
<td>19</td>
<td>E, L</td>
<td>(85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>UL87</td>
<td>1</td>
<td>L</td>
<td>(85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>UL107</td>
<td>2</td>
<td>IE, L</td>
<td>(28, 185)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>RL5</td>
<td>150</td>
<td>E, L</td>
<td>(85)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>RL8</td>
<td>15</td>
<td>IE, E, L</td>
<td>(37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>RL9</td>
<td>3</td>
<td>E, L</td>
<td>(37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>RL10</td>
<td>1</td>
<td>L</td>
<td>(226)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>RL11</td>
<td>2</td>
<td>L</td>
<td>(8, 141)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>US29</td>
<td>3</td>
<td>E, L</td>
<td>(37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>US30</td>
<td>2</td>
<td>E, L</td>
<td>(37)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Bold indicates sequence analysis confirmed a genuine poly (A) tail of the AS member of the pair.

**Table 2.4. Genes with Predicted S-AS Pairs**

<table>
<thead>
<tr>
<th>Group</th>
<th>Gene</th>
<th># AS clones</th>
<th>Library</th>
<th>Reference for S gene product</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>UL28</td>
<td>1</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>UL79</td>
<td>2</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>UL90</td>
<td>1</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>US1</td>
<td>3</td>
<td>IE, E</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>US4</td>
<td>2</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>US5</td>
<td>2</td>
<td>L</td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>US31</td>
<td>1</td>
<td>E</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>US35</td>
<td>4</td>
<td>IE, E</td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>J1S</td>
<td>1</td>
<td>L</td>
<td></td>
</tr>
</tbody>
</table>

**Table 2.5. Orphaned AS Transcripts**
Finally, we classified S-AS pairs according to functional class of known or predicted gene products (Table 2.6). Several S-AS pairs overlap genes known to be involved in DNA replication and packaging including those genes (*UL70* and *UL102*) that encode for the components of the HCMV helicase-primase complex.

We also found S-AS pairs for the viral inhibitors of apoptosis encoded by *UL36* and *UL37* genes, and the recently described noncoding β2.7 transcript overlapping the *RL4* gene (197). Additionally, we identified S-AS pairs that overlap genes encoding tegument proteins, glycoproteins, and proteins involved in subversion of immune responses (*UL111A* and *RL11*) (8, 141) or cellular antiviral defense mechanisms (*TRS1, IRS1*) (4, 35). Finally, we identified S-AS pairs for at least 32 genes of unknown function, most of which are predicted to be noncoding. This aside, these findings indicate that there is not clear bias of S-AS pairs for genes of specific functional classes, and that S-AS pairs exist for both coding and noncoding genes.
<table>
<thead>
<tr>
<th>Functional Class</th>
<th>Genes for which S-AS pairs were identified</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA replication/packaging</td>
<td>UL27, UL52, UL70, UL72, UL89, UL102, UL112</td>
</tr>
<tr>
<td>Gene expression</td>
<td>IRS1, TRS1</td>
</tr>
<tr>
<td>Anti-Apoptotic</td>
<td>UL36, UL37, RL4</td>
</tr>
<tr>
<td>Glycoproteins</td>
<td>UL11, UL13, UL22A, UL73, UL74, UL115, UL116, RL10, RL11</td>
</tr>
<tr>
<td>Subversion of immune or antiviral defense</td>
<td>UL111A, IRS1/TRS1, RL11</td>
</tr>
<tr>
<td>Tegument proteins</td>
<td>UL23, UL24, UL25, UL47, UL48, UL88</td>
</tr>
<tr>
<td>Unknown</td>
<td>UL2, UL3, UL11, UL12, UL13, UL21, UL27, UL29, UL30, UL39, UL61, UL62, UL63, UL64, UL65, UL66, UL67, UL68, UL87, UL92, UL107, UL108, UL111, RL2, RL3, RL5, RL8, RL9, RL10, US29, US32, US33</td>
</tr>
</tbody>
</table>

* Bold indicates sequence analysis confirmed a genuine poly (A) tail of the AS member of the pair.

Table 2.6. Functional Classes of Genes with Verified or Predicted cis-Natural Sense-Antisense Pairs
**Verification of HCMV AS transcripts.** At least two features of a subset of AS clones suggest they cannot represent artifacts of our infection conditions or library construction. First, at least 68 of the clones harboring AS sequences, possess genuine poly (A) tails (Supplemental Table 3 (http://jvi.asm.org/cgi/content/full/81/20/11267)). Because not all clones were sequenced from the 3’ ends, it is likely additional clones with genuine poly (A) tails will be identified. These 68 clones constitute at least one AS member for 25 of the 38 genes with S-AS pairs identified in this study (Table 2.3, genes identified in bold) and 13 of 18 genes with predicted S-AS pairs (Table 2.4). AS clones with genuine poly (A) tails were also found in each of the functional classes listed in Table 2.6. A second feature is exemplified by the S-AS pair overlapping the UL36 gene. In this case, we identified 3 S clones all of which represented mature, spliced transcripts lacking the UL36 intron sequences (Supplemental Table 1, transcript group 35). The AS clone overlapping the UL36 gene region (transcript group 36) contains this intron sequence indicating that it could not represent aberrant cloning of the sense cDNA product.

Our analysis of *cis*-NATs from the HCMV genome relied upon isolation of virally-derived transcripts from human foreskin-derived fibroblasts. If this is a truly robust phenomenon, we predict that *cis*-NATs will be observed in other infected cell types and in cells infected with other strains of HCMV. To test this, we used RT-PCR to identify antisense transcripts in lung-derived human fibroblasts infected with laboratory-adapted AD169 HCMV, and endothelial cells
infected with the VHL/E clinical strain of HCMV. We designed primers specific for antisense transcripts from both coding and noncoding viral genes. We also utilized primers specific for sense-oriented viral UL55 transcripts coding for gB or cellular GAPDH transcripts. For specificity controls, RNA isolated from virus-infected and mock-infected cells was analyzed in the presence and in the absence of reverse-transcriptase.

In the first set of experiments, RT-PCR was used to amplify antisense transcripts identified in the E and L libraries using RNA isolated from AD169-infected MRC-5 fibroblasts at 48 and 72 h after infection, respectively. As shown in Figure 2.1A, we specifically amplified portions of antisense transcripts found in the early library derived from genes known or predicted to be protein coding (UL36, US30/31) and genes predicted to be noncoding (UL61, UL67, RL2/3, and RL8/9). Similarly, we specifically amplified products of expected sizes for antisense transcripts derived from genes known to be coding (UL27, UL70, UL87, UL102) and predicted to be noncoding (RL3) from RNA isolated at 72 h after infection (Figure 2.1B). Finally, we specifically amplified products of expected sizes for antisense transcripts derived from genes specifying tegument proteins (UL23, UL24, UL25, UL47, UL88) (Figure 2.1C). Except for the cellular GADPH, we could not amplify these transcripts from mock-infected cells. The absence of bands of predicted sizes in reactions conducted without reverse transcriptase insures that these results do not reflect contamination of viral DNA in our reactions. We conclude from these experiments that virally-derived
antisense transcripts are generated during lytic HCMV infections of fibroblast cells and that these AS transcripts do not reflect DNA contamination during the cDNA library construction.
Figure 2.1 Verification of virally-derived AS transcripts. Confluent MRC-5 cells in 6-well tissue culture plates were exposed to 2 PFU per cell of the AD169 strain of HCMV (top panels) or were mock-infected (bottom panels). 2.1A RT-PCR of AS transcripts identified in the early library using total RNA isolated at 48 h after infection. 2.1B RT-PCR of AS transcripts identified in the late library using total RNA isolated at 72 h after infection. 2.1C, RT-PCR of AS transcripts specific to tegument genes using total RNA isolated at 24 h after infection (UL47) or 72 h after infection (UL23, UL24, UL25 and UL85). Primers specific for the sense transcripts of the viral UL55 (gB) and the cellular GAPDH were included as positive controls. No reverse transcriptase and no template controls were run in parallel.
In the second set of experiments we analyzed two specific AS transcripts derived from the UL36/37 gene region and the UL102 gene in primary endothelial cells infected with the VHL/E strain of HCMV. Because lytic replication in endothelial cells infected with VHL/E proceeds much more slowly than in fibroblasts infected with laboratory-adapted strains, we performed RT-PCR over a 120 h time course (Figure 2.2). Specific antisense transcripts were amplified at each time point examined for these two gene regions. We conclude that generation of cis-NATs is a common feature of infection by both laboratory-adapted and clinical isolates of HCMV, and that cis-NATs occur during infection of clinically-relevant cell types. Also, while cycloheximide and ganciclovir were used to generate the IE and E libraries, all of these experiments were performed in the absence of any drug treatment. This rules out the possibility that these AS transcripts represent aberrant transcriptional products that accumulate in cells exposed to protein synthesis or viral polymerase inhibitors.
**Figure 2.2  Verification of antisense transcripts in endothelial cells infected with a clinical strain of HCMV.** Confluent HUVEC monolayers in 6-well tissue culture plates were inoculated with the VHL/E strain of HCMV (1 PFU/cell). Cells were harvested at 24, 48, 72, 96, and 120 h post-infection or were mock-infected for 120 h. Total RNA was isolated and subjected to RT-PCR using primers specific to virally-derived AS transcripts for UL36/37 (A) and UL102 (B). Primers specific for the sense transcripts of the viral UL55 (gB) and the cellular GAPDH were included as positive controls. No reverse transcriptase and no template controls were run in parallel.
To more carefully analyze the relationship between accumulation of S-AS pairs, we performed northern analyses using RNA from AD169-infected MRC-5 cells at various times after infection with probes specific to either S or AS products of the *UL24, UL36, UL102, UL61* and *RL 4/5* genes (Figure. 2.3). We identified two sense and three antisense clones overlapping the *UL24* gene in our libraries. By using probe specific to the sense transcripts of *UL24*, we identified a major band between the 1.8 and 2.6 kb markers beginning at 24 h after infection, and increasing in abundance through 72 h after infection, and a larger, weaker band just over 3kb expressed with the same temporal profile (Figure. 2.3A). These bands correspond to predicted sizes of clones pIE811 and pL5312 described in Supplemental Table 1. RACE was performed to verify the 5’ ends of pL5312. We identified genomic positions of 30238, 30029 as initiation positions for transcripts that are 432 and 223 bp larger than clone pL5312. Using probe specific for the antisense transcript of *UL24*, we identified 2 major bands between 3.0 and 4.0 kb at 48 and 72 h after infection. The largest antisense *UL24* clone isolated was 3.2 kb (pL8212). By RACE analysis, we identified two transcript initiation sites at genomic positions 28411 and 28023 predicting transcripts of 3755 and 4143 bp in this region, that correspond well with the two bands observed by northern blot. Together these findings confirm the presence of *cis*-NATs derived from the *UL24* gene in infected cells and validate our prediction of concordant expression and our finding of similar abundance of the *UL24 cis*-NATs.
We identified 3 sense clones and 1 antisense clone overlapping the UL36 gene. As shown in Figure 2.3B, we identified a band of the expected size (1.6kb) for the spliced sense product of the UL36 gene that corresponds to the clones listed in transcript group 35 in Supplemental Table 1. Also as expected, we could identify this band throughout the 72 h infection period, though it was most abundant at 24 h after infection. In our E library, we isolated one clone (pE104114) of approximately 1.9kb that was antisense to UL36. When we used a probe specific for the antisense transcript, we identified only a weak band of 1.9kb at 48 and 72 h after infection. We also observed a larger, more abundant band at 48 and 72 h. RACE was performed in attempt to identify the boundaries of this larger transcript. In this analysis we confirmed the boundaries of the antisense clone pE104114, and provide evidence for a transcript initiating at genomic position 48214 predicting a 2880 bp transcript antisense to UL36. This may represent the larger more abundant antisense transcript identified on the blot. However, because of the limitation in resolving the bands in this size range, we cannot exclude the possibility that this represents the faint band observed between the two major bands on this blot. In either case, we predict that there exists an additional antisense transcript derived from the UL36 gene. This analysis also confirms our prediction of concordant and inverse expression for the UL36-derived cis-NATs.

In contrast to that observed for the UL24 and UL36 AS transcripts, we found that expression of the AS-UL102 transcript precedes expression of the S-UL102
transcript (Figure. 2.3C). We identified a 2.3 kb sense transcript in our library represented by clone pL2211. Additional larger S transcripts are also observed by northern analysis consistent with that reported previously (221). RACE analysis confirmed the 5’ end of AS clone pL222 and identified another, larger AS transcript of 2.1 kb initiating at genomic position 151035. A third, even larger band of approximately 3kb was observed by northern analysis that has yet to be characterized. We predicted that cis-NATs of UL102 should be expressed with similar abundance and concordantly. While concordant expression was confirmed, the difference in exposure times required to visualize these transcripts indicate inverse accumulation of the S-AS transcripts.
Figure 2.3 **Northern blot analysis of S and AS transcripts.** Film images of S and AS transcripts analyzed by northern blot and schematic depictions of select S and AS transcripts. Confluent MRC-5 cells in 6-well tissue culture plates were exposed to 2 PFU per cell of AD169 strain of HCMV or were mock-infected (M). Cells were harvested at 24, 48, and 72 hr after infection. Total RNA was isolated, subjected to denaturing agarose gel electrophoresis, and transferred to nylon membranes. Pre-labeled RNA molecular mass markers were loaded for each group (L). Membranes were incubated with probes specific for the sense and antisense transcripts of (A) UL24, 3µg RNA/lane; (B) UL36, 3µg RNA/lane; (C) UL102, 4µg RNA/lane; (D) UL61, 5µg RNA/lane and (E) RL4, 4µg RNA/lane as described in Materials and Methods. Exposure times are indicated at the top of each panel. A schematic of transcripts represented by select cDNA clones isolated in our library relative to the genome is shown in the right panels. Gene regions and intergenic regions are depicted by thick arrows and white boxes, respectively. Transcripts cloned in this study are represented as thin arrows below the gene regions. 5’ ends of transcripts are depicted with filled circles. Clones in which we identified the poly (A) tails are indicated (AAA). The genomic positions of the 5’ and 3’ ends of the clones in the libraries are shown. Underlined genomic positions are those verified by RACE. Dashed lines represent presumptive transcript sequences based on RACE analysis. Tentative assignment of bands corresponding to clones identified in this study are indicated with circled numbers.
Figure 2.3 Northern blot analysis of S and AS transcripts.
Finally, we used a northern blot approach to ascertain whether the abundance of clones in our libraries overlapping the *RL4/5* and *UL61/UL62* genes reasonably reflects the abundance of these transcripts in infected cells (Figures 2.3D, 2.3E, and 2.4). First we set out to verify the existence of S and AS clones from these gene regions. Clone pL537 was the longest S clone overlapping *UL61* in our library predicting a band of 3.9 kb which corresponds well with the largest band observed on this northern blot (Figure 2.3D). Numerous clones overlapped *UL61* in the AS orientation, the longest of which are listed in transcript group 59 (Supplemental Table 1) that predict a band of 4.6 kb. These clones correspond well with the largest band identified with the AS-specific probe for UL61 (Figure 2.3D). The smaller bands identified by both S and AS probes specific for *UL61* represent may represent distinct smaller transcripts or degradation products of the larger transcripts isolated in this study. The major 2.7 kb transcript overlapping the RL4/5 genes has been described previously (48) and represented the major band of identified by the S-specific probe (Figure 2.3E). However, most or possibly all of the clones overlapping the RL4 genes isolated in this study appear to be initiated from genomically-encoded poly (A) tracts and thus do not represent the full length 2.7 kb transcript (represented by clone pE103121 depicted in the right panel of Figure 2.3E). We identified a band of approximately 3 kb using a probe that would recognize transcripts that are in antisense orientation relative to the 2.7 kb transcript. Although the precise boundaries of the AS RL4/5 transcripts (represented by clone pE103210 in the right panel) as well as the
UL61-derived transcripts have yet to be verified, the different exposures times required to visualize the S and AS transcripts clearly demonstrate inverse expression of the UL61-derived AS transcripts relative the UL61-derived S transcripts, and inverse expression of RL4-derived S transcripts relative to the RL4-derived AS transcripts. Finally, we compared the abundance of clones overlapping the UL61/UL62 and RL4/5 gene regions to that derived from the UL102 gene region (Figure 2.4). In our libraries, we isolated two S-UL102 clones, 5× as many AS-UL61 clones, and 85× as many S-RL4 clones. The abundance of these clones in our libraries correlated with the signal intensities of transcripts that bound the S-specific RL4 probe and the AS-specific UL61 probe relative to the S-specific UL102 probe. Taken together, these experiments prove the existence of cis-NATs derived from coding (UL24, UL36, and UL102) and predicted noncoding (UL61 and RL4) genes. Furthermore, these studies indicate that the representation of S and AS transcripts in our libraries reasonably reflect the abundance and temporal expression patterns of S and AS transcripts generated in HCMV-infected fibroblasts.
Fig 2.4 Northern blot analysis to examine the relative abundance of transcripts. The relative abundance of transcripts overlapping the RL4 and UL61 gene regions are compared to transcripts derived from the UL102 gene region. White circles indicate the position of the 2.6 kb marker band. No images were altered.
Discussion

The HCMV transcriptome. In this study, we isolated and characterized transcripts generated during lytic infection of HCMV-infected fibroblasts. By selecting for cDNA clones representing viral genes, and obtaining sequence data on each isolated clone, this was a *de facto* analysis of the HCMV transcriptome. This analysis revealed two novel findings. First, there was an unexpectedly high percentage of clones derived from genomic regions currently considered to be noncoding (57). While we expected a high prevalence of RL4 transcripts based on reports from other studies (88, 107, 152), additional noncoding transcripts were also represented at high frequencies, especially those from the *UL61-UL67* gene region. Altogether, 45% of the clones isolated in this study were derived from gene regions predicted to be noncoding (57), suggesting further studies of the function and regulation of these gene regions are needed. Second, we identified a striking number of transcripts, predominantly antisense, that were not predicted by the STD database annotation of the HCMV genome. We therefore conclude that genomic maps based on open reading frame analyses and comparisons to related viral genomes drastically underestimate the true complexity of viral gene products.

The discovery of widespread antisense transcription also bears upon the use of viral gene arrays for transcriptome analyses and recombinant viruses for gene function studies. Specifically, these findings suggest that tiling arrays using overlapping probes from both DNA strands, rather than probes based upon ORF
predictions, will be required to capture an accurate representation of viral transcripts. Additionally, these findings raise the possibility that AS as well as S gene products could be affected when generating recombinant viruses harboring mutations or deletions.

There are several factors that could influence the composition of this library that are relevant to the interpretation of these data. First, because highly abundant transcripts are expected to be preferentially represented, we infer that transcripts from genes not represented in our library are of relatively low abundance. While abundance of these transcripts is likely a key factor, we cannot exclude the possibility that artifactual pressures might also influence library composition. For example, it is possible that there were selective advantages for isolating transcripts from gene regions with long or repetitive tracts of adenosines, such as the RL gene region. These transcripts may have been preferentially enriched during the purification of polyadenylated RNAs prior to cDNA library construction. Similarly, genomically encoded poly(A) tracts may have served as primer binding sites during the reverse transcription step of the cDNA library construction. While many cDNA clones appear to have genuine poly(A) tracts, we also found a number of clones, especially from predicted noncoding regions, in which the 3’ end sequence corresponded to a genomically-encoded poly(A) tract. Therefore, further studies will be necessary to define precise ends of a subset of transcripts, as well as their relative abundance in infected cells. Despite these caveats, northern analyses of transcripts derived from single copy and repeat
region genes suggest that composition of our library reasonably reflects the abundance and temporal expression patterns of transcripts in infected fibroblasts.

**Antisense Transcripts of the HCMV genome.** Individual AS transcripts have been described for many herpes viruses, including the betaherpesviruses, (12, 122, 254), the gammaherpesviruses, (188, 189, 211), and especially the alphaherpesviruses, (21, 22, 31, 34, 40, 47, 48, 103, 125, 129, 139, 192, 258, 274, 275, 277). In fact, Roizman and colleagues predicted that genes in AS orientation to known herpes virus genes could be common (34). However, to our knowledge, our study is the first to systematically document cis-NATs derived from a herpes virus. We found that at least 55% of the clones analyzed in this study contain sequences in AS orientation overlapping 56 known or predicted genes. These figures are dramatic inasmuch as they suggest that more than half of all virally-derived transcripts harbor AS sequences. Nevertheless, three factors suggest that our study may have underestimated AS transcription in HCMV. First our libraries do not contain all of the S transcripts expressed by HCMV, and thus it is likely that they do not contain all of the AS transcripts that accumulate during infection. Indeed, at least two previously reported AS transcripts overlapping the HCMV UL82 and UL123 genes were not isolated in our libraries (12, 122). Second, we selected for polyadenylated transcripts during construction of our libraries, and studies suggest that a large fraction of AS transcripts are poly(A) negative (119). Finally, we classified only cis-NATs identified in our libraries. Considering that the HCMV genome includes numerous repeat elements, it is
likely that trans-derived NATs are also generated during infection.

Relative to eukaryotic genomes, herpes virus genomes are small and densely populated with genes. In fact, many genes are directly adjacent to one another and a number of viral genes overlap with each other, often in opposite orientations. These characteristics suggest that the occurrence of S-AS pairs may only reflect the dense organization of viral genes and the relative paucity of intergenic regions relative to larger genomes. Nevertheless, S-AS pairs derived as a consequence of this gene organization may be functionally relevant with respect to regulation of viral genes. This raises the interesting possibility that regulatory consequences of S-AS pairs may constitute one evolutionary pressure influencing function and orientation of adjacent viral genes.

Given these properties of the HCMV genome, it seems reasonable to predict that higher proportions of viral genes would be associated with S-AS pairs compared to genes of eukaryotic genomes. In this study, we found that at least 29% of the currently annotated 191 HCMV genes are associated with S-AS pairs. While this figure is similar to estimates of 22-26% (46), (283) of human genes associated with S-AS pairs, it is lower than the estimate of 72% of mouse genes (118) predicted to be influenced by S-AS pairs. These preliminary findings suggest that the proportion of HCMV genes involved in the generation of S-AS pairs is similar to that observed for the human genome, despite the striking differences in sizes and structures of these genomes. Another common feature between mammalian and viral S-AS pairs is the nature of the complementary sequences. Our analyses
indicate that most S-AS pairs overlap either in 5’ or 3’ regions, with intronic organization as the least frequent class of S-AS pairs, and this is similar to reports by other groups for the human genome (134, 278, 283). One prediction from this type of overlap is that the potential regulatory consequences of these S-AS pairs relate to function, accessibility or processing of the UTRs of viral transcripts (238). Since post-transcriptional processing of HCMV viral transcripts has been described (17, 32, 74-76, 81, 149, 232, 276), it will be of interest to determine whether such events are influenced by AS transcripts. Also, in mammalian genomes, S-AS pairs appear to be overrepresented among genes involved in genomic imprinting (118), metabolic, catalytic, and cell organization functions (283), and in translational regulation (46). However, our findings indicate that there is not a clear functional bias among HCMV genes for which S-AS pairs were identified.

A key question raised by our findings is whether the virally-derived S-AS pairs have regulatory consequences during lytic or latent viral infections. AS transcripts can impact viral gene expression though multiple mechanisms, such as influencing splicing, editing, stability, localization and translation of transcripts (130, 163). In addition, double-stranded intermediates generated from direct interaction of S-AS pairs may lead to gene regulation by RNA silencing or chromatin remodeling (130, 163). S-AS pairs involved in regulatory functions are predicted to be expressed concordantly and to accumulate in an inverse manner (45). Indeed co-expressed and inversely expressed S-AS pairs are not only more
frequent in the human genome than would be expected by chance, but are also evolutionarily conserved (238). Our results predict that the majority of HCMV S-AS pairs are expressed with just such a profile, and provide experimental evidence for concordant expression and inverse accumulation for S-AS pairs derived from the *UL36, UL102, UL61/UL62* and *RL4/5* gene regions.

Another possibility is that AS transcripts may serve as primary transcripts for microRNAs (miRNA), as has been recently suggested for the antisense latency associated transcripts (LAT) of HSV and Marek’s disease virus (30, 92). Several studies have identified putative or validated miRNA’s of HCMV (63, 89, 184). Interestingly, a number of miRNA’s that have been predicted or validated are derived from the AS strand of viral genes including the *UL31, UL53, UL70, UL102, UL114, UL150* and *US29* genes (89, 184). In support of this possibility, we identified AS transcripts from a several of these genes including *UL70, UL102* and *US29*. While further studies are required to establish regulatory roles for these AS transcripts, we predict that these AS transcripts may dramatically alter our understanding of viral gene regulation during both lytic and latent infections.

To summarize, the remarkable accumulation of noncoding and antisense transcripts during infection suggests that currently available genomic maps based on open reading frame and other *in silico* analyses may drastically underestimate the true complexity of viral gene products. These findings also raise the possibility that aspects of both the HCMV life cycle and genome organization are influenced by antisense transcription. These noncoding and antisense transcripts...
may offer new insights into HCMV pathogenesis and may serve as novel targets for developing intervention strategies and treatments for HCMV-related diseases.
CHAPTER 3

THE CARBOXY TERMINAL DOMAIN OF HUMAN CYTOMEGALOVIRUS IE1 DISRUPTS TYPE II IFN SIGNALING

Introduction

After primary infection, human cytomegalovirus (HCMV) persists for the lifetime of the host, avoiding detection and elimination by the host immune system (255). In congenitally infected infants and immunosuppressed persons, HCMV can cause serious diseases (29). Both aspects of HCMV disease and persistence in human populations are suspected to be linked to the multiple mechanisms this virus has evolved to modulate human immune responses. The earliest immune responses that HCMV and all viruses encounter are mediated by cytokines known as interferons (IFNs). IFNs are classified into three categories, namely Type I IFN (including α, β, ω, ε, κ), Type II IFN (IFNγ) and Type III IFN (IFNλ), though Type I and II interferons are the best studied with respect to their antiviral functions (23). IFNs have intrigued virologists for over 50 years not only for providing insight into viral biology and the molecular combat that occurs in infected cells, but also for the potential promise of exploiting the IFN system for therapeutic control and prevention of viral diseases.
IFNs function to limit virus replication and pathogenesis through stimulation of intrinsic cellular defense mechanisms that contribute to innate immunity (83, 206) and by promoting and regulating adaptive immune responses (131, 236).

The antiviral functions of IFNs are initiated by binding of the interferons with their receptors. The Type I receptor is composed of two proteins, IFNAR1 (IFN Alpha Receptor 1) and IFNAR2 (IFN Alpha Receptor 2). The cytoplasmic tail of the receptors are associated with protein tyrosine kinases, JAK1 (Janus Kinase 1) and TYK2 (Tyrosine Kinase 2). Binding of Type I IFN to its cognate receptor leads to phosphorylation of JAK1 and TYK2 and tyrosine residues on the cytoplasmic tails of the receptors. This leads to docking of STAT1 (Signal Transducer and Activator of Transcription 1) and STAT2 (Signal Transducer and Activator of Transcription 2) to the receptor cytoplasmic tails via SH2 (Src Homology 2) domains leading to tyrosine phosphorylation and heterodimerization of the STATs. The STAT1-STAT2 heterodimers then forms a complex with IRF9 called ISGF3 (Interferon Stimulated Gene Factor 3) that translocates to the nucleus and binds ISRE (Interferon Stimulated Response Element) in the promoters of type I IFN induced genes (186, 208). Type III IFN induce the same signaling cascade as the Type I IFNs although they use different receptors. Type II IFNs bind to a different receptor composed of Interferon Gamma Receptor 1 and 2 (IFNGR1 and IFNGR2) proteins associated with the tyrosine kinases JAK1 and JAK2. IFNγ signaling results in formation of phosphorylated STAT1 homodimers that translocate to the nucleus and bind GAS (Gamma Activated
sequence) elements in the promoters of type II IFN inducible genes (23, 208, 233).

Interferons can stimulate the expression of over 300 cellular genes known as ISGs (Interferon Stimulated Genes) (59). The proteins encoded by ISGs have antiviral and immunomodulatory activity. The affect of these gene products include interference with the viral life cycle by degrading viral nucleic acids, inhibiting viral protein synthesis, interference with viral trafficking or inducing apoptosis leading to death of virally infected cells. In addition to their direct antiviral role, IFNs can control viral pathogenesis by enhancing adaptive immune responses, enhancing antigen presentation to cells of the immune system and recruiting immune cells to the sites of infection. In particular, type II IFN is important in stimulating the activity of NK cells, macrophages, T-cells and DCs.

To counteract the effects of IFNs many viruses have evolved various mechanisms to interfere with IFN signaling (271). HCMV in particular has numerous strategies to modulate both innate and adaptive immune response of the host including the IFN signaling pathway (147, 271). In the case of HCMV, it has been demonstrated that IFN γ signaling in infected cells is disrupted by modulating the JAK-STAT signaling pathway through degradation of JAK1(156). It has also been reported that there is a block in IFN γ induced CIITA induction downstream of STAT1 nuclear translocation in HCMV infected cells (132). Miller et.al (157), have also reported a blockade in binding of STAT1 to GAS elements at very early times in HCMV infected cells. Similarly in murine cytomegalovirus (MCMV)
infected macrophages there is a blockade in IFN γ mediated regulation of several genes (187). In addition, the MCMV M27 protein has been shown to disrupt IFNγ signaling through a novel, STAT2-dependent mechanism(290).

The HCMV genome of 230 Kbp is among the largest genomes of the herpes family. Wild-type strains of HCMV are predicted to encode more than 165 protein coding genes (57) and many non-protein coding genes including many specifying antisense transcripts (282). In order to identify the HCMV genes involved in altering host interferon mediated signaling, we constructed a cDNA library of HCMV laboratory strain AD169 and screened it using a reporter cell line that expresses CD2 at the cell surface when stimulated by either type I or II interferons (265). We found that expression of the HCMV \textit{UL123} gene that codes for the immediate early transactivator protein IE1 (a.k.a. IE72) led to a decrease in accumulation of CD2 on the surface of 2C4 cells.

IE1/IE72 is one of the major immediate early transcriptional transactivators of HCMV. It is a multifunctional protein with functions including promiscuous transactivation of viral and cellular genes, PML targeting, ND10 disruption, p107 binding and protein phosphorylation, and inhibition of histone deacetylase activity (123, 241, 248, 284). Recently it has also been demonstrated that IE antagonizes Type I interferon signaling (181). We report here that HCMV IE1 can disrupt signaling by both type I and Type II interferons. Furthermore we have determined that the carboxyl-terminal residues of IE1 that include the acidic domain is required for this function. We found no defect in these initial events in IFN γ
signaling in IE1 overexpressing cells, nor did we observe an association between
disruption of IFNγ signaling and ND10 disruption. Moreover, IE1 does not
interfere with nuclear accumulation of STAT1. However there is reduced binding
of STAT1 homodimers to target GAS elements in the presence of IE1. This
activity does not appear to require a direct interaction of IE1 and STAT1 and
suggests that IE1 disrupts IFN signaling in the nucleus through a novel
mechanism.
Materials and Methods

Cells and Viruses: Human embryonic lung fibroblasts (MRC5) were obtained from American type culture collection (ATCC) and maintained in ATCC complete media. Human fibrosarcoma cell line 2C4 was a kind gift from George Stark (265). 2C4 cells were maintained in DMEM supplemented with 10% FBS, Glutamine and Sodium Pyruvate. HCMV strain AD169 was obtained from ATCC and propagated in MRC5 cells. Viral titers were determined by standard plaque assay (268).

Plasmids and vectors: Construction of the AD169 cDNA library is described elsewhere (282). cDNA clones containing different truncated versions of the IE1 gene were selected from the cDNA library. The following clones were used, pie639 (exon2-3 deletion mutant), pie535 (Leucine Zipper deletion mutant), pie836 (Zinc finger deletion mutant). The following are the genomic coordinates of these IE1 clones – pie 639 (173157-171814), pie535 (172658-171814), pie836 (172436-171814). The sequence specifying the FLAG tag from pCMV3tag vector (Stratagene) along with CMVIE1 promoter sequence was taken as an NdeI – EcoRI fragment and ligated with NdeI –EcoRI digested IE1 clones. The resultant vectors had 5’ FLAG tags in frame with different truncated versions of IE1. For the C-terminal acidic domain (AD) deletion mutant, pie622 (full length IE1 clone), was digested with BglII to release IE1 without the AD, and was ligated into a vector containing a FLAG tag at the 5’ end in the same coding frame. The genomic coordinates for the AD mutant are (173705-173626+173511-173327+...
173156-172377). The full length IE1 was PCR amplified from pie622 using the following primers forward- 5’AGAGTGACTCAACCAAGCTTGACACGATG 3’ and reverse – 5’CGAGGCTGATCAGCTTAATTAACC3’. The IE1 without exon 2 was PCR amplified using the following forward primer- 5’CCTTCCTCAAGCTTCCACGGCC 3’ and the reverse primer used was the same as that used for amplifying the full length IE1. The PCR products were digested with HindIII and PacI and cloned into HindIII and PacI digested vector containing the flag tag in the 5’ end in frame.

**Transfection, Nucleofection and Infection:** 40,000 2C4 cells were plated per well of a 12 well cluster. Within 24 hrs of plating cells were transfected with the cDNA clones using Fugene HD transfection reagent (Roche) according to manufacturer’s recommendations. 1ug of plasmid DNA was used per well with 3µl of transfection reagent. For nucleofection 1.5x10^6 MRC5 cells were resuspended in 100µl of basic nucleofector solution for primary fibroblasts with 3ug of plasmid DNA and nucleofected using U23 program in the nucleofector device (Amaxa). After nucleofection cells were immediately distributed equally in 3 wells of a six well cluster. For infections confluent MRC5 cells in 75 cm² tissue culture flasks were exposed to 3 pfu per cell of HCMV strain AD169 and harvested 24hrs post infection. Wherever indicated, cells were treated with 100U/ml of human recombinant IFNβ or IFNγ (Chemicon). All interferon treatments were done 24hrs after transfection or nucleofection.

**Immunoblotting and Immunoprecipitation:** Infected or nucleofected cells were
rinsed with cold versene (PBS+ 1mM EDTA) and cells were harvested by scraping in versene. Harvested cells were pelleted by centrifugation at 1500 rpm for 5 minutes and resuspended in lysis buffer (25mM Tris, 150mM NaCl, 1% NP40, 10mM NaF, 0.1mM NaVO₄, 1mM EDTA and 1% (vol/vol) protease inhibitor cocktail (Sigma)). The cells were then sonicated in a Misonix Cuphorn sonicator for 100 seconds (5s on, 5s off) setting 5. Sonicated cells were centrifuged at 13,000 rpm for 5 minutes to pellet insoluble material. Protein was quantified using Bradford dye (Biorad) according to manufacturer recommendations. For immunoprecipitation, cell lysates were precleared with protein A agarose beads for 30 minutes at 4°C. The precleared lysates were incubated with primary antibody at 4°C overnight followed by incubation with Protein A beads for 2 hrs at 4°C. The immunoprecipitated complexes were washed 4 times in lysis buffer and eluted by boiling in 2x laemmli buffer with 5% β-ME. Proteins were separated by SDS-PAGE, transferred to a nitocellulose membrane (Whatman), probed with primary antibodies overnight, exposed to HRP-conjugated secondary antibody and detected by ECL (Amersham). The following antibodies were used – JAK1 (Upstate), JAK2 (Santacruz), STAT1 (Santacruz), STAT2 (cell signaling), phosphoSTAT1 (cell signaling), FLAG (Sigma), GAPDH (Chemicon), IE1 exon 2 antibody (Chemicon), IE1 exon 4 antibody p63-27 (kind gift of William Brit).

Flow Cytometry: 2C4 cells were washed with versene and harvested by pippeting up and down, 24hrs after IFN treatment. Cells were then resuspended in flow
buffer (PBS with 1% FBS) and stained with FITC conjugated anti CD2 antibody (Dako) according to manufacturers recommendations. After staining the cells were washed with flow buffer and CD2 fluorescence was measured in a Facscalibur Flow cytometer (Becton Dickenson).

Real Time RT-PCR: Total RNA was isolated using trizol (invitrogen) according to the manufacturer protocol. RNA was quantified using spectronic 20 spectrophotometer (Shimadzu). 1µg of total RNA from each sample was used to perform first strand cDNA synthesis using the RT kit (invitrogen) as per manufacturer’s recommendations. The cDNA was diluted 1: 4 and 8µl was used per real time reaction of 25µl total volume. For realtime using SYBR green the primers used for CIITA and GAPDH (endogenous control) are described elsewhere (16). For real time using Taqman the following primer and probe sets were used- CIITA-Hs00172106_m1, IRF1-Hs00233698_m1 and 18SrRNA-4310893E (Applied Biosystems).All Real Time reactions were set up in triplicate in 96 well format in an ABI prism 7900HT real time instrument. Fold change in CIITA or IRF1 expression in IFN treated samples compared to untreated samples was calculated using the $2^{-ΔΔCt}$ method after normalizing to the endogenous control either GAPDH or 18SrRNA.

Microscopy: Nucleofected or transfected cells were plated on 12mm coverslips in 24 well plates. Wherever indicated cells were exposed to 100U/ml IFN for 30 minutes. Following this the cells were rinsed with versene and fixed in methanol for 1 hr on ice. The cells were then air dried and stored at 4°C till staining. Before
exposure to the primary antibody the cells were first exposed to PBS for 5 minutes and blocked in blocking buffer (PBS +10% Human serum + 1% BSA) for 1 hr. This was followed by incubation with primary antibody in blocking buffer overnight at 4\(^0\)C. Then the cells were washed 3 times x 20 minutes with PBS. This was followed by exposure to fluorescently labeled secondary antibody for 1 hr at 4\(^0\)C. Excess secondary antibody was washed off with 3 washes in PBS as before. The coverslips were allowed to air dry and then they were mounted on glass slides using the Prolong antifade kit (Molecular probes). Microscopy was done using a Zeiss LSM 510 instrument. The following primary antibodies were used IE1 exon 2 antibody (Chemicon), IE1 exon 4 antibody p63-27 (kind gift of William Brit), PML (Santacruz).

**Electrophoretic Mobility Shift Assay (EMSA):** MRC5 cells were plated at a density of 1x \(10^6\) cells per 75cm\(^2\) flask post nucleofection. Following this the cells were exposed to 100U/ml IFN\(\gamma\) for 30 minutes or left untreated. Cells were washed with PBS, harvested, and nuclear extracts were prepared using Nuclear extract kit (Activ Motif). A 22 mer IRF1 GAS element (5\(^{\prime}\)GATCGATTTCGGAAATCATG 3\(^{\prime}\)) was end labeled with \(\gamma\) [P]\(^{32}\) ATP using T4 polynucleotide kinase enzyme (Invitrogen) according to the manufacturer’s recommendations. This radiolabelled IRF1 GAS element was used as the probe. 4\(\mu\)g of nuclear extract was incubated with 1x binding buffer, 0.5 \(\mu\)g poly dI:C, 0.1 \(\mu\)g poly L-lysine and 2ng of labeled probe for 15 minutes at room temperature. For competition assays 200 fold excess of unlabeled probe was
added to the reaction mix. For supershifts 5μg of rabbit polyclonal anti STAT1 antibody (Santacruz) was added to the reaction and incubated for 15 minutes at room temperature prior to adding the labeled probe. The reactions were resolved on a 4.5% native polyacrylamide gel in 0.5x TBE buffer.

More detailed protocols are provided in the appendix.
Results

**HCMV IE1 gene disrupts signaling by Type II interferon.**

In an effort to identify the HCMV genes involved in disruption of IFN signaling a cDNA library of the HCMV laboratory strain AD169 was constructed. We used the human fibrosarcoma cell line 2C4 to screen for HCMV cDNA clones that have a role in interfering with IFN signaling. The 2C4 is a fibrosarcoma cell line engineered to express the T-cell antigen CD2 under the control of the interferon inducible promoter element of the IFITM1 gene (265). Stimulation with IFN, activates the transcription of the CD2 gene leading to increased cell surface levels of CD2 in 2C4 cells. 2C4 cells transfected with any HCMV cDNA clone involved in disruption of IFN signaling would be expected to have reduced IFN- induced cell surface levels of CD2 as compared to empty vector transfected cells. Of the first 80 cDNA clones from our HCMV cDNA library that were tested, we identified 1 clone in which CD2 cell surface levels were diminished relative to empty vector-transfected cells after exposure to IFNβ, and to a lesser extent, IFNγ (Figure. 3.1). This cDNA clone, pIE622, encodes for the HCMV *UL123* gene coding for IE1/IE72 protein.

To confirm that IE1 is actually affecting the IFN signaling pathway and not interfering with the translocation of CD2 to the cell surface we examined localization of CD2 in IE1 transfected 2C4 cells using immunofluorescence microscopy. We found that IE1 and CD2 did not colocalize as IE1 is confined to the nucleus and CD2 is restricted to the cytoplasmic compartment (data not
shown). As the CD2 gene is under the control of the promoter of the IFITM1 gene in 2C4 cells, this series of studies indicates that IE1 is able to interfere with both type I and type II IFN-induced signal transduction.

Because the role of IE1 is regulating the IFNγ signaling pathway has not been investigated, we sought to further characterize this observation. First, we wanted to study the effect of IE1 in a more physiologically relevant cell type for HCMV infection, namely, fibroblasts. The basal level of MHC II Transcriptional Activator (CIITA) is very low in fibroblasts and is highly induced upon IFNγ treatment. CIITA transcript levels were quantified by Real Time RT-PCR, in MRC5 cells nucleofected with IE1. Nucleofection was selected as a method of gene delivery as the efficiency of transfection was 95%. At 24 hrs after nucleofections, cells were treated with IFNγ for 6 hrs at which peak expression of CIITA transcripts is observed. The CIITA mRNA levels in IFNγ treated cells were compared to untreated cells to quantify the fold increase in CIITA upon IFN treatment. As an additional control, we also examined CIITA expression in cells transfected with cDNA clone pL759 harboring full gene sequence of UL99 coding for the HCMV pp28 tegument protein. We found that the fold increase in CIITA induction by IFNγ in IE1 expressing cells was only 39% of that observed in empty vector nucleofected cells (Figure.3.2A). We also examined the IFNγ induced expression of IRF1 in IE1 expressing fibroblasts. We found that the fold increase in expression of IRF1 in response to IFNγ is reduced by 52% in IE1 expressing cells as compared to empty vector nucleofected cells (Figure.3.2B).
These findings confirm that expression of IE1 alone is sufficient to attenuate IFNγ signaling in cell types that are known targets of HCMV infection.
Figure 3.1: Reduced CD2 cell surface levels in cells expressing IE1.

Histograms of cell surface CD2 levels in 2C4 cells transfected with pIE622 harboring full length UL123 gene sequences and empty vector (FIN2) control. The thin grey lines show the population of cells transfected with pIE622, and the bold black lines are cells transfected with the empty vector. At 24 h after transfection cells were exposed to 100U/mL of IFNβ (middle panel) or IFNγ (bottom panel). At 24 h after IFN treatment, cells were dislodged from the plates, reacted with FITC-conjugated antibody to CD2, and analyzed by flow cytometry.
Figure 3.2. **Diminished interferon induced gene expression in human fibroblasts expressing IE1.** MRC-5 fibroblasts were nucleofected with plasmid pL759 harboring the UL99 gene sequence specifying pp28, plasmid pIE622 harboring the UL123 gene specifying IE1, or the pFIN2 empty vector. 24 hrs post nucleofection the cells were exposed to 100U/mL of IFN \( \gamma \) or were left untreated. At 6hrs after treatment total RNA was isolated and subjected to Real Time RT-PCR analysis as described in Materials and Methods. The fold increase in CIITA (A) or IRF1 (B) transcript levels in IFN treated cells relative to untreated cells was determined by the \( \Delta \Delta Ct \) method. Shown is the average of 3 independent experiments.
Mapping of the IE1 protein region involved in disruption of IFN signaling.

The HCMV IE1 protein is a multifunctional protein with different functions attributed to different domains of the protein (Figure 3.3). We generated a series of FLAG epitope-tagged N-terminal truncation mutants to identify the protein region required for IFNγ signal disruption. To accomplish this, we utilized truncated UL123 gene sequences identified in our cDNA library that mapped to different structural and functional domains of IE1. To generate an exon 2 deletion mutant we utilized PCR to amplify the sequence from exon 3 to exon 4 of IE1 and inserting this fragment it into the pCMV TAG vector. In addition, we constructed one C-terminal truncation mutant as described in Materials and Methods. In all we generated 5 different FLAG-tagged truncation mutants plus a full length FLAG-tagged version of IE1 (Figure. 3.3). These 6 different IE1 clones were nucleofected into MRC5 cells, treated with IFNγ for 6 hrs and levels of CIITA transcript induction was measured by Real Time RT PCR. We found that there was a reduction in the IFNγ induced expression of CIITA as compared to empty vector nucleofected cells in cells expressing full length and truncated versions of IE1, with the exception of the C-terminal truncation missing residues 345-491 that includes the acidic domain (∆AD) (Figure.3.4). We consistently observed higher levels of CIITA induction in cells expressing the ∆AD relative to our control transfections. Although the reason for this is not clear, it may be that the IE1 protein without the C-terminal domain provides a signal that amplifies IFNγ-induced CIITA gene expression.
Figure 3.3 Schematic of the IE1 protein and truncation mutants generated in this study. Top line, a schematic of the IE1 protein indicating the amino acid positions comprising the coding exons (below) and the positions of the known functional domains (above) including the leucine zipper (L), the zinc finger (ZF) and the acidic domain (A). The IE1 N-terminal truncation mutants generated in this study are shown below with starting amino acid positions indicated. The sixth mutant is a C-terminal truncation mutant with the last residue indicated. The designation of the plasmids harboring the truncation mutants is based on the deletion of relevant exons or functional domains shown on the right. All the gene sequences were inserted in frame to sequences specifying a FLAG epitope (depicted as a flag).
3.4 ISG induction in fibroblasts expressing truncated IE1 proteins. The indicated plasmids were nucleoporated into MRC5 cells and 24 hrs after nucleofection the cells were treated with 100U/mL IFNγ. At 6hrs after treatment total RNA was isolated. Real Time RT-PCR analysis was carried out using Taqman™ probes for CIITA and 18SrRNA (endogenous control). Fold increase in CIITA transcript levels in IFN treated cells relative to untreated cells was determined by the ΔΔCt method. Shown is the average of 2 independent experiments.
Disruption of IFNγ signaling by IE1 does not involve interaction with PML

We next considered the possibility that our observations are linked to the ability of IE1 to localize to ND10 structures in the nucleus and disperse them (3, 123, 272). Promyelocytic leukemia protein (PML) is involved in transcriptional repression and a master organizer of ND10 structures. HCMV IE1 disrupts ND10s leading to dispersal of PML. Because PML is known to mediate antiviral activities of IFNs in HSV-infected cells (41) and PML confers intrinsic immunity against CMV (123, 241), we set out to test the hypothesis that the function of IE1 in disrupting IFNγ signaling is linked to its role in disrupting ND10 structures. We examined ND10 structures by immunofluorescence microscopy. We found that in fibroblasts expressing the full length IE1 and AD deletion mutant of IE1, ND10 structures become dispersed all over the nucleus, as PML staining was seen throughout the nucleus (Figure 3.5). In cells that did not express IE1 the PML staining was in the form of punctate dots in the nucleus. Therefore, the ΔAD deletion mutant retains the ability to target PML and disrupt ND10s. We conclude from these studies that dispersal of ND10 structures is not required for IE1-mediated disruption of IFNγ signaling.
Figure 3.5: Disruption of ND10 structures in cells expressing IE1.
Immunofluorescence images of IE1 and PML in fibroblast cells. MRC5 cells were nucleofected with the indicated constructs. At 24 hrs after nucleofection cells were fixed in methanol and reacted with mouse monoclonal anti-IE1 antibody (Mab810) followed by anti mouse alexafluor 488 as secondary antibody. The cells were then exposed to rhodamine conjugated anti PML antibody (PGM3). IE1 reactivity is visualized in green (left column), PML reactivity is visualized in red (middle column) and merged images are shown in the right column.
Initial events in IFNγ signaling are not affected in IE1 expressing fibroblasts

Next we wanted to determine which steps of the IFN signaling pathway are disrupted by HCMV IE1. To this end we examined the state and abundance of proteins involved in IFNγ signaling in IE1 expressing cells. We found that there was no difference in the levels of total JAK1, JAK2 and STAT1 in IE1 expressing cells as compared to empty vector harboring cells. The activation of STAT1 by IFNγ was not affected by IE1 since similar levels of Stat1 phosphorylation could be achieved after 30 minutes of IFNγ treatment in IE1 overexpressing cells and empty vector nucleofected cells (Figure 3.6). We also examined the nuclear translocation of STAT1 after 30 minutes of IFNγ treatment in IE1 expressing cells. We found that the translocation of STAT1 to the nucleus was not affected by overexpression of IE1 (Figure 3.7). Thus, IE1 does not act by disrupting the key IFNγ signal transduction molecules, nor does it disrupt the early events in IFNγ signaling including STAT1 tyrosine phosphorylation and nuclear translocation.
Figure 3.6: IE1 does not affect the state or abundance of the initial signaling molecules of IFNγ pathway. Film images of electrophoretically separated cell lysates reacted with antibodies to IFNγ pathway signaling components. MRC5 cells were nucleofected with plasmid harboring full length IE1 (FL), the C-terminal truncation (∆AD) or empty vector (FIN2). At 24 hours after nucleofection, cells were exposed to IFNγ for 30 minutes or left untreated. Cells were solublized and equal amounts of protein from each lysates were subjected to electrophoresis in a denaturing polyacrylamide gel. Proteins were transferred to nitrocellulose sheets and reacted with the indicated antibodies. Anti-FLAG antibody was used to detect the presence of FLAG tagged IE1 proteins and antibody to GAPDH was used to evaluate protein loading.
Figure 3.7: Nuclear translocation of STAT1 is not obstructed by IE1. Immunofluorescence images of IE1 and STAT1 with and without IFNγ treatment. MRC5 cells were nucleofected with the full length IE1 (FL) or empty vector (FIN2). After 24 hours cells were exposed to IFNγ for 30 minutes or left untreated. Cells were fixed in methanol and reacted with mouse monoclonal anti-IE1 antibody (Mab810) and rabbit polyclonal anti STAT1 antibody followed by anti mouse alexafluor 543 and anti rabbit alexafluor 488 as secondary antibodies. STAT1 reactivity is visualized in green (left column), IE1 reactivity is visualized in red (middle column) and merged images are shown in the right column.
Type II interferon-induced binding of STAT1 to GAS elements is reduced in the presence of IE1

As there was no defect in the ability of STAT1 to be activated and translocate to the nucleus in response to IFNγ treatment in IE1 expressing cells, we next determined whether STAT1 molecules in the nucleus were competent to bind the GAS element derived from the promoter of the IRF1 gene by EMSA. The full length IE1, the ΔAD mutant, and empty vector were nucleofected into MRC5 cells. These cells were either treated with IFNγ for 30 minutes or left untreated. Nuclear extracts were prepared from these cells and incubated with a [P]$^{32}$ -labeled 22mer IRF1 GAS element probe. The amount of probe shifted in nuclear extracts from IE1 overexpressing cells is diminished relative to the amount of shifted probe incubated with nuclear extracts from empty vector- and ΔAD-nucleofected cells (Figure 3.8 A). The shifted bands could be supershifted by a STAT1 antibody indicating that the shift was caused by a complex containing STAT1. This indicates that there is reduced binding of STAT1 to target GAS elements in the presence of IE1. We also carried out the EMSA using nuclear extracts from HCMV infected cells at 12 hours of infection. We could detect only a minor STAT1-shifted probe band in HCMV infected cells compared to uninfected cells (Figure 3.8 B). Under these conditions, the difference in the amount of shifted probe between cells ectopically expressing IE1 and cells infected with HCMV is that much higher levels of IE1 are expressed upon
infection (data not shown). Together, these data indicate that expression of HCMV IE1 is sufficient to impair STAT1 binding to GAS elements. These data also imply that the impairment of functional STAT1 dimers binding to GAS elements at 12 hours after infection is a result of IE1 expression.
Figure 3.8: Reduced binding of STAT1 to GAS elements in the presence of IE1. Film image of an Electrophoretic Mobility Shift Assay shows a reduction in the amount of IRF1 GAS element probe that is shifted upon mixture with nuclear extracts from MRC 5 cells ectopically expressing IE1 (A) and HCMV infected cells (B). In A, MRC5 cells were nucleofected with plasmid harboring full length IE1 (FL), the C-terminal truncation (∆AD) or empty vector (FIN2). At 24 hours after nucleofection, cells were exposed to 100U/mL of IFNγ or left untreated. At 30 minutes after treatment, cells were solublized and nuclear extracts were isolated as described in Materials and Methods. Nuclear extracts were mixed with the GAS element of the IRF1 gene with or without addition of unlabelled probe and with or without anti-STAT1 antibody. Resulting complexes were resolved by electrophoresis on a 4.5% native polyacrylamide gel. In a second experiment (B), MRC5 cells were exposed to 3 PFU per cell of the AD169 strain of HCMV or left uninfected. At 12 hours after infection, cells were solubilized and processed as described above. Black arrows indicate the shifted bands and grey arrows indicate the supershifted bands.
IE1 interaction with STAT1

In the EMSA studies, the size of the shifted band in the presence or absence of STAT antibody is the same whether or not IE1 is present. This argues against a scenario in which IE1 acts by directly binding to STAT1-bound GAS elements. In order to test the possibility that IE1 directly interacts with STAT1 to reduce its binding to GAS elements, we carried out co-immunoprecipitation studies. MRC-5 fibroblast cells were exposed to 3 PFU per cell of the AD169 strain of HCMV. At 12 hrs after infection, cells were solubilized and IE1 was isolated using an antibody that recognizes an epitope derived from Exon 2. Under these conditions, we found no evidence of an interaction between STAT1 and IE1 (Figure 3.9A). Similarly, in nucleofected cells overexpressing the full length IE1 or the ΔAD mutant IE1 there was STAT1 failed to co-immunoprecipitate with either the full length or the AD mutant IE1(Figure 3.9B). However we did observe an interaction of STAT2 with the full length IE1 in nucleofected cells and in HCMV infected cells. The ΔAD mutant does not appear to interact with STAT2 under these conditions (Figure 3.9 B).
Figure 3.9. Coimmunoprecipitation studies fail to reveal a stable interaction between IE1 and STAT1. Film images of electrophoretically separated cell lysates and proteins isolated by immunoprecipitation. (A), MRC5 cells were exposed to HCMV AD169 at 3 PFU per cell or left untreated. 12hrs after infection cells were treated with 100U/ml of IFNγ for 30 minutes. The cells were harvested and solubilized. 350μg of total protein from each sample was reacted with anti-IE1 antibody (mouse monoclonal Mab810). The isolated proteins along with the 10μg total lysates were resolved by denaturing PAGE and transferred to a nitrocellulose membrane. The membrane was reacted with antibodies for STAT1, IE1 and STAT2 and the antibody reactive bands were visualized by chemiluminescence. The left panel shows films exposed for 10 seconds and the right panel shows films exposed for 5 minutes. The arrow indicates the position of the immunoprecipitated IE1 (B), MRC5 cells were nucleofected with full length IE1, ΔAD IE1 and FIN2. 48hrs after nucleofection cells were exposed to 100U/ml of IFNγ for 30 minutes. The preparation of cell lysates, immunoprecipitations, PAGE and western hybridization and detection were carried out as above. The left panel shows films exposed for 10 seconds and the right panel shows films exposed for 5 minutes. The short and long arrows indicate the position of the ΔAD IE1 in the total lysate and immunoprecipitates respectively.
**Figure 3.9 Coimmunoprecipitation studies fail to reveal a stable interaction between IE1 and STAT1**
Discussion:

The body’s first line of defense against viruses is IFN. Therefore understanding how viral proteins like HCMV IE1 can disrupt IFN mediated signaling is important for developing new antiviral therapies and vaccines, and may yield new insights into the molecular mechanism of this important defense mechanism.

We screened 80 cDNA clones from our HCMV cDNA library for genes involved in disruption of IFN signaling and identified UL123 as a candidate viral gene. The UL123 gene codes for the HCMV major transcriptional transactivator protein, IE1. In addition to its role as a promiscuous transactivator of viral and cellular genes, IE1 is known to have multiple functions including disruption of ND10 nuclear bodies (123), antagonism of histone deacetylase (HDAC3) (168) and interference with signaling by type I interferons (181). We discovered that IE1 is able to reduce IFNγ-induced upregulation of CIITA and IRF1 transcript levels in human fibroblasts. In order to gain insight into the molecular mechanism by which IE1 disrupts IFN signaling, we set out to map the region of the IE1 protein responsible for this activity. To this end we tested different truncation mutants of IE1 for their ability to reduce IFN induced expression of CIITA in MRC5 cells. We found that the IE1 mutant lacking the 147 C-terminal residues that includes the acidic domain is unable to reduce IFN mediated induction of CIITA expression. In fact this mutant showed slightly enhanced expression of CIITA in response to IFN stimulation as compared to the empty vector control. This suggests that the IE1 protein lacking the AD may actually promote IFN
signaling, and further study of this phenomenon may be of interest. Since IE1 has been reported to be involved in the disruption of PML bodies in the nucleus (3, 123, 272) and PML is important in IFN signaling and antiviral responses against HCMV (241) we investigated whether the AD deletion mutant could disrupt PML bodies. Earlier there have been conflicting reports, one suggesting that the AD is important for PML targeting and disruption of ND10 structures(3) and the other suggesting that the AD is not involved in this function of IE1(272). Yet another study revealed an intermediate role of AD deleted IE1 wherein it could target ND10 structures but could not disrupt them (133). Our results indicate that the AD deletion mutant is able to disrupt PML bodies similar to the full length IE1 protein in MRC5 cells. Our results are in agreement the findings of Wilkinson et.al (272), wherein the loss of the AD does not inhibit the ability of IE1 to target and disperse PML bodies. Although these results to do not preclude a role for PML in IE1-mediated IFN signal disruption, they do demonstrate that this function of IE1 does not involve disruption of PML bodies.

IE1 has been reported to antagonize HDAC activity (168). Although HDAC activity is typically associated with transcriptional repression, it is associated with stimulation of ISGs (171). It is tempting to speculate that antagonism of HDAC activity is by IE1 is linked to disruption of interferon signaling. However, this is difficult to reconcile with the observation that IE1 disrupts STAT1-binding to a GAS element probe used in our EMSA assays, which is not dependent on chromatin structure. We therefore propose that yet an additional function of IE1
accounts for reduced STAT1 binding to its target GAS elements.

To dissect which aspect of the IFNγ signaling pathway was disturbed by IE1, we examined the proteins in the JAK-STAT signal transduction pathway known to be involved in IFNγ mediated signaling. It was revealed that IE1 does not affect total levels of JAK1, JAK2, and STAT1 or the phosphorylation of STAT1 in response to IFNγ treatment. Furthermore the nuclear translocation of STAT1 after IFNγ treatment was not blocked in IE1 overexpressing fibroblasts. Similar results have been reported by Paulus et.al (181) in IE1 expressing fibroblasts treated with Type I interferon. We hypothesized that IE1 may somehow inhibit STAT1 from binding the promoters of IFNγ induced genes. To test this we examined the binding of STAT1 to an IRF1 GAS element in nuclear extracts from IE1 overexpressing cells. We found that there is indeed a reduced binding of STAT1 to GAS elements in extracts from IE1 expressing cells in comparison to empty vector containing cells and this could account for the dampening of IFN induced gene expression in IE1 expressing cells. Furthermore there was no reduction of STAT1 binding to GAS elements in the presence of the AD deleted IE1. In infected cells the binding of STAT1 to the IRF1 GAS element was further reduced. This may be due to the fact that the level of IE1 expression in infected cells at 12hrs of infection was higher than in nucleofected cells. Miller et.al (157) have reported a reduction in STAT1 homodimer binding to GAS elements starting at 12 hrs of infection and continuing upto 72 hrs of infection. Our results suggest that this phenomenon is mediated by IE1. Our data may also explain the defect in
CIITA induction by IFNγ that occurs downstream of STAT1 nuclear translocation, reported by LeRoy et.al (132).

Our findings indicate that the C terminal domain of IE1 comprising of the AD and the chromatin tethering domain can bring about reduced binding of STAT1 to GAS elements in the promoters of IFNγ inducible genes. It is proposed that the AD of IE1 is involved in transcriptional activation based on the acidic nature similar to other acidic transcriptional activators like HSV1 VP16 (55) and yeast Gal4 (80). It is plausible that IE1 may modulate cellular factors that interfere with STAT1 binding to GAS elements. These factors could bring about post translational modification of STAT1 that reduces its affinity to the GAS elements. The only caveat is that there is no direct evidence for the role of the AD of IE1 in transcriptional activation. The chromatin tethering domain is reported to play role in the association of IE1 with metaphase chromatin (127, 200). At present it is not clear how the chromatin tethering domain could disrupt binding of STAT1 to target promoter elements in response to IFNγ.

Although only cells of the immune system including NK cells and T cells produce IFNγ, all nucleated cells in the body have IFNγ receptors and can thus elaborate antiviral responses when stimulated by IFNγ. It is known that during a primary infection CD8 T cells play an important role in controlling the spread of the virus. During this time it would be advantageous for the virus to be able to dampen signaling induced by IFN γ, which is produced abundantly by CD8 T cells. A dampening of IFN γ signaling would allow the infected cells to escape
from the antiviral signaling pathways. Also reduced IFNγ induced expression of MHC class I and MHC class II molecules would downregulate the presentation of viral antigens at the cell surface. This would allow the virus to evade adaptive immune responses. It has been reported that treating monocytes with IFNγ and TNFα makes them differentiate into macrophages that are permissive to HCMV infection. These macrophages seem to be resistant to the antiviral affects of IFNγ (224). It is possible that IE1 confers this resistance to the effects of IFNγ. It would be of interest to overexpress IE1 in macrophages and determine if this leads to a disruption in IFNγ signaling in these cells. Moreover, one of the first genes produced during reactivation is IE1. Under these circumstances the ability of IE1 to downregulate IFNγ mediated signaling would provide a survival advantage to the virus upon reactivation.

To our knowledge this is the first report of the role of HCMV IE1 protein in disruption of Type II interferon signaling. Further study to determine how IE1 brings about reduced type II IFN induced binding of STAT1 to GAS elements is warranted. Also we have identified that the C-terminal domain of IE1 consisting of the acidic domain and the chromatin tethering domain is important for this function. It would be interesting to construct a mutant virus, which is deleted in this C terminal domain of IE1 and study whether it is more sensitive to IFN treatment. This could be a likely vaccine candidate.
CHAPTER 4

Conclusions and Future Directions

The research described in this thesis covered two important aspects of HCMV biology. The first aspect was the investigation of the transcriptional profile of HCMV during a lytic infection. We analyzed the transcriptome of HCMV during lytic infection of cultured human fibroblasts. Generally the study of the genome and transcriptome of the virus has focused on protein coding potential and protein coding transcripts respectively. To determine the protein coding potential various in-silico approaches have been used with a view to identify ORFs that are likely to encode proteins. Based on this HCMV cDNA arrays have been synthesized. HCMV cDNA libraries have been made in the past and the transcripts were characterized by hybridization to the ORF based arrays. Such approaches preclude the identification of non protein coding transcripts. To the best of our knowledge this is the first study wherein an attempt was made to clone and sequence all the transcripts of HCMV during a lytic infection. Since our approach did not introduce a bias towards protein coding transcripts we were able to identify non protein coding transcripts and an abundance of antisense transcripts. Based on the 604 HCMV cDNA clones that we analyzed we found that approximately 45% originated from genomic regions which were predicted to be
noncoding. Also, 55% of the transcripts were completely or partially antisense to known or predicted HCMV genes. These findings raise the important question of the significance of these noncoding and antisense transcripts in HCMV biology. In recent years there has been a burgeoning of research on noncoding RNAs. It is now being increasingly recognized that noncoding RNAs including siRNAs, microRNAs and antisense RNAs play an important role in regulation of gene expression. This would suggest that the noncoding transcripts of HCMV are involved in regulation of viral and possibly host gene expression. Our studies pave the way for more research into the role of noncoding RNAs in lytic and latent HCMV infections.

The second aspect of HCMV biology that was explored in this dissertation was the disruption of IFN signaling by HCMV. We screened the HCMV cDNA library that we generated in order to identify genes that interfere with IFN signaling. We found that the \textit{UL123} gene product that codes for IE1, the immediate early transcription transactivator of HCMV, is able to interfere with both type I and type II IFN signaling. During the course of our studies, IE1 was reported to be involved in disruption of type I IFN signaling. Therefore, we focused on the affect of IE1 on type II IFN signaling. We found that IE1 does not affect the abundance of the classical IFN\(\gamma\) signaling pathway proteins including JAK1, JAK2 and STAT1. IE1 also did not affect activation of STAT1 or nuclear translocation of activated STAT1 in response to IFN\(\gamma\). However we did see a defect in STAT1 binding to target promoter elements in the presence of IE1 as
evidenced by EMSA. At present we do not know the precise mechanism by which IE1 affects STAT1 homodimers to reduce their affinity to target promoter elements. Nevertheless our findings point to a novel mechanism by which IE1 influences IFN signaling. It also highlights the fact that HCMV has devised various mechanisms to target the IFN $\gamma$ signaling pathway. At immediate early times HCMV IE1 disrupts IFN signaling by interfering with STAT1 homodimers binding to promoters, at early times there is a blockade in the phosphorylation of STAT1 in response to IFN$\gamma$ (see below), and starting at early times and continuing to late times there is degradation of JAK1. We infer from these findings that regulation of IFN$\gamma$-induced signaling seems to be important for the survival of the virus during a lytic infection.

**Future Directions:** The research work described in the last two chapters opens the doors to new areas of investigation- (a) to understand the full coding potential of HCMV and (b) to uncover novel mechanisms by which the virus influences cellular antiviral signaling in order to persist for the lifetime of the host. This chapter will discuss some of the possible future directions to continue this work. It should be noted that some of this work is already in progress.

**Determine the full coding potential of HCMV**

We analyzed 604 cDNA clones from our cDNA library and found an abundance of transcripts from genomic regions previously considered to be noncoding and a
high proportion of antisense transcripts. Several known transcripts were not found amongst these 604 clones. This implies that our library does not account for all the sense and antisense transcripts expressed during a lytic infection. In order to understand the full coding potential of HCMV, synthesis of a tiling array consisting of oligonucleotides representing both sense and antisense strands of the entire HCMV genome is warranted. Hybridization of this array with labeled transcripts from infected cells would give a true picture of all the transcribed genes during a lytic infection. Furthermore, the tiling array could be used to identify the viral transcriptome of a latently infected cell, or to analyze the viral transcriptome in specific cell types of individuals with HCMV associated disease. This has far reaching implications in understanding viral gene expression under various conditions that would advance our knowledge of HCMV pathogenesis and may reveal new targets for antiviral drug development.

**Determine biological significance of antisense transcripts of HCMV**

As described in chapter 2, characterization of our HCMV cDNA library revealed that 55% of the clones were partially or completely in antisense orientation to known or predicted HCMV genes. This raises an important question about the significance of these transcripts in the viral life cycle. It is possible that the antisense transcripts are regulatory molecules that influence the stability of the sense transcripts. Testing this hypothesis requires detailed knowledge of the temporal profile of accumulation of S-AS pairs. With this knowledge, it is
possible to design experiments to alter the time or duration of expression one
member of the pair in an infected cell, and analyze the accumulation of the
 corresponding transcript. For example, a northern hybridization based approach
could be used to monitor the expression level of the sense transcript at IE, E and L
times of infection in cells overexpressing the corresponding antisense transcripts
or in which the expression of the AS transcript is reduced or eliminated. The
affect of the antisense transcript on the sense transcript could also be analyzed by
a more sensitive Real Time PCR based approach using specific primers for the
sense and antisense transcripts.

It is also possible that the antisense transcripts influence translation of the sense
 transcripts. To address this possibility, the levels of protein expression from the
sense transcript could be examined by western blot analysis in cells
overexpressing the corresponding antisense. As an alternative to overexpressing
the antisense transcripts, an RNAi based knockdown approach could be used to
partially or completely inhibit the expression of the antisense. This would reveal
whether loss of the antisense transcript has any influence on the sense transcript
or the protein encoded by the sense transcript. Furthermore, the antisense
transcripts could serve as precursors for miRNAs. At least two approaches can be
used to explore this possibility. To study this, the antisense sequences could be
analyzed using bioinformatics tools to identify potential miRNA precursors.
These putative primary transcripts for miRNAs could then be overexpressed in
cells followed by northern analysis to determine whether 21-25 bp miRNAs are
produced.

An alternative approach would be to directly isolate and sequence miRNAs from virus infected cells. Total RNA can be extracted from HCMV infected cells and small RNAs (<200bp) can be isolated using the MirVanaTM miRNA isolation kit (Ambion). To increase the probability of analyzing only mature miRNAs from infected cells, the small RNAs will be further fractionated using the flashPAGE Fractionator and reagents (Ambion). These miRNA fractions can be used to generate cDNA to which adaptors can be added and used to perform massively parallel sequencing or Solexa sequencing. The miRNA sequences obtained can be searched against our HCMV cDNA library to identify all possible virally-derived miRNAs. This analysis will also reveal the miRNAs mapping to the antisense and noncoding transcripts. An additional advantage of the Solexa sequencing approach is that it will simultaneously generate profiles of cellular miRNAs in infected cells. A recent report indicated that the cellular miRNA profile is altered in HCMV infected cells (263). As a different strain of HCMV was used in this study it would be of interest to see how different strains of HCMV affect cellular miRNAs and this information may provide key insights into viral gene regulation.

**Determine the molecular mechanism of IE1 mediated disruption of IFN signaling**

Upon stimulation with IFNγ STAT1 is phosphorylated at tyrosine residue 701 and forms homodimers in parallel orientation through the interaction of the
SH2 domains in the carboxy terminal of the STAT1 molecule (286). These STAT1 homodimers interact with importin-α5 through an NLS that includes Leucine 407 and are translocated to the nucleus (151). Here they bind specific GAS elements in the promoters of IFN induced genes and activate their transcription. In IE1 expressing cells there was no defect in tyrosine phosphorylation of STAT1 or translocation of the STAT1 homodimers to the nucleus. However we did find that inside the nucleus the ability of STAT1 homodimers to bind to GAS elements is reduced in the presence of IE1. So how does IE1 bring about reduced binding of STAT1 to GAS elements?

One possibility is that IE1 directly interacts with STAT1 and sequesters it away from the GAS elements. We studied the association of IE1 with STAT1 by examining the co-immunoprecipitation of STAT1 with IE1. We found a very weak association of STAT1 with IE1. Based on other studies during the course of this work, it is very likely that this represents nonspecific binding of STAT1 to protein A beads rather than a specific IE1-STAT1 association. If it is a specific association, we can conclude that IE1 binds to only a fraction of the STAT1 in the nucleus, or binds in a transient manner that is difficult to capture by this method. In any event, it does not seem likely that sequestering such a small proportion of STAT1 could account for the large reduction in GAS element binding seen in the presence of IE1.
An alternative hypothesis could be that IE1 activates factors that inhibit STAT1 homodimer binding to GAS elements. It is well known that PIAS1 and PIASxα inhibit STAT1 binding to target promoter elements (143, 202). IE1 is a transcriptional transactivator which could activate the transcription of PIAS1 or PIASxα. The increased levels of PIAS1 or PIASxα would bring about a reduction in STAT1 binding to GAS elements. To determine whether this hypothesis is valid we could compare the protein levels of PIAS1 and PIASxα in cells expressing IE1 and cells carrying the empty vector. Furthermore knocking down expression of PIAS1 and PIASxα by RNAi in IE1 expressing cells should restore the ability of STAT1 to bind GAS elements to the levels in empty vector carrying cells. One of the probable modes of action of the PIAS proteins that reduces STAT1 binding to target promoters is by enhancing SUMO conjugation to STAT1 (252). Although there are reports that SUMO modification of STAT1 by PIAS proteins does not affect its ability to bind target promoters (202). The contradictory nature of these reports may be due to the different cell lines used in the different studies. Therefore it may be worthwhile to examine whether there is increased sumoylation of STAT1 in the presence of IE1 in our system (human fibroblasts). It is also possible that some as yet unknown modification of STAT1 is brought about by IE1. A 2D gel followed by mass spectrometry approach could be used to determine whether IE1 indeed brings about a post translation modification of STAT1. For this purpose the full length IE1, the AD mutant and the empty vector could be nucleofected into fibroblasts followed by IFN γ
treatment. STAT1 could be immunoprecipitated from these cells and subjected to
2D gel electrophoresis to separate the various modified forms of STAT1. If
different forms of STAT1 are seen on the 2D gel in IE1 expressing cells relative
to empty vector containing cells and AD mutant it would indicate that IE1 does
affect posttranslational modification of STAT1. Mass spectrometric analysis of
the spots on the 2D gel that correspond to forms of STAT1 that are different
between IE1 expressing cells and empty vector or AD mutant would reveal the
specific modification of STAT1 affected by IE1.

Yet another possibility is that IE1 activates a phosphatase like TC45 that
dephosphorylates STAT1 in the nucleus leading to inhibition of DNA binding.
However we did not see any difference in the phosphorylation status of STAT1 in
IE1 expressing cells as compared to empty vector carrying cells as evidenced by
western hybridization. It remains likely that at 30 mins of IFN γ treatment when
there is maximal phosphorylation of STAT1, any enhanced dephosphorylation
cannot be detected by western analysis. In order to dissect these processes it may
be necessary to treat the cells with staurosporine (a kinase inhibitor) after 10
minutes of IFNγ treatment and harvest the cells after 30 minutes of IFNγ
treatment. In the absence of kinase activity from 10 minutes to 30 minutes of
IFNγ treatment it would be clear if there is enhanced dephosphorylation of
STAT1 in IE1 expressing cells compared to empty vector carrying cells.

Another important aspect of the STAT1 dimer that is important for DNA binding
is its conformation. There is increasing evidence for the existence of STAT1 dimers both in IFN stimulated and unstimulated cells. In unstimulated cells the bulk of the STAT1 dimers are in antiparallel orientation wherein the dimer is formed via interaction of the N terminal domain and the DNA binding domain of the STAT1 monomers. In IFN stimulated cells there is a predominance of parallel dimers wherein the SH2 domains of the monomers interact to form the dimer. In the parallel conformation the DNA binding ability of STAT1 is 200 fold higher than in the antiparallel conformation (267). It is possible that upon translocating to the nucleus of IE1 expressing cells the parallel STAT1 dimer conformation is disrupted rendering it less able or unable to bind DNA. It would be interesting to study the conformation of STAT1 dimers from the nuclei of IE1 expressing cells and empty vector nucleofected cells to determine if there is a difference in the conformation of the STAT1 dimer in the presence of IE1. Disruption of STAT1 dimers would also induce enhanced nuclear export of STAT1. This could be examined by monitoring the levels of STAT1 protein in the nucleus and cytoplasm of IE1 expressing cells as compared to empty vector nucleofected cells. It should be noted that our immunofluorescence data does not reveal increased cytoplasmic accumulation of STAT1 in IFNγ treated IE1 expressing cells in comparison to empty vector carrying cells. However it is possible that the immunofluorescence technique is not sensitive enough to reveal a small increase in cytoplasmic accumulation of STAT1. A western blot approach may be more appropriate for determine if there is indeed increase nuclear export of STAT1.
resulting in enhanced cytoplasmic accumulation. Deciphering the precise molecular mechanism used by IE1 to bring about reduced binding of STAT1 to target GAS elements may reveal a new level of regulation of IFN signaling. This may be yet another example of how the study of viruses enhances our knowledge of fundamental cellular processes.

**Construction of a C-terminal Domain deleted IE1 mutant HCMV**

We identified that the C-terminal domain of IE1 containing the acidic domain and the chromatin tethering domain is responsible for disruption of IFNγ mediated signaling. It would be of interest to construct a mutant virus deleted in just this C-terminal domain of IE1. For this purpose a plasmid construct has been made containing a kanamycin resistance gene flanked on the 5’ by the sequence 500bp upstream of the C-terminal domain and on the 3’ by the sequence 500 bp down stream of the C-terminal domain (which includes 3’ UTR of IE1 and intronic regions). This will be used for homologous recombination with an AD169 BAC (280). The resultant mutant BAC will have the IE1 gene with a kanamycin cassette in place of the C-terminal domain. This mutant will be transfected into fibroblasts to generate mutant virus. Studies with this mutant virus can inform us about the significance of the C-terminal domain during a viral infection. We would expect this mutant virus to be more susceptible to IFN than the wild type AD169 virus. If so, we may predict that this mutant may be suitable for testing as a vaccine candidate.
Screening of HCMV cDNA library for other genes involved in disruption of IFN signaling

We screened 80 HCMV cDNA clones from our library and identified IE1 as a candidate gene involved in disruption of IFN signaling. Our results indicated that IE1 disrupts IFN signaling by a novel mechanism, which does not involve JAK1 degradation. This indicates that other viral genes are responsible for degradation of JAK1. Screening of additional cDNA clones using the 2C4 reporter cell line or other reporter cell lines could result in the identification of these genes. Additional screening could also identify other HCMV genes, which play a role in disruption of IFN signaling. This research may not only shed new light on HCMV viral biology and evolution, but may also reveal novel insights into the components and regulation of the cellular antiviral response.

Examining reduced STAT1 phosphorylation at early times of HCMV infection

Using the 2C4 reporter cell line system described in chapter 3, we identified another HCMV cDNA clone that could reduce IFN stimulated expression of CD2 on 2C4 cells (Figure 4.1). This cDNA clone encoded the singly spliced US3 (ssUS3) transcript, which is expressed at immediate early times of infection. As a preliminary study we wanted to determine whether loss of the US3 gene affects IFN signaling in HCMV infected cells. To this end we studied the RV7186 mutant of HCMV in comparison to the AD169 wild type strain. The genomic
region from IRS1- US11 is deleted in the RV7186 mutant. Specifically RV7186 is missing IRS1, US1, US2, US3, US4, US5, US6, US7, US8, US9, US10 and US11. HFF cells were plated at a density of $1.5 \times 10^5$ cells per well of a six well plate. These cells were exposed to 1 PFU per cell of AD169 or RV7186 virus. Infections were allowed for 12hrs, 30hrs, 48hrs and 72hrs. At these time points the cells were exposed to 100U/ml of IFN $\gamma$ for 10 minutes, 30 minutes or 2hrs as indicated (Figure 4.2). After the IFN treatment cells were harvested and lysed by sonication. Equal amounts of total protein were electrophoresed on a polyacrylamide gel and transferred to a nitrocellulose membrane. The membrane was probed with anti-phospho STAT1 antibody and anti-GAPDH antibody (loading control). The results indicate that there is no difference in STAT1 phosphorylation between RV7186-infected cells and AD169-infected cells. We conclude that at least in human fibroblasts, ssUS3 does not inhibit STAT1 phosphorylation in cultured cells infected with HCMV. In the course of these studies, we observed an unexpected pattern of STAT1 phosphorylation in HCMV infected cells. It is evident from Figure. 4.3 that at early times (30hrs and 48hrs) of infection there is a defect in STAT1 phosphorylation. However at immediate early and late times (12hrs and 72 hrs respectively) STAT 1 phosphorylation is still observed, though the time course is altered relative to uninfected cells. It has been reported that JAK1 degradation takes place in infected cells starting at early times of infection and is completely degraded by late times (157). This could explain the reduced STAT1 phosphorylation at early times. However when JAK1
is almost absent at late times STAT1 phosphorylation is still observed, although the time course of phosphorylation is different than that of uninfected cells. Specifically, there appears to be a delay at 10 minutes after IFN treatment relevant to uninfected cells. At 30 minutes after IFN treatment the levels of phosphorylated STAT1 appear equivalent in infected and uninfected cells. Last, downregulation of signaling appears to occur more rapidly in infected cells such that by 2 hours after IFN treatment, less phosphorylated STAT1 is observed relative to uninfected cells. One possible explanation is that while JAK1 accumulation is substantially reduced in infected cells, sufficient levels of JAK1 is present to allow STAT1 phosphorylation at this time. Also the ability of JAK2 to phosphorylate STAT1 could compensate for the loss of JAK1. An alternative possibility is that a mechanism other than JAK1 degradation is responsible for reduced STAT1 phosphorylation in HCMV infected cells. It is possible that HCMV encodes a gene expressed strictly at early times that inhibits STAT1 phosphorylation. On the other hand HCMV could activate a cellular inhibitor of STAT1 phosphorylation. If so, why is JAK1 degraded in infected cells? One possibility is that the strategy of HCMV to target JAK1 degradation is related to inhibition of other cytokine signaling pathways that utilize JAK1 as a signaling intermediate. Further study is warranted to examine this phenomenon and identify the viral gene/s responsible for it. To identify the viral genes that interfere with STAT1 phosphorylation individual cDNA clones from our cDNA library could be transfected into cells, followed by IFN γ treatment and a cell based
ELISA approach to determine STAT1 phosphorylation status. Those clones which show reduced STAT1 phosphorylation in response to IFN γ as compared to empty vector transfected control should be further analyzed.
Figure 4.1: Reduced CD2 cell surface levels in cells expressing ssUS3. Histograms of cell surface CD2 levels in 2C4 cells transfected with pIE936 harboring the gene encoding the ssUS3 transcript and empty vector (FIN2) control. The thin grey lines show the population of cells transfected with pIE936, and the bold black lines are cells transfected with the empty vector. At 24 h after transfection cells were exposed to 100U/mL of IFNβ (middle panel) or IFNγ (bottom panel). At 24 h after IFN treatment, cells were dislodged from the plates, reacted with FITC-conjugated antibody to CD2, and analyzed by flow cytometry.
**Figure 4.2: STAT1 phosphorylation at immediate early and late times of infection.** Western blot for phospho-STAT1 in cells infected with AD169 strain of HCMV or the RV7186 strain of HCMV or mock infected cells. At 12 hrs (top panel) or 72 hrs (bottom panel) after infection cells were treated with 100U/ml of IFNγ for 10 minutes, 30 minutes or 2hrs as indicated. The same blots were probed with anti GAPDH antibody as a loading control.
Figure 4.3: STAT1 phosphorylation at early times of infection. Western blot for phospho-STAT1 in cells infected with AD169 strain of HCMV or the RV7186 strain of HCMV or mock infected cells. At 30 hrs (top panel) or 48 hrs (bottom panel) of infection, cells were treated with 100U/ml of IFNg for 30 minutes or 2hrs as indicated. The same blots were probed with anti GAPDH antibody as a loading control.
Concluding Remarks

HCMV is a supreme pathogen that has the ability to persist for the lifetime of the host. Upon primary infection it is able to modulate the antiviral defense mechanisms of the host and establish latency. It has managed to maintain a delicate balance with host cell processes such that multiple reactivations of the virus during the life of the host on the one hand does not lead to elimination of the virus by host immune responses and, on the other hand it does not cause serious morbidity of the host. The ability of this virus to achieve such equilibrium, in order to coexist with its host is the primary reason for the success of this obligate parasite. The research presented in this thesis reveals two novel strategies used by HCMV that may contribute to the ability of this virus to persist in the host. The discovery of the abundance of noncoding transcripts and antisense transcripts in the HCMV transcriptome suggests that viral gene expression and regulation is more sophisticated than was earlier believed. These transcripts may also function to regulate host gene expression to the advantage of the virus. Furthermore, the discovery of the role of C-terminal domain of the HCMV immediate early transactivator IE1 in disruption of IFN signaling suggests that it is important for the virus to block IFN signaling very early in infection. The C-terminal domain of IE1 contains its activation domain implicated in transcriptional transactivation. This may suggest that this C-terminal domain of IE1 evolved to carry out the function of transcriptional transactivation of viral genes and attenuating IFN mediated signaling as both of these processes are important for the survival of the
virus during a primary infection and during reactivation from latency. The research presented in this thesis also points to a novel mechanism of disruption of IFN signaling that preserves the initial events in signaling but specifically interferes with transcription factor binding to target DNA. In addition to this attenuation of IFN signaling by IE1 at immediate early times of infection, HCMV blocks IFN signaling at early and late times by targeting various components of the IFN signaling pathway. It can be inferred from these findings that such a temporally regulated interference with IFN signaling is an important aspect of HCMV biology that allows it to co-exist with its host and establish life long persistence.
APPENDIX

DETAILED PROTOCOLS

**Nucleofection**: For all nucleofections MRC5 cells between passage 2 to passage 10 were used. Nucleofections were carried out within 3 to 5 days of subculturing the cells. To harvest cells for nucleofection 0.0125% trypsin was used unlike the 0.025% trypsin used for regular cell culture. The trypsin was inactivated by adding B medium. The cells in B media were transferred to 50ml falcon tubes and centrifuged at 500rpm for 10 minutes. The supernatant was decanted and excess media was removed with a 1ml pipette. The cell pellet was resuspended in basic nucleofector solution for primary fibroblasts. 100µl of the nucleofector solution was used for each nucleofection. A minimum of 1 million cells were used per nucleofection. For every million cells 2µg of plasmid DNA was used. The U23 program was used for nucleofection. Immediately after nucleofection warm B media was added to the cells in the nucleofection cuvette and the cells along with the media were removed (using the pipettes provided with the kit) and transferred to a falcon tube containing warm B media. For a six well plate format (used for real time pcr) 1.5 million cells were used per nucleofection with 3µg of plasmid.
DNA. After nucleofection the cells were evenly distributed into 3 or 4 wells of a six well plate. For a T-75 format (used for IPs) 3 million cells were used with 6ug of plasmid DNA and plated in one T-75 flask. After nucleofection the cells were maintained in 2ml of media per well of a six well plate or 10ml of media per T-75 flask. For all real time pcr experiments the cells were harvested in trizol, 24hrs after nucleofection. For the IP experiments, cells were harvested by scraping the cells in cold versene at 48 or 72 hrs after nucleofection.

**Immunofluorescence:** Cells were nucleofected as described above for the 6 well plate format. However the cells were distributed evenly into 12 wells of a 24 well plate. Prior to plating the cells, one poly-L-lysne coated cover slip was placed in each well of the 24 well plate. 24hrs after nucleofection and plating cells were rinsed twice with warm B media. Following this the cells were fixed in methanol for 2hrs on ice. The methanol was removed and the coverslips in the wells were allowed to air dry for one hour. Following this the cells were rehydrated in PBS for 10 minutes. This was followed by blocking for 1 hour in blocking solution (10% human serum+ 1% BSA in PBS) at room temperature. Primary antibody was diluted in blocking solution and added to the wells and incubated overnight at 4°C. The coverslips were rinsed 3x 20 minutes with PBS at room temperature. This was followed by addition of 1: 1000 dilution of secondary antibody and incubation for 1 hour at 4°C. Excess of secondary antibody was rinsed off by 3x
20 minute washes in PBS. The coverslips were then allowed to air dry for 20 minutes. They were mounted on glass slides using the Prolong antifade kit (Molecular probes). Microscopy was done using a Zeiss LSM 510 instrument. For STAT1 and IE1 double staining 1:1000 dilution of Mab810 (mouse monoclonal IE1 antibody (Millipore)) and 1:100 dilution of rabbit polyclonal STAT1 (Santa Cruz) antibody were simultaneously added to each of the wells and incubated overnight at 4°C. The secondary antibodies were also added simultaneously; 1:1000 dilution, of anti mouse alexafluor 543 and anti rabbit alexafluor 488. For PML and IE1 double staining, the cells on the coverslips were first exposed to 1:1000 dilution of Mab810 (IE1 antibody) overnight at 4°C , rinsed as described above and reacted with 1:1000 dilution of anti-mouse alexafluor 488. Following this the cells were rinsed as before and exposed to 1:100 dilution of rhodamine conjugated PGM3 (PML antibody) and incubated overnight at 4°C. The cells were rinsed again as described earlier, air dried and mounted onto glass slides.

Electrophoretic mobility Shift Assay (EMSA): MRC5 cells were plated at a density of 1.5 million cells per 75cm² flask post nucleofection. (Nucleofecting 3 million cells would result in 1.5 million cells post nucleofection, assuming 50% survival of cells). At 48 hrs after nucleofection the cells were exposed to 100U/ml IFNγ for 30 minutes or left untreated. Cells were washed with PBS, harvested, and nuclear extracts were prepared using Nuclear extract kit (Activ
Motif). The nuclear extraction protocol provided with the kit was followed with the addition of an extra wash of the nuclear pellet with hypotonic buffer after removal of the cytoplasmic fraction.

A 22 mer IRF1 GAS element (5’GATCGATTTCCCCGAAATCATG 3’) was end labeled with $\gamma [P]^{32}$ ATP using T4 polynucleotide kinase enzyme (Invitrogen) according to the manufacturer’s recommendations. The reaction mix contained 100ng ds IRF1GAS element, 5µl forward reaction buffer, 1µl T4 polynucleotide kinase, 5µl $\gamma [P]^{32}$ ATP (250uC stock) and 13µl sterile water. After the labeling reaction was stopped, 65µl of TEN buffer was added to this 25µl reaction mix. This 80µl mix containing the labeled probe was purified using the NucTrap Probe purification column (Stratagene) according to the manufacturer’s instructions. The amount of radioactivity incorporated was measured by adding 1µl of the labeled probe to 25µl of microscint scintillation fluid and measured in the scintillation counter using protocol 25. Readings between 5000 and 20,000cpm were considered optimal for use as probe in EMSA.

4µg of nuclear extract was incubated with 1x binding buffer (Roche), 0.5 µg poly d I:C (Roche), 0.1 µg poly L-lysine and 2ng of labeled probe for 15minutes at room temperature. For competition assays 200 fold excess of unlabeled probe was added to the reaction mix. For supershifts 5µg of rabbit polyclonal anti STAT1 antibody (Santacruz) was added to the reaction and incubated for 15 minutes at room temperature prior to adding the labeled probe. The reactions were resolved on a 4.5% native polyacrylamide gel (40% acrylamide/bisacrylamide -2.25mL, 10x
Chromatin Immunoprecipitation Protocol

Treat 4x 10^6 cells in one T150 tissue culture flask. (Treat means to infect, treat with interferon, transfect etc). Cross link by adding formaldehyde directly to culture medium to a final concentration of 1% (Add 270µl of 37% formaldehyde to 10 ml growth medium on plate/flask). Incubate at RT for 10 minutes with shaking. Add 0.125M Glycine, incubate for 5 mins at RT with shaking. Aspirate medium removing as much medium as possible and wash 2 x with cold TBS (20mM Tris pH7.4, 150 mM NaCl). Harvest cells by scraping in 10 ml SDS Buffer (50mM Tris Ph 8.1, 0.5% SDS, 100mM NaCl, 5mM EDTA, protease inhibitor). Centrifuge at 2000 rpm for 4 minutes at 4°C. Resuspend cell pellet in 800µl IP buffer (100mM Tris Ph 8.6, 0.3% SDS, 1.7% Triton X-100, 5mM EDTA, protease inhibitors). Distribute 200µl each to 4 flat bottomed 2ml eppendorf tubes. Disrupt cells and shear chromatin by sonication for 15 minutes (30 s on, 30 s off) at setting 5 in Misonix cuphorn sonicator to yield DNA fragments with a bulk size of 100-500bp. Dilute the lysate 10 fold to 2ml (in each tube) with IP buffer. Preclear the lysate by adding 40µl of blocked protein A beads (50% slurry protein A Sepharose, 0.5mg/ml fatty acid free BSA, 0.2mg/mL salmon sperm DNA in TE). Immunoprecipitate overnight with the specific antibody at 4°C. Recover immune complexes by incubating with 40µl of preblocked Protein A beads at 4°C for 2hrs. Wash and elute beads as follows-
wash once with 300µl of mixed micelle buffer (20mM Tris-HCl pH8.1, 150mM NaCl, 5mM EDTA, 5%(w/v) sucrose, 0.2% SDS, 0.2% Triton-X-100), Buffer 250 (50mM HEPES pH7.5, 250mM NaCl, 1mM EDTA, 0.1% deoxycholine, 0.2% Triton X-100), LiCl detergent buffer (10mM TrisHCl pH8.0, 250mM LiCl, 1mM EDTA, 0.5% deoxycholine, 0.25% NP-40) and Tris-EDTA (pH7.5). Elute with 250µl of elution buffer (1% SDS, 0.1M NaHCO3). Repeat this process and combine eluates to make a total of 500µl. Add 20µl 5M NaCl to each tube and reverse crosslink at 65°C for 4hrs. Add 10µl 0.5M EDTA, 20µl Tris pH6.8, 1µl of 20mg/ml proteinase K and incubate at 45°C for 2hrs. Extract the eluted DNA with phenol chloroform and precipitate with ethanol. Wash with 70% ethanol and air dry. Resuspend in 50µl sterile distilled water. Use 5µl of the eluted DNA for each real time PCR reaction.
LIST OF REFERENCES


15. **Bentz, G. L., Yurochko, A.D.** 2008. Human CMV infection of endothelial cells induces an angiogenic response through viral binding to EGF receptor and beta 1 and beta 3 intergrins. PNAS **105**:5531-36.


48. Cheung, A. K. 1990. The BamHI J fragment (0.706 to 0.737 map units) of pseudorabies virus is transcriptionally active during viral replication. J Virol 64:977-83.


210. Segikawa, I., Nawata, M., Seta, N., Yamada, M., Iida, N., Hashimoto,


231. **Staeheli, P., O. Haller, W. Boll, J. Lindenmann, and C. Weissmann.** 1986. Mx protein: constitutive expression in 3T3 cells transformed with


cytopathogenicity of cytomegalovirus by propagation in endothelial cells. Arch Virol 117:143-64.


