MOLECULAR INVESTIGATION OF THE MCMV 7.2 KB lncRNA: A VIRAL PERSISTENCE FACTOR

by

TONI MARIE SCHWARZ

B.S., B.A., University of Minnesota, 2006

A thesis submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

Microbiology

2014

This thesis for the Doctor of Philosophy degree by

Toni Marie Schwarz

has been approved for the

Microbiology Program

by

Linda F. van Dyk, Chair

Caroline A. Kulesza, Advisor

David J. Barton

Jerome Schaack

Richard E. Davis

Date 6/10/14

Schwarz, Toni Marie (Ph.D., Microbiology)

Molecular Investigation of the MCMV 7.2 kb lncRNA: A Viral Persistence Factor.

Thesis directed by Professor Caroline A. Kulesza

ABSTRACT

HCMV encodes a stable 5 kb RNA (RNA5.0) of unknown function that is conserved across cytomegalovirus species. *In vivo* studies of the MCMV orthologue, a 7.2 kb RNA (RNA7.2), demonstrated that viruses that do not express the RNA fail to establish efficient persistent replication in the salivary glands of mice. Current analysis demonstrates that RNA7.2 is expressed with late viral gene kinetics during productive infection of mouse fibroblasts. The termini of the precursor RNA that is processed to produce the intron were identified and we demonstrate that the m106 open reading frame, which resides on the spliced mRNA derived from precursor processing, can be translated during infection. Mapping the 5' end of the primary transcript revealed minimal promoter elements located upstream that contribute to transcript expression. Analysis of recombinant viruses with deletions in the putative promoter elements, however, revealed these elements exert only minor effects on intron expression and viral persistence *in vivo*. Low transcriptional output by the putative promoter element(s) is compensated by the long half-life of RNA7.2 of approximately 28.8 hours. This extraordinarily long half-life is caused by RNA sequence elements that reside primarily within the 3' end of the processed intron. These sequence elements are postulated to prevent debranching of a RNA7.2 intron lariat structure and subsequent exonuclease digestion. Our data indicates that RNA7.2 is protected from debranching and 5' to 3' exonuclease degradation, but

future studies will need to validate the hypothesized lariat conformation. Stability analysis of the HCMV RNA5.0 also revealed a long half-life of greater than 32 hours. Detailed analysis of viral spread prior to the establishment of persistence also showed that the intron is not likely required for efficient spread to the salivary gland, but rather enhances persistent replication in this tissue site. The data presented in this dissertation provides a comprehensive transcriptional analysis of the MCMV RNA7.2 locus. Our studies indicate that both RNA7.2 and RNA5.0 are extremely long-lived RNAs, a feature which is likely to be important in their role promoting viral persistent replication.

The form and content of this abstract are approved. I recommend its publication.

Approved: Caroline A. Kulesza

ACKNOWLEDGEMENTS

This dissertation would not have been possible without the patience, unconditional support, and encouragement to discover from my advisor Caroline Kulesza. My dissertation committee members also deserve recognition for their support, mentorship, and constructive criticism throughout these past six years. Overall, this has been an exceptional training opportunity allowing me to mature not only scientifically but also personally. I hope the work presented in this dissertation not only satisfies the requirements for a doctoral degree in microbiology, but also is truly deserving of it.

TABLE OF CONTENTS

CHAPTER

LIST OF TABLES

TABLE

LIST OF FIGURES

FIGURE

CHAPTER I

INTRODUCTION

THE HERPESVIRIDAE FAMILY[1]

The Herpesviridae family represents a ubiquitous group of viruses that infect a wide range of taxa. Since these viruses have adapted to infect a diverse array of eukaryotic organisms, it is speculated that the amount that exist in nature far outnumber those that have been identified and investigated. Current investigations of herpesviruses focus on those that infect higher order taxa including mammals, birds, and reptiles. However, in order to be classified as a Herpesvirus, all viruses must meet specific structural and biological criteria [1].

Evidence for a common origin between distantly related herpesviruses is observed by a shared virion architecture [2]. Each virion is composed of four layers; the core, the capsid, the tegument, and the envelope (Figure 1). At the center, or the core, of all herpesvirus particles is the linear dsDNA genome that varies in size from 120 kb to 240 kb depending on the herpesvirus species [2]. The viral genome is encased by the capsid that is composed of a specific ratio of pentameric and hexameric capsomeres and measures 100nm to 130nm in diameter depending on the size of the viral genome [1]. Together, the viral genome and capsid are called the nucleocapsid [3]. This nucleocapsid is surrounded by a protein-tegument layer and enveloped [4]. Viral glycoproteins are

[1] The work presented in this chapter has been accepted for publication and is used with permission. Schwarz, TM and Kulesza, CA. Long non-coding RNAs expressed during Human Cytomegalovirus Infections. A review. *Future Virology*. In press, June 2014.

embedded within this lipid bilayer envelope and are necessary for attachment and entry into a naïve host cell [5]. In total, the Herpesvirus particle may range in size from 120nm to 300nm depending on the genome size, the thickness of the tegument layer, and the integrity of the envelope [1].

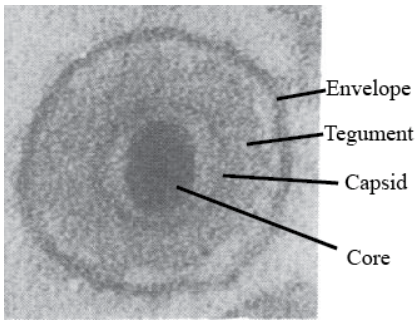


Figure 1. Representative Structure of a Herpesvirus Virion.
Electron micrograph of HHV-7 isolated from human CD4+ T cells [6]. The envelope, tegument, capsid and core are indicated.

The Herpesvirus Lifecycle

In general, Herpesvirus infection of a permissive cell results in either a lytic or latent infection program [1]. During lytic infection, the virus proceeds through the entire replication cycle and may result in the destruction of the host cell. In contrast to lytic replication, the viral replication cycle is halted during latent infection with limited viral gene expression occurring that cannot support viral genome replication [1]. Active viral replication and production of infectious virus during the lytic phase leads to the range of disease outcomes associated with herpesvirus infection. Transmission of herpesviruses also occurs primarily during the lytic phase of replication when infectious virus is generated and secreted via mucosal secretions to an unsuspecting host [1]. Few herpesviruses can, however, be transmitted during latency [7].

Lytic Infection

Productive infection of a Herpesvirus is initiated after the nucleocapsid traffics to the nuclear membrane of the host cell and expels the viral DNA inside the nucleus [8]. Viral gene expression then commences in an ordered cascade of immediate early (IE), early (E), and late (L) genes [9]. IE genes are expressed first and do not require prior viral protein expression for their transcription [9]. This class of genes is largely required for creating an environment conducive for viral replication to occur [1]. Included in this class of genes are transcriptional transactivators necessary for the induction of E and L gene expression. The E class of genes primarily encodes the viral proteins and enzymes critical for viral replication to commence. E genes also encode transactivators needed for the induction L genes. L genes are not only reliant on IE and E transactivators for their transcription, but also require viral DNA replication to occur before they can be expressed [9]. The L class of genes encode proteins essential for assembling the virion and egress from the host cell [1].

Although the exact array of enzymes Herpesviruses encode may vary based on the subfamily, all members of this family contain enzymes that are involved in nucleic acid metabolism, DNA synthesis, and the processing of proteins [1]. These virally encoded enzymes are used for viral genomic replication that occurs exclusively in the nucleus of infected host cells. This viral process takes over the host cell nucleus in distinct replication compartments created by the viral replication machinery [10]. Replication results in multi-genomic length concatemers that are cleaved at specific sites within the genome before packaging into the capsid [1]. After the nucleocapsid is fully assembled

within the nucleus of the infected cell, it traffics through the cytoplasm where the tegument and envelope layers are acquired [11].

Latent Infection

During latency, the lytic transcription program is largely suppressed and only a small subset of viral genes are expressed [1]. Although the purpose of these genes is not fully understood in each Herpesvirus examined, evidence suggests that some may promote the maintenance of the latent state of the circular viral genome or they are critical for reactivation. Progeny virus is not generated during latency and the virus is considered to be in a quiescent state of infection. Consequently, the viral genome is maintained in the host cell in the absence of productive infection. Virus can, however, reactivate from latency after receiving certain stimuli from the host cell allowing it to undergo a lytic replication state [1]. The exact stimuli that cause reactivation may vary between different herpesviruses.

The Herpesviridae Subfamilies

The Herpesviridae family contains a diverse range of viruses that can be distinguished by their cellular tropism, host range, disease outcomes, genomic sequence, and biological properties. These distinguishing features have caused the creation of three different Herpesvirus subfamilies in order to further classify these viruses; the alphaherpesviruses, betaherpesviruses, and gammaherpesviruses.

Alphaherpesviruses are characterized by their ability to latently infect sensory neuronal cells and reactivate to cause lesions at or near the site of initial infection [1]. Alphaherpesvirus are also known to have a variable host range and short replication

cycles. Commonly known alphaherpesviruses are Herpes Simplex 1 and 2 (HSV-1 and HSV-2) as well as Varicella Zoster Virus (VZV) [1]. Betaherpesviruses are known for their slow reproductive cycle, species specificity, and opportunistic disease outcomes [1, 7]. Betaherpesviruses establish latency within glandular tissue of different organs where viral transmission can also occur. Sequencing analyses have demonstrated that betaherpesviruses have the largest genomes among the herpesviruses making them among the largest viruses known to infect and cause disease in humans [2, 12]. The most commonly researched betaherpesvirus is Cytomegalovirus (CMV). Lastly, the gammaherpesviruses are known for their oncogenic potential resulting in carcinomas, lymphomas, and sarcomas [1]. These viruses have a highly restricted host cell range and establish latency within lymphocytes. Gammaherpesviruses that are a major concern worldwide include Epstein Barr Virus (EBV) and Kaposi Sarcoma Virus (KSHV) [1].

THE BETAHERPESVIRUSES

The betaherpesvirus subfamily consists of four genera: *Roseolavirus*, *Muromegalovirus*, *Proboscivirus*, and *Cytomegalovirus* [1]. Only two genera include viruses that infect humans, Roseolaviruses and Cytomegaloviruses. Cytomegaloviruses are the prototypic betaherpesvirus since their characteristic cellular enlargements were first observed over a century ago in the kidneys of a stillborn infant [13]. Cytomegaloviruses have since been classified as an opportunistic pathogen associated with an array of immunological disorders stemming from premature birth to HIV/AIDS and even bone marrow allograft recipients [14-18]. The strict species specificity, tropism, and slow growth in culture differentiate *Cytomegalovirus* from the other herpesviruses. Cytomegaloviruses are widespread and have co-evolved with their hosts causing CMV

speciation [19]. The co-evolution of CMV with its respective host has allowed it to efficiently adapt to long-term infection by encoding an arsenal of genes to subvert clearance by the host's immune system [19]. This highly specific adaptation to the host allows CMV to persist throughout the host's lifetime.

HCMV Virion Structure

The HCMV genome is the largest of all herpesviruses with a high G+C content of nearly 60% and approximately 240 kb of genomic sequence that encodes for over 200 viral proteins [19]. This genome is encased by a capsid structure that is composed of four HCMV proteins; the minor capsid protein (mCP), pUL46, pUL48.5, and the major capsid protein encoded by UL85 and UL86 [19, 20]. Together, these capsid proteins form the common, icosohedral structure that all herpesviruses posses. Surrounding the nucleocapsid is the poorly defined tegument layer. Several HCMV proteins within this layer are known to play critical roles during the early stages of viral infection. Viral RNAs have also been detected in preparations of highly purified, infectious HCMV particles, however, the significance of these RNAs packaged into the virion remains unknown [21]. The outer layer of the HCMV particle consists of a phospholipid envelope that contains 6 virus-encoded glycoproteins including gpUL55, gpUL73, gpUL74, gpUL75, UL100, and gpUL115 [20]. These glycoproteins play essential roles in virus entry into host cells, cell-to-cell spread, and virion maturation [7, 19].

Once the HCMV genome is expelled into the nucleus of an infected cell, the ordered cascade of viral gene expression commences. Within 1 hour post infection, the major IE (MIE) genes UL122/123 (IE1 and IE2) are transcribed without *de novo* protein synthesis [20]. Instead, MIE gene expression is activated by cellular transcription factors

and tegument proteins (pp71 encoded by UL82) that recognize the CMV major immediate early promoter (MIEP) [19, 20]. These MIE genes are critical for regulating host cell gene expression and also act as transactivators for subsequent E and L gene expression [7, 19]. Expression of E genes is reliant on IE gene activation and this class of genes can be divided into two subclasses: the E genes (β1) and the E-L (β2) genes according to their time of expression [7, 19]. Neither β1 nor β2 E genes are reliant on viral DNA replication. The E class of genes encodes nonstructural proteins necessary for viral replication and the transactivation of true L genes. Three recently characterized E genes are important for L gene activation along with viral DNA replication: UL79, UL87, and UL95 [22-25].

HCMV genome replication initiates at the lytic replication origin (oriLyt) located between UL57 and UL69 [19, 20]. In addition to the core set of Herpesvirus DNA synthesis enzymes needed for productive infection, CMV encodes additional genes that form the DNA replication complex and assist with replication itself. Viral replication of HCMV genomes results in the generation of four genomic isomers (Class E structure) that contains varied arrangements of the unique long (UL), unique short (US), and repeat regions [7, 19]. Before encapsidation, HCMV genomes are cleaved at specific cis-acting packaging elements located within the repeat regions [19]. While the nucleocapsid traffics outside of the host cell, it acquires the tegument layer and lipid membrane. It is hypothesized that the virion experiences several envelopments while moving through different cellular compartments and most likely retains its final envelope from the Golgi apparatus [19]. In vitro, a full HCMV replication cycle to produce an infectious virion takes approximately 72 hours.

HCMV Transmission and Persistence

HCMV is a widespread pathogen that infects 50-90% of the world's population. Exposure to HCMV typically occurs early in life and susceptibility increases with age with nearly 80% of the US population seropositive for CMV by age 40 [26, 27]. Socioeconomic status also influences the risk of acquiring HCMV and studies suggest that individuals with less education, lower income, and of non-white race are exposed to HCMV earlier [19, 28, 29]. In particular, nearly 90% of pre-school age children in sub-Saharan Africa, South America, and India are seropositive for CMV, and by early adulthood, studies have shown that 100% of this age group is seropositive [30]. This high exposure rate to HCMV in resource poor settings can lead to severe consequences on adult health [31]. Overall, the ubiquitous presence of HCMV observed in the population can be directly attributed to its transmission and persistence strategies *in vivo*.

HCMV is transmitted by direct contact either vertically or horizontally. Vertical HCMV transmission occurs between mother and child either *in utero*, during delivery, or post partum by the ingestion of infected breast milk [19]. A pregnant mother experiencing either a primary infection or recurrent infection can pass HCMV transplacentally to her developing fetus [19, 32]. Although the risk of primary HCMV infection in a pregnant mother is low (0.7% to 4%), the rate of congenital HCMV infection is higher in women undergoing primary infection at approximately 20 to 40% [19, 29]. Recurrent infection does not carry as high of a congenital transmission rate since it may not result in systemic infection. Intrapartum transmission occurs if the mother is shedding virus vaginally or cervically during delivery in which case the transmission rate to the infant is nearly 50% [33]. Approximately 2 to 28% of women locally shed HCMV

during delivery [33]. Nearly all seropositive women will reactivate latent HCMV while lactating and reactivation typically is localized to mammary tissue [26]. The rate of HCMV transmission to a nursing infant is dependent on the concentration of HCMV in breast milk, the length of time the infant is breast-fed, and if the infant is premature. Of women who secrete HCMV into their breast milk, approximately 59% of breast-fed infants acquire the virus [26, 32, 34]. Vertical transmission, therefore, is a significant contributor to the HCMV burden worldwide.

Horizontal transmission occurs from direct contact between individuals by mucosal secretions containing infectious HCMV. Common body fluids that transmit the virus to an unsuspecting host include saliva, urine, blood products, allografts, and genital secretions [19]. The differences in age-related acquisition and prevalence worldwide likely reflects differences in sexual behaviors, child rearing practices, and living conditions. In addition, the duration of viral shedding from a host can be linked to age increasing the probability of exposure to HCMV negative individuals [19].

Two of the most common types of exposures that have consistently been linked with horizontal transmission of HCMV are sexual activity and contact with young children [26, 35]. The prevalence of HCMV infection is higher among adolescents and adults that not only have greater number of sexual partners, but also are carriers of other sexually transmitted diseases. HCMV can be shed unnoticed into cervicovaginal secretions as well as semen allowing viral acquisition by sexual contact [19].

HCMV transmission rates are high in young children after infancy. Common settings for HCMV transmission in young U.S. children are day care facilities and preschools [36, 37]. Evidence highly suggests that horizontal HCMV transmission occurs

between children since viral shedding may occur for years at a time from saliva and urine once acquired in this population. Multiple studies have concluded that greater than 50% of seronegative children at the time of enrollment, will seroconvert and shed virus after 6 to 12 months in a day care setting [38]. This high exposure rate to HCMV between young children also accounts for a high child to adult transmission rate. HCMV negative day care providers and parents alike are at risk of acquiring the virus due to their level of contact with young, infected children [36, 39]. HCMV, therefore, has evolved a unique replication program that allows for transmission either vertically or horizontally from different mucosal sites.

In vivo, HCMV undergoes three different phases of infection, namely, the acute phase, the persistent phase, and a latent phase [19]. During primary infection, HCMV acutely replicates in a wide range of cell and tissue types. Epithelial cells, endothelial cells, fibroblasts, and smooth muscle cells are the predominant target cells for HCMV replication that can be detected in almost every organ during acute infection [40]. Leukocytes and vascular endothelial cells spread the virus throughout the body from the initial site of infection to those sites that are critical for the secondary stage of viral infection. Typically, the virus acutely replicates until the immune system limits viral replication to tissues composed of glandular epithelium such as the salivary glands, kidneys, and breast tissue [19]. It is within these specialized tissues where the virus can replicate for up to months (adults) or years (children) allowing for the generation and shedding of low concentrations of infectious virus within saliva, urine, genital secretions, and breast milk.

The prolonged replication and shedding of virus within glandular epithelium is the persistent phase of infection. Chronic virus shedding may stem from the acute phase during primary infection or may result following reactivation of latent virus. This infection phase is unique in that, despite a robust immune response, HCMV has adapted to balancing long-term productive infection within glandular tissue with clearance by the host. The virus has adapted to these sites by encoding an arsenal of immune evasion proteins that can subvert the immune response by affecting antigen presentation to CD8+ and CD4+ T cells, innate immune signaling, or by modulating host cytokine responses [41]. Furthermore, HCMV does not induce cytopathic effect (CPE) within sites of viral persistence and, in fact, encodes proteins to prevent tissue damage. The absence of viral CPE in this tissue challenges the central dogma of Herpesviral replication: productive infection leads to lysis of the host cell. The virus likely has an evolutionary advantage to preserve the integrity of the host cell during persistent infection in order to maintain a low level of viral shedding for long periods of time [19, 41]. Together, these viral properties are required to achieve co-existence with the potent antiviral defense mechanisms of the host and to allow the prolonged viral shedding that ultimately results in HCMV transmission and preservation.

HCMV Latency

Like all Herpesviruses, HCMV establishes a reversibly quiescent state in which viral genomes are maintained in the absence of infectious virus [19]. While productive infection occurs in a variety of cell types, the reservoirs of HCMV latency are restricted to cells of the myeloid progenitor population, particularly, CD14+ and CD34+ cells [42-44]. Recent investigations suggest that the HCMV infection program not only depends on

the permissiveness of the host cell but also on the differentiation stage [45-47]. Undifferentiated cells do not support viral gene expression necessary for completing the full viral replication cycle, therefore, the virus remains latent until the host cell terminally differentiates. Presumably, the viral genome is maintained as a chromatinized episome in latent reservoirs until a stimulus causes reactivation [48, 49].

The mechanisms controlling HCMV latency and reactivation remain unknown. Although not surprising, multiple transcriptome and DNA analyses using an *in vitro* CD34+ latency model has revealed an expression profile of HCMV genes that is unique to the latent, compared to the lytic, state of infection [50]. In addition to the limited expression of viral protein coding genes during latency, several long noncoding RNAs are also implicated in the maintenance of the viral genome during latency [50]. It remains largely unknown how these viral genes influence genome maintenance or reactivation from latency. Additionally, it remains unknown how cellular factors influence the latency program and the exact stimuli during progenitor cell differentiation that triggers viral reactivation [19]. Establishment of a state of molecular latency is another mechanism by the virus to ensure persistence throughout the lifetime of the host.

HCMV Disease Outcomes

HCMV infection in healthy, immunocompetent hosts is generally subclinical. Primary infection and persistent replication, therefore, typically do not result in any signs or symptoms in healthy individuals and virus can be transmitted unknowingly to naïve individuals [19]. Reduced or limited immune function of an individual is a risk factor for HCMV pathogenesis and disease. Immunocompromised individuals that are commonly afflicted by HCMV associated disease include premature infants, cancer patients,

individuals with HIV/AIDS, patients receiving stem cell allografts, and solid organ transplant recipients [26]. HCMV pathogenesis in these individuals leads to severe morbidity, and in some cases, mortality.

Although healthy individuals generally are not at risk of developing HCMV associated disease, rare cases do occur and recent studies demonstrate that long-term infection may be a risk factor for certain chronic diseases. In healthy adults, the virus does account for 8% of all cases of mononucleosis and may also lead to rare cases of pneumonia, myocarditis, hemolytic anemia, retinitis, hepatitis, and peripheral neuropathy primarily during the acute phase of infection [19, 51]. More recent clinical studies have demonstrated a link between HCMV infection and chronic human diseases such as coronary atherosclerotic heart disease, gastric ulcer disease, rheumatologic disorders, and some human cancers [19, 52]. The mechanism behind this link to chronic human disease remains unknown but draws attention to the importance of this pathogen and the need for more effective therapy options.

HCMV infection of immunocompromised individuals is life threatening and may be caused by reactivation of latent virus, primary infection, or re-infection with a different HCMV strain [19]. Reactivation of latent HCMV is a major opportunistic infection complicating the outcome of hematopoietic stem cell transplants (HSCT) and solid organ transplants [26]. Although antiviral therapy has been effective in decreasing the incidence of HCMV disease, predictors of HCMV reactivation remain elusive prohibiting effective prophylaxis [53]. AIDS patients and cancer patients receiving immunosuppressive chemotherapy are also at risk of HCMV disease that may result in a range of clinical syndromes [26].

Nearly 5-10% of all congenital infections leads to HCMV associated complications affecting over 40,000 infants per year in the United States making HCMV the primary cause of viral birth defects [54]. Congenital infection may result in symptoms including intrauterine growth retardation, jaundice, hepatosplenomegaly, thrombocytopenia, hepatitis, and a range of central nervous system complications including hearing loss, microcephaly and encephalitis all characterized as cytomegalic inclusion disease (CID) [55-57]. Although acquisition of HCMV through breastmilk is normally asymptomatic in healthy newborn infants, low-birth weight, preterm infants are at risk of symptomatic HCMV infection [58]. Preterm infants and congenitally infected infants all carry a higher risk of long-term neurodevelopment sequelae.

Treatment and Vaccines

A vaccine to prevent HCMV infection does not exist and only a few antiviral drugs are available that can be administered in attempt to quell HCMV pathogenesis. Three, FDA approved drugs commonly used are foscarnet, cidofovir, and ganciclovir [19]. All three drugs target the viral polymerase UL54 either as nucleotide analogs or by directly inactivating polymerase activity [59]. Unfortunately, drug resistance can be observed to all three drugs and toxicity is high in the patient. Long term studies carried out primarily in AIDS patients, demonstrated that the duration of exposure to a drug, host immune competence, and the amount and duration of ongoing replication all are risk factors contributing to HCMV drug resistance [59]. In addition, it is known that UL97 and UL54 confer resistance to these drugs [59-62]. In order to become biologically active, gancyclovir needs to become phosphorylated [63]. Incidentally, gancyclovir is a substrate for the viral UL97 phosphotransferase. UL97 can develop resistance mutations

at several gancyclovir-binding locations contributing to the range of resistance observed to this drug. Mutations also occur in the UL54 polymerase gene that can develop resistance to any of the three drugs.

Currently, an approved, anti-HCMV vaccine has not been developed. Early attempts to develop a vaccine relied on passaging the Towne and AD169 clinical HCMV strains in order to attenuate the virus in cell culture [64-67]. These vaccines were immunogenic but failed to show clinical efficacy in preventing infection or disease in all but one of the target populations; high risk kidney transplant recipients that were seronegative but received organs from seropositive donors [68, 69]. Currently, subunit and vectored vaccine approaches are predominantly pursued and under investigation in clinical trials. Several vaccine candidates currently exist that target a conserved envelope glycoprotein, gB, and the highly abundant tegument protein pp65 [70]. Because current antiviral therapies have adverse effects on the health of the patient and long-term use can instigate viral resistance, an effective vaccine is a priority to prevent transmission and the development of HCMV associated disease.

DETERMINANTS OF CMV PERSISTENCE

In order for HCMV to successfully persist, it has evolved to replicate in cell types where the full replication cycle elicits little to no CPE, such as glandular epithelial cells and some types of endothelial cells [26, 71, 72]. The ability to persistently replicate in the immunocompetent host likely depends on reduced immune recognition of virus-infected cells at these specialized sites [41, 73]. Reduced immune recognition from the immunocompetent host is critical since sustained virus replication in glandular tissue represents the main source for transmission and acquisition of CMV.

Few viral determinants that mediate cytomegalovirus persistence have been identified and little is known about the specific molecular functions that facilitate persistence. This gap in knowledge is directly attributed to the host restriction for CMVs and the lack of an animal model for HCMV infection studies. Murine cytomegalovirus (MCMV) infection of the mouse is widely used as an outstanding small-animal model of HCMV infection for several reasons. HCMV and MCMV share similar genomic sequence and organization and undergo similar replication cycles [74]. Like HCMV, MCMV acutely infects multiple tissues in the mouse, persistently replicates in the salivary gland and establishes a lifelong latent infection of the host [75]. Using MCMV as a model, few viral determinants of persistence have been described. These include virus-encoded micro-RNAs and a conserved, virus-encoded G-protein-coupled receptor [76-78]. In addition, we previously identified a long, non-coding RNA (lncRNA), expressed by all cytomegaloviruses, that we showed to also be an important viral determinant of persistence [79].

Non-coding RNAs of Herpesviruses

Herpesviruses have a tremendous coding capacity within their double stranded DNA genomes that have allowed them to establish a life-long infection of their host. Transcription of their viral genomes is not limited to protein-coding genes; in fact, recent high-throughput transcriptome analyses of several herpesviruses during lytic infection demonstrate that viral non-coding RNAs (ncRNA) account for most of the transcription that occurs [50, 80, 81]. It is intriguing that all Herpesviruses examined produce an array of ncRNAs, and of particular interest, they all produce long ncRNAs (lncRNA) [82]. Since viruses have a limited real estate, there must be a selective advantage for including

ncRNAs into their genome that either assists with completing their replication cycle or maintains persistence throughout the lifetime of their host.

Generally, ncRNAs encompass a structurally and functionally diverse class of molecules that includes transfer RNAs, ribosomal RNAs, spliceosomal RNAs, small nucleolar RNAs, micro RNAs (miRNAs), and long non-coding RNAs (lncRNAs), among others [83]. These ncRNAs can exhibit cell type-specific expression and localize to specific subcellular compartments where they carry out their respective functions.

The diversity of ncRNA functions stems from the dynamic multi-functionality possessed by RNA. As a polymer, RNA has several molecular features that allow it to do more than just serve as an intermediate in translation [84, 85]. Through base-pairing with other nucleic acids, RNA can regulate transcription, RNA-processing, translation, and the stability of other RNA molecules. Functional RNAs require less genomic space to specify sequence-specific binding than proteins that perform the same function. Self-base pairing directs complex tertiary structures that allow for alterations in function and responsiveness to environmental signals. Importantly, these structures are tolerant of mutations, yet quickly evolve new functions since sequence conservation is less restrictive. Viral RNAs that look and function like cellular RNAs are also less immunogenic than proteins, allowing a virus to escape further immunosurveillance. It is unsurprising, then, that many viruses have adopted ncRNAs to expand the functional repertoire of determinants that can quickly serve to subvert host responses.

Long ncRNAs are typically defined as RNA transcripts longer than 200 nucleotides that are not translated into a functional protein [84]. Many lncRNAs do interact with polysomes, however, resulting in the translation of short, nonfunctional

peptides [85, 86]. Only a limited number of cellular and viral lncRNAs have been fully characterized and recent evidence indicates they can regulate gene expression at multiple different levels. In each Herpesvirus subfamily, several viruses encode for lncRNAs. Although function has not been ascribed to all of these lncRNAs, they further demonstrate the biological relevance of these transcripts.

Alphaherpesvirus lncRNAs: HSV-1 Latency Associated Transcript

Extensive research has been devoted to characterizing and uncovering the function of the Latency Associated Transcript (LAT) produced by HSV-1. During HSV-1 neuronal latency, LAT is the only abundantly produced viral transcript [87, 88]. This lncRNA is processed from a larger, 8.3 kb precursor transcript as an intron [88]. Depending on the cell type, it is spliced as either a 2 kb intron or a 1.5 kb intron [87]. This intron, regardless of size, is suspected of having several roles critical for maintaining the molecular state of latency and for efficient reactivation of quiescent HSV-1 genomes in sensory ganglia [89]. LAT is observed to have antiapoptotic activity that allows successful reactivation of the virus [90]. Although a mechanism ascribing function to the LATs remains elusive, this lncRNA clearly has a critical role during the latent phase of HSV-1 infection.

Gammaherpesvirus lncRNAs: KSHV PAN

Kaposi's Sarcoma-Associated Herpesvirus (KSHV) is the causative agent of several human cancers and lymphoproliferative disorders, including Kaposi's Sarcoma, Multicentric Castleman's Disease, and Primary Effusion Lymphoma [1]. This virus transcribes the lncRNA called polyadenylated nuclear RNA or PAN [91-93]. PAN

resembles a host, cellular RNA since it is transcribed by RNA polymerase II, is polyadenylated, and is capped at its 5'end. PAN does not, however, export to the cytoplasm nor is it spliced to yield alternative transcripts. It is the most abundant transcript accounting for approximately 80% of total RNA isolated from an infected cell [91].

Betaherpesvirus lncRNAs

Four long ncRNAs (RNA5.0, RNA4.9, RNA1.2, RNA2.7) account for the majority of HCMV transcription during lytic replication and many questions remain regarding the biological role these ncRNAs play during pathogenesis (Figure 2) [50, 80].

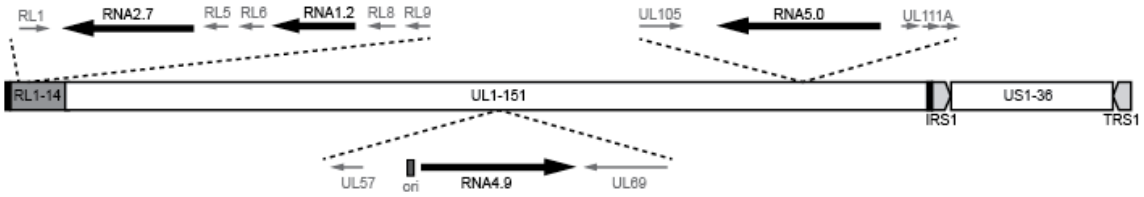


Figure 2. **Schematic diagram of the location of genomic loci that encode lncRNAs in a conventional clinical isolate of HCMV.**
Short repeats flanking the UL region in clinical isolates are indicated by black bars. The lncRNAs are indicated by thick black arrows. Protein-coding genes flanking the lncRNAs are indicated as gray arrows for orientation purposes only; diagram is not to scale.

RNA4.9

RNA4.9 was recently identified in transcriptome analysis of HCMV-infected fibroblasts and monocytes and bone marrow progenitors [50, 80]. The 5' end of this lncRNA initiates adjacent to the viral origin of replication (*ori*) (Figure 2). Two smaller, lncRNA molecules have been identified that form RNA-DNA hybrid structures with the adjacent *ori* sequences [94]. These RNAs appear to be transcribed in the opposite

direction of RNA4.9 and are produced independently. Recent studies have shown that RNA4.9 is also transcribed during latent infection of CD14+ monocytes and CD34+ bone marrow progenitors and might play a regulatory role in cellular and viral gene expression. Chromatin isolation by RNA purification (ChIRP) demonstrated that RNA4.9 interacts with the HCMV major immediate promoter (MIEP) region and RNA crosslinking immunoprecipitation (rCLIP) confirmed RNA4.9 interaction with the viral protein UL84 and components of the Polycomb Repressive Complex 2 (PRC2) [50]. These interactions are hypothesized to result in the deposition of a repressive chromatin structure observed on the MIEP, thereby suppressing IE gene expression during latency. Further studies examining the requirement for RNA4.9 in the establishment and maintenance of latent infections are necessary to define the role of RNA4.9 in HCMV infections.

Table 1. Long Non-Coding RNAs Expressed by Human Cytomegalovirus.

Non-coding RNA	Orthologs in other CMVs	Description	Refs.
RNA2.7	None predicted	Capped and polyadenylated RNA. Inhibits apoptosis.	[95-99]
RNA1.2	None predicted	Capped and polyadenylated RNA. Function unknown.	[95, 100-102]
RNA5.0	MCMV RNA7.2; RhCMV RNA3.5; predicted for all β-herpesviruses	Large stable intron RNA; mouse RNA involved in viral persistence in the host	[79, 103-107]
RNA4.9	None predicted	Capped and polyadenylated RNA. Function unknown, may interact with cellular chromatin modulating complexes.	[50, 80]

RNA2.7

RNA2.7 was first identified as an abundant early transcript derived from repeat sequences in the laboratory-adapted strain of HCMV, AD169 (Figure 2) [95]. Unlike laboratory-adapted strains, clinical isolates of HCMV have not undergone extensive recombination and rearrangement with passage in culture. The sequence from which RNA2.7 is expressed in clinical isolates of HCMV is not duplicated and, as a result, the locus encoding RNA2.7 is present as a single copy [96]. Despite the duplication observed within laboratory-adapted HCMV strains, the genomic sequence of RNA2.7 is well conserved between laboratory strains and clinical isolates of HCMV suggesting a critical role for this RNA *in vitro*. Like many cellular lncRNAs, RNA2.7 resembles an mRNA, as it is transcribed by RNA polymerase II, it is polyadenylated, and it is transported to the cytoplasm of infected cells [97]. Although short open reading frames are encoded by RNA2.7 and it interacts with polysomes, until recently, no protein products originating from this locus had been detected. Evidence of translation of several short ORFs encoded by RNA2.7 was observed in a high resolution RNA-seq/mass spectrometry analysis of HCMV-infected cells [108]. It is unknown whether these short translated ORFs are functional or production of these short peptides may simply be a secondary feature of lncRNAs. Regardless of its protein-coding functions, recent studies described below have firmly identified a function for RNA2.7 as a lncRNA during infection.

RNA2.7 is the most studied lncRNA of CMV. It is highly expressed, as observed by northern blot analysis and high-throughput sequencing methods [50, 80, 98]. Transcription of RNA2.7 exhibits early expression kinetics and the locus remains transcriptionally active throughout the virus replication cycle, with peak accumulation

occurring at late times post infection [95, 99]. Recent transcriptome analysis demonstrated that RNA2.7 alone accounts for over 40% of all virally encoded polyadenylated transcripts from a productive infection in fibroblasts [50]. This high expression level implies that this RNA may play an important role in the viral replication cycle.

A function of RNA2.7 was elucidated in studies that used the RNA itself as a probe to identify target binding proteins [98]. Targets of RNA2.7 were identified by Northwestern screen of a human cDNA library and physical interactions with the potential interactors were validated by immunoprecipitation. The initial screen led to the identification of an interaction between RNA2.7 and the GRIM-19 (*G*enes associated with *R*etinoid/*I*nterferon-induced *M*ortality-*19*) subunit of complex I of the mitochondrial respiratory chain (MRC). The MRC is responsible for the formation of the electrochemical gradient across the mitochondrial inner membrane needed for ATP production. GRIM-19 is important for the assembly and electron transfer activity of MRC complex I. GRIM-19 also has a secondary function as a tumor suppressor gene and shuttles to the nucleus under cellular stress to promote apoptosis by inhibiting STAT-3 activation of proto-oncogenes [109, 110]. Initial investigation of the interaction between RNA2.7 and GRIM-19 demonstrated that in the presence of a specific MRC complex I inhibitor, RNA2.7 prevents disruption of ATP production that would normally lead to cellular apoptosis. During productive HCMV infection, RNA2.7 is observed to be cytoplasmic and GRIM-19 does not translocate to the nucleus consistent with the idea that RNA2.7 might be actively involved in sequestration of GRIM-19 in the cytoplasm, thereby keeping the cell alive during virus infection [110, 111]. In cells infected with a

recombinant virus lacking RNA2.7, GRIM-19 is detected in the nucleus during viral infection, supporting this hypothesis [98]. While these studies suggest a critical role for RNA2.7 in the prevention of apoptosis, this RNA, surprisingly, is not essential for viral replication *in vitro.* This may reflect redundancy in the mechanisms that HCMV has evolved to prevent cellular apoptosis. Interestingly, RNA2.7-orthologous loci are not clearly evident in other cytomegaloviruses. Examination of the chimpanzee CMV genome sequence has identified a homologous region to RNA2.7, however, transcription of this RNA has not been investigated in other CMV species [96]. Where in the virus life cycle RNA2.7 is most critical, remains to be determined. RNA2.7 is highly expressed in experimental and natural latent HCMV infections of CD14+ monocytes and CD34+ bone marrow progenitors and is one of only a handful of transcripts expressed [50]. Perhaps the anti-apoptotic function of RNA2.7 is especially critical to HCMV infection of the latent reservoir.

RNA1.2

RNA1.2 is located approximately two kilobases from RNA2.7 on the same DNA strand (Figure 2) [95]. Similar to RNA2.7, RNA1.2 was also originally identified in the laboratory-adapted strain AD169 within the repeat sequences. RNA1.2 is transcribed with early expression kinetics, with peak accumulation occurring after the onset of viral genome replication [95, 101, 102]. The RNA1.2 transcript is polyadenylated and may undergo a small, internal splicing reaction since a set of splice sequences were identified by next generation sequencing analysis [80]. Open reading frames have been identified throughout the transcript and a 30-kDa protein is predicted to be expressed [102]. Like RNA2.7, evidence of translation of several short ORFs encoded by RNA1.2 was detected

in a high resolution RNA-seq/mass spectrometry analysis of HCMV-infected cells but it is unknown whether these short translated ORFs are functional [108]. Although independent transcriptional analyses have consistently demonstrated this RNA to be highly abundant, little work has been done to establish the function of RNA1.2.

RNA5.0

RNA5.0 was first identified as an immediate early transcript of HCMV [103-105]. Like other lncRNAs expressed by HCMV, numerous short open reading frames are encoded by RNA5.0 but no protein products, not even short peptides, originating from this locus have been detected during HCMV infection. Further characterization of this RNA revealed that it is a nuclear-localized, intron derived from the processing of a precursor RNA spanning the UL105-UL111A region of the genome (Figure 2) [79]. Very few lncRNAs that are introns have been described from eukaryotic cells and this is the first intron observed to accumulate during a betaherpesvirus infection. Recombinant HCMV that is unable to express RNA5.0 replicates in fibroblasts with wild-type kinetics, indicating that RNA5.0 is dispensable for replication in cultured cells.

In all cytomegaloviruses, orthologous genomic loci have been identified spanning the UL105-UL112 region of the genome and this region is conserved among all β-herpesviruses (Figure 3). Each locus shares common features, including a consensus splice donor sequence that defines the 5' end of the RNA, high AT sequence content (~60%), and the presence of many homopolymeric stretches of A or T residues. However, the loci vary considerably in size and overall sequence content is not well-conserved; the conservation of intron length and sequence across CMV species might not be as significant as maintaining the high AT nucleotide content since this feature could

directly contribute to the structural flexibility of RNA and ultimately to the function of ncRNAs.

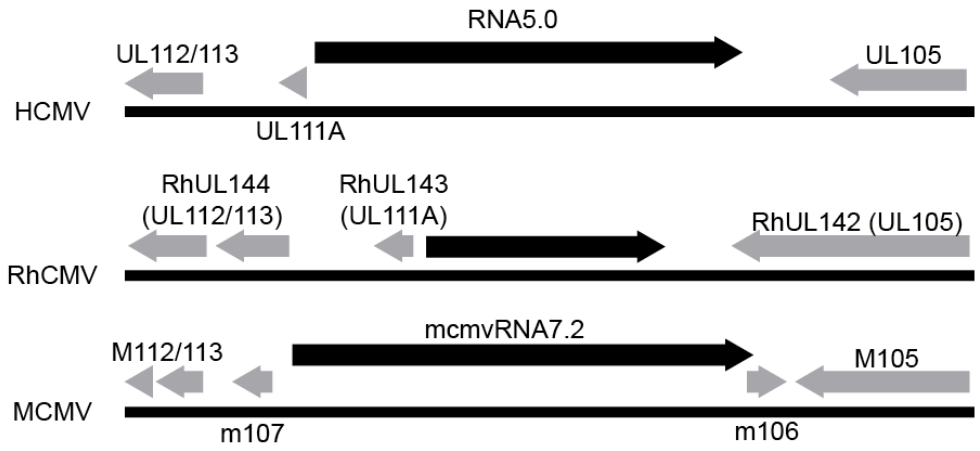


Figure 3. **RNA5.0 Orthologs in Other Cytomegaloviruses.**
The genomic locus spanning UL105-UL112/113 is shown in the reverse orientation to highlight RNA5.0 and its orthologs. Long ncRNA expression from the locus has been confirmed in HCMV, rhesus cytomegalovirus (RhCMV; a 3.5 kb RNA) and murine cytomegalovirus (MCMV), and are indicated by black arrows. Protein coding genes in the locus are indicated as gray arrows for orientation purposes.

Murine Cytomegalovirus (MCMV) expresses a 7.2 kb ortholog of RNA5.0 (RNA7.2) and studies of MCMV infection of the mouse have been carried out to examine the function of this lncRNA during infection *in vivo* [106, 107]. These studies demonstrated a role for the intron during the persistent phase of replication in the salivary glands of mice. Recombinant viruses that are unable to express RNA7.2 replicate with wild-type kinetics in cultured cells and initiate a robust acute infection in immunocompetent mice. Interestingly, these recombinant viruses fail to establish a robust persistent infection at this site, suggesting that the intron might play a critical role in evading the immune response in this tissue [106]. It is unknown if there is a function for RNA7.2 during latency in mice.

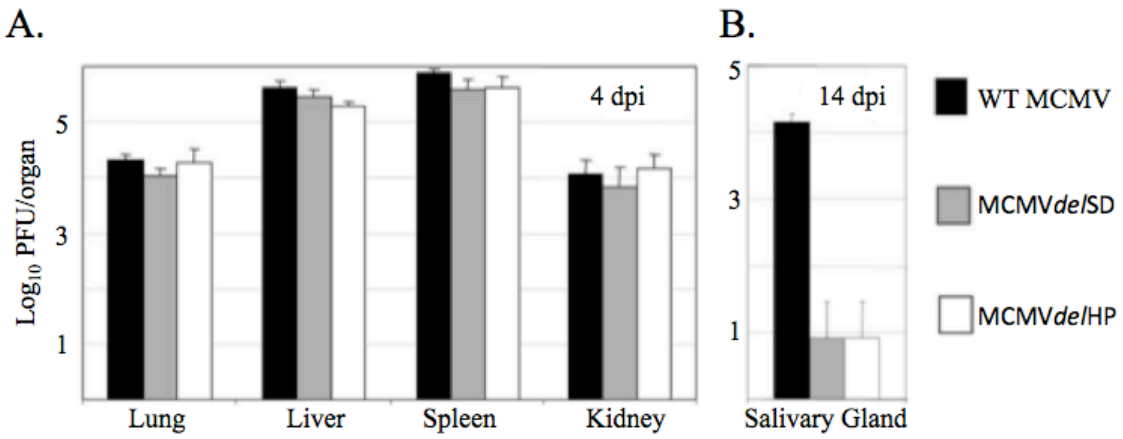


Figure 4. Recombinant Viruses Unable to Process or Accumulate RNA7.2 Do Not Establish Persistent Replication *In Vivo* (modified from [106]).
Recombinant viruses were generated for a splice donor site (MCMV*del*SD) that prevents processing of the RNA7.2 primary transcript or for a deletion in a stem loop structure (MCMV*del*HP) located in the 3' end of the intron that prevents RNA7.2 accumulation but not splicing of the primary transcript. *In vivo* analysis of recombinant virus replication. Six week old Balb/C mice were inoculated i.p. with $1x10^6$ pfu of the indicated virus. Total yield of infectious virus is reported by day post infection (dpi) in the organ homogenates indicated by plaque assay. A) Both recombinant viruses replicate similarly to WT MCMV during the acute phase of infection. B) Both recombinant viruses are below the limit of detection for the plaque assay or are attenuated at 14 dpi.

Accumulation of the RNA5.0 and RNA7.2 intron molecules is hypothesized to be due to retention of the lariat structure of the intron in order to avoid cellular RNA decay machinery. It is hypothesized that the intron remains in the form of a lariat, similar to the Latency Associated Transcript (LAT) of HSV-1, thereby protecting it from degradation. A stem-loop structure located between the polypyrimidine track at the 3' end of the intron and the putative intron branch point was identified using the structural prediction software mFold [106]. Although deletion of this sequence does not impact processing of the precursor transcript, it does prevent accumulation of the intron during infection [106].

Dissertation Research Objectives

A major gap in our knowledge surrounding Cytomegalovirus infection is how this virus maintains persistent replication in glandular epithelium. Persistent replication allows for the global burden of this potentially lethal pathogen by the continual shedding of virus in mucous secretions that can occur for months and even years at a time without notice. The overall research goals of this dissertation, therefore, are to contribute more information on this CMV transmission strategy that will eventually lead to answers on how to curb the persistent replication cycle and prevent associated CMV disease. Specifically, the goal of this dissertation is to further define a previously established CMV persistence factor, the MCMV RNA7.2. We hypothesize that RNA7.2 is a CMV virulence factor critical for the maintenance of the sustained persistent replication phase of viral infection *in vivo*. Evidence supporting this overall hypothesis is explained throughout this dissertation. We demonstrate that RNA7.2 is transcribed with late gene kinetics and processed from a larger precursor transcript. This larger precursor transcript also encodes a small protein, m106, that is expressed during productive infection. Translation of m106 is reliant on the canonical splicing signals within the primary transcript. RNA7.2 has a long half-life that may prove essential for maintaining a replication program in glandular tissues that remains undetected from the host. Preliminary stability analysis of the HCMV RNA5.0 also reveals a long half-life suggesting that this RNA7.2 trait is conserved among the intron homologues across CMV species. The data presented in this thesis will provide a foundation for future studies that seek to examine function of RNA7.2 during persistent replication.

CHAPTER II

MATERIALS AND METHODS

VIRUSES AND TISSUE CULTURE

The BAC clone of the wild type Smith MCMV strain, pSM3fr, was used as the parent strain throughout this dissertation [112, 113]. Recombinant viruses were generated by seamless, red-mediated recombination in the DH10B *Escherichia coli* strain GS1783 [106, 114]. A kanamycin cassette containing an Isce I restriction site was PCR amplified from the pEPKan-S plasmid using primers that have 50 bases of homology flanking the MCMV targeted region (Table 2). Kanamycin cassette PCR products were digested with DpnI, gel purified, and transformed by electroporation into recombinogenic GS1783 *E. coli*. Kanamycin positive recombinants were selected for growth on LB agar plates containing 12.5 ug/ml chloramphenicol and 50 ug/ml kanamycin. 80-bp complementary oligonucleotides designed with the targeted mutation to replace the kanamycin cassette were annealed and transformed into the GS1783 *E. coli* containing the Smith-Kan$^+$ BAC. After an hour-long outgrowth, the cells were treated with 1% L-Arabinose to induce Isce-I enzyme production. This enzyme digests the Isce-I site located on the kanamycin cassette that facilitates the recruitment of the DNA recombination machinery to the site of recombination. BAC recombinants were then screened from kanamycin negative GS1783 *E. coli* by PCR and sequencing. Recombinant BAC DNA was isolated and electroporated into 10.1 fibroblasts to produce viral stocks.

Viral stocks were concentrated and purified by ultracentrifugation over a 20% sorbitol cushion. Concentrated virus was titrated by plaque assay on 10.1 mouse

Table 2. Primers Used For Recombinant Virus Production.

Target	Sequence 5' to 3'
MCMV*del* 20	GGAGTGTAGGTATTCACCGTCAGACGCAACCTGACGCATCCCGGCTAGAATCGATTTATTCAACAAAGCCACG
	CACCTGAGCCTGCTCGGCCGTTCGCTCAGGTGTGATAATGCACCTTTCAGCGCGTATATCTGGCCCGTACATCG
	TATTCACCGTCAGACGCAACCTGACGCATCCCGGCTAGAACTGAAAGGTGCATTATCACACCTGAGCGAACGGCCGAGCA
	TGCTCGGCCGTTCGCTCAGGTGTGATAATGCACCTTTCAGTTCTAGCCGGGATGCGTCAGGTTGCGTCTGACGGTGAATA
MCMV*del* 100	GATCACGCTACCACCGTGTGTCTCCGTACTCCGCTATTATACTTTGCGGCTCGATTTATTCAACAAAGCCACG
	CGCTACCACCGTGTGTCTCCGTACTCCGCTATTATACTTTGCGGCCTGAAAGGTGCATTATCACACCTGAGCGAACGGCCGAGCAGGCTC
	GAGCCTGCTCGGCGTTCGCTCAGGTGTGATAATGCACCTTTCAGGCCGCAAAGTATAATAGCGGAGTACGGAGACACACGGTGGTAGCG
MCMV*del* 135	CGGCACGGGGAAATAAAATGATCACGCTACCACCGTGTGTCTCCGTACTCTCGATTTATTCAACAAAGCCACG
	GATGCGTCCGCCGCCTCACCTGAGCCTGCTCGGCCGTTCGCTCAGGTGTGCGCGTATATCTGGCCCGTACATCG
	AAATAAAATGATCACGCTACCACCGTGTGTCTCCGTACTCCACACCTGAGCGAACGGCCGAGCAGGCTCAGGTGAGGCGG
	CCGCCTCACCTGAGCCTGCTCGGCCGTTCGCTCAGGTGTGGAGTACGGAGACACACGGTGGTAGCGTGATCATTTTATTT
M106-GFP	TCCACCAACACGATCCCCGAGATACCCAGAATCGTGGTCGAGGTGGTAGACGCCGGAAGAAGATGGAAAAAG
	GTTTTCTGACATGAGTCTGTGTGTTTATTTATTAATTATCTGTCAGTTTACGTCGTGGAATGCCTTCG
Kan-BP	AGTATACCTATTTTTCTGCAAAAATAAGGATTACTATATTCTAACCACCCTCGATTTATTCAA
	GAGAACGGGTGGGAGCCGAGGCCGCCGTGAGACCTCGACTCCCGTGAGACCGCGTATATC

fibroblasts. The replication kinetics of recombinant and wild type viruses was determined by infecting 10.1 fibroblasts at a multiplicity of infection (MOI) of 1 PFU per cell, collecting supernatant every 24 hours for five days starting at time zero, and titrating the supernatant by plaque assay on 10.1 fibroblasts.

10.1-mouse embryonic fibroblasts and human foreskin fibroblasts were propagated in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% newborn calf serum, 100 U/ml penicillin, and 100 ug/ml streptomycin. Cells were maintained at 37°C with 5% CO_2.

RNA ISOLATION

To determine expression kinetics, 10.1 fibroblasts were pretreated with 100 ug/mL cyclohexamide or 200 ug/mL PAA 1 hour before MCMV infection. Total RNA was harvested from 10.1 fibroblasts at either 24 hours post infection (h p.i.) or 48 h p.i. with TRIzol LS (Life Technologies) according to the manufacturer's protocol.

RNA7.2 half-life analysis was performed by infecting 10.1 fibroblasts with WT MCMV or MCMV*del*HP (Hairpin deletion virus) at an MOI of 1.0. At 30 hours post infection, 4 ug/mL of Actinomycin D was added to the infected cells and RNA was harvested over a time course starting at time 0 and ending at 32 hours post infection. Transcript levels were quantified by qRT-PCR at the different time points relative to time zero. To isolate RNA for stability analysis, total RNA was extracted from either infected or transfected cell monolayers using the RNeasy Mini Kit (Qiagen). RNA concentrations were quantified using a Nano Drop. Stability analysis of the transfected pCDNA3.1 + (Life Technologies) constructs was carried out in a similar manner to the infection

analysis except Actinomycin D was added to transfected cells at 24 hours post transfection [115].

5' AND 3' RACE

Total RNA harvested from mock or WT MCMV infected 10.1 mouse fibroblasts at 48 hours post infection was analyzed by the First Choice RNA ligase-mediated rapid amplification of cDNA ends kit (RLM-RACE, Ambion). Briefly, 10ug of total RNA was DNase treated (Ambion) then processed according to the RLM-RACE protocol. Before reverse transcription, the ligated RNA was denatured at 90°C for 5 minutes then incubated immediately on ice. Reverse transcription proceeded at 50°C for 30 minutes to generate cDNA from the capped mRNA species. The cDNA was amplified by nested PCR using the FirstChoice 5' outer primers and gene specific primers outer 541 and inner 542 (Table 3). T4 DNA Polymerase (New England Biolabs), 400 nM primers, 200 nM deoxynucleoside triphosphates (dNTPs), 5% DMSO, and 1x Standard Buffer was used to amplify cDNA and cycling conditions were followed according to the FirstChoice protocol. Amplification products were gel extracted then TA cloned into pGEM-T-Easy and sequenced. As a negative control, RNA was not treated with the tobacco acid phosphatase (TAP). For 3' RACE, RNA was reverse transcribed using a poly(A)adapter. cDNA was amplified using the FirstChoice primers specific for the adapter and gene specific primers outer 572 and inner 573 (Table 3). Amplification products were cloned and sequence as previously mentioned.

Table 3. RNA Analysis Primer Sequences.

Target	Purpose	Sequence 5' to 3'
7.2kb Intron (a, a')	qPCR	Fwd: GAGTCAGTTCTAACCCATCACG Rev: AGCTCGAAAGTTGAACGGG Probe: ACGAACGGGTAAAACGGGTAAGGG
Exon2 (b, b')	qPCR	Fwd: CCACTACCTCTCGATGACAAC Rev: AGCGAATTCTAGCGTTACCG Probe: CGGAGCCTGCGACTTGTCTGC
Spliced mRNA (c, c')	qPCR	Fwd: TTATCACACCTGAGCGAACG Rev: GCAGAGTTCGATGTGTCCG Probe: AGGATGCGAGATGGCGACGG
M54	qPCR	Fwd: AACATATCCCTGCCGATCTTG Rev: CAACGCTTTCTACGGTTTCAC Probe: ATGCTCCCGTGTCTCCCCATC
GAPDH	qPCR	Fwd: GTGGAGTCATACTGGAACATGTAG Rev: AATGGTGAAGGTCGGTGTG Probe: TGCAAATGGCAGCCCTGGTG
Actin B	qPCR	Fwd: CTTGATCTTCATGGTGCTAGGAG Rev: CGTTGACATCCGTAAAGACCT Probe: ACCATGTACCCAGGCATTGCTGA
18S rRNA	qPCR	Fwd: GTTGATTAAGTCCCTGCCCTT Rev: ATAGTCAAGTTCGACCGTCTTC Probe: ACCGATTGGATGGTTTAGTGAGGCC
Exon1	Primer Extension	497:GGCCTTCGGGACGCCGTCACCTCCGCCGCCGC
Exon2	3'RACE	541: GATCGTTGTCGTCTCTGTCGTGTT
	3'RACE	542: TGTCATCGAGAGGTAGTGGAGGAT
	5'RACE	571: ATCCTCCACTACCTCTCGATGACA
	5'RACE	572: AACACGACAGAGACGACAACGATC
	Northern Blot Probe	253: GTCGACATGGCGACGGCGAGCCAGCAA 263: GCGGCCGCGTCTACCACCTCGACCACGATT
RNA7.2	Northern Blot Probe	262: CTCCAATCGGCCTAGGAATCCTGGCTAGGT 263: AGCAACACGATGCTCTGTGTCGTCGGTCGG

PRIMER EXTENSION

Oligonucleotides were end labeled with [γ-^{32}P]ATP using T4 Polynucleotide Kinase (New England Biolabs) and purified over a sephadex G-50 column. 1,000,000 cpm of end-labeled primer was hybridized to 40 ug of total RNA prepared from either mock infected or WT MCMV infected 10.1 fibroblast. The RNA and oligonucleotide probe mixture was denatured at 95°C for 3 minutes to disrupt secondary structure before incubation at 55°C with superscript III reverse transcriptase for 1 hour (Table 3) [116]. The primer extension products were analyzed by denaturing 10% urea-polyacrylamide gel electrophoresis followed by phosphorimager analysis.

PLASMID DNA ISOLATION AND ANALYSIS

All plasmid DNA was isolated from overnight cultures of transformed DH5α *E. coli* using the Maxi Prep Kit (Qiagen) per the manufacturer's protocol.

NORTHERN BLOT ANALYSIS

5 ug of total RNA was glyoxalated and resolved on either a 0.7% or 1.4% glyoxal gel. The RNA was transferred to a Nytran SuPerCharged membrane using a Turboblotter kit (Whatman). Antisense riboprobes were radiolabeled with [α-^{32}P]CTP by in vitro transcription from linearized pGEM-T-Easy plasmids (Promega) and purified over sephadex G-50 columns (Roche). Membranes were prehybridized with 10 mls of formamide buffer (1% SDS, 1x Denhardts, 4xSSC, 50% formamide) containing blocking reagents 10ug tRNA and 10ug salmon sperm for 1 hour at 60°C. 1,000,000 cpm/ml of a radiolabeled probe was hybridized overnight at 60°C with in 10 mls of formamide hybridization buffer and blocking reagents. Membranes were washed the following day twice with a low stringency buffer (2xSSC + 0.1% SDS) at room temperature (RT) for 10

minutes and twice with a high stringency buffer (0.1xSSC + 0.1% SDS) at RT for 10 minutes and once with the same high stringency buffer for 1 hour at 65°C [116]. Membranes were exposed to a K-screen for 1 hour (intron probe) or overnight (m106 probe) (Table 3).

TRANSFECTIONS

Mouse fibroblasts were seeded in a 24 well dish. At 60% confluency, 10.1 mouse fibroblasts were cotransfected with a pGL3 construct and the phRL-TK normalization vector using polyethyleimine (PEI) at a 6:1 ratio of PEI to plasmid DNA. The plasmids were cotransfected at a 1:1 ratio of pRL-TK to pGL3 or 3 x 10^{10} copies per well of each plasmid. Three biological replicates were carried out to examine each pGL3 construct. Within a biological replicate, three technical replicates were performed. Protein lysates were harvested 48 hours post transfection and assayed following the Promega Duo-Glow Luciferase assay kit. Luciferase activity was normalized to renilla activity in each well and the data is expressed as the fold change of luciferase induction relative to the luciferase induction from the pGL3 promoterless vector.

pCDNA3.1^+ constructs were transfected into subconfluent 10.1 mouse fibroblasts using a 6:1 ratio of PEI to plasmid DNA. 3 x 10^{10} copies of each plasmid was transfected into mouse fibroblasts. Following Actinomycin D treatment, cells were harvested at the indicated time points by washing with PBS, trypsinized with 0.25% Trypsin, concentrated by centrifugation, then frozen by liquid nitrogen. Transfected cells were stored at -80°C until all time points were collected. Once all time points were collected for the Actinomycin D experiment, RNA was isolated using the RNeasy Kit (Qiagen).

FISH AND IMMUNOFLUORESCENCE

Flourescently labeled RNA probes antisense to the 7.2kb intron were generated using the FISH Tag kit (Invitrogen). Briefly, probes were in vitro transcribed from linearized pGEM-T-Easy constructs using an amino allyl modified base in which an alexa flour can be chemically attached to. Following in vitro transcription of the probes, the DNA templates are digested using DNase I and the amino modified RNA is purified over a column then ethanol precipitated. The purified probes are fluorescently labeled according to the manufacturer's instructions then column purified and subsequent ethanol precipitation. Cells were fixed for 20 minutes in 4% paraformaldehyde, 10% acetic acid in 1x PBS. The fixation was quenched for 20 minutes in PBS with 0.1M glycine. Cells were washed twice with PBS then permeabilized with 70% ethanol overnight at 4°C. Cells were rehydrated by washing twice with 50% formamide/2x SSC. The probe was denatured by heating at 65°C for 10 minutes in probe buffer then cells were incubated overnight with the denatured probe at 37°C. The following day, cells were washed twice with 0.1X SSC/50%formamide at 50°C then washed once with PBST.

PLASMID CONSTRUCTS

RACE products were cloned into pGEM-T-Easy (Promega) and sequenced. Reporter constructs were generated by PCR amplifying sequence upstream of the MCMV RNA7.2 RNA splice donor site.

Primers used to generate the pGL3 luciferase reporter constructs are indicated in (Table 4). Amplicons were resolved by gel electrophoresis and gel purified using the Qiaquick Gel Extraction Kit (Qiagen). Amplicons were TA-cloned into pGEM-T-Easy. After insertion into the pGEM-T-Easy plasmid, the inserts were digested from the

plasmid using the flanking EcoRI sequences then subcloned into the pGL3 Basic vector at a newly generated EcoRI site using the site directed mutatgenesis kit (Stratagene). Orientation of the cloned insert was determined by sequencing the pGL3 plasmid. The pGL3-SV40 control plasmid and the renilla phRL-TK normalization control plasmid were used in the luciferase assays (Promega).

Primers used to generate the RNA7.2 locus pCDNA3.1$^+$ construct are indicated in Table 5. Briefly, the RNA7.2 locus was cloned into pCDNA3.1 in two fragments. Fragments were amplified using the Phusion high-fidelity DNA polymerase with in 2.5% DMSO, 10 ul Phusion Buffer, 10 mM dNTPs, 0.5 uM of each primer, 1:100 dilution of prepared pSM3fr DNA, 0.5 ul of Phusion DNA (0.02U/ul), and up to 50 ul with water. The cycling conditions followed the 3-step protocol per manufacturer's instructions: Initial denaturation 98°C for 30 seconds for one cycle; 35 cycles of denaturation 10 seconds at 98°C, annealing 30 seconds at 64°C, and extension at 72°C for 2 minutes; 1 cycle of 72°C for 10 minutes, and hold at 4°C.Amplified products were gel purified using the Qia Quick Gel Extraction Kit (Qiagen) then digested with either BamHI and NotI or NotI. Prior to ligation with the PCR amplified products, pCDNA3.1$^+$ was digested with either BamHI followed by NotI or just NotI. The enzymatic reactions were halted by incubating at 65°C. Following digestion, the pCDNA3.1$^+$ was dephosphorylated with CIP for 1 hour at 37°C then gel purified using the Qia Quick Gel Extraction Kit (Qiagen). 100 ng of the digested, amplified product was ligated to 50 ng of the prepared pCDNA3.1$^+$ using T4 DNA ligase over night in an ice water bath to create the pIntron constuct. Substitution or deletion mutations were generated within pIntron using the QuickChange Site-Directed Mutagenesis Kit (Stratagene) per manufacturer's protocol.

Table 4. pGL3 Construct Primer Sequences.

M112/113	pGL3 Construct	459: ACGAAGGTCTTTTCACCGGT 435: ACCATCTGCTAGGCGGGTCC
PR1	pGL3 Construct	28: AGATAGCGCGGCGTCCGTCG 349: CTGAGAGCTCCGGGCCTTCGG
PR2	pGL3 Construct	133: AAAAGAAAGTCCGTGACCGGGTCG 349: CTGAGAGCTCCGGGCCTTCGG
PR3	pGL3 Construct	28: AGATAGCGCGGCGTCCGTCG 29: AAGCGGACCTGAAAACGGGG
PR4	pGL3 Construct	439:GATGGTCTATAACCTCACCGCGGACC 440: GTAACGAGGCGGAAGACCCACATTTC
PR5	pGL3 Construct	439:GATGGTCTATAACCTCACCGCGGACC 29: AAGCGGACCTGAAAACGGGG
PR6	pGL3 Construct	439:GATGGTCTATAACCTCACCGCGGACC 76: TGCTCGCGTCGAGTGACCGC
PR7	pGL3 Construct	28: AGATAGCGCGGCGTCCGTCG 76: TGCTCGCGTCGAGTGACCGC
PR8	pGL3 Construct	351: ATACGGCGTACGGAGCTCCG 76: TGCTCGCGTCGAGTGACCGC
PR9	pGL3 Construct	459: ACGAAGGTCTTTTCACCGGT 79:TGCTCGCGTCGAGTGACCGC
PR10	pGL3 Construct	459: ACGAAGGTCTTTTCACCGGT 29: AAGCGGACCTGAAAACGGGG
PR11	pGL3 Construct	459: ACGAAGGTCTTTTCACCGGT 440: GTAACGAGGCGGAAGACCCACATTTC
PR12	pGL3 Construct	459: ACGAAGGTCTTTTCACCGGT 77: ACGGACGCCGCGCTATCTCGA

Table 5. Primer Sequences for Generating pCDNA3.1^{+} Constructs.

pI Seg.1	577: GATCGGATCCACCTGGAGTGTAGGTATTCACC 580:GATCGCGGCCGCTGTCCCCTCACACGCTTTGTTACGGTTCGGA
pI Seg.2	581: GATCGCGGCCGCACCACAGATTCGGCATCCT 582: GATCCTCGAGTGGCGGTCTGGGTATAGG
3bp-sub	600:CTATTTTTCTGCAAAAATAAGGATTACTATATTCGTGCCACCCG CATCTGTACGCAAT 601: ATTGCGTACAGATGCGGGTGGCACGAATATAGTAATCCTTATTTTT GCAGAAAAATAG
Top HP	598:GAGAGAACGGGTGGGAGCGTTAACCGCCGTGAGACCTCGAC 599: GTCGAGGTCTCACGGCGGTTAACGCTCCCACCCGTTCTCTC
4bp	596: CAATTTTCTGGTCTCACCCCACTCGAGGTCTCACGGCGG 597: CCGCCGTGAGACCTCGAGTGGGGTGAGACCAGAAAATTG
7BP Loop	606:GGAGCCGAGGCCGACCTCGACTCC 607:GGAGTCGAGGTCGGCCTCGGCTCC
3BP Loop	608:AGTCGAGGTCTCGCGGCCTCGGCT 609: AGCCGAGGCCGCGAGACCTCGACT

QRT-PCR

Total RNA was DNase treated and reverse transcribed using the Quantitect Reverse Transcription kit (Qiagen). Quantitative PCR was performed using the LightCycler 480 Probes Master Mix (Roche) along with IDT hydrolysis probes specific for the intron locus RNAs and selected housekeeping genes (Table 3). Ct values were determined using the LightCycler 480 (Roche) software and the Basic Relative Quantification analysis. Primer-probe efficiencies were determined by three biological replicates of 10-fold dilutions. The 18S rRNA was used as a reference gene and the relative target levels were quantified by the pfaffl method that incorporates the calculated primer-probe efficiencies [117].

WESTERN BLOT ANALYSIS

Mock-infected and MCMV-infected cells were trypsinized, centrifuged, and collected in PBS. Cells were lysed in RIPA buffer (150 mM NaCl, 1% v/v Nonidet P-40, 0.5% w/v deoxycholate, 0.1% w/v SDS, 5 mM EDTA, 50 mM Tris; pH 8.0) containing protease inhibitor cocktail (Roche). The cell lysate was briefly sonicated to facilitate nuclear protein release and insoluble debris was centrifuged. The soluble lysate was assayed for protein content using Bradford reagent (Bio-Rad). 30 µg of protein from each sample was separated by SDS-PAGE on a 10% gel and transferred to a nitrocellulose membrane (Whatman). Following the transfer, membranes were blocked with 5% milk in TBST for 1 hour then incubated with the primary antibody for 1 hour in 5% milk TBST. After washing, the membrane was incubated with secondary antibody in 5% milk TBST. GFP tagged m106 protein was immunobloted using a rabbit polyclonal antibody and detected with a fluorescently conjugated secondary antibody using the SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific). HP1 was detected similarly as a loading control (Santa Cruz).

ANIMAL MODELING

BALB/c mice were intraperitoneally innoculated with $5x10^6$ pfu of tissue culture derived wild type or recombinant MCMV in 300ul DMEM. At designated times mice were sacrificed and liver, spleen, lungs, kidneys, and salivary glands were removed. Part of the tissue was homogenized and titrated on mouse fibroblasts. When infectious virus could not be titrated in a particular organ, a titer corresponding to the limit of detection of the assay was assigned to that particular organ in order to calculate the median values.

EXONUCLEASE AND DEBRANCHING ASSAYS

Total RNA was isolated from either MCMV infected or pCDNA3.1+ transfected 10.1 fibroblasts. Total RNA was treated with 1 unit of Terminator (Epicentre) per 1 ug of total RNA for 60 minutes at 30°C. After incubation with Terminator, the reaction was halted using 1 ul of 100 mM EDTA pH 8.0. The RNA was then purified using an RNeasy column and prepared for qRT-PCR. Treated RNA was compared relative to untreated RNA. GAPDH was used to normalize the RNA samples; 18S rRNA was used as a positive control for degradation; total RNA isolated from pCDNA3.1 + LAT transfected cells was used as a negative control for degradation.

5 ug of total RNA was used for the debranching assay. Briefly, total RNA was isolated from MCMV infected or pCDNA3.1 transfected cells. 5 ug of total RNA was incubated at room temperature in the presence or absence of 1 uM yeast debranching enzyme (Dbr1) in a reaction mixture containing 50 mM MOPS pH 7.0, 10 mM $MnCl_2$, and 150 mM NaCl. Following debranching, the RNA samples were purified using an RNeasy column. Subsequent exonuclease degradation was carried using the protocol mentioned above. Dbr1 treated RNA was compared relative to Dbr1 untreated RNA.

RNA FOLDING PREDICTION SOFTWARE

All RNA secondary structure predictions were generated using the mFold web server [118].

CHAPTER III

MAPPING THE MCMV RNA7.2 LOCUS[2]

INTRODUCTION

Viral determinants of persistence remain poorly characterized for their contribution to the establishment and maintenance of persistent replication *in vivo*. Previous work identified the requirement for the MCMV RNA7.2 in promoting persistent replication using a mouse model of infection; recombinant MCMV lacking production of RNA7.2 is attenuated in the salivary gland of mice during the persistent phase of infection. Molecular studies of RNA7.2 defined this RNA as an intron processed from a larger primary transcript [106]. How the RNA7.2 locus gene products contribute to persistence remains elucidated. This chapter will further illustrate the topography of the MCMV RNA7.2 locus and attempt to explain the contribution of the RNA7.2 locus products to persistent replication over the course of infection *in vivo*.

The MCMV RNA7.2 is a unique lncRNA that is processed from its primary transcript as an intron. Introns are commonly thought of as a byproduct of the pre-mRNA splicing reaction because they are excised from the mature mRNA and are quickly degraded. Splicing occurs co-transcriptionally and is catalyzed by the cellular spliceosome machinery. The cellular splicesome machinery is composed of five snRNPs and numerous proteins that are drawn together by specific sequence elements located on

---

[2] The work presented in this chapter has been previously published and is used with permission. Schwarz TM, et al.: Molecular investigation of the 7.2 kb RNA of murine cytomegalovirus. *Virol J* 10(1), 348 (2013).

the pre-mRNA [119]. The different spliceosome components are assembled either onto the exon-intron junction splice sequences or onto the intron branch point sequence. The 5' splice donor sequence is located on the exon-intron boundary and typically is composed of the canonical eukaryotic sequence (C/A)AGGT(A/G) [120-122]. The 3' splice acceptor sequence is located on the opposite exon-intron boundary and consists of the canonical AGG sequence. The exons flanking an intron are combined by a two-step transesterification reaction resulting in the juxtaposition of the exons and excision of the intron from the primary transcript as a lariat structure. The 5' end of the intron typically starts with the GU from the splice donor sequence and the 3' end is composed of the AG sequence from the splice acceptor site [120]. Previous work defined RNA7.2 as an intron because it contains a canonical eukaryotic splice donor site that can be mutated to prevent RNA7.2 production and accumulation [106]. A consensus splice donor site nearly identical to that of the RNA7.2 is found in loci across CMV species indicating that this lncRNA is conserved as an intron [79, 106].

Processing of the primary transcript not only yields RNA7.2, but also a spliced mRNA from the exons flanking RNA7.2. This spliced mRNA encodes an open reading frame (ORF), m106, located entirely on exon two. m106 is predicted to be translated into a small, highly basic protein during productive viral infection [123]. A recombinant virus generated with a 5 base pair substitution mutation for the splice donor sequence was not observed to undergo processing of the intron locus primary transcript thereby precluding intron production, generation of the spliced mRNA and, possibly preventing translation of the m106 ORF. Although work presented in this chapter further explores m106

translational dependence on proper splicing of the RNA7.2 primary transcript, its functional relevance to viral infection and persistence remains unknown.

Recent analysis of the MCMV Smith strain identified two clusters of miRNAs. These miRNAs are located upstream of the splice donor site by approximately 500 to 700 nucleotides [124]. The function of these miRNAs remains unknown and they are not conserved across CMV species suggesting that there is not a selective pressure to maintain these ncRNAs. It is unclear what the relationship of these miRNAs is to the RNA7.2 locus and if they are produced from the same primary transcript. Therefore, this chapter will illustrate the boundaries of the RNA7.2 locus to resolve the transcripts that emanate from this region.

The work presented in this chapter represents a comprehensive analysis of the RNA7.2 locus. Current analyses focused on mapping the RNA7.2 locus for the temini of the primary transcript. Additional work describes the viral gene kinetic class to which these gene products are expressed that might assist in predicting their potential function. Although there is no evidence for translation of the intron itself, we discovered that the spliced mRNA encodes a small protein that co-localizes with the RNA within the nuclei of infected cells. Importantly, we show that the RNA is not required for trafficking of virus to the salivary gland *in vivo,* supporting our hypothesis that the 7.2 kb RNA functions to either evade the host response or maintain viral replication at sites of persistence.

RESULTS

The MCMV RNA7.2 Locus is Transcribed With True Late Kinetics

To determine the transcription kinetics of the RNA7.2 locus during productive MCMV infection, northern blot analysis was performed on total RNA prepared from cells treated with the translation inhibitor cycloheximide (CHX) or the DNA replication inhibitor phosphonoacetic acid (PAA) prior to MCMV infection. Cycloheximide pre-treatment of cells inhibits translation of immediate early (IE) genes blocking subsequent transcription of both early and late classes of viral genes. PAA treatment blocks DNA replication, on which expression of late (L) genes are dependent [7, 19]. Transcription of the intron (Figure 5) and the spliced mRNA (Figure) was inhibited by both cycloheximide and PAA treatment [107]. This data indicates that the intron locus RNAs are transcribed with the late class of viral genes during productive infection in fibroblasts.

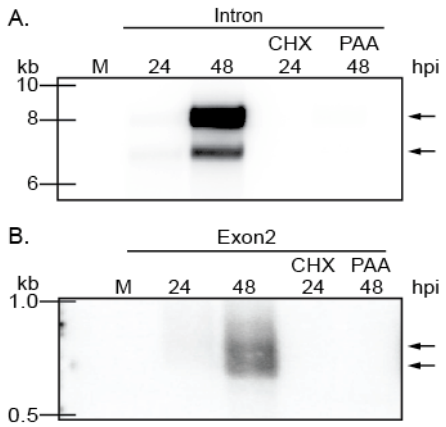


Figure 5. Expression Kinetics of Intron Locus Transcripts.
Total RNA was harvested from infected mouse fibroblasts (MOI=1.0) at the indicated times and analyzed by northern blot analysis using radio-labeled, antisense RNA probes specific for (A) the intron or (B) exon 2 of the spliced mRNA. To determine the kinetics of intron locus expression, cells were pre-treated with either CHX or PAA for 1 hour prior to infection to block either E or L gene expression respectively. Total RNA was prepared at the indicated times.

In high resolution northern blot analyses specific for the intron RNA, we routinely observed a doublet of bands near 7.2 kb: a major species at approximately 8.0 kb and a minor species migrating faster at 7.2 kb (Figure 5). These observations were made with multiple intron-specific probes (data not shown). We have been unable to ascertain the basis for this difference in size, although we hypothesize it may be due to effects of lariat secondary structure on RNA migration during electrophoresis resulting in slower migration (data not shown). Likewise, we also observed a doublet of closely migrating bands in northern blot analyses of the spliced RNA product of the locus (Figure 5). We cannot account for the difference in size based on sequencing of the 5' and 3' Rapid Amplification of cDNA Ends (RACE) products (see below). It is possible that we did not capture both species in the RACE reactions but we think it is likely the differences in size reflect variability in 3' end processing and poly-adenylation that we cannot assess.

Location of Transcriptional Start Sites and RNA Processing Signals

While the splice donor and acceptor sites used in the processing of RNA7.2 from the primary transcript were previously mapped, the termini of the primary transcript were not determined [80, 106]. We therefore sought to identify the termini of the primary transcript to further understand biogenesis of RNA7.2. To identify the 5' and 3' ends of the primary transcript, we cloned and sequenced PCR products generated by RACE (summarized in Figure 6). To identify the 5' end of the precursor RNA and capture the predicted intron-exon junctions in the 5' RACE reaction, we used nested PCR primers specific for the predicted second exon located 3' of the intron (primers 541 and 542 Table 1, Figure 6). Sequencing of cloned RACE products identified two transcriptional start sites located three nucleotides apart at positions 161,738 and 161,735 in the MCMV

genome (sequence coordinates based on the MCMV Smith strain, Genbank accession #NC004065). Sequence alignment of the 5'-RACE products to MCMV genomic sequence also confirmed the splice donor (SD) and splice acceptor (SA) sequences at nucleotide positions 161,622 and 154,366, respectively, as previously annotated. We also performed primer extension analysis to confirm the transcriptional start sites identified by 5'-RACE. We observed two primer extension products of 101 and 104 nucleotides in length, consistent with the location of the 5' ends of the spliced RNA as defined by RACE (primer 497, Figure 6C).

A single 3' end was identified at nucleotide position 153,872 by sequencing of 3'-RACE products (Figure 6A). This end is located downstream of a putative polyadenylation signal at position 153,898 (Figure 6B). We also examined the polyadenylation status of the spliced RNA by northern blot analysis of oligo(dT)-selected RNA prepared from MCMV-infected cells. The majority of the spliced mRNA from the intron locus was detected in the poly A+ fraction of RNA (Figure 6D). 18S rRNA can only be detected in the non-polyadenylated fraction demonstrating that our fractionation protocol efficiently captured polyadenylated mRNA only (Figure 6D lanes A^+ and A^-). Taken together, our data suggest that a large precursor RNA is transcribed from the intron locus at late times of infection and processed to yield a single, 7.2 kb stable intron and a spliced poly-adenylated mRNA consisting of two exons.

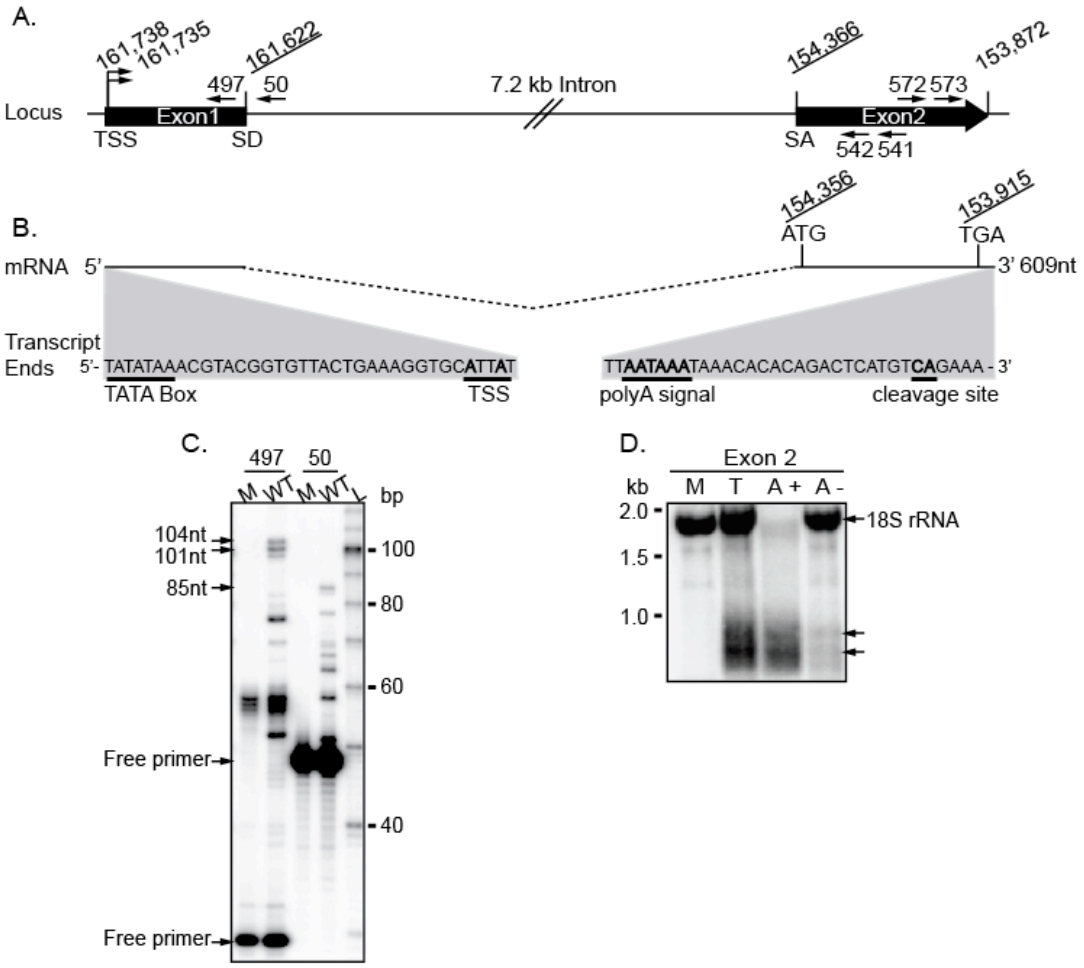


Figure 6. Mapping of the MCMV RNA7.2 Locus.
A) Diagram of the genomic region encompassing the primary transcript of the 7.2 kb intron, illustrating the transcriptional start sites (TSS) and the splice donor (SD) and splice acceptor sites (SA). The location of primers used for primer extension or RLM-RACE are indicated. B) Diagram of the spliced mRNA and the 5' and 3' ends that were identified by RLM-RACE using total RNA harvested from infected mouse fibroblasts at 48 hours post infection (hpi). Putative TATA box directly upstream of the transcriptional start sites is indicated. Also shown are the start and stop codons of the m106 ORF encoded on the second exon of the mRNA as well as the poly-adenylation and cleavage sites used in processing this mRNA. C) Primer extension analysis was performed on total RNA harvested from either mock-infected (M) or MCMV (WT) infected mouse fibroblasts at 48 hpi. Radiolabeled primers were used to validate the 5'RLM-RACE products (primer 497) and confirm the known 7.2 kb splice donor site (primer 50) as a control. D) Northern blot analysis of total RNA (T) harvested from infected mouse fibroblasts and fractionated for either polyadenylated (A+) or non-polyadenylated (A-) RNA. The blot was hybridized with a radiolabeled, antisense RNA probe specific for exon 2.

The m106 Open Reading Frame is Translated During Infection

The second exon of the spliced mRNA processed from the primary transcript that produces RNA7.2 spans a previously annotated open reading frame (ORF) called m106 [123]. Positional orthologues of m106 have been identified in all CMVs, including the UL106 ORF of HCMV, yet there is little sequence homology among the group [12, 125-127]. In general, UL106 orthologues score poorly with algorithms designed to predict the potential of an ORF to encode a protein [128, 129]. To determine if m106 protein is translated during MCMV replication we constructed two recombinant viruses engineered to express m106 as a GFP fusion at the carboxy-terminus (Figure 7A). The first recombinant virus expresses the m106-GFP fusion from the wild-type MCMV genome (MCMV:m106GFP). The second recombinant virus expressing the m106-GFP fusion also contains a five-nucleotide substitution at the splice donor site that defines the 5' end of RNA7.2 (MCMV*del*SD:m106GFP). This substitution prevents processing of the intron from the primary transcript and we predicted that it would also prevent translation of m106-GFP protein. Both recombinant viruses replicate with wild-type kinetics in multi-step growth analysis in mouse fibroblasts (Figure 7B). Immunoblotting for the m106-GFP fusion protein with antibody specific for GFP only detected protein expression during MCMV:m106GFP infection and not MCMV*del*SD:m106GFP infection, indicating that splicing of the mRNA is necessary for translation of m106 (Figure 7C)[107]. Furthermore, this data indicates that cryptic transcriptional initiation does not appear to occur within the unspliced transcript produced by MCMV*del*SD:m106GFP at any significant level.

Although we previously demonstrated in fractionation studies that the HCMV 5-kb intron localizes to the nuclear compartment of infected cells, the specific sub-nuclear localization of the RNA was not examined [79]. Fluorescent in situ hybridization (FISH) was used to visualize the 7.2 kb intron in infected mouse fibroblasts. The FISH staining revealed an even, granular distribution of the 7.2 kb intron throughout the nuclear compartment of infected fibroblasts (Figure 7D). Co-staining for the 7.2 kb RNA and m106-GFP revealed that the RNA and m106 protein are found co-localizing in the nucleus late during infection (Figure 7D). We also observed some m106-GFP protein was localized to the cytoplasm of infected cells [107].

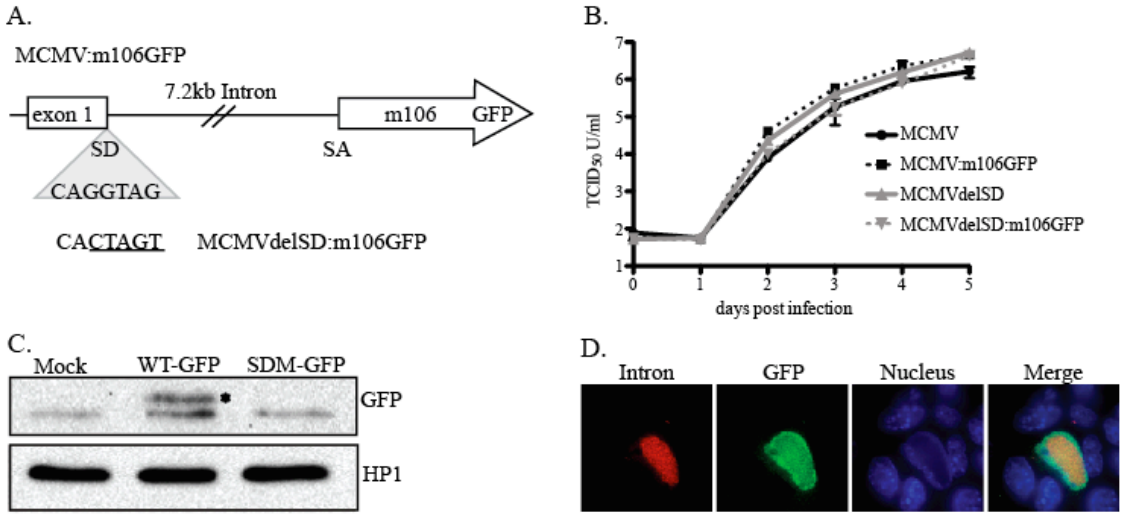


Figure 7. The m106 Open Reading Frame is Translated During MCMV Infection. A) Diagram of the GFP cassette insertion within the MCMV Smith BAC clone. B) Analysis of recombinant virus replication in mouse fibroblasts. Cells were infected at a multiplicity of 0.05 PFU/cell and viral supernatants were collected daily and titrated. Graph represents two biological replicates. C) Western blot analysis of protein lysates prepared from mock-, MCMV:m106GFP (WT-GFP), or MCMV*del*SD:m106GFP (SDM-GFP) infected cells at 48 h p.i.. m106-GFP expression was detected using a polyclonal anti-GFP antibody. Asterisk denotes m106-GFP. D) Mouse fibroblasts were infected with MCMV:m106-GFP (MOI = 0.05). Cells were fixed at 72 h p.i. and m106-GFP protein expression and 7.2-kb intron production was detected by by combined immunofluorescence assay and FISH.

Intron Locus Products Do Not Influence Dissemination To The Salivary Gland

Although we have shown that recombinant viruses that fail to express the intron replicate poorly in the salivary glands, it was unclear if this attenuation was caused by a lack of dissemination to or a failure to replicate within the salivary gland [106]. To examine whether the intron is required for dissemination to the salivary gland, mice were inoculated with wild-type MCMV or MCMV*del*SD and viral yields in various tissues were measured at 4, 6, 8, and 14 days post infection (Figure 8). MCMV*del*SD replicated to similar levels as wild-type MCMV until 6 days post infection in all tissues examined. At 8 days post infection, levels of the splice donor mutant virus were significantly reduced in the liver, kidney, and spleen but were unchanged in comparison to wild-type MCMV within the salivary gland and lung. By 14 days post infection, replication of the splice donor mutant virus was severely attenuated in all organs assayed and infectious virus was below the limit of detection by plaque assay.

Interestingly, the relative number of MCMV*del*SD genomes was reduced 100-fold in salivary glands at 14 days post infection relative to wild-type MCMV, suggesting that the virus was effectively cleared from this tissue (Figure 9) [107]. This data indicates that the intron does not influence viral dissemination to the salivary gland over a time course of infection but may function to promote viral persistence in the glandular epithelial tissue.

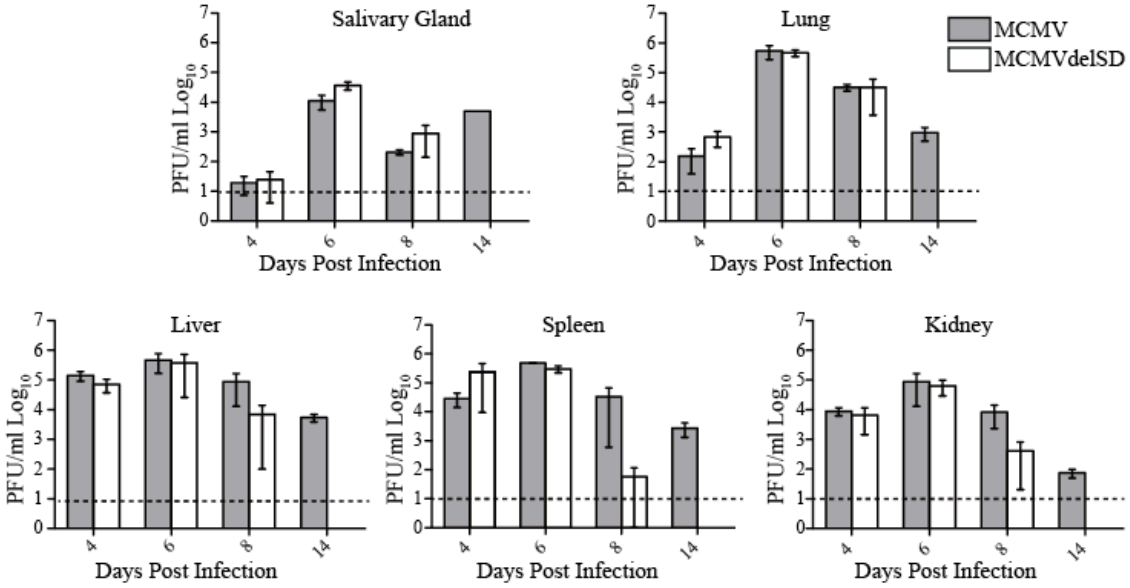


Figure 8. Intron Locus RNAs Are Dispensable For Virus Dissemination to the Salivary Glands During Acute Infection.
Three-month-old female BALB/c mice were inoculated intraperitoneally (i.p.) with $5x10^5$ PFU of WT MCMV or MCMV*del*SD. At the indicated days post infection, organs were harvested from three mice per infection group to quantify infectious virus by plaque assay. Organs were homogenized in a specific volume of DMEM treated with penicillin and streptomycin depending on mass (spleen 2 mLs, lung 2 mLs, liver 5 mLs, kidneys 3 mLs, salivary glands 2 mLs). Data is presented as the amount of infectious virus per mL of media the respective organ was originally homogenized in. Bars represent the mean and error bars represent the standard error of the mean (SEM). The dashed line indicates the limit of detection.

DISCUSSION

In this chapter, we characterized the MCMV RNA7.2 locus in finer detail. We demonstrated that during productive infection in fibroblasts, the RNA7.2 locus RNAs are transcribed with true late gene kinetics. The termini of the precursor RNA that is processed to produce the RNA7.2 and spliced mRNA were identified. Identification of the transcriptional start sites rules out the possibility that a cluster of miRNAs mapped upstream of the RNA7.2 splice donor site originate from the same primary transcript. It remains unknown, however, what functional relationship the miRNAs may have with the MCMV RNA7.2 locus, if any, during virus replication and pathogenesis.

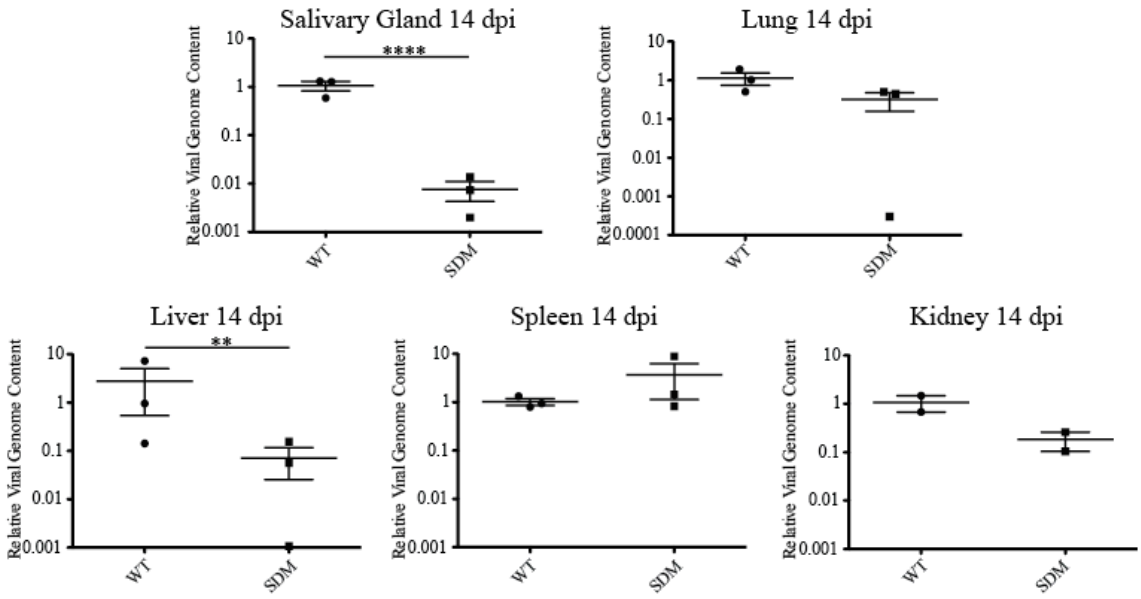


Figure 9. Detection of Viral Genomes is Significantly Reduced in Salivary Glands During Persistence.
Three-month-old female BALB/c mice were inoculated i.p. with $5x10^5$ PFU of WT MCMV (WT) or MCMV*del*SD (SDM). DNA was harvested from the indicated organs of three mice per infection group at 14 days post infection and viral genomes were quantified by qPCR using a primer probe set specific for the M54 MCMV gene and normalized to the beta-actin cellular gene. Bars represent the mean and error bars represent the standard error of the mean (SEM). *p* values represent the Student's T Test result between WT MCMV and MCMV*del*SD infected mice (* $p<0.05$ ** $p<0.01$ *** $p<0.001$ **** $p<0.0001$).

We observed a doublet of bands with probes specific for RNA7.2 in northern blot analysis at 8.0 and 7.2 kb. While it is formally possible that these represent two different intron species, we did not detect evidence of alternative splicing that would produce a larger intron from the precursor transcript by sequencing of 5'-RACE products. Instead, it is likely that RNA7.2 remains in the form of a branched lariat after processing and therefore migrates more slowly during electrophoresis representing the 8.0 kb band. The 7.2 kb band likely corresponds to a linear form of RNA7.2 resulting from RNA nicking during the RNA extraction process or partial digestion from endoribonucleases. Similar

observations have been made for the LAT of HSV-1 when resolving this lncRNA, characterized as a lariat intron, by denaturing gel electrophoresis [130]. A doublet of bands corresponding to the spliced mRNA is also detected by northern blot analysis. Again, sequencing of 5'-RACE products did not reveal any splicing variations that could account for the spliced mRNA size differences. Alternatively, polyadenylation chain lengths could differ for the individual spliced mRNA molecules representing the mRNA doublet.

The spliced mRNA produced by processing of RNA7.2 spans the m106 ORF. Using an epitope-tagging strategy, we showed that this ORF could be translated during MCMV infection. The GFP-tagged protein co-localized to the nucleus of infected cells with RNA7.2. This may reflect a related function of the intron and the m106 protein. Previous studies suggest that m106 is not critical for establishment of persistent replication however. Using a recombinant virus that disrupts the stability of RNA7.2 but not splicing of the mature mRNA or translation of the m106 ORF demonstrated a persistent replication defect phenotype similar to MCMV*del*SD in mice. Recombinant viruses that specifically disrupt m106 expression without impacting intron production will be useful reagents to investigate the function for this viral protein further.

The m106 protein and its orthologues encoded by other CMVs, including UL106 of HCMV, have some unusual properties. Despite not sharing significant sequence homology, all UL106 orthologues are small (<150 amino acids), highly basic, arginine-rich peptides. It is unknown if other UL106 orthologues are expressed during infection, but given the conservation of the genomic organization of the intron locus among CMVs, it is a distinct possibility to be explored.

Production of RNA7.2 is required for the establishment of persistence in the salivary glands of mice. By analysis of multiple time points between the acute and persistent phases of infection in mice, we showed that recombinant virus lacking the intron appears to disseminate to the salivary gland as efficiently as wild-type MCMV. However, it is unable to maintain a highly productive replication program in the salivary glands as observed at 14 days post infection. In addition, we did not detect infectious MCMV*del*SD in any organs at 14 dpi and genome copy number of the mutant virus was substantially reduced in liver and kidney. It is possible that the mechanisms that prevent establishment of intron-mutant virus persistence in the salivary gland may also promote accelerated clearance of that virus from liver and kidney. At this time, the adaptive immune response acts to limit viral replication and it is possible that RNA7.2 is involved in modulating immune surveillance in some way. Some cellular lncRNAs are involved in transcriptional regulatory processes, therefore, a possible mechanism for evading the immune response could be to regulate cellular or viral genes that are involved in this host pathogen relationship [84]. Although a function has yet to be determined for RNA7.2, this current analysis has provided the framework for investigating its function during viral persistence.

CHAPTER IV

TRANSCRIPTIONAL ANALYSIS OF THE MCMV RNA7.2 LOCUS[3]

INTRODUCTION

Previously, it remained unknown why MCMV RNA7.2 accumulated to detectable levels during productive infection *in vitro* [106]. Introns are degraded rapidly after excision from the primary transcript and MCMV RNA7.2, therefore, would not be expected to accumulate. The level of an mRNA within a cell is dependent on both its rate of synthesis and rate of decay. Although these two processes are not mutually exclusive, we sought to identify the putative MCMV RNA7.2 locus core promoter elements and queried their potential contribution to the transcriptional output of the MCMV RNA7.2 locus RNA products.

Genomic DNA serves two major purposes; as a repository of information that can be transcribed and/or translated into functional material for cell maintenance and survival and to provide regulatory sequences critical for processes such as replication and transcription. Transcriptional regulatory sequences are structurally and functionally distinct and can provide information on the type of gene that is being expressed [131-133]. As a general guide, tightly regulated genes have focused core promoters whereas constitutively expressed housekeeping genes have dispersed core promoters. The core promoter is a stretch of genomic DNA that consists of a variety of sequence elements that

[3] The work presented in this chapter has been previously published and is used with permission. Schwarz TM, Volpe LA, Abraham CG, Kulesza CA: Molecular investigation of the 7.2 kb RNA of murine cytomegalovirus. *Virol J* 10, 348 (2013). *Virol J* 10, 348 (2013).

are recognized and bound by the RNA polymerase II preinitiation complex (PIC) proteins (Figure 10A). In a focused promoter, transcription initiation typically begins at a single nucleotide whereas in dispersed there are multiple transcriptional initiation sites over a range of sequence from 50 to 100 nucleotides [134-137]. Surrounding the transcriptional start site are the core promoter sequence motifs that are recognized and bound by transcription factors responsible for recruiting RNA polymerase II to form the PIC.

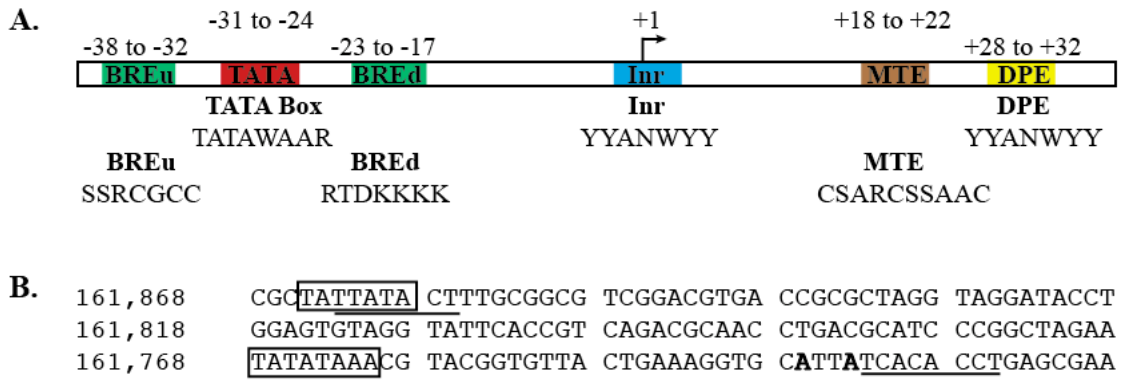


Figure 10. Core Promoter Elements.
A) Common eukaryotic core promoter elements with their consensus sequences and relative positions to the transcription start site (+1). B) Predicted core promoter sequences of the MCMV RNA7.2 locus relative to the annotated transcriptional start sites (bold). Boxed sequence = putative TATA boxes; underlined sequence = putative Inr elements.

There are a variety of core promoter sequence motifs and not all are included in each promoter region. Common sequence motifs include the Initiator (Inr), the TATA box and TFIIB recognition element (BRE), the downstream core promoter element (DPE) and the motif ten element (MTE). All of these elements have canonical sequences that are commonly found in a specific location relative to the transcriptional start site (Figure 10A). Genes that are tightly regulated commonly have strong core promoter elements consisting of a canonical TATA box and/or initiator sequence whereas a dispersed core

promoter might have several, weaker non-consensus motifs [132, 138, 139]. The principle behind this sequence structure is explained by the ease at which focused promoters can control gene expression of one transcriptional start site versus controlling multiple transcriptional start sites over a range of nucleotide sequence. Given that the primary transcriptional start sites for MCMV RNA7.2 are located within two nucleotides of one another, it is reasonable to assume that this locus is tightly regulated and has a focused core promoter region.

The CMV genome contains transcriptional regulatory sequences controlling viral gene expression similar to its host. The structure and function of a CMV promoter region is generally dependent on the kinetic class the gene belongs to. Viral genes are transcribed by cellular RNA polymerase II in the nucleus of infected cells and are expressed in a temporally controlled cascade of three classes of viral genes; the immediate-early (IE), early (E), and late (L) genes. Expression of the three viral gene classes are dependent on the stage of infection; IE genes are transcribed soon after entry into a host cell, the E genes are reliant on the expression of IE transactivators for their transcription, and L genes require viral genomic replication in addition to IE and E viral transactivator expression [19]. CMV gene expression initiates from a few IE proteins shortly after viral entry into a host cell without de novo protein expression. These IE genes are under control of the highly responsive and strong, major immediate early promoter (MIEP). This promoter spans a 500-bp stretch of sequence and contains many eukaryotic core promoter motifs in addition to repeat elements with transcription factor binding sites for NF-kB, AP-1, Sp1, and CREB/ATF [140]. The MIEP contains an

abundance of transcriptional activation sequences since IE gene expression is required for the viral replication cycle to begin.

Once the IE transactivators IE86 and IE72 are expressed, they can stimulate E and E-L gene expression either synergistically or individually [141-143]. E gene promoter regions have focused core promoters typically containing IE transactivator-binding sites upstream from a TATA box sequence [143]. The IE transactivator proteins interact with factors of the basal-transcription machinery and promoter-specific transcription factors. It is hypothesized that IE86 in particular may act as an adapter protein that stabilizes basal and specific transcription factors on promoter regions [7, 19]. Structure and function of L gene promoters are less understood. It is known that several IE and E transactivators are needed for their transcription. IE86 and IE72 both are known to be critical for activation of the UL83 L gene promoter, however, the IE genes alone are not sufficient for activation of all true L genes.

Recently, three E genes were implicated in the activation of L gene transcription: UL79, UL87, and UL95 [22-25, 144, 145]. These genes have homology to ORFs 18, 24, 30, and 34 of gammaherpesvirus 68 (MHV-68) for late gene transcription [20, 146-148]. All three HCMV genes have no affect on viral replication but are essential for a subset of L genes. There is no evidence for these genes in controlling HCMV RNA5.0 transcription nor is their evidence for their MCMV homologues in controling MCMV RNA7.2 [144]. There still remains a paucity of information describing true L gene promoters and it is unknown what sequence elements and factors drive transcription of the MCMV RNA7.2 locus.

This chapter will attempt to illustrate the putative core promoter elements that initiate transcription of the MCMV RNA7.2 locus. Several eukaryotic core promoter consensus sequences were identified within the vicinity of the transcriptional start sites. The transcriptional activity of these sequences was queried by either cloning these regions into a reporter plasmid or by generating recombinant viruses with deletions in these regions. The data remains inconclusive for identifying definitive transcriptional control elements but validates the difficulty of investigating L gene transcription.

RESULTS

Identification of Minimal Promoter Elements

Little data is available regarding the sequence elements driving late transcriptional units of Cytomegaloviruses. Therefore, the DNA sequence within the vicinity of the recently identified transcriptional start sites of the intron locus was examined for transcriptional regulatory sequences. A putative TATA box element was identified starting at 31 nucleotides upstream of the transcriptional start site, as well as an additional TATA box-like sequence located 127 nucleotides upstream (Figure 10B). In addition, two canonical initiator sequences are located either within the upstream TATA box sequence or just downstream from the transcriptional start sites (Figure 10B). In order to examine the transcriptional activity of these putative minimal promoter elements and the nucleotide sequence surrounding them, DNA sequences from the region between the intron splice donor site and the M112/113 locus were cloned into the pGL3-Basic luciferase reporter plasmid to quantify transcriptional promoter activity (Figure 11A).

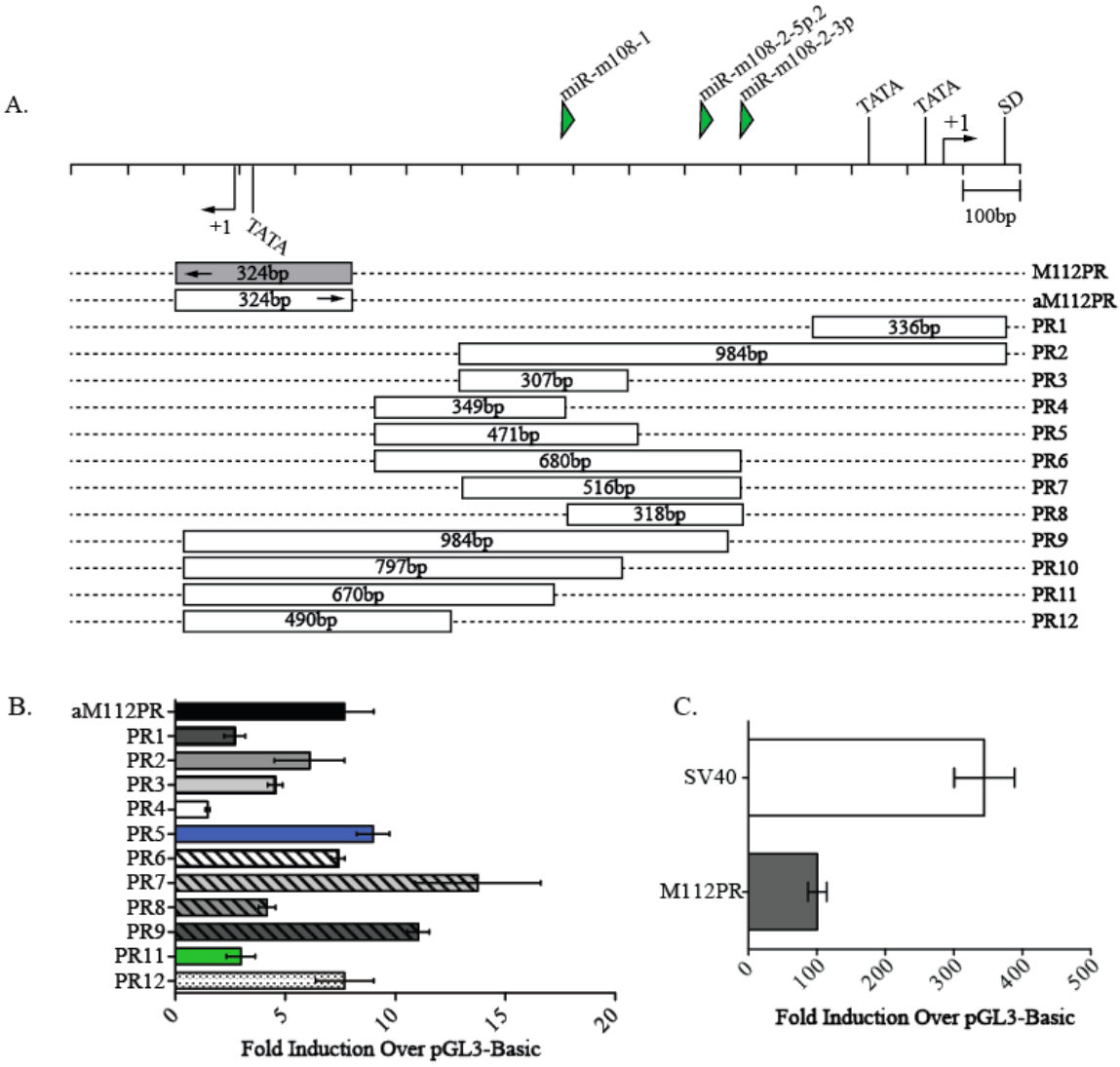


Figure 11. Analysis of Transcriptional Activity of Putative Intron Locus Promoter Elements.

A) Diagram of genomic location of the viral sequences cloned into pGL3-reporter plasmids. Schematic representation of the genomic region spanning upstream of the MCMV RNA7.2 splice donor sequence to the M112/113 promoter region that lays on the opposite DNA strand. M112PR represents the genomic sequence cloned into the pGL3-reporter construct that includes the M112/113 promoter region is the sense orientation whereas aM112PR represents the same genomic sequence cloned in the antisense orientation respective the M112/113 locus. The PR1- PR12 regions indicate genomic sequences that were cloned into the pGL3-reporter construct in the sense orientation respective to the RNA7.2 locus. B) pGL3-reporter constructs were co-transfected into mouse fibroblasts and luciferase induction was assayed 24 h p.i. C) Analysis of control pGL3 constructs transfected into mouse fibroblasts.

The M112/113 locus is located upstream of the intron locus and on the opposite strand. The transcriptional start site and promoter elements controlling the M112/113 locus have been defined [149, 150]. The promoter element consists of a canonical TATA box sequence that has demonstrated strong transcriptional activity in the forward direction. Therefore, the M112/113 early-late promoter was cloned into the reporter vector in both sense and antisense orientations, M112PR and aM112PR respectively (Figure 11A).To serve as a positive control for reporter induction, only the M112/113 promoter sequence in the sense orientation induced luciferase activity similar to the SV40 promoter control pGL3 plasmid (Figure 11B-C). Surprisingly, no appreciable induction of luciferase activity was observed for any of the cloned pGL3 constructs relative to the aM112PR control vector (Figure 11B-C). We also found that MCMV co-infection of reporter-tranfected cells did not induce luciferase activity from the minimal PR1 promoter construct that encompasses the sequence within the vicinity of the 5' transcriptional start site (Figure 12).

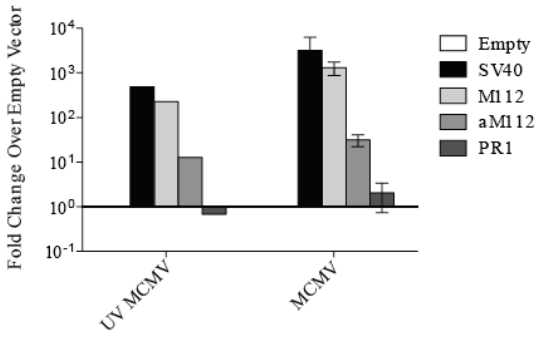


Figure 12. Analysis of Transcriptional Activity of Putative Intron Locus Promoter Elements.
A) pGL3-reporter constructs were co-transfected into mouse fibroblasts and luciferase induction was assayed 24 h p.i. following either mock co-infection, B) UV inactivated co-infection, or MCMV co-infection (MOI = 0.05).

In Vitro Examination of Minimal Promoter Elements

To examine the contribution of the putative minimal promoter sequences to transcription of the intron locus RNAs in the context of virus infection, three recombinant viruses were made with deletions in this region (Figure 13A). We constructed recombinant viruses with (1) a 20 bp deletion spanning the TSS and proximal TATA box-like sequence, (2) a 100 bp deletion including the proximal TATA box-like sequence and upstream sequence, and (3) a 135 bp deletion spanning both the proximal and distal TATA boxes in the putative minimal promoter sequence. All three recombinant viruses replicated similar to wild-type MCMV in multi-step growth analysis (Figure 13B). We quantified RNA expression in cells infected with our panel of recombinant viruses, including previously characterized recombinants with mutations that result in a failure to express significant levels of the intron (MCMV*del*HP and MCMV*del*SD) [106]. The MCMV*del*HP contains a 28 bp deletion spanning a predicted stem loop structure at the 3'end of the intron that is hypothesized to confer stability. Without this stem loop structure, processing of the primary transcript still occurs since the mRNA is detected by qRT-PCR and northern blot analysis, but accumulation of the intron is significantly reduced (Figure 13B-D). While the splice donor site mutation impacts processing of the precursor transcript and is not expected to affect transcriptional output of the promoter, we predicted that MCMV*del*135 would reduce overall transcript production from the locus. Measured reductions in levels of RNA7.2 were only significant for cells infected with the MCMV*del*135 mutant in addition to the MCMV*del*SD and MCMV*del*HP recombinant viruses (Figure 13B). In cells infected with MCMV*del*135, RNA7.2 abundance was reduced approximately 5-fold while levels of the mature mRNA

transcript were reduced by 10-fold (Figure 14B-D). Despite different predicted consequences of the mutations, the reduction of the spliced mRNA transcript abundance is similar between the MCMV*del*SD and the MCMV*del*135 suggesting that the half-life of the spliced mRNA is shorter than the half-life of RNA7.2.

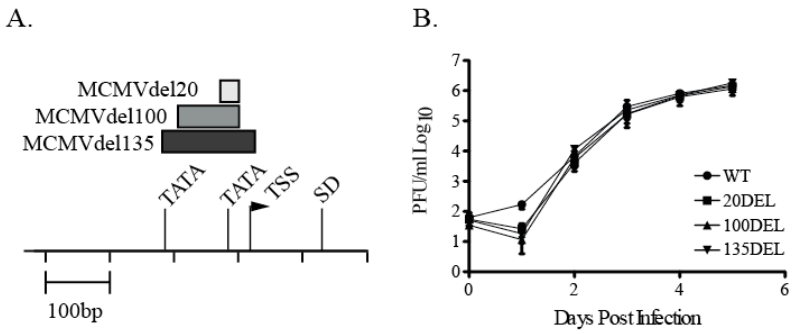


Figure 13. Recombinant Viruses For Putative Promoter Deletions.
A) Diagram of the genomic location of intron locus promoter region deletions. B) Multi-step growth analysis of replication of all three recombinant viruses in comparison to WT MCMV. Mouse fibroblasts were infected (MOI = 0.05) and culture supernatants were collected every 24 hours and titrated by plaque assay. Graph represents three biological replicates.

In Vivo Examination of Minimal Promoter Elements

To determine if reductions in intron and mRNA expression have an effect on the establishment of persistence *in vivo*, mice were inoculated with a subset of our panel of recombinant viruses and viral yields measured in the salivary gland at 14 days post-infection (Figure 14E). We observed a slight reduction in viral yield in the salivary glands of mice infected with MCMV*del*20 and a ten-fold reduction of viral yield in mice infected with MCMV*del*135. Viral genome quantification corroborated the measure of

infectious virus within the salivary gland (Figure 14F). However, despite 5-10 fold reductions of intron and mRNA production, neither promoter deletion mutant fully attenuated persistent replication to the levels observed in mice infected with MCMV*del*SD.

DISCUSSION

Our studies did not identify sequence elements that robustly contribute to the transcriptional control of the MCMV RNA7.2 late transcriptional unit. In attempt to identify the MCMV RNA7.2 locus core promoter elements, we analyzed sequence near the identified transcriptional start sites for reporter activity using a transient transfection assay. There was not an appreciable induction of reporter activity in comparison to the m112/113 promoter region or the CMV MIEP control plasmid. Surprisingly, induction of reporter activity was not observed from any of the transfected constructs with subsequent viral infection. This information lead us to believe that relying on transient transfection assays to study promoter regulatory elements might not account for the wide range of temporal variation the late class of genes experience within a CMV-infected cell. In addition, the transfection-infection experiments might not be providing us with accurate information regarding the regulatory events that occur during a normal viral infection. There may also be template-specific differences between plasmid DNA and viral DNA that would prohibit activation of late gene transcription.

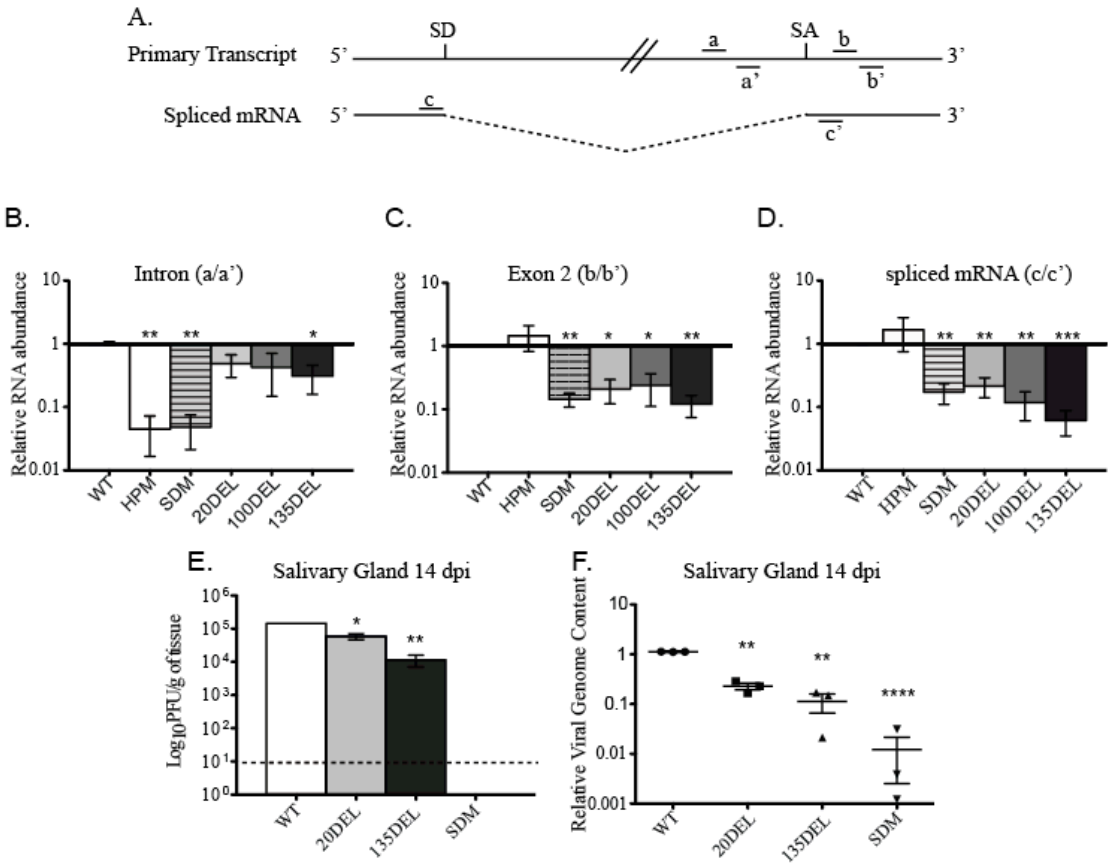


Figure 14. Deletion Mutations in Putative Viral Promoter Elements Reveal Reduction in Transcriptional Output in Cell Culture and Decreased Recovery of Infectious Virus *In Vivo*.

A) Diagram of qPCR primer-probe sequence locations used to quantify the RNA7.2 locus transcripts. B-D) Quantification of intron locus RNAs in cells infected with promoter mutant viruses. Mouse fibroblasts were infected (MOI=1.0) and total RNA was harvested 48 hours post infection. Intron locus transcript levels are quantified relative to WT MCMV transcript levels by qRT-PCR. Graphs represent three biological replicates. E-F) Three-month-old female BALB/c mice were infected with an i.p. dose of 5 x 10^5 PFU with the indicated viruses. At 14 days post infection, animals were euthanized and tissues collected for analysis of infectious virus yield (E) and viral genome number (F). (E) Salivary gland homogenates were analyzed by plaque assay on mouse fibroblasts. *p* values represent the Student's T Test result between WT MCMV infected cells or mice and cells or mice infected with the given recombinant viral mutant for each transcript analyzed(* $p<0.05$ ** $p<0.01$ *** $p<0.001$ **** $p<0.0001$). WT MCMV = WT; MCMV*del*20 = 20DEL; MCMV*del*100 = 100DEL; MCMV*del135* = 135DEL; MCMV*del*HP = HPM; MCMV*del*SD = SDM.

Because the reporter assay results remained inconclusive, we generated recombinant viruses for the suspected promoter elements to query transcriptional output from these sequences and determine if the persistence phenotype would be disrupted in virally infected mice. By deleting a stretch of sequence including both TATA box-like elements (MCMV*del*135), RNA7.2 levels were modestly reduced in comparison to the spliced mRNA. This data suggests that the half-lives differ between RNA7.2 and the spliced mRNA. From this data, we can also infer that transcriptional output does not contribute to the accumulation of RNA7.2; although transcription was not dramatically reduced by the 135 base pair deletion, basal transcription still allowed for the highly stable RNA7.2 to accumulate. This revelation can explain why this recombinant virus was unable to produce a similar persistence replication defect as MCMV*del*SD in mice since RNA7.2 is still produced and accumulates. The *in vivo* data provides additional evidence that m106 does not play a critical role for persistent replication in the salivary gland of mice.

Our difficulty in identifying a core promoter element largely responsible for transcription of the MCMV RNA7.2 locus is likely due to the nature of late gene expression. Current data in the literature does not allow any general conclusions regarding CMV true late gene promoter regulation to be drawn. Structurally, those L gene promoters that have been investigated are reported as simple, or focused, likely rendering them less competitive for cellular and viral transcription factors. Therefore, it is hypothesized that only through increased concentration of promoter sequences by DNA replication that late genes are capable of successfully competing for the factors necessary to initiate transcription [7]. DNA replication also allows for viral DNA-associated

histones to be modified causing late gene promoters to become accessible for transcription factors. Current evidence therefore suggests that multiple mechanisms influence the strict temporal activity of true late gene transcription.

Recently, it has been demonstrated that viral replication and L gene expression also relies on a distinct set of five genes conserved across beta and gamma herpesviruses: UL79, UL87, UL91, UL92, and UL95. It is hypothesized that an RNA polymerase II transcriptional complex including one or more of these gene products is assembled to drive transcription of L genes. MCMV homologs of HCMV UL87, UL91, UL92, and UL95 have been annotated, but not tested for transcriptional activating functions. M79, the MCMV homolog of HCMV UL79, has been shown to regulate L gene expression, although it does not appear to promote transcription of the intron locus [144]. The complete array of viral proteins necessary for activation of a late viral promoter requires further investigation. Clearly, transcriptional regulation of L genes remains largely unexplored and is significantly different from the activation of immediate early and early transcriptional units.

CHAPTER V

STABILITY ANALYSIS OF THE MCMV RNA7.2[4]

INTRODUCTION

Transcriptional regulation studies did not provide a mechanism that accounts for the accumulation of the MCMV RNA7.2. Although regulation of transcriptional output is largely responsible for changes in gene expression levels, RNA stability also provides a mechanism to regulate gene expression [151-153]. Therefore, the rate of RNA7.2 decay was investigated to determine if its accumulation is due to a long half-life. Until recently, little was known about post-transcriptional regulation, metabolism and function of lncRNAs let alone stable introns. Similar to cellular proteins, lncRNA half-lives vary and may be a reflection of lncRNA function and cellular location [151, 154-157]. The aim of this chapter, therefore, is to examine the basis for the accumulation of the MCMV RNA7.2 that will provide additional information of its functionality over the course of infection.

RNA7.2 is excised from its primary transcript as an intron [106]. RNA splicing proceeds through two successive transesterifications followed by the release of a branched, lariat-shaped RNA molecule (Figure 15). The 5' end of the intron is linked to an internal adenosine residue (branch point) by a 2'-5' phosphodiester bond forming the lariat shape [158]. Intron lariats are produced in an equal quantity to the spliced exons,

---

[4] The work presented in this chapter has been previously published and is used with permission. Schwarz TM, et al.: Molecular investigation of the 7.2 kb RNA of murine cytomegalovirus. *Virol J* 10, 348 (2013).

but normally they are degraded rapidly. Almost immediately following excision from the primary transcript, the cellular debranching enzyme hydrolyzes the 2'-5' phosphodiester bond of lariat introns leaving a 5' monophosphate vulnerable to exonucleic decay [159-161]. It remains unknown if RNA7.2 maintains this lariat form to avoid degradation.

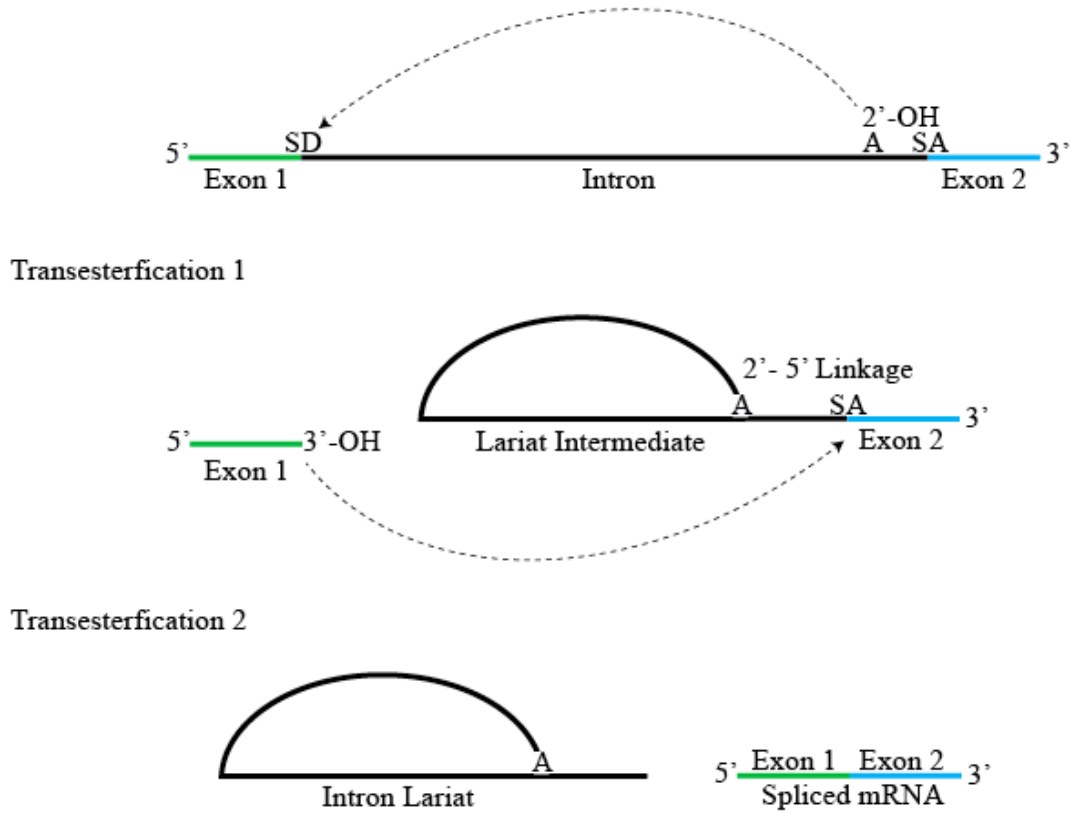


Figure 15. Biogenesis of an Intron Lariat and Spliced mRNA From a Primary Transcript.
RNA splicing proceeds through two successive transesterifications. The first transesterification is initiated by the nucleophillic attack on the splice donor sequence (SD) by the 2' hydroxyl group of an adenosine residue (A) within the branch point sequence. This leads to the formation of the 2'-5' phosphodiester linkage of the lariat intermediate structure and the excision of Exon 1 from the primary transcript. The second transesterification reaction is caused by the nucleophillic attack on the splice acceptor site (SA) by the 3' hydoxyl group of exon 1. This second reaction results in the release of the intron as a lariat structure and the juxtoposition of the two exons.

In addition to being characterized as an intron, other intrinsic properties of RNA7.2 would generalize it as an unstable RNA at first glance. RNA7.2 is composed of approximately ~61% A and T residues. High AT content typically leads to weaker secondary structures than high GC content thereby rendering RNA more susceptible to degradation. In addition to sequence content, both the MCMV RNA7.2 and HCMV RNA5.0 are retained in the nucleus of infected cells. Recent high-throughput analysis demonstrated that nuclear lncRNAs are typically less stable than cytoplasmic lncRNAs [155]. This general instability of nuclear lncRNAs can be attributed to possible functionality since it is known that many proteins that have regulatory roles in the nucleus generally have short half-lives at either the RNA or protein level. For example, the nuclear-localized lncRNA Neat1 is one of the least stable lncRNAs known [155]. The instability of Neat1 is suggested to contribute to the dynamic nature of the subcellular domain where it functions; paraspeckles [162]. This raises the possibility that the accumulation and stability of the CMV intron reflects its function. Unlike most nuclear lncRNAs, the CMV intron accumulates to detectable levels despite its biological features. Identifying determinants of stability for the CMV intron, therefore, is of critical importance for understanding functionality of this RNA over the course of infection.

Several sequence elements identified within the 3' region of RNA7.2 were hypothesized to confer stability of this transcript allowing for its observed accumulation. A 33-bp stem-loop structure located near the 3' end of RNA7.2 between the polypyrimidine track and putative branch point was identified using the structural prediction software mFold [106]. Deletion of this stem-loop does not impact processing of the precursor transcript, but does prevent accumulation of the intron during infection in

cultured fibroblasts [106]. In infected mice, the hairpin deletion causes a similar persistence phenotype as MCMV*del*SD further signifying that this structure is critical for the integrity of the RNA (Review Figure 4). Additional evidence that this hairpin structure is necessary for stable retention of the intron can be observed by its conservation across CMV species (unpublished data). Sequence analysis by mFold also suggests that a similar hairpin loop structure forms in the 3' region of the HCMV RNA5.0. The hairpin structure within the 3' region of RNA5.0 is considerably larger with a predicted 77 bp sequence (data not shown). It remains unknown how this hairpin structure confers stability to the MCMV RNA7.2.

The stability elements of the HSV-1 and HSV-2 LATs also reside in their 3' regions [163]. These transcripts are excised from their primary transcript as introns and remain in lariat conformation [88, 130, 164]. The lariat conformation of the LATs is predicted to account for their observed accumulation in host cells during lytic or latent infection as well as the long-half life of these ncRNAs [165-170]. Sequence elements located within the last 100 nucleotides of the LAT introns include a nonconsensus branch point sequence as well as a hairpin structure. A hairpin structure in the LAT 3' region is predicted to confer stability by influencing the selection of a nonconsensus branch point sequence. By mutating this hairpin sequence in the HSV-1 LAT, not only is the long LAT intron half-life reduced, but splicing efficiency is also negatively affected [163]. Since the nonconsensus branch point is located at the base of the hairpin structure, it is further predicted that the 2'-5- phosphodiester bond is protected from enzymatic debranching activity. In fact, studies suggest that the HSV-1 LAT is not sensitive to

debranching activity when exposed to S100 extract from HeLa cells indicating the important role this 3' region in conferring stability to LAT [171].

Similar to the HSV LATs, the CMV hairpin structure may serve several functions in protecting the CMV intron from degradation. Because this structure is located in the 3' region, it may be critical for preventing 3' to 5' exoribonucleic decay. If the lariat conformation is retained, the hairpin may also prevent debranching by providing steric hindrance. This steric hindrance would prohibit the debranching enzyme from accessing and binding to the 2'-5' phosphodiester bond thereby preventing hydrolysis of the branch. Lastly, the hairpin structure may confer stability to the intron by directing the spliceosome machinery to the preferred branch point sequence thereby ensuring the formation of a favorable conformation that will resist decay. Current data, however, does not indicate that this hairpin structure is important for efficient splicing since spliced mRNA levels are not compromised without this RNA structure as observed by northern blot or qRT-PCR analysis [106, 107].

The work presented in this chapter demonstrates that the accumulation and predicted stability of the CMV intron is due to a long, RNA half-life. This half-life is due primarily to sequence elements that are located within the 3' end of the MCMV RNA7.2 and is predicted to be similar across CMV species for this intron. From this data, we hypothesize that the intron remains in the form of a lariat, similar to the Latency Associated Transcript (LAT) of Herpes Simplex Virus-1 (HSV), thereby protecting it from degradation. Evidence for this nonlinear structure is supported by the insensitivity of the CMV RNAs to XRN1 and debranching enzyme. Together, this data demonstrates

that the stability of the CMV intron is due to RNA structure and sequence elements located within its 3' region.

RESULTS

The CMV Intron is Highly Stable

The MCMV RNA7.2 accumulates to high levels during infection as detected by northern blot analysis, suggesting it is unusually stable for an intron [106, 107]. To quantify transcript stability, we measured RNA decay rates of intron-locus transcripts during MCMV infection. RNA half-lives were quantified by measuring RNA abundance by quantitative RT-PCR at several time points after treatment of infected cells with Actinomycin D. This compound inhibits RNA Polymerase II by intercalating between GC residues thereby blocking processivity of the enzyme [115]. Actinomycin D treatment effectively arrests transcription of RNA pol II-dependent RNAs and allows us to measure relative decay rates over a time course. Using this strategy, we calculated the half-life of RNA7.2 to be ~28.8 hours (Figure 16B). In general, the half-life of low-stability RNAs is typically less than 2 hours whereas long-lived RNAs with high stability possess a half-life greater than 12 hours [155]. The half-life of the spliced mRNA derived from processing of the intron was measured using two different primer probe sets: b/b' targets the second exon and c/c' spans the splice junction. We measured a half-life of ~6.8 hours using the primer-probe set targeting the second exon, while a half-life of ~7.8 hours was measured using the primer-probe set that spans the splice junction (Figure 16C-D) [107]. This difference may be accounted for by differences in primer-probe efficiency, processing of the primary transcript, or location of the primer-probe sets in relation to protective secondary structures within the mRNA from degradation

machinery. Both measurements, however, are consistent with the average half-life for a protein coding RNA. As a control, we determined the half-life of GAPDH mRNA to be ~15 hours, similar to published estimates (Figure 16E) [155]. Our data demonstrates that the MCMV RNA7.2 is, in fact, unusually stable for an intron RNA.

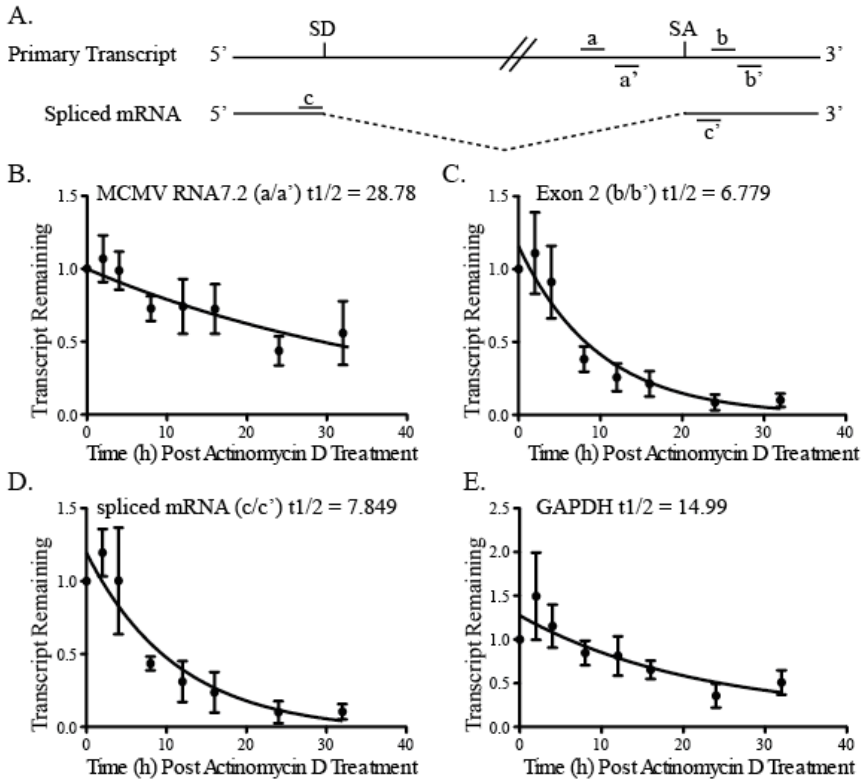


Figure 16. Half-life Analysis of the MCMV RNA7.2 Locus Transcripts.
Mouse fibroblasts were infected with MCMV (MOI = 1.0). At 30 hours post infection, cells were treated with 4ug/ml of Actinomycin D. Total RNA was harvested over the indicated time course. RNA7.2, mRNA, and 18S rRNA transcript levels were quantified by qRT-PCR. Intron and spliced mRNA transcripts were normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on four biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value. Bars represent the mean and error bars represent the standard error of the mean (SEM). (A) Schematic representation of primer probe sets used for qRT-PCR analysis. SD = splice donor sequence, SA = splice acceptor sequence. Half-life decay curves for the (B) RNA7.2 using primer probe set a/a', (C) the second exon of the spliced mRNA (b/b'), (D) the spliced mRNA using primer probe set c/c', and (E) GAPDH.

Similar to the MCMV RNA7.2, half-life analysis for the HCMV RNA5.0 also demonstrated high stability indicating that this feature is conserved across CMV species. The same transcriptional inhibition strategy was used to determine the half-life of RNA5.0. However, since RNA5.0 is transcribed with IE and L viral gene kinetics, transcription was inhibited by Actinomycin D treatment within infected human fibroblasts at two different time points; 12 h p.i. and 48 h p.i.. These two times points reflect the times that IE and L genes are transcribed during *in vitro* infection. Using this method, the half-life of the HCMV RNA5.0 is greater than 28 hours regardless of the viral kinetic class from which RNA5.0 is transcribed (Figure 17B). At this last time point of the assay, HCMV RNA5.0 has only been reduced by approximately 30%. The spliced RNA transcript that is predicted to encode UL106 has a shorter half-life, similar to the MCMV m106 spliced transcript (Figure 17C). Together, the half-life analysis of the CMV intron across CMV species demonstrates that this transcript is long-lived and its abundance is not due to transcriptional activity levels.

Determinants of CMV Intron Stability

Because we demonstrated that the CMV intron accumulates as a consequence of a slow decay rate, we investigated if specific RNA sequence elements contribute to the stability of the intron. Published work demonstrated that a stem-loop structure located near the 3' end of the intron between the polypyrimidine track and putative branch point of the MCMV RNA7.2 contributes to stability of this RNA [106, 107]. A recombinant virus containing a deletion of this structure does (MCMV*del*HP) not impact processing of the precursor transcript, but does prevent accumulation of the intron during infection as observed by northern blot analysis and qRT-PCR. To investigate this further, we

quantified the MCMV RNA7.2 over time post Actinomycin D treatment to determine a potential change in half-life. As anticipated, the half-life of the intron is reduced substantially to approximately 0.89 hours while that of the spliced mRNA remains unchanged in comparison to Wild-Type MCMV (Figure 18C-D). This data set confirmed that this RNA structure is important for stable retention of the MCMV RNA7.2.

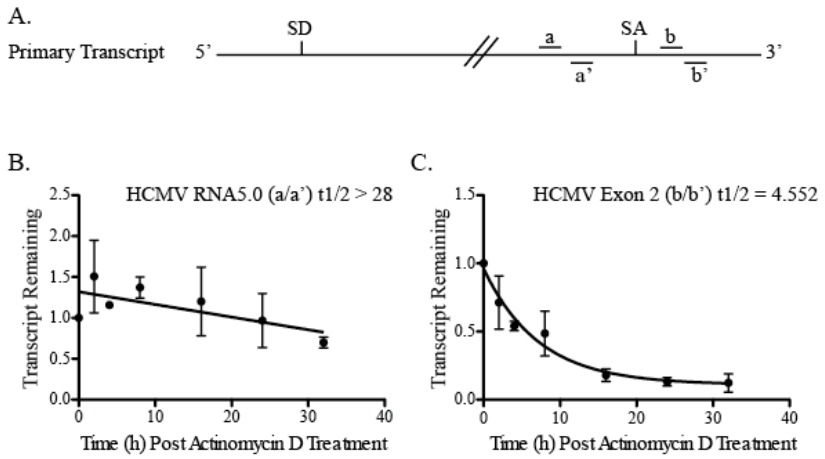


Figure 17. Half-life Analysis of the HCMV RNA5.0 Intron Locus Transcripts. Human fibroblasts were infected with HCMV (MOI = 2.0). At 24 and 48 hours post infection, cells were treated with 4ug/ml of Actinomycin D. Total RNA was harvested over the indicated time course. RNA5.0, the second exon (Exon2), and 18S rRNA transcript levels were quantified by qRT-PCR. RNA5.0 and Exon 2 were normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on three biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value or annotated as greater than the RNA7.2 half-life (28 hours). The 24 and 48 hour time points were indistinguishable from one another for all transcripts analyzed. Bars represent the mean and error bars represent the standard error of the mean (SEM). (A) Schematic representation of primer probe sets used for qRT-PCR analysis. SD = splice donor sequence, SA = splice acceptor sequence. Half-life decay curves for the (B) RNA5.0 using primer probe set a/a', (C) the second exon of the mRNA (b/b').

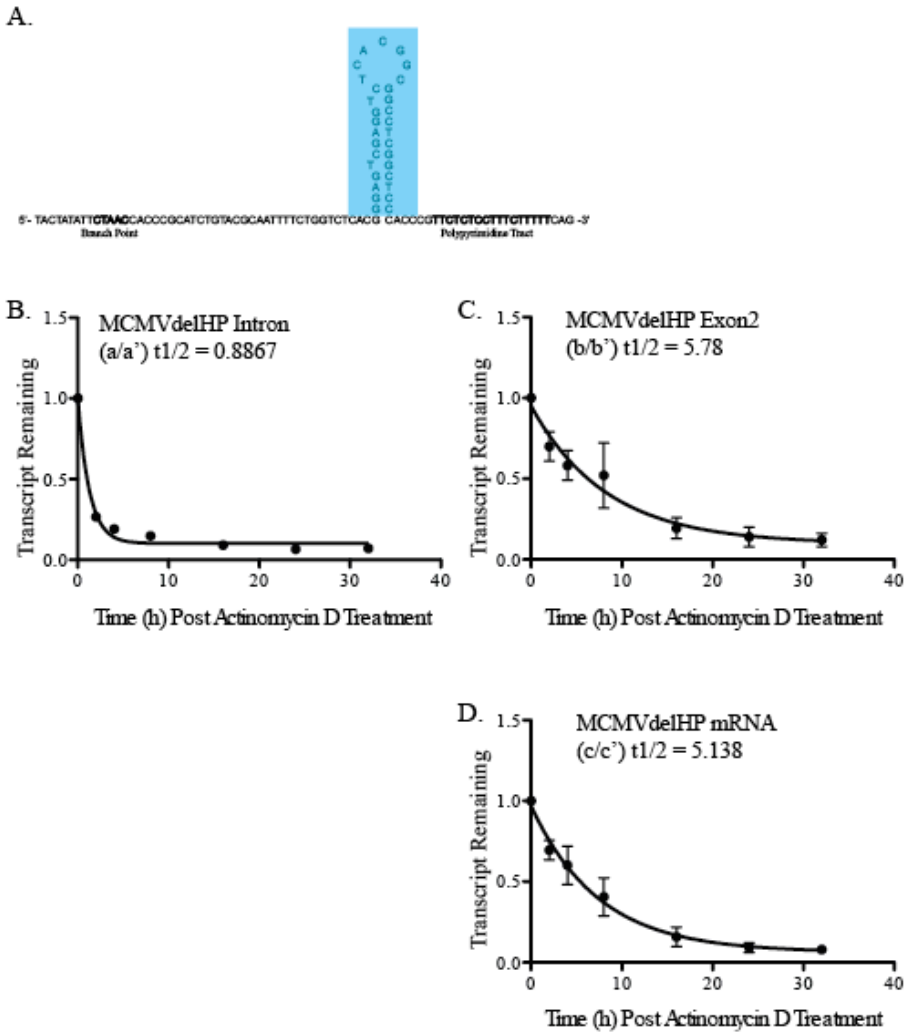


Figure 18. Half-life Analysis of the MCMV RNA7.2 Locus Transcripts From Cells Infected With MCMVdelHP.
Mouse fibroblasts were infected with MCMV*del*HP (MOI = 1.0). At 30 hours post infection, cells were treated with 4ug/ml of Actinomycin D. Total RNA was harvested over the indicated time course. RNA7.2, mRNA, and 18S rRNA transcript levels were quantified by qRT-PCR. Intron and spliced mRNA transcripts were normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on two biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value. Bars represent the mean and error bars represent the standard error of the mean (SEM). A) Schematic representation of the MCMV RNA7.2 hairpin deletion, B) Half-life of RNA7.2 using primer probe set a/a', C) Half-life of the second exon of the spliced mRNA (b/b'), D) Half-life of the spliced mRNA using primer probe set c/c'.

To investigate if other sequences within RNA7.2 contribute to its stability, the RNA7.2 locus spanning the transcriptional start site and into the second exon was cloned into the pCDNA3.1^{+} expression vector (Figure 19). Using this cloned, RNA7.2 plasmid pIntron, we were able to easily query additional sequence elements that contribute to stability. To determine if internal sequence was critical for the stable retention of RNA7.2, sequence was deleted to create either a smaller 3.58-kb RNA (pInt3.58) or a 1.8-kb RNA (pInt1.8) by utilizing the EcoRV or PstI and EcoRV (PstI-EcoRV construct) restriction sites respectively (Figure 20). An additional construct was generated to determine if sequence orientation would affect RNA7.2 stability by reversing the EcoRV fragment (pIntRVrev). After transfecting these constructs into mouse fibroblasts and inhibiting transcription using Actinomycin D, we found that relative transcript levels did not change between the deletion mutants in comparison to the wild type RNA7.2 construct over time. Unexpectedly, pIntRVrev reduced the half-life of RNA7.2 to approximately 8.9 hours (Figure 20). Although still highly stable for an intron, this reduced half-life could be due to structural instability of reversing the sequence. Regardless, this data allowed us to conclude that stability determinants of RNA7.2 reside primarily within the 3' region of this RNA.

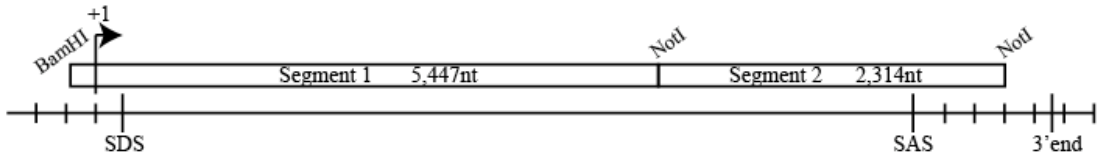


Figure 19. Cloning Strategy for the MCMV RNA7.2 Locus into pCDNA3.1+.
The MCMV RNA7.2 locus was cloned into pCDNA3.1^{+} as two segments as indicated. The first segment includes sequence upstream of the transcriptional start site and the second segment includes sequence into the second exon.

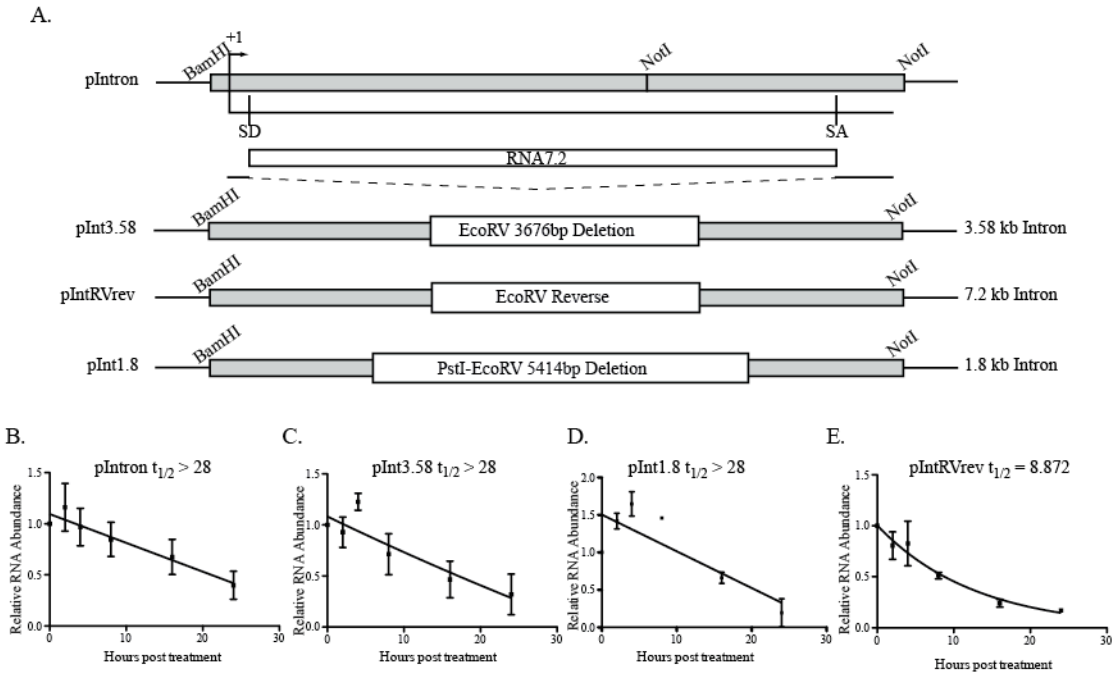


Figure 20. Internal Deletions of RNA7.2 Do Not Reduce Its Half-Life.
Mouse fibroblasts were transfected with the indicated pCDNA3.1+ (or pIntron) constructs. 24 hours post transfection, cells were treated with 4ug/mL of Actinomycin D. Total RNA was harvested over the indicated time course. RNA7.2, and 18S rRNA transcript levels were quantified by qRT-PCR. RNA7.2 and spliced mRNA transcripts were normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on two biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value or annotated as greater than the RNA7.2 half-life (28 hours). Bars represent the mean and error bars represent the standard error of the mean (SEM). (A) Schematic representation deletions made to pIntron. SD = splice donor sequence, SA = splice acceptor sequence. Half-life decay curves for the RNA7.2 using primer probe set a/a' B) pIntron, C) pInt3.58, D) pInt1.8, E) pIntRVrev.

To further investigate how a small RNA structure in the 3' region of the RNA7.2 accounts for such a drastic change in transcript half-life, we hypothesized that this stem loop is important for preserving the predicted intron-lariat conformation. The stem of the hairpin consists of nine GC base pairs that likely contribute to the favorable, free energy of this RNA structure to be at -26.8kcal/mol. To determine if minor modifications to the

stem loop structure would impact stability, nucleotide deletions to the loop or substitutions to the stem of the hairpin were generated within the pIntron construct (Figure 21). The four base-pair substitution to the base of the hairpin caused the most severe reduction in RNA7.2 half-life likely due to the disruption of the G-C bonds (Figure 21E). The 7 base-pair and 3 base-pair deletion mutants had no affect on RNA7.2 stability (Figure 21C-D).

To determine if the intron-lariat formation is dependent on the usage of a consensus branch point sequence (YURAY), this predicted sequence was disrupted (Figure 22A). 34 nucleotides upstream of the base of the hairpin structure is the consensus branch point sequence; CTAAC. Typically, the adenosine 2' hydroxyl from this sequence attacks the splice donor site (SDS) to form the 2'-5' phosphodiester bond generating the intron-lariat structure. We disrupted the TAA residues in this putative branch point sequence by introducing a minimal, 3 base-pair substitution to GTG within pIntron. This substitution reduced the half-life of RNA7.2 to approximately 4.5 hours (Figure 22B). Splicing was not altered, however, from mutating this sequence as (Figure 22C). This observation indicates that an alternative, less stable branch point is used for the splicing reaction to occur. Because splicing efficiency does not appear to be affected by deleting the hairpin structure, we hypothesize that this structure is critical for maintaining the intron-lariat conformation by preventing debranching of the 2'-5' phosphodiester bond.

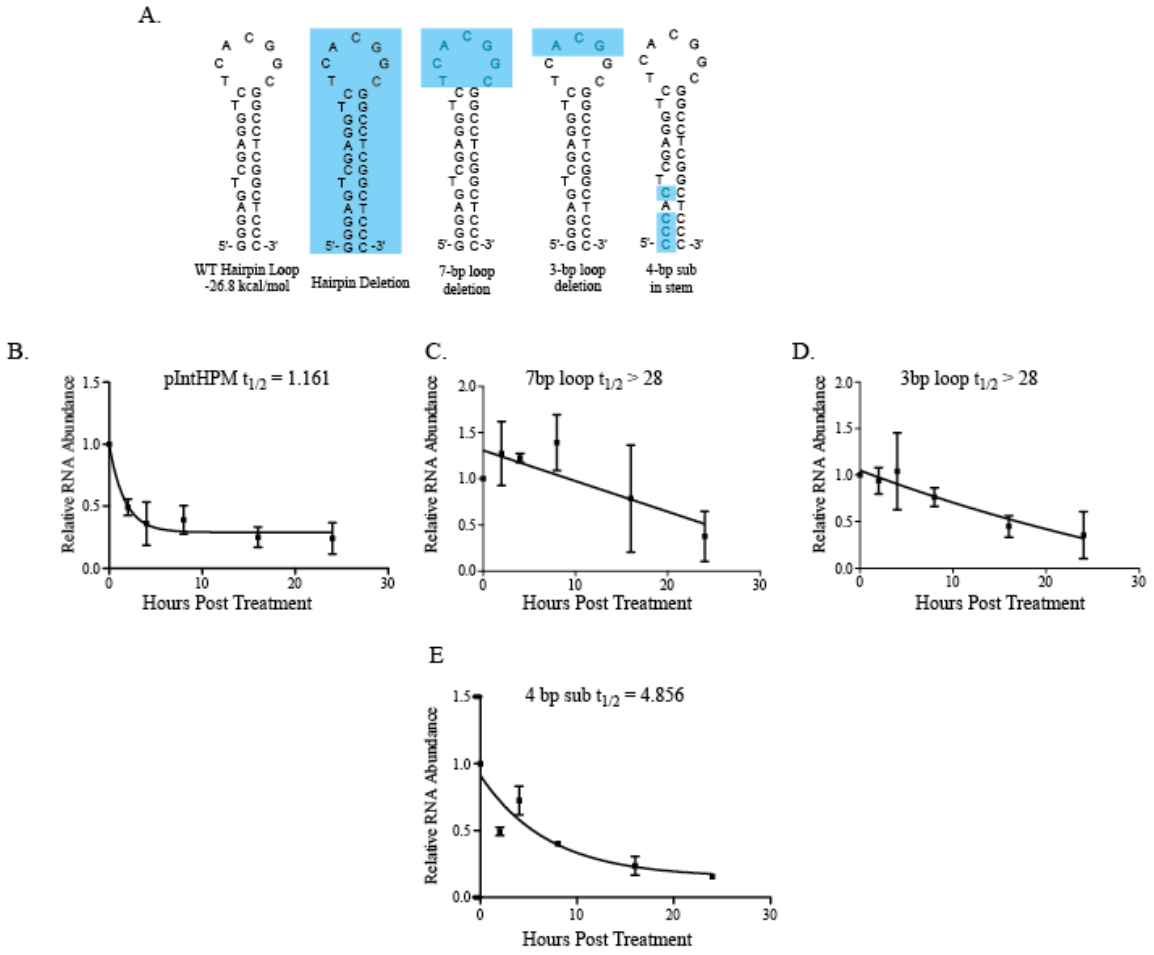
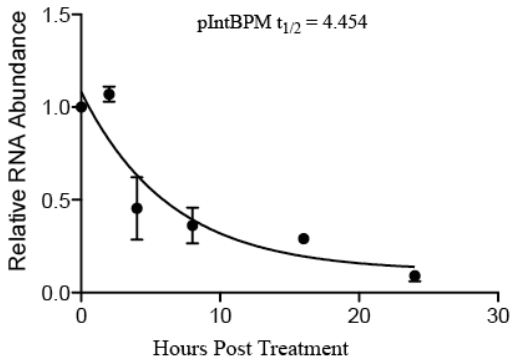


Figure 21. Mutagenesis to the Hairpin Reduces Stability of RNA7.2.
Mouse fibroblasts were transfected with the indicated pCDNA3.1+ constructs (representative pIntron in Figure 21). 24 hours post transfection, cells were treated with 4ug/mL of Actinomycin D. Total RNA was harvested over the indicated time course. RNA7.2 and 18S rRNA transcript levels were quantified by qRT-PCR. RNA7.2 was normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on two biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value or annotated as greater than the RNA7.2 half-life (28 hours). Bars represent the mean and error bars represent the standard error of the mean (SEM). Half-life decay curves for RNA7.2 using primer probe set a/a' B) pIntHPM, C) 7bp loop, D) 3bp loop, E) 4 bp sub.

A.

	Branch Point Substitution YURAY
WT Branch Point	TACTATATT**CTAAC**CACCCG
3bp Substitution	TACTATATT**CGTGC**CACCCG

B.



C.

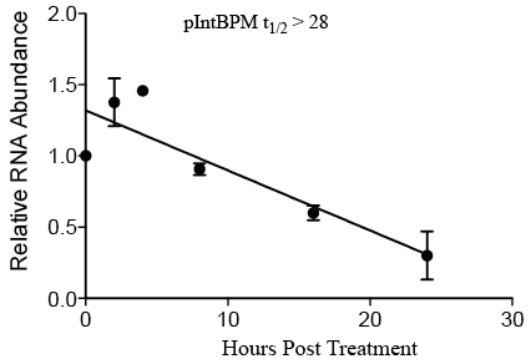


Figure 22. The Branch Point Sequence is Critical for Stability of RNA7.2. Mouse fibroblasts were transfected with the pIntBPM (representative pIntron in Figure 21). 24 hours post transfection, cells were treated with 4ug/mL of Actinomycin D. Total RNA was harvested over the indicated time course. RNA7.2, spliced mRNA, and 18S rRNA transcript levels were quantified by qRT-PCR. RNA7.2 was normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on two biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value or annotated as greater than the RNA7.2 half-life (28 hours). Bars represent the mean and error bars represent the standard error of the mean (SEM). (A) Schematic representation of Branch Point Mutation. SD = splice donor sequence, SA = splice acceptor sequence. Half-life decay curves for B) RNA7.2 and the C) spliced mRNA.

To test the possibility that hairpin proximity to the consensus branch point sequence is imperative for preventing access of debranching enzyme to the lariat branch site, we generated a recombinant virus that has a kanamycin cassette inserted between the branch point and hairpin structure thereby increasing the distance between the two sequences by approximately 1-kb (MCMV*Kan*BH) (Figure 23). There was no observable difference in replication kinetics between the MCMV*Kan*BH virus and WT MCMV in fibroblasts. Stability analysis did not reveal any difference in half-life between the WT

MCMV and MCMV*Kan*BH virus suggesting that spacing between the hairpin and branch does not contribute to overall stability of RNA7.2 (Figure 23). A caveat to this experiment is the possibility that additional structures might form within the kanamycin cassette that could skew the results towards a population of stable RNA7.2s.

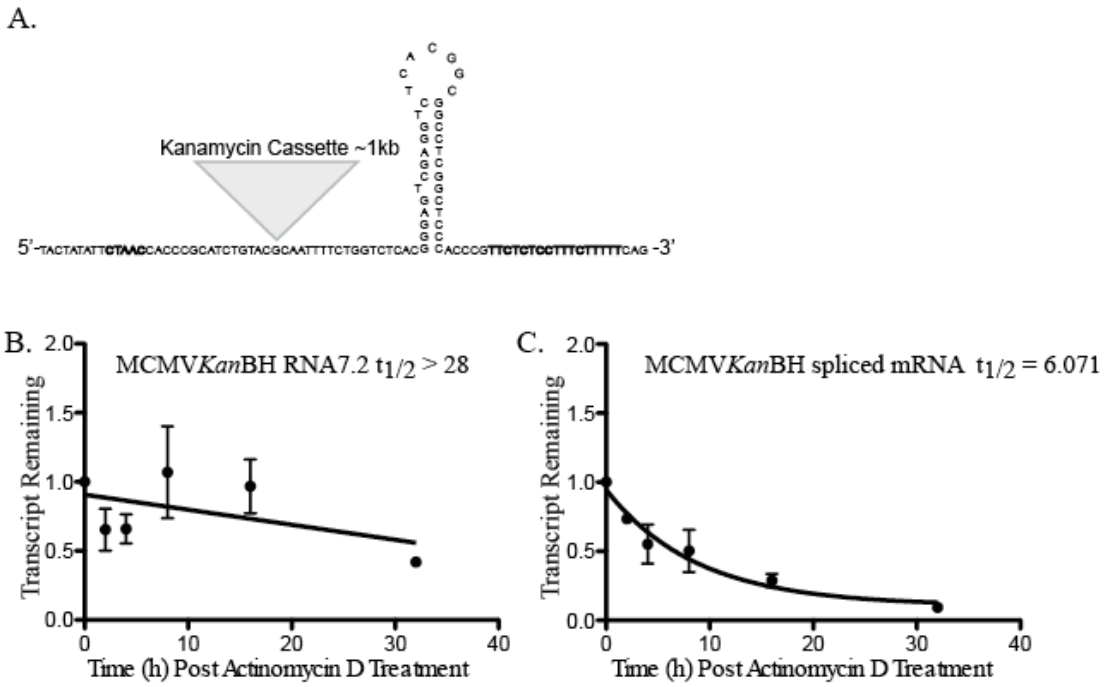


Figure 23. Insertion of Sequence Between the Branch Point and Stem Loop Structure Does Not Alter RNA7.2 Stability.
Mouse fibroblasts were infected with MCMV*del*HP (MOI = 1.0). At 30 hours post infection, cells were treated with 4ug/ml of Actinomycin D. Total RNA was harvested over the indicated time course. RNA7.2, mRNA, and 18S rRNA transcript levels were quantified by qRT-PCR. RNA7.2 and spliced mRNA transcripts were normalized to 18S rRNA. The relative quantitative values at time zero hours were adjusted to 100% and transcript remaining was compared relative to time zero. The fitted curve was modeled by one phase decay using a non-linear regression analysis on two biological replicates for each time point. The half-life (t1/2) shown for each transcript is the best-fit value or annotated as greater than the RNA7.2 half-life (28 hours). Bars represent the mean and error bars represent the standard error of the mean (SEM). A) Schematic representation of the MCMV*Kan*BH, B) Half-life of RNA7.2 using primer probe set a/a', C) Half-life of the spliced mRNA using primer probe set c/c'.

Structural Conformation of CMV Intron

In order to determine if RNA7.2 is nonlinear, we tested its susceptibility to exoribonucleolytic degradation. RNA harvested from mouse fibroblasts that were either transfected with pIntron or infected with the wild type MCMV strain was treated with the 5' to 3' exonuclease Terminator (Epicentre) (Figure 26). This enzyme recognizes and digests RNA containing 5' monophosphates. Therefore, if RNA7.2 is in the conformation of a lariat, it will be protected from digestion by this enzyme. 18S rRNA serves as a positive control for Terminator degradation because ribosomal RNA lacks a 5'terminal m^7G cap, and instead, has a 5'-monophate that is targeted by Terminator. GAPDH mRNA served as the normalization control in this experiment because it has a 5'terminal m^7G cap protecting it from Terminator recognition. RNA harvested from cells transfected with the cloned LAT was also used as a negative control in this experiment since studies demonstrate it as nonlinear and therefore, it should not be digested (Figure 26D). RNA was treated with terminator then quantified by qRT-PCR. In comparison to the 18S rRNA and LAT, RNA7.2 was protected from 5' to 3' digestion providing support that it is not linear (Figure 26A-E). Although this data does not definitively establish that RNA7.2 maintains a lariat structure, it provides evidence that its 5' end is somehow protected from 5' to 3' exoribonucleolytic degradation.

To provide further evidence that RNA7.2 remains in a lariat conformation, we examined its ability to resist debranching activity. Total RNA harvested from MCMV infected fibroblasts was treated with debranching enzyme (Dbr1) prior to digestion with Terminator (Figure 26F). To control for efficient debranching sensitivity, RNA harvested from *Saccharomyces cerevisiae* cells lacking Dbr1 was also examined. Because the Dbr1

gene has been knocked out in these cells, intron lariats are stabilized and accumulate [172]. Specifically, we chose to observe degradation of the *S. cerevisiae* Act1 intron from the Dbr1$^{-/-}$ yeast cells. Debranching activity was measured by comparing RNA that had been Dbr1 treated relative to Dbr1 untreated RNA. Using this method, RNA7.2 was not observed to be sensitive to debranching activity in comparison to the *S. cerevisiae* Act1 intron that demonstrated a 30% decrease (Figure 26F). Although a higher reduction in the Act1 intron was expected due to published results that illustrate complete degradation by DBR1, this result may be accounted for by the sensitivity of qPCR.

To compare debranching sensitivity between MCMV RNA7.2, and the HSV-1 LAT, the debranching assay was repeated with RNA isolated from fibroblasts transfected with either pIntron7.2 or pLAT (Figure 26F). Analysis revealed that all plasmid RNAs were resistant to debranching activity and subsequent Terminator degradation. This result demonstrates that DBR1 is unable to provide a free, 5'-monophosphate for Terminator degradation. Additional studies will need to be performed in order to definitively provide evidence that the MCMV intron remains in the conformation of a lariat structure similar to the HSV LATs.

DISCUSSION

In this chapter, we demonstrated that the MCMV RNA7.2 and HCMV RNA5.0 both accumulate as a consequence of a slow decay rate. This data confirmed that the RNA accumulation observed is due to the unusual stability of this viral intron and not the transcriptional output of the locus. Although it remains unknown why this intron is highly stable, we predict that it is a reflection of its function during persistent infection.

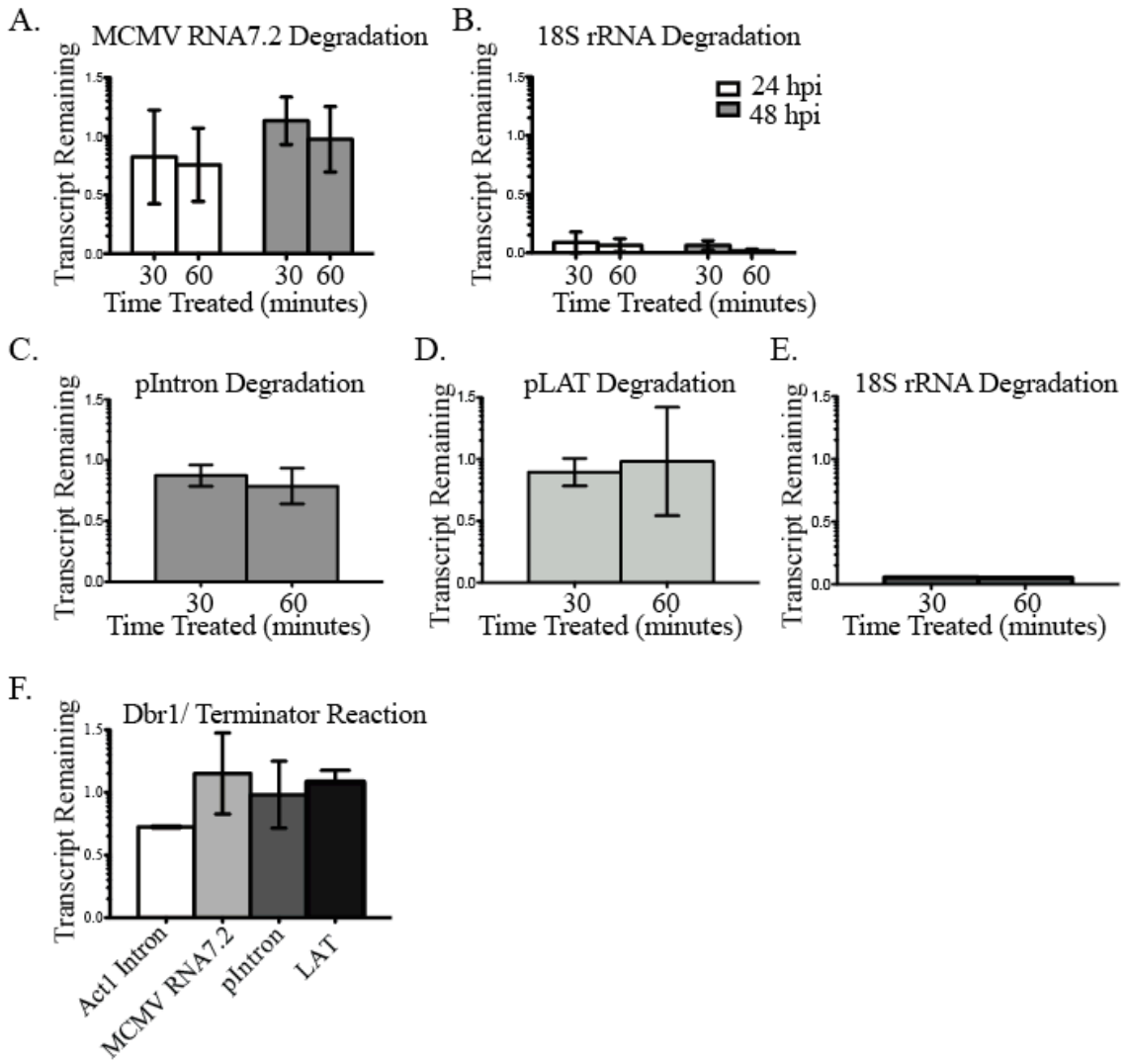


Figure 26. RNA7.2 is Protected From 5' to 3' Degradation and Dbr1 Activity. Total RNA was isolated from either MCMV infected mouse fibroblasts (MOI = 1) at 24 or 48 hpi, fibroblasts transfected with pIntron, or fibroblasts transfected with pLAT. 3ug of Total RNA was treated with Terminator for either 30 minutes or 60 minutes. Following Terminator treatment, Total RNA from infected fibroblasts was analyzed by qRT-PCR for A) RNA7.2 degradation (primer set a/a') or B) 18S rRNA degradation, and normalized to GAPDH. Following Terminator treatment, transfected cells were analyzed by qRT-PCR for C) RNA7.2 degradation, D) LAT degradation, or E) 18S rRNA degradation (representative figure from pIntron transfected cells), and normalized to GAPDH. F) RNA was treated with debranching enzyme prior to Terminator treatment. RNA was analyzed for the Act 1 intron from DBR1-/- yeast cells; RNA7.2 from MCMV infected fibroblasts; RNA7.2 from pIntron transfected cells; or LAT from pLAT transfected cells. Bars on all graphs represent the mean, error bars represent standard error of the mean.

Evidence for the MCMV RNA7.2 maintaining a lariat conformation is illustrated by previous work that examined its migration by high resolution northern blot analysis [107]. We routinely observe a doublet of bands near 7.2 kb; a major species at approximately 8.0 kb and a minor species migrating faster at 7.2 kb. These observations were made with multiple intron-specific probes. We have been unable to ascertain the basis for this difference in size, although we hypothesize it may be due to effects of lariat secondary structure on RNA migration during electrophoresis resulting in slower migration [160, 173]. When present in a lariat conformation, introns are less susceptible to degradation, presumably because of the low level of endonucleolytic activity in the cell nucleus. Therefore, to maintain this conformation, the RNA must have mechanisms to prevent debranching.

In line with the prediction that the MCMV RNA7.2 maintains a lariat structure, stability determinants map primarily to RNA7.2s 3' end suggesting that debranching and subsequent exoribonucleolytic decay might be prevented by specific sequences. To identify the sequence elements required for stability, we analyzed a panel of mutated pIntron constructs expressed in tissue culture. We demonstrated that internal sequence does not contribute to the long half-life of RNA7.2. Instead, RNA stability was compromised by site directed mutagenesis to the predicted branch point sequence and disruption of a hairpin structure. Both the predicted branch point sequence and hairpin structure are located entirely within the last 100 nucleotides of RNA7.2 similar to the HSV LATs. This data emphasizes the importance of this 3' region on stable retention of RNA7.2.

Previous studies demonstrated that a recombinant virus for the predicted hairpin structure within the 3' end of RNA7.2 is critical for accumulation of the intron. Disruption of the hairpin does not influence splicing of the primary transcript since accumulation and stability of the spliced mRNA transcript remains unaffected. This data indicates that the hairpin structure confers stability to RNA7.2 differently than the hairpin structure within the HSV LATs; the LATs hairpin is critical for directing the spliceosome machinery to a stable, nonconsensus branch point sequence at the base of the hairpin. Instead, we predict that this RNA structure is important for preventing debranching and 3' to 5' exoribonucleolytic decay to preserve the branched lariat structure. Comparative sequence analysis across CMV species suggests that the hairpin sequence is conserved. Although a recombinant HCMV has not been generated with the predicted hairpin structure deleted, we hypothesize that a similar reduction in RNA5.0 accumulation and stability would be observed as compared to the MCMV RNA7.2.

If the lariat conformation is maintained, it should resist debranching and 5' to 3' exoribonucleolytic decay. The MCMV RNA7.2 and HSV-1 LAT both resisted degradation to treatment with an XRN1-like enzyme Terminator indicating that their 5' ends are protected from enzymatic, exoribonucleolytic degradation. Although this data does not conclude that RNA7.2 remains in lariat conformation, it provides evidence for the protection of the 5' end leading to a mechanism of how this RNA maintains a long half-life. By treating RNA7.2 with debranching enzyme and subsequent Terminator digestion, we examined the RNAs susceptibility to cleavage of the 2'-5' phosphodiester bond which forms the branch point. Again, the RNA was unaffected by debranching

enzyme suggesting that it contains RNA structures at its terminal 5' end for protecting itself from exoribonucleic degradation or RNA7.2 does remain in the form of a lariat.

Additional support for retention of the lariat conformation is provided by RNA7.2 stability analysis after disrupting the predicted branch point sequence. Disruption of this consensus branch point sequence within the pIntron construct correlates with intron stability since the half-life was reduced to approximately 4 hours. Splicing was not compromised with the branch point mutation as observed by the accumulation of the spliced transcript. Since intron accumulation was not completely abolished immediately after Actinomycin D treatment, alternative sequences most likely can be used that do not offer as strong of a 2'-5' phosphodiester bond likely due to structural constraints.

Several approaches were attempted to distinguish between a circular, lariat conformation and a linear form of RNA7.2. RT-PCR across the predicted branch point was unsuccessful likely due to the strong secondary structure in this region as well as the 2'-5' backbone linkage thereby preventing read-through of reverse transcriptase. Primer extension analysis using a radiolabeled probe was also unsuccessful. In order to identify the branch by primer extension, a complementary radiolabeled probe must be hybridized 3' to the branch point. Reverse transcription from this radiolabeled probe will continue only up until the branch since reverse transcriptase is unable to proceed through this region. However, difficulty with this method is rooted in the limited sequence 3' of the branch that is available due to its vulnerability to 3' to 5' exoribonuclease degradation. Future methods to distinguish the conformation of RNA7.2 as circular or linear might involve RNAse R digestion or nicking the predicted loop portion of the RNA to collapse the intron doublet into a single band for northern blot analysis. Both methods are not

without additional limitations, however, indicating the difficulty of identifying the structural conformation of a predicted intron lariat.

In conclusion, our results indicate that stability of the MCMV RNA7.2 in tissue culture is due to sequences in the 3' region of the intron. A strong, secondary structure is critical for accumulation and stability of this RNA and we hypothesize that it is important for preventing debranching and 3' to 5' exoribonucleolytic decay. The predicted branch point sequence also confers stability to RNA7.2 likely due to its consensus sequence that is easily recognized by splicesomal machinery as well as its location within the 3' region. Although the function of the RNA remains unknown, its stable retention is likely to be a reflection of function.

CHAPTER VI

DISCUSSION

INTRODUCTION

This dissertation sought to further characterize a lncRNA that is required for the establishment of MCMV persistent replication in mice [106]. Persistent replication is defined as the sustained replication that occurs in glandular tissue for days and even months at a time. This replication occurs asymptomatically in the immunocompetent host and the only method for recognizing persistent replication is by directly quantifying infectious virus from mucosal secretions. A major challenge to controlling HCMV infection is to define the mechanisms that control viral persistence and transmission. Although clues as to how the virus can establish replication for an extended period of time are beginning to unfold, few viral determinants of persistence have been fully characterized. Because this phase of CMV infection is critical for CMV persistence in the human population, a research priority within the CMV field is to investigate the viral mechanisms of persistence for identifying potential therapeutic and vaccine targets.

Our lab studies a large, stable non-coding RNA encoded by HCMV that has been linked to viral persistence [106]. This lncRNA is spliced from its primary transcript as an intron. This intron is conserved across all species of cytomegaloviruses signifying its necessity to the replication cycle of the virus. Because cytomegaloviruses are species specific, we cannot cross infect a host cell with different CMV species. We therefore use Murine Cytomegalovirus (MCMV) as a model system to study the function of the intron in viral replication *in vitro* and *in vivo*. MCMV mutants lacking the stable intron replicate

normally during the acute phase of infection, yet fail to establish persistent replication in the salivary glands of the mouse [106].

Current work provides a finer map of the MCMV RNA7.2 locus. We demonstrate that during productive infection in fibroblasts, the intron locus RNAs are transcribed with true late gene kinetics. These RNAs are derived from a common precursor RNA and are produced as the result of splicing; two exons are joined together forming the spliced mRNA and RNA7.2 is excised from the primary transcript as an intron. The spliced mRNA encodes the open reading frame m106 that is translated into a protein during productive infection. Translation of m106 is reliant on proper splicing of the primary transcript likely because it requires the 5' untranslated region located on the first exon. Identification of the transcriptional start sites rules out the possibility that a cluster of miRNAs mapped near to the RNA7.2 splice donor site originate from the same primary transcript [124]. It remains unknown, however, what functional relationship the miRNAs may have with the MCMV 7.2kb intron locus, if any, during virus replication and pathogenesis.

Mapping the 5' end of the primary transcript revealed minimal promoter elements located upstream that contribute to transcript expression. Analysis of recombinant viruses with deletions in the putative promoter elements, however, revealed these elements exert only minor effects on intron production and viral persistence *in vivo*. Because RNA7.2 accumulates to detectable levels observed by several methods proposes that it is extremely stable in comparison to most introns studied. Half-life analysis of RNA7.2 demonstrated that accumulation is due to a long half-life. Low transcriptional output by the putative promoter element(s) is compensated by the long half-life of RNA7.2. The

long half-life is due to elements located primarily within the 3'end of RNA7.2. Elements shown to confer stability include a canonical branch point sequence and a stem loop structure located between the branch point sequence and polypyrimidine tract. We speculate that RNA7.2 remains in the conformation of a lariat structure similar to the HSV-1 LATs because of its migration pattern on a denaturing agarose gel and it is resistant to debranching enzyme activity and subsequent degradation by a 5' to 3' exonuclease. Half-life analysis of RNA5.0 from HCMV demonstrated that stability of this RNA is conserved across CMV species.

We further examined the role that the MCMV RNA7.2 plays over the course of infection. Previously, it was concluded that RNA7.2 is a viral persistence factor because without it, MCMV is attenuated during the persistent phase of infection within the salivary glands of mice [106]. Work presented in this dissertation further defines the role for RNA7.2 in maintaining viral replication during persistence since viral dissemination to glandular tissues was unimpeded in virus unable to produce this intron during the acute phase of infection. Additionally, detection of CMV genomic DNA was reduced 100-fold in virus lacking RNA7.2 production in comparison to wild type virus. This data suggests that RNA7.2 might serve to prevent clearance from this tissue during the persistent phase of infection.

Together, the data presented within this dissertation sheds new light onto the possible roles RNA7.2 may play during persistent replication in glandular tissue. Although we have not described a mechanism of action for the MCMV RNA7.2, these studies open up discussion points for how to proceed in identifying a specific function for this lncRNA over the course of infection. New technologies allow for specialized

analyses to query lncRNA interactions *in vitro* and *in vivo*. Because RNA5.0 and RNA7.2 alike do not display a replication phenotype *in vitro*, it will be necessary to investigate function of RNA7.2 using the mouse model of infection.

CONCLUSIONS AND FUTURE DIRECTIONS

Transcriptional Regulation of the RNA7.2 Locus

Transcriptional regulation of the RNA7.2 locus remains unclear. We were unable to successfully identify a minimal promoter element that is largely responsible for the transcriptional output of this region. It is possible that transcription from the sequence elements identified do initiate there, but due to the viral gene kinetic class this locus belongs to, there might be other confounding factors that contribute to its transcriptional regulation.

Because our *in vitro* and *in vivo* studies indicate that RNA7.2 exerts its functional role exclusively during the persistent replication phase, it is possible that studying its transcriptional control might be more relevant *in vivo*. The viral gene expression profile could be different during persistent replication due to differences in gene regulatory mechanisms that are exerted either from the host or the virus itself. These gene expression differences could be accounted for by the necessity to maintain the host-pathogen relationship within this tissue to allow viral shedding. To examine viral gene regulation during persistent replication, RNA-Seq experiments could be designed to observe transcriptionally active and repressive marks at promoter regions representing all three viral gene classes during acute and persistent replication phases. It is possible that if the regulation of the kinetic classes is more pronounced during this time, the putative

promoter sequences we identified become relevant to regulating transcription of the RNA7.2 locus.

RNA7.2 Structure

RNA7.2 has an extremely long half-life that is predicted to be ascribed to its structural conformation. Evidence for RNA7.2 maintaining a lariat structure was presented not only by the migration pattern observed by denaturing gel electrophoresis (Figure 5), but also by the protection from the cellular debranching enzyme and subsequent exonucleic degradation. The 5' to 3' Terminator activity could stall if there is structure near the 5' end of the intron preventing degradation from its potential linear form. Therefore, these studies do not definitively support a lariat structure and additional experiments could be carried out including: RT-PCR across the branch point junction, labeling free 5' phosphate groups using T4 polynucleotide kinase (circular RNAs should not be labeled), RNase H digestion to excise the branch from the lariat structure, and RNase protection to determine the branch point [174-176]. Because our Terminator degradation experiments do not directly reveal if there is a 5' end, the most straight for experiment to consider trying is to label any potentially free 5' monophosphate groups with polynucleotide kinase then visualize labeled RNA after gel electrophoresis. Unfortunately, all of these experiments come with major caveats since it is reasonably easy to obtain false positives due to an array of technical issues associated with each method [177].

Additional experiments to determine if the 3' end of the intron provides stability to RNA7.2 might be considered. First, the ability of the RNA7.2 stem loop structure to allow accumulation of a normally unstable intron could be examined. This structure

would be inserted between the branch point sequence and polypyrimidine tract of an unstable intron and the half-life subsequently would be analyzed. This experiment would offer validation to our hypothesis that the stem loop structure is the primary determinant of RNA7.2 stability. Additional experiments might include substituting the stem loop structure of RNA7.2 for the Latency Associated Transcript (LAT) 3' stem loop structure. There is evidence that the branch point sequence is located at the base of the stem loop structure in LAT and, consequently, it is predicted that this stem loop structure acts to direct the spliceosome machinery to the branch point sequence facilitating splicing and ultimately formation of the lariat. From this experiment, we could ultimately determine if there are functional differences between the stem loop structures from these two viral introns.

PREDICTED FUNCTIONAL ROLE OF RNA7.2

lncRNA Function in the Literature

Many lncRNAs have been discovered in recent years yet ascribing function to them is a difficult task. lncRNA function has largely been associated with gene regulation throughout the developmental and cell cycle processes [84, 178, 179]. For example, the earliest function identified for a lncRNA illustrated the mechanism behind genomic imprinting of X-chromosome inactivation [180, 181]. X-chromosome inactivation is the process by which one of the two X chromosome copies present in females is transcriptionally silenced in order to balance out the X-chromosome gene expression between males and females. This silencing is mediated by genes within the X-inactivation center (Xic) that is enriched for lncRNAs. Five lncRNAs produced from this region are known to have a role in X-inactivation . These lncRNAs span approximately

200 nucleotides of sequence within the Xic and exemplify the variety of functions lncRNAs may possess. Each of the Xic lncRNAs examined have separate functions regulating the inactivation of the second X-chromosome copy by either directly targeting repressive epigenetic complexes to the X-chromosome, acting as antisense inhibitors, or by activating transcription from the Xic.

The lncRNA responsible for coating the X-chromosome in cis and transcriptionally inactivating it is called *Xist*. *Xist* is a 17- to 20- kb RNA that is transcribed only from the X-chromosome to be inactivated [182]. *Xist* directly binds to the polycomb repressive complex 2 (PRC2) and targets this complex to the X-chromosome [183]. PRC2 is an epigenetic complex responsible for depositing a repressive mark onto histones that will cause the formation of heterochromatin thereby compacting the X-chromosome DNA. By compacting the X-chromosome, the DNA is no longer accessible for transcription to occur and gene expression is therefore silenced. Once *Xist* targets the PRC2 complex to the X-chromosome, the Xic 1.6 kb lncRNA RepA loads the PRC2 complex onto the *Xist* promoter to nucleate inactivation of the X-chromosome [183].

To ensure that only one X-chromosome is silenced, *Xist* function is tightly regulated both negatively and positively. The Xic lncRNA *Tsix* is transcribed antisense to *Xist* and it determines allelic choice by repressing *Xist* transcription on one allele through several methods including the recruitment of DNA methyltransferase to silence *Xist* and also by blocking the recruitment of PRC2 to the Xic by duplexing with the Xist-RepA complex [184-189]. *Tsix* transcription is regulated in turn by another Xic lncRNA locus, *Xite*. Without *Xite* production, *Tsix* expression is down regulated allowing for

inactivation of an X-chromosome [190, 191]. Xist RNA is positively regulated by another Xic lncRNA, Jpx. By deleting *Jpx* from the Xic, *Xist* function is precluded [183, 192]. It remains unknown how *Jpx* induces X-inactivation, but evidence suggests it antagonizes *Tsix* function. The Xic lncRNAs, therefore, demonstrate a range of possible functions that RNA7.2 may possess in regulating gene expression during persistent replication in glandular tissue.

RNA7.2 Predicted Function

Based on current data, we speculate that RNA7.2 plays a gene regulatory role within infected glandular tissues. In cell culture, RNA7.2 does not does not recapitulate the attenuated phenotype observed during the persistent phase of infection in mice. This indicates that RNA7.2 does not influence lytic infection, but is required for maintaining a different replication program in specialized tissue such as the salivary glands.

Persistent viral replication and shedding from glandular tissue must remain undetected, or at the very least, must have mechanisms for reducing immune recognition by the host to allow for horizontal transmission. It has been established that MCMV expresses several gene products that impair antigen presentation for MHC class I and II molecules within infected cells [73]. These gene products are just a few of the many immuno-evasion genes that MCMV encodes. Genetic deletion of these immune-evasion genes results in increased MCMV immune control in a tissue dependent as well as a CD8+ T cell dependent manner within BALB/c mice. Salivary gland tissues are particularly sensitive to the deletion of these immune-evasion genes allowing for the activation of MCMV specific CD8 T cells that can clear MCMV from this tissue. This data suggests that MCMV expressed immune-evasins are only effective in the acinar

glandular epithelial cells (AGECs) of salivary glands in comparison to cells of other tissues. Despite the effective down regulation of MHC II molecules within infected cells, recent data supports a role for CD4 + T cells in controlling MCMV persistent replication in salivary gland tissue. MCMV antigens released by infected cells are taken up by local antigen presenting cells and presented on MHCII stimulating MCMV specific CD4+ T cells to secrete IFNγ. IFNγ secretion by these MCMV specific CD4+ T cells contributes to controlling infection and horizontal transmission over time.

The delicate host-pathogen balance in glandular tissue must be maintained by the virus in order to allow for horizontal transmission. It is possible that RNA7.2 is contributing to the persistent replication cycle by directly regulating expression of the immune evasion genes, modulating host gene expression in a tissue specific manner, or both. I predict that RNA7.2 regulates gene expression during persistent replication in glandular tissue by re-targeting cellular transcriptional repression complexes from viral genomes to the host genome. Specifically, I anticipate that RNA7.2 interacts with PRC2 and targets it to occupy specific genomic locations for host genes responsible for mounting an immune response against the virus. Host genes that are targeted for transcriptional repression by PRC2 might include genes involved in the type I interferon pathway as well as genes involved in antigen presentation. By re-targeting PRC2 to the host genome, viral transcription could precede unhindered allowing progression of the full viral replication cycle that could ultimately lead to the secretion of infectious virus into mucosal secretions.

To determine if RNA7.2 is contributing to the regulation of viral or cellular gene expression, experiments could be designed to examine the abundance of RNA7.2 in

salivary gland AGECs by high-throughput transcriptome analysis. The viral and cellular transcriptome has not been analyzed during this phase of MCMV infection *in vivo* and would provide valuable information for illustrating gene expression profiles over the course of infection. To address the question if RNA7.2 physically interacts with specific promoter regions of viral or cellular genes of infected glandular tissue, chromatin isolation by RNA purification (ChIRP) experiments could be designed.

Subnuclear localization between tissues that are acutely infected versus persistently infected may also provide valuable information on the role RNA7.2 plays over the course of infection in mice. The nucleus is composed of different subnuclear compartments that are associated with different steps of gene regulation. For example, there exists specific compartments that are dedicated to transcriptional repression, activation, and processing of primary RNA transcripts. Although there is a possibility these cellular compartments might be disrupted during persistent replication in glandular tissue, observing the staining pattern of RNA7.2 within this tissue has the potential of narrowing down a function for RNA7.2 by association with a particular nuclear compartment. Since an *in vitro* system modeling persistent replication does not exist, the RNA7.2 staining pattern we observe in fibroblasts likely does not represent that within infected salivary gland acinar cells providing further impetus for executing these experiments [107].

If RNA7.2 functions to regulate viral gene expression, its extremely long half-life is predicted to be useful for a persistent infection that can last for months and even years at a time. Data suggests that RNA7.2 half-life is linked to viral persistent replication. When mice are infected with a viral mutant lacking the hairpin, we observe a similar

phenotype as the viral mutant that cannot excise RNA7.2 from its primary transcript. As demonstrated in chapter 5, this hairpin mutant virus reduces the intron half-life to approximately 55 minutes. This half-life is still unusually stable for an intron but might not allow sufficient accumulation for RNA7.2 to function. Therefore, additional *in vivo* infection experiments to be considered will include infecting mice with a mutant virus that only disrupts hairpin folding thereby causing a less severe intron half-life reduction.

The role for RNA7.2 in latency remains unknown. Transcriptome analysis of an HCMV latency infection model indicates that RNA5.0 is not transcribed at this time of infection [50]. Instead, two alternative viral lncRNAs were detected during latency: RNA2.7 and RNA4.9. RNA4.9 was demonstrated to directly interact with the viral major immediate promoter and components of the polycome repression complex (PRC) suggesting that this lncRNA plays a role in suppressing the activation of immediate early gene expression. Latency studies using the mouse model, however, have not been attempted to determine the role RNA7.2 might play.

Latency Associated Transcripts As a Model For RNA7.2 Function

Little is known about lncRNA function and even less is known about functional introns. Only a few stable introns have been identified that have functional roles during cellular or viral processes. As mentioned previously, the HSV-1 LAT region encodes for multiple transcripts including an 8.3 kb primary transcript that is processed to produce either a stable 1.5 kb intron or a stable 2.0 kb intron [88]. The only abundant transcripts produced during HSV-1 latency in mice, rabbits, and humans are the LATs [193]. Although RNA7.2 and LATs share similar RNA sequence features, it is unknown if RNA7.2 has similar functional roles during persistence as the LATs.

The LATs were first described in 1984 and since then, many studies describe their roles in several HSV-1 infection stages including establishing latency, maintaining latency, reactivation from latency, and protecting infected neurons from apoptosis. During acute infection, HSV-1 infects sensory nerve endings and travels up to the nerve bodies where it establishes latency within the nucleus of ganglionic sensory neurons. Sporadic reactivations can occur with the virus traveling down the neuronal axon to the original peripheral site where the virus was acquired. It is at this same location where the virus can actively replicate and shed itself. The LATs have been implicated in the establishment and reactivation of virus within sensory neurons. The main role of LAT during the latency-reactivation cycle appears to be its antiapoptotic function [194]. To determine if LATs antiapoptotic activity is required for its involvement in the latency-reactivation cycle, LAT was replaced by an antiapoptosis gene within HSV-1. This antiapoptosis gene remained under the control of the LAT promoter to ensure transcription during the latent phase of infection within a rabbit ocular model of infection. The results of this study concluded that HSV-1 that is lacking LAT could restore reactivation levels similar to wild-type with the antiapoptotic gene substitution [194].

Additional work has further defined a role for LAT in modulating the immune response to HSV-1 latent infection. Recently, it was demonstrated that LAT upregulates expression levels of Herpesvirus entry mediator (HVEM) [195]. HVEM is a member of the tumor necrosis factor (TNF) receptor superfamily and is expressed on a number of cell types. HSV-1 can bind HVEM using its envelope glycoprotein gD allowing the virus to access NFkB dependent cell survival pathways. Interestingly, LAT upregulates HVEM expression during latency providing an additional mechanism to promote cell survival

within latently infected neurons. Other studies suggest that LAT plays an immune evasion role during latency by promoting CD8+ T cell exhaustion resulting in increased HSV-1 reactivation [195].

Although a mechanism that describes LAT function has not been identified, it could be assumed that RNA7.2 plays similar roles to LAT during the persistent phase of infection. Although latency and persistent replication are two very different programs during the virus life cycle, the virus must maintain a close relationship with the host in order to prevent clearance from the tissue and cell types involved. Like cellular lncRNAs that have emerged as significant regulators of gene expression in human cells, RNA7.2 and LAT most likely affect viral and cellular gene expression patterns to maintain this host-pathogen relationship. Although a mechanism has not been ascribed to LAT, it does have a gene regulatory function, which is more pronounced with its role in upregulating HVEM. In addition to gene regulation, LAT seems has an antiapoptotic role during latency to allow for the eventual reactivation of virus. It could be possible that RNA7.2 is involved in preventing apoptosis in AGECs of salivary glands as well since apoptosis would disrupt the sustained replication program and horizontal transmission of virus from this tissue. Although we can speculate on the functional roles that might be similar between the LATs and RNA7.2 based on their phenotypes during infection, we will not be able to make conclusions until further experiments are done. The LAT studies, however, provides a useful model for us to evaluate the role of RNA7.2 during MCMV persistence.

Current Model of RNA7.2s Role in Viral Persistence

Based on the data that I have generated over the course of this dissertation, I hypothesize that RNA7.2 plays a role in regulating the suppressing the host immune response to virus infection in glandular tissue. The long half-life likely offers an advantage to the RNA for allowing it to function despite cellular attempts to downregulate its transcription and production from its locus. Experiments to directly examine RNA7.2s role in regulating the immune response would involve infecting immunosuppressed mice, such as Rag 1 deficient mice that lack mature B an T cells, with MCMVdelHP and observing if the viral attenuation could be restored to wild-type levels.

IMPLICATIONS

Like the eukaryotic genome, the CMV genome displays the same level of transcriptional complexity. Numerous transcriptome studies identified the existence of a high level of RNA splicing, antisense transcription, and noncoding RNA production throughout the entire viral genome [80]. Although these findings are recent, it is no surprise this DNA virus has similar genomic complexities to its host especially since CMV has co-evolved with its host over millennia. Mimicking the transcriptional potential of the host genome has undoubtedly contributed to the viral mechanisms allowing for a life-long infection of its eukaryotic host and the discovery that several viral lncRNAs are critical for viral persistence strategies is testament to this.

REFERENCES

1. Pellett Pe, Roizman B: *The Family Herpesviridae: A Brief Introduction.* In: *Fields Virology*. (5th) Wolters Kluwer Lippincott Williams & Wilkins, Philadelphia. 2, 2479-2499 (2007).

2. Davison Aj, Eberle R, Ehlers B *et al.*: The order Herpesvirales. *Arch Virol* 154(1), 171-177 (2009).

3. Baines Jd, Duffy C: *Nucleocapsid assembly and envelopment of herpes simplex virus*. In: *Alpha Herpesviruses: Molecular and Cellular Biology*. Caister Academic Press, Norfolk, UK. 1, (2006).

4. Mettenleiter Tc, Klupp Bg, Granzow H: Herpesvirus assembly: an update. *Virus Res* 143(2), 222-234 (2009).

5. Asher Y, Heller M, Becker Y: Incorporation of lipids into herpes simplex virus particles. *J Gen Virol* 4(1), 65-76 (1969).

6. Frenkel N, Schirmer Ec, Wyatt Ls *et al.*: Isolation of a new herpesvirus from human CD4+ T cells. *Proc Natl Acad Sci U S A* 87(2), 748-752 (1990).

7. Landolfo S, Gariglio M, Gribaudo G, Lembo D: The human cytomegalovirus. *Pharmacol Ther* 98(3), 269-297 (2003).

8. Dohner K, Sodeik B: The role of the cytoskeleton during viral infection. *Curr Top Microbiol Immunol* 285, 67-108 (2005).

9. Honess Rw, Roizman B: Regulation of herpesvirus macromolecular synthesis. I. Cascade regulation of the synthesis of three groups of viral proteins. *J Virol* 14(1), 8-19 (1974).

10. Peng L, Ryazantsev S, Sun R, Zhou Zh: Three-dimensional visualization of gammaherpesvirus life cycle in host cells by electron tomography. *Structure* 18(1), 47-58 (2010).

11. Mettenleiter Tc, Klupp Bg, Granzow H: Herpesvirus assembly: a tale of two membranes. *Curr Opin Microbiol* 9(4), 423-429 (2006).

12. Davison Aj, Dolan A, Akter P *et al.*: The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* 84(1), 17-28 (2003).

13. Farber S, Wolbach Sb: Intranuclear and Cytoplasmic Inclusions ("Protozoan-Like Bodies") in the Salivary Glands and Other Organs of Infants. *Am J Pathol* 8(2), 123-136 123 (1932).

14. Rifkind D: Cytomegalovirus infection after renal transplantation. *Arch Intern Med* 116(4), 554-558 (1965).

15. Myers Jd, Spencer Hc, Jr., Watts Jc *et al.*: Cytomegalovirus pneumonia after human marrow transplantation. *Ann Intern Med* 82(2), 181-188 (1975).

16. Ho M: Virus infections after transplantation in man. Brief review. *Arch Virol* 55(1-2), 1-24 (1977).

17. Rubin Rh: Impact of cytomegalovirus infection on organ transplant recipients. *Rev Infect Dis* 12 Suppl 7, S754-766 (1990).

18. Klatt Ec, Shibata D: Cytomegalovirus infection in the acquired immunodeficiency syndrome. Clinical and autopsy findings. *Arch Pathol Lab Med* 112(5), 540-544 (1988).

19. Mocarski Es, Shenk T, Pass Rf: *Cytomegaloviruses*. In: *Fields Virology*. (5th). Lippincott Williams & Wilkins, Philadelphia. II, 2701-2272 (2007).

20. Mocarski Es: *Betaherpes viral genes and their functions*. In: *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.*, Arvin A, Campadelli-Fiume G, Mocarski Es (Eds). Cambridge University Press, Cambridge (2007).

21. Terhune Ss, Schroer J, Shenk T: RNAs are packaged into human cytomegalovirus virions in proportion to their intracellular concentration. *J Virol* 78(19), 10390-10398 (2004).

22. Dunn W, Chou C, Li H *et al.*: Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A* 100(24), 14223-14228 (2003).

23. Yu D, Silva Mc, Shenk T: Functional map of human cytomegalovirus AD169 defined by global mutational analysis. *Proc Natl Acad Sci U S A* 100(21), 12396-12401 (2003).

24. Omoto S, Mocarski Es: Cytomegalovirus UL91 is essential for transcription of viral true late (gamma2) genes. *J Virol* 87(15), 8651-8664 (2013).

25. Omoto S, Mocarski Es: Transcription of True Late (gamma2) Cytomegalovirus Genes Requires Betagamma-conserved UL92 Function. *J Virol*, 88(1), 120-130 (2013).

26. Boppana Sb, Fowler Kb: *Persistence in the population: epidemiology and transmisson.* In: *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.*, Arvin A, Campadelli-Fiume G, Mocarski Es (Eds). Cambridge University Press, Cambridge (2007).

27. Bate Sl, Dollard Sc, Cannon Mj: Cytomegalovirus seroprevalence in the United States: the national health and nutrition examination surveys, 1988-2004. *Clin Infect Dis* 50(11), 1439-1447 (2010).

28. Gratacap-Cavallier B, Bosson Jl, Morand P *et al.*: Cytomegalovirus seroprevalence in French pregnant women: parity and place of birth as major predictive factors. *Eur J Epidemiol* 14(2), 147-152 (1998).

29. Stagno S, Pass Rf, Cloud G *et al.*: Primary cytomegalovirus infection in pregnancy. Incidence, transmission to fetus, and clinical outcome. *JAMA* 256(14), 1904-1908 (1986).

30. Gold E, Nankervis: Cytomegalovirus. In viral infections of humans: Epidemiology and Control. (2nd). Plenum Press, New York. 143-161 (1982).

31. Dowd Jb, Aiello Ae, Alley De: Socioeconomic disparities in the seroprevalence of cytomegalovirus infection in the US population: NHANES III. *Epidemiol Infect* 137(1), 58-65 (2009).

32. Kerrey Bt, Morrow A, Geraghty S, Huey N, Sapsford A, Schleiss Mr: Breast milk as a source for acquisition of cytomegalovirus (HCMV) in a premature infant with sepsis syndrome: detection by real-time PCR. *J Clin Virol* 35(3), 313-316 (2006).

33. Reynolds Dw, Stagno S, Hosty Ts, Tiller M, Alford Ca, Jr.: Maternal cytomegalovirus excretion and perinatal infection. *N Engl J Med* 289(1), 1-5 (1973).

34. Schleiss Mr: Acquisition of human cytomegalovirus infection in infants via breast milk: natural immunization or cause for concern? *Rev Med Virol* 16(2), 73-82 (2006).

35. Drew Wl, Mintz L, Miner Rc, Sands M, Ketterer B: Prevalence of cytomegalovirus infection in homosexual men. *J Infect Dis* 143(2), 188-192 (1981).

36. Adler Sp: The molecular epidemiology of cytomegalovirus transmission among children attending a day care center. *J Infect Dis* 152(4), 760-768 (1985).

37. Hutto C, Ricks R, Garvie M, Pass Rf: Epidemiology of cytomegalovirus infections in young children: day care vs. home care. *Pediatr Infect Dis* 4(2), 149-152 (1985).

38. De Mello Al, Ferreira Ec, Vilas Boas Ls, Pannuti Cs: Cytomegalovirus infection in a day-care center in the municipality of Sao Paulo. *Rev Inst Med Trop Sao Paulo* 38(3), 165-169 (1996).

39. Pass Rf, Hutto C, Ricks R, Cloud Ga: Increased rate of cytomegalovirus infection among parents of children attending day-care centers. *N Engl J Med* 314(22), 1414-1418 (1986).

40. Sinzger C, Digel M, Jahn G: Cytomegalovirus cell tropism. *Curr Top Microbiol Immunol* 325, 63-83 (2008).

41. Campbell Ae, Cavanaugh Vj, Slater Js: The salivary glands as a privileged site of cytomegalovirus immune evasion and persistence. *Med Microbiol Immunol* 197(2), 205-213 (2008).

42. Goodrum Fd, Jordan Ct, High K, Shenk T: Human cytomegalovirus gene expression during infection of primary hematopoietic progenitor cells: a model for latency. *Proc Natl Acad Sci U S A* 99(25), 16255-16260 (2002).

43. Mendelson M, Monard S, Sissons P, Sinclair J: Detection of endogenous human cytomegalovirus in CD34+ bone marrow progenitors. *J Gen Virol* 77 (12), 3099-3102 (1996).

44. Sissons Jg, Bain M, Wills Mr: Latency and reactivation of human cytomegalovirus. *J Infect* 44(2), 73-77 (2002).

45. Reeves Mb, Macary Pa, Lehner Pj, Sissons Jg, Sinclair Jh: Latency, chromatin remodeling, and reactivation of human cytomegalovirus in the dendritic cells of healthy carriers. *Proc Natl Acad Sci U S A* 102(11), 4140-4145 (2005).

46. Reeves Mb, Lehner Pj, Sissons Jg, Sinclair Jh: An in vitro model for the regulation of human cytomegalovirus latency and reactivation in dendritic cells by chromatin remodelling. *J Gen Virol* 86(11), 2949-2954 (2005).

47. Sinclair J, Sissons P: Latency and reactivation of human cytomegalovirus. *J Gen Virol* 87(7), 1763-1779 (2006).

48. Goodrum F, Caviness K, Zagallo P: Human cytomegalovirus persistence. *Cell Microbiol* 14(5), 644-655 (2012).

49. Knipe Dm, Cliffe A: Chromatin control of herpes simplex virus lytic and latent infection. *Nat Rev Microbiol* 6(3), 211-221 (2008).

50. Rossetto Cc, Tarrant-Elorza M, Pari Gs: Cis and trans acting factors involved in human cytomegalovirus experimental and natural latent infection of CD14 (+) monocytes and CD34 (+) cells. *PLoS Pathog* 9(5), e1003366 (2013).

51. Klemola E, Von Essen R, Henle G, Henle W: Infectious-mononucleosis-like disease with negative heterophil agglutination test. Clinical features in relation to Epstein-Barr virus and cytomegalovirus antibodies. *J Infect Dis* 121(6), 608-614 (1970).

52. Melnick Jl, Adam E, Debakey Me: Cytomegalovirus and atherosclerosis. *Eur Heart J* 14 Suppl K, 30-38 (1993).

53. Singh N: Preemptive therapy versus universal prophylaxis with ganciclovir for cytomegalovirus in solid organ transplant recipients. *Clin Infect Dis* 32(5), 742-751 (2001).

54. Demmler Gj: Congenital cytomegalovirus infection and disease. *Adv Pediatr Infect Dis* 11, 135-162 (1996).

55. Bale Jf, Miner L, Petheram Sj: Congenital Cytomegalovirus Infection. *Curr Treat Options Neurol* 4(3), 225-230 (2002).

56. Noyola De, Demmler Gj, Nelson Ct *et al.*: Early predictors of neurodevelopmental outcome in symptomatic congenital cytomegalovirus infection. *J Pediatr* 138(3), 325-331 (2001).

57. Pass Rf, Stagno S, Myers Gj, Alford Ca: Outcome of symptomatic congenital cytomegalovirus infection: results of long-term longitudinal follow-up. *Pediatrics* 66(5), 758-762 (1980).

58. Stagno S, Reynolds Dw, Pass Rf, Alford Ca: Breast milk and the risk of cytomegalovirus infection. *N Engl J Med* 302(19), 1073-1076 (1980).

59. Lurain Ns, Chou S: Antiviral drug resistance of human cytomegalovirus. *Clin Microbiol Rev* 23(4), 689-712 (2010).

60. Chou Sw: Cytomegalovirus drug resistance and clinical implications. *Transpl Infect Dis* 3 Suppl 2, 20-24 (2001).

61. Chou S, Waldemer Rh, Senters Ae *et al.*: Cytomegalovirus UL97 phosphotransferase mutations that affect susceptibility to ganciclovir. *J Infect Dis* 185(2), 162-169 (2002).

62. Lurain Ns, Weinberg A, Crumpacker Cs, Chou S: Sequencing of cytomegalovirus UL97 gene for genotypic antiviral resistance testing. *Antimicrob Agents Chemother* 45(10), 2775-2780 (2001).

63. Gentry Bg, Drach Jc: Metabolism of cyclopropavir and ganciclovir in human cytomegalovirus-infected cells. *Antimicrob Agents Chemother* 58(4), 2329-2333 (2014).

64. Plotkin Sa, Starr Se, Friedman Hm, Gonczol E, Brayman K: Vaccines for the prevention of human cytomegalovirus infection. *Rev Infect Dis* 12 Suppl 7, S827-838 (1990).

65. Marshall Gs, Ricciardi Rp, Rando Rf *et al.*: An adenovirus recombinant that expresses the human cytomegalovirus major envelope glycoprotein and induces neutralizing antibodies. *J Infect Dis* 162(5), 1177-1181 (1990).

66. Elek Sd, Stern H: Development of a vaccine against mental retardation caused by cytomegalovirus infection in utero. *Lancet* 1(7845), 1-5 (1974).

67. Plotkin Sa, Furukawa T, Zygraich N, Huygelen C: Candidate cytomegalovirus strain for human vaccination. *Infect Immun* 12(3), 521-527 (1975).

68. Adler Sp, Hempfling Sh, Starr Se, Plotkin Sa, Riddell S: Safety and immunogenicity of the Towne strain cytomegalovirus vaccine. *Pediatr Infect Dis J* 17(3), 200-206 (1998).

69. Heineman Tc, Schleiss M, Bernstein Di *et al.*: A phase 1 study of 4 live, recombinant human cytomegalovirus Towne/Toledo chimeric vaccines. *J Infect Dis* 193(10), 1350-1360 (2006).

70. Lilja Ae, Mason Pw: The next generation recombinant human cytomegalovirus vaccine candidates-beyond gB. *Vaccine* 30(49), 6980-6990 (2012).

71. Henson D, Strano Aj: Mouse cytomegalovirus. Necrosis of infected and morphologically normal submaxillary gland acinar cells during termination of chronic infection. *Am J Pathol* 68(1), 183-202 (1972).

72. Pereira L, Maidji E, Fisher Sj, Mcdonagh S, Tabata T: *HCMV persistence in the population: potential transplacental transmission.* In: *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.*, Arvin A, Campadelli-Fiume G, Mocarski Es (Eds). Cambridge University Press, Cambridge (2007)

73. Walton Sm, Mandaric S, Torti N, Zimmermann A, Hengel H, Oxenius A: Absence of cross-presenting cells in the salivary gland and viral immune evasion confine cytomegalovirus immune control to effector CD4 T cells. *PLoS Pathog* 7(8), e1002214 (2011).

74. Rawlinson Wd, Farrell He, Barrell Bg: Analysis of the complete DNA sequence of murine cytomegalovirus. *J Virol* 70(12), 8833-8849 (1996).

75. Ho M: Cytomegalovirus : biology and infection. (2nd ed.). Plenum Medical Book Co., New York. (1991).

76. Cardin Rd, Schaefer Gc, Allen Jr, Davis-Poynter Nj, Farrell He: The M33 chemokine receptor homolog of murine cytomegalovirus exhibits a differential tissue-specific role during in vivo replication and latency. *J Virol* 83(15), 7590-7601 (2009).

77. Dolken L, Krmpotic A, Kothe S *et al.*: Cytomegalovirus microRNAs facilitate persistent virus infection in salivary glands. *PLoS Pathog* 6(10), e1001150 (2010).

78. Davis-Poynter Nj, Lynch Dm, Vally H *et al.*: Identification and Characterization of a G Protein-Coupled Receptor Homolog Encoded by Murine Cytomegalovirus. *J Virol* 71(2), 1521-1529 (1996).

79. Kulesza Ca, Shenk T: Human cytomegalovirus 5-kilobase immediate-early RNA is a stable intron. *J Virol* 78(23), 13182-13189 (2004).

80. Gatherer D, Seirafian S, Cunningham C *et al.*: High-resolution human cytomegalovirus transcriptome. *Proc Natl Acad Sci U S A* 108(49), 19755-19760 (2011).

81. Zhang G, Raghavan B, Kotur M *et al.*: Antisense transcription in the human cytomegalovirus transcriptome. *J Virol* 81(20), 11267-11281 (2007).

82. Steitz J, Borah S, Cazalla D *et al.*: Noncoding RNPs of viral origin. *Cold Spring Harb Perspect Biol* 3(3), (2011).

83. Eddy Sr: Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* 2(12), 919-929 (2001).

84. Geisler S, Coller J: RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat Rev Mol Cell Biol* 14(11), 699-712 (2013).

85. Guttman M, Rinn Jl: Modular regulatory principles of large non-coding RNAs. *Nature* 482(7385), 339-346 (2012).

86. Ingolia Nt, Lareau Lf, Weissman Js: Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147(4), 789-802 (2011).

87. Spivack Jg, Fraser Nw: Detection of herpes simplex virus type 1 transcripts during latent infection in mice. *J Virol* 61(12), 3841-3847 (1987).

88. Farrell Mj, Dobson At, Feldman Lt: Herpes simplex virus latency-associated transcript is a stable intron. *Proc Natl Acad Sci U S A* 88(3), 790-794 (1991).

89. Ahmed M, Lock M, Miller Cg, Fraser Nw: Regions of the herpes simplex virus type 1 latency-associated transcript that protect cells from apoptosis in vitro and protect neuronal cells in vivo. *J Virol* 76(2), 717-729 (2002).

90. Inman M, Perng Gc, Henderson G *et al.*: Region of herpes simplex virus type 1 latency-associated transcript sufficient for wild-type spontaneous reactivation promotes cell survival in tissue culture. *J Virol* 75(8), 3636-3646 (2001).

91. Sun R, Lin Sf, Gradoville L, Miller G: Polyadenylylated nuclear RNA encoded by Kaposi sarcoma-associated herpesvirus. *Proc Natl Acad Sci U S A* 93(21), 11883-11888 (1996).

92. Zhong W, Ganem D: Characterization of ribonucleoprotein complexes containing an abundant polyadenylated nuclear RNA encoded by Kaposi's sarcoma-associated herpesvirus (human herpesvirus 8). *J Virol* 71(2), 1207-1212 (1997).

93. Zhong W, Wang H, Herndier B, Ganem D: Restricted expression of Kaposi sarcoma-associated herpesvirus (human herpesvirus 8) genes in Kaposi sarcoma. *Proc Natl Acad Sci U S A* 93(13), 6641-6646 (1996).

94. Prichard Mn, Jairath S, Penfold Me, St Jeor S, Bohlman Mc, Pari Gs: Identification of persistent RNA-DNA hybrid structures within the origin of replication of human cytomegalovirus. *J Virol* 72(9), 6997-7004 (1998).

95. Mcdonough Sh, Staprans Si, Spector Dh: Analysis of the major transcripts encoded by the long repeat of human cytomegalovirus strain AD169. *J Virol* 53(3), 711-718 (1985).

96. Mcsharry Bp, Tomasec P, Neale Ml, Wilkinson Gw: The most abundantly transcribed human cytomegalovirus gene (beta 2.7) is non-essential for growth in vitro. *J Gen Virol* 84(9), 2511-2516 (2003).

97. Greenaway Pj, Wilkinson Gw: Nucleotide sequence of the most abundantly transcribed early gene of human cytomegalovirus strain AD169. *Virus Res* 7(1), 17-31 (1987).

98. Reeves Mb, Davies Aa, Mcsharry Bp, Wilkinson Gw, Sinclair Jh: Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* 316(5829), 1345-1348 (2007).

99. Klucher Km, Rabert Dk, Spector Dh: Sequences in the human cytomegalovirus 2.7-kilobase RNA promoter which mediate its regulation as an early gene. *J Virol* 63(12), 5334-5343 (1989).

100. Mcdonough Sh, Spector Dh: Transcription in human fibroblasts permissively infected by human cytomegalovirus strain AD169. *Virology* 125(1), 31-46 (1983).

101. Wade Ej, Klucher Km, Spector Dh: An AP-1 binding site is the predominant cis-acting regulatory element in the 1.2-kilobase early RNA promoter of human cytomegalovirus. *J Virol* 66(4), 2407-2417 (1992).

102. Hutchinson Ni, Sondermeyer Rt, Tocci Mj: Organization and expression of the major genes from the long inverted repeat of the human cytomegalovirus genome. *Virology* 155(1), 160-171 (1986).

103. Jahn G, Knust E, Schmolla H *et al.*: Predominant immediate-early transcripts of human cytomegalovirus AD 169. *J Virol* 49(2), 363-370 (1984).

104. Demarchi Jm: Post-transcriptional control of human cytomegalovirus gene expression. *Virology* 124(2), 390-402 (1983).

105. Plachter B, Traupe B, Albrecht J, Jahn G: Abundant 5 kb RNA of human cytomegalovirus without a major translational reading frame. *J Gen Virol* 69 (9), 2251-2266 (1988).

106. Kulesza Ca, Shenk T: Murine cytomegalovirus encodes a stable intron that facilitates persistent replication in the mouse. *Proc Natl Acad Sci U S A* 103(48), 18302-18307 (2006).

107. Schwarz Tm, Volpe La, Abraham Cg, Kulesza Ca: Molecular investigation of the 7.2 kb RNA of murine cytomegalovirus. *Virol J* 10, 348 (2013).

108. Stern-Ginossar N, Weisburd B, Michalski A *et al.*: Decoding human cytomegalovirus. *Science* 338(6110), 1088-1093 (2012).

109. Moreira S, Correia M, Soares P, Maximo V: GRIM-19 function in cancer development. *Mitochondrion* 11(5), 693-699 (2011).

110. Wu Tc, Lee Wa, Pizzorno Mc *et al.*: Localization of the human cytomegalovirus 2.7-kb major early beta-gene transcripts by RNA in situ hybridization in permissive and nonpermissive infections. *Am J Pathol* 141(5), 1247-1254 (1992).

111. Gawn Jm, Greaves Rf: Absence of IE1 p72 protein function during low-multiplicity infection by human cytomegalovirus results in a broad block to viral delayed-early gene expression. *J Virol* 76(9), 4441-4455 (2002).

112. Wagner M, Jonjic S, Koszinowski Uh, Messerle M: Systematic excision of vector sequences from the BAC-cloned herpesvirus genome during virus reconstitution. *J Virol* 73(8), 7056-7060 (1999).

113. Smith Mg: Propagation of salivary gland virus of the mouse in tissue cultures. *Proc Soc Exp Biol Med* 86(3), 435-440 (1954).

114. Tischer Bk, Von Einem J, Kaufer B, Osterrieder N: Two-step red-mediated recombination for versatile high-efficiency markerless DNA manipulation in Escherichia coli. *Biotechniques* 40(2), 191-197 (2006).

115. Hurwitz J, Furth Jj, Malamy M, Alexander M: The role of deoxyribonucleic acid in ribonucleic acid synthesis. III. The inhibition of the enzymatic synthesis of ribonucleic acid and deoxyribonucleic acid by actinomycin D and proflavin. *Proc Natl Acad Sci U S A* 48, 1222-1230 (1962).

116. Sambrook J, Russel Dw: Molecular Cloning: A Laboratory Manual. (3rd). Cold Spring Harbor Press, Cold Spring Harbor. 1, (2001).

117. Pfaffl Mw: A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* 29(9), e45 (2001).

118. Zuker M: Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* 31(13), 3406-3415 (2003).

119. Matera Ag, Wang Z: A day in the life of the spliceosome. *Nat Rev Mol Cell Biol* 15(2), 108-121 (2014).

120. Wang Z, Burge Cb: Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* 14(5), 802-813 (2008).

121. Berget Sm: Exon recognition in vertebrate splicing. *J Biol Chem* 270(6), 2411-2414 (1995).

122. Matlin Aj, Clark F, Smith Cw: Understanding alternative splicing: towards a cellular code. *Nat Rev Mol Cell Biol* 6(5), 386-398 (2005).

123. Rawlinson Wd, Zeng F, Farrell He *et al.*: The murine cytomegalovirus (MCMV) homolog of the HCMV phosphotransferase (UL97(pk)) gene. *Virology* 233(2), 358-363 (1997).

124. Buck Ah, Santoyo-Lopez J, Robertson Ka, Kumar Ds, Reczko M, Ghazal P: Discrete clusters of virus-encoded micrornas are associated with complementary strands of the genome and the 7.2-kilobase stable intron in murine cytomegalovirus. *J Virol* 81(24), 13761-13770 (2007).

125. Schleiss Mr, Mcgregor A, Choi Ky, Date Sv, Cui X, Mcvoy Ma: Analysis of the nucleotide sequence of the guinea pig cytomegalovirus (GPCMV) genome. *Virol J* 5, 139 (2008).

126. Chee Ms, Bankier At, Beck S *et al.*: Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* 154, 125-169 (1990).

127. Hansen Sg, Strelow Li, Franchi Dc, Anders Dg, Wong Sw: Complete sequence and genomic analysis of rhesus cytomegalovirus. *J Virol* 77(12), 6620-6636 (2003).

128. Murphy E, Rigoutsos I, Shibuya T, Shenk Te: Reevaluation of human cytomegalovirus coding potential. *Proc Natl Acad Sci U S A* 100(23), 13585-13590 (2003).

129. Murphy E, Yu D, Grimwood J *et al.*: Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A* 100(25), 14976-14981 (2003).

130. Wu Tt, Su Yh, Block Tm, Taylor Jm: Atypical splicing of the latency-associated transcripts of herpes simplex type 1. *Virology* 243(1), 140-149 (1998).

131. Smale St, Baltimore D: The "initiator" as a transcription control element. *Cell* 57(1), 103-113 (1989).

132. Smale St: Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim Biophys Acta* 1351(1-2), 73-88 (1997).

133. Smale St: Core promoters: active contributors to combinatorial gene regulation. *Genes Dev* 15(19), 2503-2508 (2001).

134. Smale St, Kadonaga Jt: The RNA polymerase II core promoter. *Annu Rev Biochem* 72, 449-479 (2003).

135. Carninci P, Sandelin A, Lenhard B *et al.*: Genome-wide analysis of mammalian promoter architecture and evolution. *Nat Genet* 38(6), 626-635 (2006).

136. Juven-Gershon T, Cheng S, Kadonaga Jt: Rational design of a super core promoter that enhances gene expression. *Nat Methods* 3(11), 917-922 (2006).

137. Juven-Gershon T, Hsu Jy, Kadonaga Jt: Perspectives on the RNA polymerase II core promoter. *Biochem Soc Trans* 34(Pt 6), 1047-1050 (2006).

138. Zhou T, Chiang Cm: The intronless and TATA-less human TAF(II)55 gene contains a functional initiator and a downstream promoter element. *J Biol Chem* 276(27), 25503-25511 (2001).

139. Sandelin A, Carninci P, Lenhard B, Ponjavic J, Hayashizaki Y, Hume Da: Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat Rev Genet* 8(6), 424-436 (2007).

140. Boshart M, Weber F, Jahn G, Dorsch-Hasler K, Fleckenstein B, Schaffner W: A very strong enhancer is located upstream of an immediate early gene of human cytomegalovirus. *Cell* 41(2), 521-530 (1985).

141. Kerry Ja, Priddy Ma, Jervey Ty *et al.*: Multiple regulatory events influence human cytomegalovirus DNA polymerase (UL54) expression during viral infection. *J Virol* 70(1), 373-382 (1996).

142. Kerry Ja, Priddy Ma, Stenberg Rm: Identification of sequence elements in the human cytomegalovirus DNA polymerase gene promoter required for activation by viral gene products. *J Virol* 68(7), 4167-4176 (1994).

143. Fortunato Ea, Spector Dh: Regulation of human cytomegalovirus gene expression. *Adv Virus Res* 54, 61-128 (1999).

144. Chapa Tj, Johnson Ls, Affolter C *et al.*: Murine Cytomegalovirus Protein pM79 Is a Key Regulator for Viral Late Transcription. *J Virol* 87(16), 9135-9147 (2013).

145. Isomura H, Stinski Mf, Murata T *et al.*: The human cytomegalovirus gene products essential for late viral gene expression assemble into prereplication complexes before viral DNA replication. *J Virol* 85(13), 6629-6644 (2011).

146. Arumugaswami V, Wu Tt, Martinez-Guzman D *et al.*: ORF18 is a transfactor that is essential for late gene transcription of a gammaherpesvirus. *J Virol* 80(19), 9730-9740 (2006).

147. Wong E, Wu Tt, Reyes N, Deng H, Sun R: Murine gammaherpesvirus 68 open reading frame 24 is required for late gene expression after DNA replication. *J Virol* 81(12), 6761-6764 (2007).

148. Wu Tt, Park T, Kim H *et al.*: ORF30 and ORF34 are essential for expression of late genes in murine gammaherpesvirus 68. *J Virol* 83(5), 2265-2273 (2009).

149. Spector Dh: Activation and regulation of human cytomegalovirus early genes. *Intervirology* 39(5-6), 361-377 (1996).

150. Perez Kj, Martinez Fp, Cosme-Cruz R, Perez-Crespo Nm, Tang Q: A short cis-acting motif in the M112-113 promoter region is essential for IE3 to activate M112-113 gene expression and is important for murine cytomegalovirus replication. *J Virol* 87(5), 2639-2647 (2013).

151. Rabani M, Levin Jz, Fan L *et al.*: Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* 29(5), 436-442 (2011).

152. Sharova Lv, Sharov Aa, Nedorezov T, Piao Y, Shaik N, Ko Ms: Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. *DNA Res* 16(1), 45-58 (2009).

153. Blattner C, Kannouche P, Litfin M *et al.*: UV-Induced stabilization of c-fos and other short-lived mRNAs. *Mol Cell Biol* 20(10), 3616-3625 (2000).

154. Friedel Cc, Dolken L, Ruzsics Z, Koszinowski Uh, Zimmer R: Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* 37(17), e115 (2009).

155. Clark Mb, Johnston Rl, Inostroza-Ponta M *et al.*: Genome-wide analysis of long noncoding RNA stability. *Genome Res* 22(5), 885-898 (2012).

156. Lam Lt, Pickeral Ok, Peng Ac *et al.*: Genomic-scale measurement of mRNA turnover and the mechanisms of action of the anti-cancer drug flavopiridol. *Genome Biol* 2(10), 1-11 (2001).

157. Schwanhausser B, Busse D, Li N *et al.*: Global quantification of mammalian gene expression control. *Nature* 473(7347), 337-342 (2011).

158. Wallace Jc, Edmonds M: Polyadenylylated nuclear RNA contains branches. *Proc Natl Acad Sci U S A* 80(4), 950-954 (1983).

159. Moore Mj, Query Cc, Sharp Pa: Splicing of precursors to messenger RNAs by the spliceosome. Cold Spring Harbor Press, Cold Spring Harbor. 1-30 (1993).

160. Ruskin B, Green Mr: An RNA processing activity that debranches RNA lariats. *Science* 229(4709), 135-140 (1985).

161. Chapman Kb, Boeke Jd: Isolation and characterization of the gene encoding yeast debranching enzyme. *Cell* 65(3), 483-492 (1991).

162. Clemson Cm, Hutchinson Jn, Sara Sa *et al.*: An architectural role for a nuclear noncoding RNA: NEAT1 RNA is essential for the structure of paraspeckles. *Mol Cell* 33(6), 717-726 (2009).

163. Krummenacher C, Zabolotny Jm, Fraser Nw: Selection of a nonconsensus branch point is influenced by an RNA stem-loop structure and is important to confer stability to the herpes simplex virus 2-kilobase latency-associated transcript. *J Virol* 71(8), 5849-5860 (1997).

164. Zabolotny Jm, Krummenacher C, Fraser Nw: The herpes simplex virus type 1 2.0-kilobase latency-associated transcript is a stable intron which branches at a guanosine. *J Virol* 71(6), 4199-4208 (1997).

165. Mukerjee R, Kang W, Suri V, Fraser Nw: A non-consensus branch point plays an important role in determining the stability of the 2-kb LAT intron during acute and latent infections of herpes simplex virus type-1. *Virology* 324(2), 340-349 (2004).

166. Thomas Dl, Lock M, Zabolotny Jm, Mohan Br, Fraser Nw: The 2-kilobase intron of the herpes simplex virus type 1 latency-associated transcript has a half-life of approximately 24 hours in SY5Y and COS-1 cells. *J Virol* 76(2), 532-540 (2002).

167. Brinkman Kk, Mishra P, Fraser Nw: The half-life of the HSV-1 1.5-kb LAT intron is similar to the half-life of the 2.0-kb LAT intron. *J Neurovirol* 19(1), 102-108 (2013).

168. Dobson At, Sederati F, Devi-Rao G *et al.*: Identification of the latency-associated transcript promoter by expression of rabbit beta-globin mRNA in mouse sensory nerve ganglia latently infected with a recombinant herpes simplex virus. *J Virol* 63(9), 3844-3851 (1989).

169. Rodahl E, Haarr L: Analysis of the 2-kilobase latency-associated transcript expressed in PC12 cells productively infected with herpes simplex virus type 1: evidence for a stable, nonlinear structure. *J Virol* 71(2), 1703-1707 (1997).

170. Wu Tt, Su Yh, Block Tm, Taylor Jm: Evidence that two latency-associated transcripts of herpes simplex virus type 1 are nonlinear. *J Virol* 70(9), 5962-5967 (1996).

171. Rodahl E, Haarr L: Analysis of the 2-Kilobase Latency-Associated Transcript Expressed in PC12 Cells Productively Infected with Herpes Simplex Virus Type 1: Evidence for a Stable, Nonlinear Structure. *J Virol* 71(2), 1703-1707 (1996).

172. Nam K, Lee G, Trambley J, Devine Se, Boeke Jd: Severe growth defect in a Schizosaccharomyces pombe mutant defective in intron lariat degradation. *Mol Cell Biol* 17(2), 809-818 (1997).

173. Ruskin B, Krainer Ar, Maniatis T, Green Mr: Excision of an intact intron as a novel lariat structure during pre-mRNA splicing in vitro. *Cell* 38(1), 317-331 (1984).

174. Vogel J, Hess Wr, Borner T: Precise branch point mapping and quantification of splicing intermediates. *Nucleic Acids Res* 25(10), 2030-2031 (1997).

175. Odom Ow, Shenkenberg Dl, Garcia Ja, Herrin Dl: A horizontally acquired group II intron in the chloroplast psbA gene of a psychrophilic Chlamydomonas: in vitro self-splicing and genetic evidence for maturase activity. *RNA* 10(7), 1097-1107 (2004).

176. Salgia Sr, Singh Sk, Gurha P, Gupta R: Two reactions of Haloferax volcanii RNA splicing enzymes: joining of exons and circularization of introns. *RNA* 9(3), 319-330 (2003).

177. Coombes Ce, Boeke Jd: An evaluation of detection methods for large lariat RNAs. *RNA* 11(3), 323-331 (2005).

178. Hung T, Chang Hy: Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol* 7(5), 582-585 (2010).

179. Wilusz Je, Sunwoo H, Spector Dl: Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev* 23(13), 1494-1504 (2009).

180. Lee Jt: Epigenetic regulation by long noncoding RNAs. *Science* 338(6113), 1435-1439 (2012).

181. Lee Jt: Gracefully ageing at 50, X-chromosome inactivation becomes a paradigm for RNA and chromatin control. *Nat Rev Mol Cell Biol* 12(12), 815-826 (2011).

182. Brown Cj, Hendrich Bd, Rupert Jl *et al.*: The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71(3), 527-542 (1992).

183. Zhao J, Sun Bk, Erwin Ja, Song Jj, Lee Jt: Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322(5902), 750-756 (2008).

184. Lee Jt, Davidow Ls, Warshawsky D: Tsix, a gene antisense to Xist at the X-inactivation centre. *Nat Genet* 21(4), 400-404 (1999).

185. Lee Jt, Lu N: Targeted mutagenesis of Tsix leads to nonrandom X inactivation. *Cell* 99(1), 47-57 (1999).

186. Donohoe Me, Silva Ss, Pinter Sf, Xu N, Lee Jt: The pluripotency factor Oct4 interacts with Ctcf and also controls X-chromosome pairing and counting. *Nature* 460(7251), 128-132 (2009).

187. Sado T, Hoki Y, Sasaki H: Tsix silences Xist through modification of chromatin structure. *Dev Cell* 9(1), 159-165 (2005).

188. Sun Bk, Deaton Am, Lee Jt: A transient heterochromatic state in Xist preempts X inactivation choice without RNA stabilization. *Mol Cell* 21(5), 617-628 (2006).

189. Ogawa Y, Sun Bk, Lee Jt: Intersection of the RNA interference and X-inactivation pathways. *Science* 320(5881), 1336-1341 (2008).

190. Ogawa Y, Lee Jt: Xite, X-inactivation intergenic transcription elements that regulate the probability of choice. *Mol Cell* 11(3), 731-743 (2003).

191. Stavropoulos N, Rowntree Rk, Lee Jt: Identification of developmentally specific enhancers for Tsix in the regulation of X chromosome inactivation. *Mol Cell Biol* 25(7), 2757-2769 (2005).

192. Tian D, Sun S, Lee Jt: The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 143(3), 390-403 (2010).

193. Kent Jr, Kang W, Miller Cg, Fraser Nw: Herpes simplex virus latency-associated transcript gene function. *J Neurovirol* 9(3), 285-290 (2003).

194. Jin L, Perng Gc, Carpenter D *et al.*: Reactivation phenotype in rabbits of a herpes simplex virus type 1 mutant containing an unrelated antiapoptosis gene in place of latency-associated transcript. *J Neurovirol* 13(1), 78-84 (2007).

195. Allen Sj, Rhode-Kurnow A, Mott Kr *et al.*: Interactions between herpesvirus entry mediator (TNFRSF14) and latency-associated transcript during herpes simplex virus 1 latency. *J Virol* 88(4), 1961-1971 (2014).