5-9-2016

Characterization of Simian Hemorrhagic Fever Virus Subgenomic RNAs and Proteins

Han Di

CHARACTERIZAITON OF SIMIAN HEMORRHAGIC FEVER VIRUS SUBGENOMIC RNAS AND PROTEINS

by

HAN DI

Under the Direction of Margo A. Brinton

ABSTRACT

Simian hemorrhagic fever virus (SHFV) is an enveloped, single stranded, positive sense RNA virus that infects monkeys. SHFV has the largest genome (15.7 kb) in the family *Arteriviridae* that encodes two polyproteins. These polyproteins are auto-cleaved into nonstructural proteins by three different types of viral proteases. Compared to other arteriviruses, SHFV encodes three, rather than one or two, functional papain-like protease (PLP) 1 domains at the 5′ end of the genome. The catalytic residues and cleavage site(s) were determined for each of the SHFV PLP1 domains. The 3′ end of the SHFV genome encodes 12 structural proteins including a second set of minor structural proteins not found in other arteriviruses. The SHFV structural proteins are translated from a nested set of 3′ and 5′ coterminal subgenomic mRNAs (sg mRNA). Syntheses of the minus strand templates for the sg mRNAs is discontinuous and regulated by transcription regulating sequences (TRSs) in the genome. The SHFV sg mRNAs produced were analyzed by Northern blot assay and 37 additional functional TRSs were

discovered in SHFV genome by cloning and sequencing the sg mRNA junction sequences. Next generation sequencing analysis of the SHFV transcriptome confirmed sg mRNA production from all of the discovered TRSs and allowed estimation of the transcription level of each sg mRNA as well as the expression level of each 3′ ORF at early and late times after infection. Some of the sg mRNAs generated from newly discovered TRSs encode new ORFs. Two of these are functionally relevant for efficient virion production in MA104 cells. The potential for constructing a recombinant SHFV expressing a foreign gene from an additional sg mRNA was explored using an SHFV LVR cDNA infectious clone. The position between the two sets of minor structural protein ORFs was found to be tolerant for insertion of a foreign gene. Three different TRSs were inserted individually after the foreign gene and each of them restored the expression of the downstream SHFV structural protein ORF and generated infectious progeny virus. However, only one of these progeny virus that maintained the full-length foreign gene in the genome after three passages.

INDEX WORDS: Simian Hemorrhagic Fever Virus, Papain-Like Protease, Discontinuous RNA Synthesis, Subgenomic mRNA, Transcription Regulating Sequence, Next Generation Sequencing, Recombinant Virus, Foreign Gene Expression Vector, Infectious Clone

CHARACTERIZAITON OF SIMIAN HEMORRHAGIC FEVER VIRUS SUBGENOMIC RNAS AND PROTEINS

by

HAN DI

A Dissertation Submitted in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in the College of Arts and Sciences

Georgia State University

2016

CHARACTERIZAITON OF SIMIAN HEMORRHAGIC FEVER VIRUS SUBGENOMIC RNAS AND PROTEINS

by

HAN DI

Committee Chair: Margo A. Brinton

Committee: Richard K. Plemper

Parjit Kaur

Electronic Version Approved:

Office of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2016

DEDICATION

This dissertation is dedicated to my parents, Weimin Jin and Jiping Di who have given me unconditional love and tremendous support in my life, who have always been great role models all around, who pushed me to aim high and encouraged me to go far. To my dearest husband Joseph Madden who cheers me up, calms me down, makes me laugh and has given me all the love, care, support and insights throughout this long journey. I could not have gone this far without every one of you.

.

ACKNOWLEDGEMENTS

I would like to extend my sincerest gratitude to my principle advisor Dr. Margo A. Brinton. Thank you very much for always being patient and giving me guidance, knowledge and help when I needed it, for training me how to think critically and write professionally, for sponsoring me to attend all the different conferences and for supporting me to explore new areas and reach out for collaborations. I would also like to thank my committee members Dr. Richard K. Plemper and Dr. Parjit Kaur for providing me with all the detailed and updated knowledge in virus entry and prokaryotic genetics, and also for your valuable suggestions, witty comments and insightful questions about my dissertation research. Lastly, I would like to extend my thanks to all the past Brinton lab members who taught me techniques and let me work with them on their projects, to all the present Brinton lab members, especially Esther Morantz, Joseph Madden and Heena Sadhwani, who have been great team members and assisted me with my experiments.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS ..... **v**

LIST OF TABLES ..... **xii**

LIST OF FIGURES ..... **xiv**

LIST OF ABBREVIATIONS ..... **xvi**

1 INTRODUCTION **1**

1.1 Nidoviruses **1**

1.2 SHFV infection in monkeys **2**

1.3 SHFV outbreaks and isolates **3**

1.4 The arterivirus life cycle **3**

1.5 Arterivirus non-structural proteins **5**

1.6 Arterivirus structural proteins **7**

1.7 Discontinuous RNA synthesis **8**

1.8 Arterivirus sg mRNAs and TRSs **11**

1.9 Goals of this dissertation **12**

1.9.1 Aim I: Analyses of the catalytic residues and cleavage sites of the three SHFV PLP1s **12**

1.9.2 Aim II: Identification of SHFV body TRSs and analyses of sg mRNA production and protein coding capacity in the SHFV genome **13**

1.9.3 Aim III: Construction of a recombinant SHFV expressing a foreign gene from an additional sg mRNA **13**

2 FUNCTIONAL ANALYSES OF THE THREE SIMIAN HEMORRHAGIC FEVER VIRUS NONSTRUCTURAL PROTEIN 1 PAPAIN-LIKE PROTEASES 17

2.1 Abstract 18

2.2 Importance 18

2.3 Introduction 19

2.4 Materials and methods 22

2.4.1 Sequence and bioinformatics analyses. 22

2.4.2 Cells and virus. 22

2.4.3 Construction of SHFV nsp1 clones. 23

2.4.4 In vitro, coupled transcription/translation reactions. 24

2.4.5 Immunoprecipitation. 24

2.4.6 MassSpec sequencing of autoprocessed peptides. 25

2.4.7 Western blotting. 26

2.5 Results: 26

2.5.1 Analysis of SHFV nsp1s by sequence alignment and homology modeling. 26

2.5.2 Analysis of the peptides produced in vitro by autoprocessing of wildtype and mutant SHFV N-terminal ORF1a polyprotein fragments. 29

2.5.3 Functional analysis of the predicted PLP1 catalytic Cys residues 31

2.5.4 Functional analysis of the cleavage sites utilized by the SHFV PLPs 34

2.5.5 MassSpec sequencing of selected in vitro autoprocessing cleavage products. 36

2.5.6 Analysis of nsp1 processing in infected cells. **37**

2.6 Discussion **38**

3 CHARACTERIZATION OF TRANSCRIPTION REGULATING SEQUENCES (TRS) AND PROTEINS ENCODED BY THE SHFV GENOME **49**

3.1 Introduction **49**

3.2 Materials and methods: **53**

3.2.1 Cells and virus **53**

3.2.2 Digoxigenin (DIG)-labeled SHFV probes **54**

3.2.3 Northern blot hybridization **55**

3.2.4 Analysis of the leader-body junction sequences **56**

3.2.5 NGS of SHFV sg mRNAs **57**

3.2.6 NGS data analysis **58**

3.2.7 Construction of mutant SHFV infectious clone virus **59**

3.2.8 Western Blot Assay **60**

3.2.9 Plaque Assay **61**

3.3 Results: **61**

3.3.1 Discovery of additional sg mRNAs generated during an SHFV infection in MA104 cells **61**

3.3.2 Additional body TRSs in the SHFV genome are used to initiate sg RNAs *63*

3.3.3 NGS analysis of SHFV sg mRNAs in MA104 cells **66**

3.3.4 NGS analysis of SHFV sg mRNAs in macaque MΦs **68**

3.3.5 Analysis of the expression level of SHFV 3′ ORFs **69**

3.3.6 Many additional body TRSs exist in both the 3′ end of the genome and in the ORF1a/1b region **71**

3.3.7 Knocking out the expression of ORF5-C or ORF6-C reduced SHFV yield in MA104 cells **73**

3.4 Discussion **74**

4 DEVELOPMENT OF STRATEGIES FOR EXPRESSING AN EXOGENOUS GENE FROM SHFV GENOME **101**

4.1 Introduction **101**

4.2 Materials and methods: **103**

4.2.1 Cells **103**

4.2.2 Construction of an M013 mammalian expression plasmid pEF6-M013-V5 ..... **103**

4.2.3 Insertion of the M013 gene into the full-length SHFVic to construct SHFVic-M013 ..... **104**

4.2.4 SHFVic-M013 virus cell passage protocol **105**

4.2.5 Reverse transcription and amplification of SHFVic-M013 genome RNA **106**

4.2.6 TA-cloning of RT-PCR products **106**

4.2.7 Western Blot Assay **107**

4.2.8 Immunofluorescent Assay (IFA) **107**

4.2.9 Plaque Assay **108**

4.3 Results: **109**

4.3.1 Overexpression of M013 does not negatively affect SHFV replication in MA104 cells **109**

4.3.2 Generation of SHFVic virus expressing M013 protein from an additional sg mRNA **109**

4.3.3 Analysis of M013 expression from the SHFVic-M013 viruses during serial virus passage **111**

4.3.4 Analysis of the M013 sequence in the genomes of SHFVic-4′/2b-M013-TRS viruses **112**

4.3.5 Analysis of the M013 gene expression pattern in infected cells **113**

4.3.6 Comparison of the growth kinetics of SHFVic-WT virus and SHFVic-M013 virus **114**

4.4 Discussion **114**

5 CONLUSIONS AND FUTURE DIRECTIONS **125**

6 ADDITIONAL WORK: CHARACTERIZATION OF HUMAN OAS1 ISOFORMS **131**

6.1 Introduction **131**

6.2 Materials and methods: **133**

6.2.1 Cells **133**

6.2.2 Protein purification **134**

6.2.3 In vitro OAS1 synthesis assay **135**

6.2.4 Functional analysis of human OAS1 isoforms in mammalian cells **135**

6.2.5 Yeast two hybrid assay **135**

6.2.6 Yeast co-transformation **136**

6.2.7 In vitro transcription/translation and pull-down assay **136**

6.2.8 Mammalian cell co-immunoprecipitation **137**

6.2.9 Immunofluorescent assay (IFA) **138**

6.2.10 Western Blot Assay **139**

6.3 Results: **139**

6.3.1 Human OAS1 p52 is functionally active in vitro **139**

6.3.2 Overexpressed human OAS1 isoforms activate RNase L cleavage after transfection of cells with poly (I:C) **140**

6.3.3 Yeast two hybrid assay identified novel binding partners for human OAS1 p42 and p44 isoforms **142**

6.3.4 Human OAS1 p42 and p44 isoforms interact with their binding partners in vitro **143**

6.3.5 Human OAS1 p44 isoform interacts with its binding partner in cells **144**

6.4 Discussion **144**

REFERENCES **153**

LIST OF TABLES

Table 2.1. Primers 48

Table 3.1. Primers used to generate probes specific for each SHFV sg mRNA by targeting the regions between each set of adjacent predicted body TRSs 82

Table 3.2. Primers used to amplify the leader-body junctions of individual sg mRNAs generated from known/predicted body TRSs in the 3′ end of the SHFV genome 83

Table 3.3. Primers used to knock out the expression of the new C-terminal in-frame ORFs of known structural proteins in the SHFV genome. 83

Table 3.4. Body TRSs identified in the SHFV genome by amplification, cloning and sequencing of known or predicted leader-body junctions. 84

Table 3.5. NGS analysis of the transcription level and relative abundance of sg mRNAs generated from each identified body TRSs at 8 and 18 hpi in MA104 cells. 85

Table 3.6. NGS analysis of the transcription level and relative abundance of sg mRNAs generated from each identified body TRSs at 7 and 16 hpi in macaque MΦs..... 87

Table 3.7. NGS analysis of the expression level and relative abundance of SHFV 3′ ORFs encoded by sg mRNAs generated from identified body TRSs at 8 and 18 hpi in MA104 cells. 89

Table 3.8. NGS analysis of the expression level and relative abundance of SHFV 3′ ORFs encoded by sg mRNAs generated from identified body TRSs at 7 and 16 hpi in macaque MΦs. 91

Table 3.9. Prediction of transmembrane domains in the C-terminal in-frame ORFs of known structural proteins..... 93

Table 4.1. Primers 118

Table 6.1. Identification of peptide candidate binding partners for the full-length human OAS1 p42 and p44 isoforms using yeast two hybrid analysis..... 149

Table 6.2. Confirmation of the peptide candidates as binding partners for the full-length human OAS1 p42 and p44 isoforms using yeast co-transformation 150

LIST OF FIGURES

Figure 1.1. Diagram of the discontinuous RNA synthesis model for arteriviruses and coronaviruses. 14

Figure 1.2. Diagram of the SHFV genome, sg mRNAs and encoded proteins. 15

Figure 2.1. Prediction of the three SHFV PLP domains. 42

Figure 2.2. SHFV nsp1 polyprotein cleavage map. 43

Figure 2.3. Cleavage products produced by nsp1 polyproteins containing mutations in one or two catalytic Cys residues. 44

Figure 2.4. Cleavage products produced by nsp1 polyproteins with a single predicted PLP cleavage site mutated. 45

Figure 2.5. MassSpec analyses of selected polyprotein cleavage products. 46

Figure 2.6. Autoprocessing of SHFV nsp1 proteins in infected cells. 47

Figure 3.1. Northern blotting analysis of sg mRNAs produced in SHFVic infected MA104 cells. 94

Figure 3.2. Amplification and cloning of sg mRNA leader-body junctions generated from the predicted TRS3′ region. 95

Figure 3.3. Amplification and cloning of sg mRNA leader-body junctions generated from the region between the identified TRS5 and TRS6. 96

Figure 3.4. Mapping the NGS reads not containing identified leader-body junction sequences to the SHFV genome. 97

Figure 3.5. Mapping the NGS reads not containing identified leader-body junction sequences to the SHFV genome. 98

Figure 3.6. Discovery of a separate sg mRNA encoding ORF5a as its 5′ proximal ORF. 99

Figure 3.7. Analysis of the effect of knocking out the expression of each new C-terminal ORF on virus replication in MA104 cells..... 100

Figure 4.1. Overexpression of M013 did not negatively affect SHFV RNA replication in MA104 cells. 119

Figure 4.2 Western blot analysis of protein expression in cells transfected with one of the three SHFVic-4′/2b-M013-TRS virus RNAs..... 119

Figure 4.3. Effect of virus passage on M013 expression from the SHFVic-M013 viruses. 120

Figure 4.4. Analysis of M013 inserts in the genomes of SHFVic-M013 viruses after passage. 121

Figure 4.5. Diagrams of internal deletions in the M013-V5-TRS sequence in recombinant SHFVic genomes. 122

Figure 4.6. Analysis of the M013 expression pattern from the SHFVic-4′/2b-M013-TRS2′ virus. 123

Figure 4.7. Growth curves of SHFVic-WT virus and SHFVic-4′/2b-M013-TRS2′ virus in MA104 cells. 124

Figure 6.1. Analysis of the synthetase activity of the human OAS1 p52 isoform in vitro 147

Figure 6.2. Analysis of RNase L cleavage in cells overexpressing individual human OAS1 isoforms in response to poly (I:C) 148

Figure 6.3. Confirmation of direct interaction between human OAS1 isoforms p42 and p44 and their respective binding partners 151

Figure 6.4. The interaction between human OAS1 p44 and the endogenous full-length SVIL in cells was confirmed by co-immunoprecipitation. 152

LIST OF ABBREVIATIONS

Simian hemorrhagic fever virus (SHFV)

Porcine reproductive and respiratory syndrome virus (PRRSV)

Equine arteritis virus (EAV)

Lactate dehydrogenase-elevating virus (LDV)

Wobbly possum disease virus (WPDV)

Foot-and-mouth disease virus (FMDV)

Open reading frame (ORF)

Next generation sequencing (NGS)

Macrophages (MΦs)

Dendritic cells (DCs)

Replication and transcription complex (RTC)

Double-membrane vesicles (DMVs)

Endoplasmic reticulum (ER)

Subgenomic RNA (sg RNA)

Subgenomic mRNA (sg mRNA)

Transcription regulating sequence (TRS)

Nucleocapsid protein (N)

Membrane protein (M)

Glycoprotein (GP)

Non-structural proteins (nsp)

Papain-like protease (PLP)

Chymotrypsin-like serine protease (SP)

Multiplicity of infection (MOI)

Hours post infection (hpi)

Cytopathic effect (CPE)

Interferon (IFN)

3′-untranslated region (3′-UTR)

SHFV infectious clone virus (SHFVic)

Peripheral blood mononuclear cells (PBMCs)

Digoxigenin (DIG)

Immunofluorescent assay (IFA)

Oligoadenylate synthetase (OAS)

Double-stranded RNA (dsRNA)

Single-stranded RNA (ssRNA),

2′-5′ linked oligoadenylates (2′-5′A).

Endoribonuclease L (RNase L),

Ribosomal RNA (rRNA)

Single nucleotide polymorphism (SNP)

Supervillin (SVIL)

Fibrillin-1 (FBN1)

Voltage-dependent anion channel2 (VDAC2)

Phosphoglucomutase 1 (PGM1)

Minimal Essential Medium (MEM)

Dulbecco's Modified Eagle Medium (DMEM)

Double dropout (DDO)

Triple dropout (TDO)

Quadruple dropout (QDO)

1 INTRODUCTION

1.1 Nidoviruses

The order *Nidovirales* consists of four virus families: *Coronaviridae*, *Toroviridae*, *Roniviridae* and *Arteriviridae*. Among the Nidoviruses, coronaviruses, toroviruses and roniviruses have the largest known RNA genomes (~26-32 kb), whereas arteriviruses have much smaller genomes (~13–16 kb) (Gorbalenya, Enjuanes et al. 2006). Nidoviruses are single-stranded, positive-sense RNA viruses that have a 5′ cap and a 3′ poly (A) tail. All Nidoviruses share the same gene arrangement. The 5′ two-thirds of the genome encodes the non-structural proteins (nsp) while the 3' one third encodes the viral structure proteins and, in some families, accessory proteins. The 3′ genes are translated from a nested set of subgenomic mRNAs (sg mRNAs) (Gorbalenya, Enjuanes et al. 2006). Coronaviruses and Arteriviruses are the two most studied groups of Nidoviruses. Coronaviruses can infect humans and a large variety of animals. A pandemic caused by a single coronavirus, SARS, killed 774 people and caused 8,096 disease cases in 30 countries in 2003 (Cheng, Lau et al. 2007). Coronaviruses are further subdivided into four genera (alpha, beta, gamma and delta) based on phylogenetic clustering (Fehr and Perlman 2015). Arteriviruses are only known to infect animals and have restricted host ranges. Porcine reproductive and respiratory syndrome virus (PRRSV), equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) and simian hemorrhagic fever virus (SHFV) were previously identified as arteriviruses. Recently, new arteriviruses, wobbly possum disease virus (WPDV), a novel arterivirus from the forest giant pouched rat and multiple new simian arteriviruses have been discovered (Dunowska, Biggs et al. 2012; Snijder, Kikkert et al. 2013; Kuhn, Lauck et al. 2016). The economic cost of PRRSV infection for the US swine industry was estimated at $664 million annually between 2005 and 2010 (Holtkamp, Kliebenstein et al. 2013).

1.2 SHFV infection in monkeys

The natural hosts of simian arteriviruses, such as SHFV, are various species of African monkeys. Infections in these monkeys establish asymptomatic persistent infections (London 1977; Lauck, Hyeroba et al. 2011; Lauck, Sibley et al. 2013; Bailey, Lauck et al. 2014). SHFV, simian hemorrhagic encephalitis virus (SHEV) and Pebjah virus (PBJV) infections in macaque monkeys trigger an acute, fatal hemorrhagic fever disease characterized by symptoms including fever, dehydration, nose bleeds, skin petechiae and retrobulbar hemorrhage (Allen, Palmer et al. 1968; Palmer, Allen et al. 1968; Johnson, Dodd et al. 2011; Lauck, Alkhovsky et al. 2015; Vatter, Donaldson et al. 2015). Death usually occurs 7 to 13 days after infection. The symptoms and disease progression in macaques induced by SHFV are similar to those induced by other hemorrhagic fever viruses such as Ebola virus (Bray 2005; Bray and Geisbert 2005). Previous studies of viral hemorrhagic fever disease induced by Ebola virus have shown that upregulated inflammatory mediators can lead to increased vascular permeability and coagulopathy (Bray 2005; Bray and Geisbert 2005). Because SHFV does not infect humans, SHFV infections in macaques can serve as a biosafety level II animal model for the study of viral hemorrhagic fever disease (Johnson, Dodd et al. 2011). Macrophages (MΦs) and dendritic cells (DCs) are the main target cells for SHFV infection in monkeys. A recent study comparing the host responses induced by SHFV infection in primary MΦs and DCs from baboons and macaques showed that SHFV infection induced the production of pro-inflammatory cytokines such as IL-1β, IL-6, IL-12/23 (p40), TNF-α and MIP-1α in macaque cells, but not in baboon cells (Vatter and Brinton 2014). MA104, an embryonic kidney epithelial cell line from African green monkey, is permissive to SHFV infection in tissue culture.

1.3 SHFV outbreaks and isolates

SHFV was first isolated during a simian hemorrhagic fever epizootic that occurred in 1964 among captive macaques in research institutes in the USSR and the United States (Allen, Palmer et al. 1968; Palmer, Allen et al. 1968). From 1964 to 1996, there were a total of 17 simian hemorrhagic fever outbreaks in captive macaques (Lauck, Alkhovsky et al. 2015). Although different SHFV isolates were reported during those historic outbreaks, all of them were destroyed except for the 1964 isolate NIH *LVR*42-0/*M*6941 (Tauraso, Shelokov et al. 1968). A complete sequence of the SHFV LVR genome was obtained in the early 1990s by "shot-gun" sequencing and updated in 2014 by 454 sequencing of the SHFV LVR genome RNA (NC_003092.2) (Vatter, Di et al. 2014). Recently, next generation sequencing (NGS) of viral RNA isolated from different species of wild African monkeys, such as red colobus and red-tailed guenons, discovered a number of highly divergent simian arterivirus sequences, suggesting expansion of simian arteriviruses (Lauck, Hyeroba et al. 2011; Lauck, Sibley et al. 2013; Bailey, Lauck et al. 2014; Bailey, Lauck et al. 2014). To date, only one of these new simian arteriviruses has been isolated in cell culture (MARC-145 cells) and was shown to only cause mild self-limiting disease in macaques (Wahl-Jensen, Johnson et al. 2015).

1.4 The arterivirus life cycle

After attachment to cell surface receptors, arteriviruses enter cells through the clathrin-mediated endocytosis pathway (Kreutz and Ackermann 1996; Nitschke, Korte et al. 2008; Cai, Postnikova et al. 2015). The cell receptors and factors used for arterivirus attachment and entry have been only partially identified. Two MΦ-specific proteins, sialoadhesin (CD169) and CD163, were shown to mediate virion attachment and entry (Vanderheijden, Delputte et al.

2003; Calvert, Slade et al. 2007; Van Breedam, Van Gorp et al. 2010; Welch and Calvert 2010; Cai, Postnikova et al. 2015). After uncoating of the genome inside cells, two large polyproteins (pp1a and pp1ab) are translated from the 5′ two thirds of the genome. Translation of pp1ab requires a -1 ribosomal frame shift just 5′ of the open reading frame (ORF) 1a stop codon (Figure 1-2A). Both pp1a and pp1ab are proteolytically processed into multiple non-structural proteins that form the replication and transcription complex (RTC) for viral RNA synthesis (Snijder, Kikkert et al. 2013). The arterivirus RTC was shown to associate with double-membrane vesicles (DMVs) that are interconnected with each other and with the endoplasmic reticulum (ER) (Pedersen, van der Meer et al. 1999; Posthuma, Pedersen et al. 2008; Knoops, Barcena et al. 2012). During replication, a full-length minus-strand RNA is generated which is used as the template to amplify the plus-strand genome. During synthesis of minus-strand RNA, a complex discontinuous transcription mechanism is used to generate a nested set of minus-strand subgenomic RNAs (sg RNAs) that are used as templates to generate plus-strand subgenomic mRNAs (sg mRNAs) from which the structural proteins are translated (Pasternak, Spaan et al. 2006) (Figure 1-2C). The nucleocapsid protein (N) forms a spherical structure that encapsidates the viral genome (Wissink, Kroese et al. 2005; Veit, Matczuk et al. 2014). Translated viral major and minor envelope structural proteins associate with the ER membrane and virions bud into the lumen of ER acquiring a lipid envelope (Wood, Tauraso et al. 1970; Knoops, Barcena et al. 2012). The enveloped virions are transported through the Golgi complex, where the glycosylation of some of the envelope structural proteins occurs, and then to the plasma membrane where they are released by exocytosis. (Veit, Matczuk et al. 2014)

1.5 Arterivirus non-structural proteins

Arterivirus non-structural proteins are released from pp1a and pp1ab via auto-cleavage mediated by protease domains encoded in the polyproteins. There are three types of arterivirus protease domains, papain-like protease (PLP) 1, PLP2 and chymotrypsin-like serine protease (SP) (Snijder, Kikkert et al. 2013). The PLP1 domains in studied arteriviruses employ a Cys-His catalytic tandem and exclusively cleave *in cis* at a single site downstream of the PLP domain (den Boon, Faaberg et al. 1995; Sun, Xue et al. 2009; Xue, Sun et al. 2010). The PLP2 domains in studied arteriviruses are more like coronavirus PLPs in that they use a Cys-His-Asp catalytic triad and can cleave both *in cis* and *in trans* (Han, Rutherford et al. 2009; Mielech, Chen et al. 2014). Unlike the coronavirus PLPs, the cleavage sites for arterivirus PLP1s and PLP2s are not well conserved (Sun, Xue et al. 2009; Xue, Sun et al. 2010; Mielech, Chen et al. 2014). The SP domain contains the His-Asp-Ser catalytic triad of the classical chymotrypsin-like protease and cleaves specifically at sites consisting of a Glu or Gln adjacent to a Gly, Ala or Ser (Snijder, Wassenaar et al. 1996; Tian, Lu et al. 2009). In all arteriviruses, nsp4 contains the SP domain which serves as the main protease that cleaves at multiple sites in the polyproteins (Snijder, Wassenaar et al. 1996; van Dinten, Rensen et al. 1999). Nsp2 contains the PLP2 domain that cleaves at a downstream site to release itself (Snijder, Wassenaar et al. 1995). The LDV and PRRSV nsp1 region contain two PLP1 domains, PLP1α and PLP1β, that each cleaves at a downstream site to release nsp1α and nsp1β, respectively (den Boon, Faaberg et al. 1995; Chen, Lawson et al. 2010). The EAV nsp1 region also contains two PLP1 domains but the catalytic Cys in the N-terminal PLP1 domain is substituted by a Lys, therefore a single nsp1 fusion protein is produced (Snijder, Wassenaar et al. 1992; Ziebuhr, Snijder et al. 2000). Three PLP1 domains were predicted in the SHFV genome (Figure 1-2A). In addition to protease activity, arterivirus

nsp2s also have deubiquitinating activity that can remove ubiquitin from innate immune components such as retinoic acid-inducible gene 1 (RIG-I) and IκB-α to antagonize type I interferon induction in infected cells (Sun, Chen et al. 2010; van Kasteren, Beugeling et al. 2012). The PRRSV nsp2 was also shown to inhibit the expression and conjugation of interferon-stimulated gene 15 (ISG15) in infected cells (Sun, Li et al. 2012). In addition to nsp2, the arterivirus nsp1α, nsp1β and nsp4 proteins were also reported to antagonize interferon (IFN) induction and signaling pathways via different mechanisms in infected cells, demonstrating the multifunctionality of arterivirus nsps (Beura, Sarkar et al. 2010; Song, Krell et al. 2010; Han, Du et al. 2013; Huang, Zhang et al. 2014; Rascon-Castelo, Burgara-Estrella et al. 2015).

Arterivirus infections were shown to induce the rearrangement of intracellular membranes and the formation of interconnected DMVs in the perinuclear region that are associated with viral RTCs (Wood, Tauraso et al. 1970; Pedersen, van der Meer et al. 1999; Knoops, Barcena et al. 2012). Arterivirus ORF1a encoded proteins, such as nsp2, nsp3 and nsp5 may function in cell membrane rearrangement and as anchors of the RTCs to the DMV membranes (van der Meer, van Tol et al. 1998; Pedersen, van der Meer et al. 1999; Snijder, van Tol et al. 2001). Major components of the viral RTCs are the RNA dependent RNA polymerase (nsp9) and the helicase (nsp10) that can unwind dsRNA in a 5′ to 3′ direction (Bautista, Faaberg et al. 2002; Beerens, Selisko et al. 2007). Another protein nsp11 that contains an endoribonuclease domain (NendoU) that is specific to Nidoviruses may also function in viral RNA synthesis (Posthuma, Nedialkova et al. 2006; Nedialkova, Ulferts et al. 2009). The N-terminal zinc finger domains of the EAV nsp1 and the PRRSV nsp1α proteins were shown to be functionally important for arterivirus sg mRNA transcription (Tijms, van Dinten et al. 2001; Tijms, Nedialkova et al. 2007; Kroese, Zevenhoven-Dobbe et al. 2008; Nedialkova, Gorbalenya

et al. 2010). The predicted metal-binding region of the EAV helicase was also reported to play a role in sg mRNA synthesis (van Dinten, van Tol et al. 2000)

1.6 Arterivirus structural proteins

Arteriviruses express their structural proteins from the 3′ one-third of their genomes via a nested set of sg mRNAs (Figure 1-2A). All arteriviruses express three major structural proteins: glycoprotein 5 (GP5), non-glycosylated membrane protein (M) and nucleocapsid protein (N) (Snijder, Kikkert et al. 2013). EAV, PRRSV and LDV also encode five minor structural proteins: GP2, GP3, GP4, E and the recently identified ORF5a (Firth, Zevenhoven-Dobbe et al. 2011; Johnson, Griggs et al. 2011). SHFV is unique in that it encodes an extra set of minor structural proteins: GP2′, GP3′, GP4′ and E′, for a total of nine minor structural proteins (Godeny, de Vries et al. 1998) (Figure 1-2A and B). The majority of the functional studies of the arterivirus structural proteins have been on EAV and PRRSV. Among the major structural proteins, homodimers of N protein interact with each other to form a spherical nucleocapsid that contains the viral genome (Doan and Dokland 2003; Dokland 2010). The formation of GP5/M heterodimers in the ER membrane of infected cells was shown to be essential for EAV assembly (Snijder, Dobbe et al. 2003; Wieringa, de Vries et al. 2004). GP5/M heterodimers present in the viral envelope were reported to function in virus attachment to host cell receptors (Delputte, Vanderheijden et al. 2002; Van Breedam, Van Gorp et al. 2010). The ectodmain of GP5 was shown to contain the primary neutralization epitope of EAV, LDV and PRRSV (Plagemann, Rowland et al. 2002; Balasuriya, Dobbe et al. 2004; Plagemann 2004). GP2, GP3 and GP4 form heterotrimers in the viral envelope (Wieringa, de Vries et al. 2003). PRRSV GP2 and GP4 were shown to interact with the cell receptor CD163 to mediate virus entry into susceptible cells (Das,

Dinh et al. 2010). The E protein was postulated to form homo-oligomers that function as ion-channels in the viral envelope and to be essential for virus infectivity through facilitating virus uncoating and release of the genome into the cytoplasm (Lee and Yoo 2006). An interesting study showed that a PRRSV chimera expressing the GP2, GP3, GP4 and E proteins from EAV acquired the broad cell tropsim that is typical of EAV, suggesting these minor structural proteins are major determinants of viral tropism in cell culture (Tian, Wei et al. 2012). The function of the recently discovered ORF5a protein is not yet known. Knocking out the expression of ORF5a was lethal for the production of infectious PRRSV but only reduced the yields of EAV (Firth, Zevenhoven-Dobbe et al. 2011; Johnson, Griggs et al. 2011; Sun, Li et al. 2013). Compared to the other arteriviruses, SHFV encodes an extra set of minor structural proteins. A set of full-length infectious clones of SHFV-LVR was constructed with the start codon of each minor structural protein inactivated. Each minor structural protein was found to be required for the production of infectious extracellular virus, indicating that the two sets of SHFV minor structural proteins are not functionally redundant (Vatter, Di et al. 2014).

1.7 Discontinuous RNA synthesis

Nidovirus RNA synthesis results in the generation of 3′-coterminal, nested sets of sg mRNAs used for structural protein translation. Arterivirus and coronavirus discontinuous RNA synthesis generates sg mRNAs that each have a leader sequence derived from the 5′ end of the genome (Baric, Stohlman et al. 1983; de Vries, Chirnside et al. 1990). In contrast, a leader sequence was not found in the sg mRNAs generated from the ronivirus, Gill-associated virus (GAV) (Cowley, Dimmock et al. 2002), and among the sg mRNAs generated by equine torovirus (EToV), only one was shown to contain a leader sequence at the 5′ end (van Vliet,

Smits et al. 2002), indicating that the majority of the ronivirus and torovirus sg mRNAs are not generated by discontinuous RNA synthesis. The arterivirus and coronavirus discontinuous RNA synthesis mechanism closely resembles similarity-assisted, copy-choice RNA recombination and is regulated by transcription regulating sequences (TRSs) in the genomes (Pasternak, Spaan et al. 2006; Sawicki, Sawicki et al. 2007). The TRSs of arterivirus, which are shorter than those of the coronaviruses, are usually 30-40 nt long and contain a 5-9 nt core sequence and two 10-15 nt immediately adjacent flanking sequences. Although there are multiple body TRSs located within the 3′ region of the genome, there is only one leader TRS at the 5′ end (de Vries, Chirnside et al. 1990). The core sequences of the various body TRSs have different degrees of homology with the core sequence of the leader TRS and both the leader and body TRSs can fold into stem-loop structures with the core sequence located in the top loop and the flanking sequences located in the stem (Van Den Born, Gultyaev et al. 2004; van den Born, Posthuma et al. 2005). The models proposed for coronavirus discontinuous RNA synthesis were controversial until the detection of minus-strand subgenomic RNAs for both arteriviruses and coronaviruses showed that discontinuous RNA synthesis occurred during the synthesis of the minus-strand, rather than the plus-strand RNA (Sethna, Hung et al. 1989; Sawicki and Sawicki 1995; den Boon, Kleijnen et al. 1996). Minus-strand RNA synthesis can either continue or attenuate at a particular body TRS. If attenuation occurs, a nascent minus-strand RNA containing an anti-body TRS sequence at its 3′ end is generated that can base-pair with the leader TRS (van Marle, Dobbe et al. 1999). After alignment of the nascent minus-strand RNA TRS with the leader TRS at the 5′ end of the genome, synthesis of the nascent minus-strand continues generating a minus-strand subgenomic RNA with a unique leader-body junction sequence (den Boon, Kleijnen et al. 1996; Pasternak,

van den Born et al. 2001). The minus-strand subgenomic RNAs are then used as templates to transcribe sg mRNAs that are used for protein expression (Figure 1-1).

Many factors regulate the efficiency of arterivirus and coronavirus discontinuous RNA synthesis. Because discontinuous RNA synthesis is guided by base-pairing between the anti-body TRS and the leader TRS, the degree of homology and the stability of this duplex are expected to affect the relative abundance of individual sg mRNAs (Pasternak, van den Born et al. 2003). However, there are sequence motifs within the genomes of arteriviruses and coronaviruses that perfectly match the sequence of the leader TRS but that are not used as body TRSs. Also, sometimes a body TRS with high homology to the leader TRS generates less sg mRNA than a body TRS with less homology, suggesting additional factors also play a role in regulation (Pasternak, Gultyaev et al. 2000). The sequences immediately flanking the core sequences of body TRSs have been reported to have an impact on the efficiency of sg mRNA synthesis (An and Makino 1998; Sola, Moreno et al. 2005). Extended base-pairing between the leader TRS region and the body TRS flanking sequences could compensate when the core sequences of the leader and body TRSs are less homologous (den Boon, Kleijnen et al. 1996). Also, the relative location of a body TRS in the genome can contribute to the relative sg mRNA abundance. Insertion of multiple adjacent copies of the same TRS core and flanking sequences into the EAV genome showed that there was an abundance gradient with the sg mRNA generated from the downstream TRS having the highest abundance (Pasternak, Spaan et al. 2004). These data support an earlier study done with a coronavirus showing that a downstream body TRS can suppress the function of an upstream TRS, especially when they are located close to each other (Joo and Makino 1995). The secondary structure formed by each body TRS region can also affect the accessibility of the core sequence and thereby the abundance of the sg mRNA

generated from that body TRS. A previous study on EAV showed that if the core sequence of a body TRS is located in the single-stranded region of a predicted secondary structure, it tends to be more active than when the core sequence is located in regions that are fully or partially base paired (Pasternak, Gultyaev et al. 2000).

1.8 Arterivirus sg mRNAs and TRSs

The sg mRNAs and TRSs of EAV and PRRSV have been the most extensively studied while those of LDV and SHFV have been less studied. Six sg mRNAs were first detected in EAV infected cells (sg mRNA2 - sg mRNA7) and six corresponding body TRSs (TRS2 - TRS7) were identified in the EAV genome (de Vries, Chirnside et al. 1990; den Boon, Kleijnen et al. 1996). The number of sg mRNAs produced in PRRSV infected cells varied from six to seven among different PRRSV isolates (Conzelmann, Visser et al. 1993; Meng, Paul et al. 1996). The body TRS of the additional sg mRNA was identified in a region located between TRS3 and TRS4. This additional sg mRNA encodes a novel small ORF of 45 amino acids (Meng, Paul et al. 1996). Studies done with LDV infected cells detected seven sg mRNAs and identified seven corresponding body TRSs in the genome (Kuo, Harty et al. 1991; Kuo, Chen et al. 1992). The additional LDV sg mRNA was larger than sg mRNA2 and encoded the C-terminal 200 amino acids of ORF1b (Chen, Kuo et al. 1993). Six strong sg mRNA bands were first detected in a Northern blot assay using SHFV infected cell extracts and corresponded to the findings in other arteriviruses (Godeny, Zeng et al. 1995; Zeng, Godeny et al. 1995). Later, sequencing results for the 3' region of the SHFV genome revealed the presence of nine ORFs, instead of the expected six (Smith, Wang et al. 1997). However, a subsequent study only identified eight body TRSs in the SHFV genome (Godeny, de Vries et al. 1998) (Figure 1-2A). Since then, additional

arterivirus structural proteins have been identified, such as the E protein, the ORF5a protein and an additional E′ protein for SHFV. Because a separate sg mRNA and body TRS for the E protein or for the ORF5a protein had not been identified, both sg mRNA2 and sg mRNA5 in all arteriviruses were thought to be bicistronical expressing GP2/E and GP5/ORF5a, respectively (Snijder, van Tol et al. 1999; Firth, Zevenhoven-Dobbe et al. 2011; Johnson, Griggs et al. 2011). The SHFV sg mRNA2′ was already assumed to be bicistronic expressing both GP2′ and GP3′ because a separate sg mRNA for SHFV GP3′ could not be identified (Godeny, de Vries et al. 1998). The discovery of E′ added the possibility that a third ORF was expressed from this sg mRNA (Figure 1-2C). Although each known structural protein had been assigned to one sg mRNA generated from one body TRS for each of the studied arteriviruses, new body TRSs generating alternative sg mRNAs encoding the same structural proteins, GP3, GP4 and GP5 were later identified in both the EAV and PRRSV genomes (Pasternak, Gultyaev et al. 2000; Lin, Chang et al. 2002). Mutagenesis of the major body TRSs in the EAV genome suggested that although the alternative sg mRNAs generated from the new TRSs were less abundant, they may be in a sufficient amount for production of infectious progeny virus. However, the virus yield was reduced when the major TRSs were mutated (Pasternak, Gultyaev et al. 2000).

1.9 Goals of this dissertation

1.9.1 Aim I: Analyses of the catalytic residues and cleavage sites of the three SHFV PLP1s

After infection, the 5′ end of the arterivirus genome is translated into two polyproteins pp1a and pp1ab that are cleaved into viral non-structural proteins by three viral encoded proteases: PLP1, PLP2 and SP (Snijder, Kikkert et al. 2013). Both EAV and PRRSV encode two

PLP1 domains at the 5′ end of the genome and one of the EAV PLP1 domains is not functionally active (Snijder, Wassenaar et al. 1992). SHFV encodes three PLP1 domains and the functionality of each SHFV PLP1 domain was partially studied by previous students in the lab. The goals of the present study were to continue the analysis of the catalytic residues and cleavage site(s) of each PLP1 using an *in vitro* transcription/translation and pull-down assay and to confirm the cleavage site(s) of each PLP1 by MassSpec analysis of the peptides generated by cleavage.

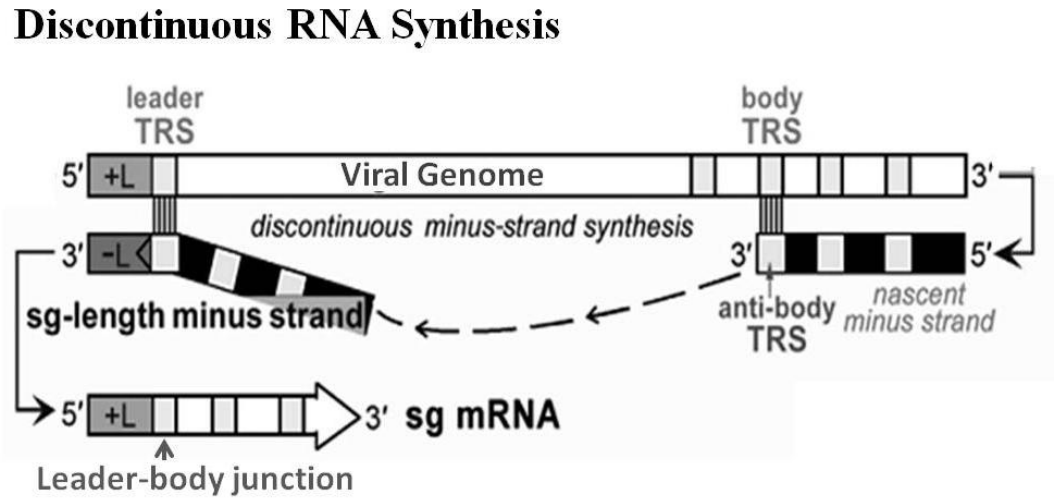
1.9.2 Aim II: Identification of SHFV body TRSs and analyses of sg mRNA production and protein coding capacity in the SHFV genome

With the expanded knowledge gained from recent studies in the arteriviruses, SHFV genome was predicted to express 12 structural proteins from sg mRNAs generated from the body TRSs at the 3′ end of the genome. However, previously published studies detected six sg mRNA bands on a Northern blot and identified eight body TRSs in the SHFV genome (Godeny, Zeng et al. 1995) (Godeny, de Vries et al. 1998). The goals of the present study were to carry out a comprehensive analysis of SHFV sg mRNA production, body TRS usage and predict the coding capacity of the SHFV genome using Northern blot assay with multiple probes, in-depth cloning of leader-body junction sequences and next-generation sequencing analysis.

1.9.3 Aim III: Construction of a recombinant SHFV expressing a foreign gene from an additional sg mRNA

The potential of arteriviruses as expression vectors has been previously explored using EAV and PRRSV infectious clones (de Vries, Glaser et al. 2001; Pei, Hodgins et al. 2009). SHFV has the largest genome among arteriviruses and it encodes one extra PLP1 domain at the

5′ end and four extra minor structural proteins at the 3′ end of the genome (Brinton, Di et al. 2015). The potential of SHFV as a vector to carry and express additional genes in the genome has not yet been explored. The goal of the present study was to construct the first recombinant SHFV expressing a foreign gene from an additional sg mRNA using a SHFV LVR cDNA infectious clone.



Modified from Pasternak, A. O., W. J. M. Spaan, et al. (2006)

Figure 1.1 Diagram of the discontinuous RNA synthesis model for arteriviruses and coronaviruses.
The 5′ leader TRS core sequence has different levels of homology with each of the body TRS core sequences at the 3′ end of the plus strand viral genome. The synthesis of a minus strand RNA can attenuate at any of the body TRSs due to the TRS secondary structure, but the base pairing between the anti-body TRS from the 3′ end of the nascent minus strand and the leader TRS at the 5′ end of the genome is needed for the synthesis to continue and generate a discontinuous minus strand sg RNA template for the transcription of sg mRNA. White boxes represent the plus strand viral genome or sg mRNA and the black boxes represent the minus strand template. The leader TRS, body TRS, anti-body TRS and the leader-body junction are represented by light grey boxes. The dark grey boxes at the genome 5′ end represent the plus strand or minus strand leader sequences. Dashed line with arrows represents the attenuation of nascent minus strand synthesis at the 3′ body TRS and the continuation of minus strand synthesis at the 5′ leader TRS.

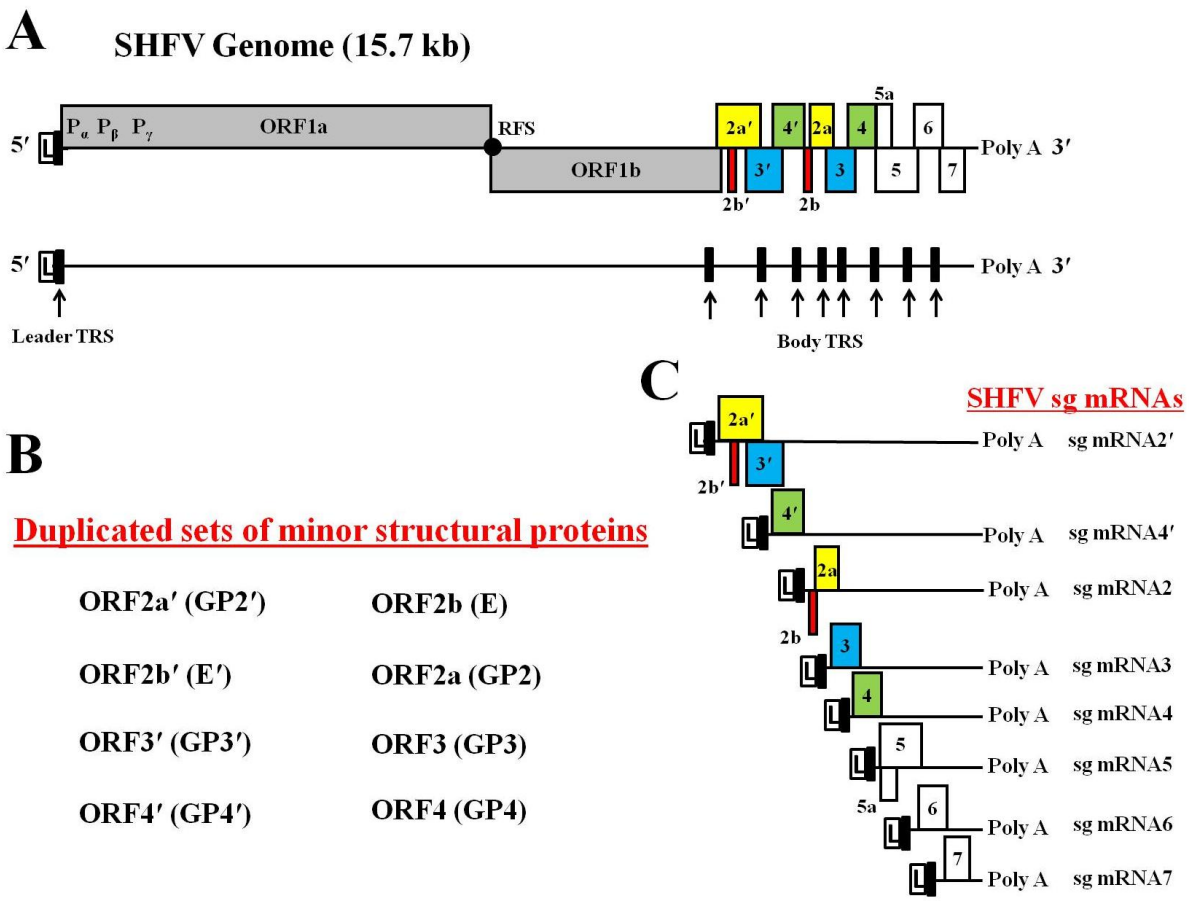


Figure 1.2. Diagram of the SHFV genome, sg mRNAs and encoded proteins.
(**A**) Diagram of the SHFV genome arrangement. The two gray boxes at the 5′ end of the genome represent the two polyprotein ORFs, ORF1a and ORF1b. The colored boxes represent the two sets of 3′ minor structural protein ORFs. Orthologs are indicated by the same color. Black vertical bars represent the positions of previously published leader TRS and body TRSs in the genome. The white box with the letter L inside represents the leader sequence. P_{α}, P_{β} and P_{γ} represent the three PLP1 domains encoded at the 5′ end of the genome. RFS, ribosomal frameshift site. (**B**) The ORFs and the corresponding proteins encoded by the two sets of minor structural proteins are listed. The order of the first two ORFs in the two sets differ. (**C**) SHFV sg mRNAs transcribed from the minus strand templates generated from each of the previously published body TRSs are shown. The ORFs that were predicted to be expressed from each sg mRNAs are indicated. Black vertical bars represent the leader-body junction in each sg mRNA.

2 FUNCTIONAL ANALYSES OF THE THREE SIMIAN HEMORRHAGIC FEVER VIRUS NONSTRUCTURAL PROTEIN 1 PAPAIN-LIKE PROTEASES

Heather A. Vatter[1Δ], Han Di[1Δ], Eric F. Donaldson[2], Gertrud U. Radu[1], Taronna R. Maines[1] and Margo A. Brinton[1]*

[1]Department of Biology, Georgia State University, Atlanta GA 30302,
[2] Department of Epidemiology, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599.

[Δ] These two authors contributed equally to the work described.

*To whom correspondence should be addressed:

Georgia State University, Department of Biology

P.O. Box 4010

Atlanta, GA, 30303

E-mail: mbrinton@gsu.edu.

Phone: 404-413-5388, Fax: 404-413-5301

Key words: papain-like protease, simian hemorrhagic fever virus, nsp1α, nsp1β, nsp1γ, polyprotein autoprocessing, PLP1 cleavage site, PLP1 catalytic residues

2.1 Abstract

The N-terminal region of the simian hemorrhagic fever virus (SHFV) nonstructural polyprotein 1a is predicted to encode three papain-like proteases (PLP1α, PLP1β and PLP1γ). Catalytic residues and cleavage sites for each of the SHFV PLP1s were predicted by alignment of the SHFV PLP1 region sequences with each other as well as with those of other arteriviruses and the predicted catalytic residues were shown to be proximal by homology modeling of the SHFV nsp1s on porcine respiratory and reproductive syndrome virus (PRRSV) nsp1 crystal structures. The functionality of the predicted catalytic Cys residues and cleavage sites were tested by analysis of the autoproteolytic products generated in *in vitro* transcription/translation reactions done with wildtype or mutant SHFV nsp1 contructs. Cleavage sites were also analyzed by mass spectroscopy analysis of selected immunoprecipitated cleavage products. The data showed that each of the three SHFV PLP1s is an active protease. Cys63 was identified as the catalytic Cys of SHFV PLP1α and is adjacent to an Ala instead of the canonical Tyr observed in other arterivirus PLP1s. SHFV PLP1γ is able to cleave at both downstream and upstream nsp1 junction sites. Although intermediate precursor polyproteins as well as alternative products generated by each of the SHFV PLP1s cleaving at sites within the N-terminal region of nsp1β were produced in the *in vitro* reactions, Western blotting of SHFV-infected, MA104 cell lysates with SHFV nsp1 protein-specific antibodies detected only the three mature nsp1 proteins.

2.2 Importance

SHFV is unique among arteriviruses in having three N-terminal papain-like protease 1 (PLP1) domains. Other arteriviruses encode one or two active PLP1s. This is the first functional study of the SHFV PLP1s. Analysis of the products of i*n vitro* autoprocessing of an N-terminal SHFV

nonstructural 1a polypeptide fragment showed that each of the three SHFV PLP1s is active and the predicted catalytic Cys residues and cleavage sites for each PLP1 were confirmed by testing mutant constructs. Several unique features of the SHFV PLP1s were discovered. The SHFV PLP1α catalytic Cys63 is unique among arterivirus PLP1s in being adjacent to an Ala instead of a Trp. Other arterivirus PLP1s cleave only in *cis* at a single downstream site but SHFV PLP1γ can cleave at both the downstream nsp1γ-nsp2 and upstream nsp1β-nsp1γ junctions. The three mature nsp1 proteins were produced in both the *in vitro* reactions and in infected cells.

2.3 Introduction

Simian hemorrhagic fever virus (SHFV) is a member of the Family *Arteriviridae.* Other members include the prototype equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) and porcine reproductive and respiratory syndrome virus (PRRSV). Based on similarities in genomic organization and replication strategy, the Families *Arteriviridae*, *Coronaviridae* and *Roniviridae* are classified in the Order *Nidovirales* (deVries, Horzinek et al. 1997; Snijder and Kikkert 2013). Arterivirus genomes are polycistronic, positive-sense, single-stranded RNAs with a 5' type I cap and a 3' poly (A) tract (Sagripanti 1985; Sagripanti, Zandomeni et al. 1986). The ~15.7 kb SHFV genome is the longest known arterivirus genome. The structural protein ORFs are located at the 3' end of the genome and expressed from a nested set of 5' and 3' co-terminal subgenomic mRNAs (sg mRNAs) (Zeng, Godeny et al. 1995; Godeny, deVries et al. 1998). The nonstructural polyproteins 1a and 1ab are expressed from overlapping 5' ORFs. Translation of polyprotein 1a terminates at the first in-frame UAA. Polyprotein 1ab is produced when a -1 ribosomal frameshift occurs on a slippery sequence located before the polyprotein 1a stop codon (Snijder and Kikkert 2013). The arterivirus ORF1a

polyprotein contains a papain-like protease (PLP) 1 domain in each of the three N-terminal nsp1s, a PLP2 domain in nsp2 and the main serine protease in nsp4 that cleaves at multiple sites in the 1a and 1ab polyproteins (Snijder and Kikkert 2013). All coronavirus PLPs (PLP1, PLP2, and PLpro) and the arterivirus PLP2 contain a catalytic Cys-His-Asp triad also found in other positive-sense RNA virus PLP sequences (van Kasteren, Bailey-Elkin et al. 2013; Mielech, Chen et al. 2014). However, the active sites of arterivirus PLP1 proteases were predicted to consist of a Cys-His tandem and this was confirmed by the recent crystal structures of PRRSV nsp1α and nsp1β proteins (Sun, Xue et al. 2009; Xue, Sun et al. 2010). The length of the sequence between the catalytic residues in arterivirus PLPs is half of that in coronavirus PLPs, making the arterivirus PLPs the smallest known.

All of the coronavirus PLPs as well as the arterivirus PLP2s also have deubiquitinating and deISGylating activities but the arterivirus PLP1s do not (Mielech, Chen et al. 2014). The coronavirus PLPs and arterivirus PLP2s both have *cis* and *trans* cleavage activities while the arterivirus PLP1s have been reported to cleave exclusively in *cis* at a single site downstream of their catalytic domains (Snijder, Wassenaar et al. 1995; Han, Rutherford et al. 2009; Sun, Xue et al. 2009; Xue, Sun et al. 2010; Kikkert 2013; Mielech, Chen et al. 2014). The crystal structures of Type II PRRSV nsp1α and nsp1β suggest that these enzymes utilize an intramolecular cleavage mechanism (Sun, Xue et al. 2009; Xue, Sun et al. 2010; Snijder and Kikkert 2013). The N-terminal EAV PLP1 domain is inactive due to a Lys substitution of the catalytic Cys residue but the downstream EAV PLP1 is active and produces an nsp1 fusion protein (Snijder, Wassenaar et al. 1992; Ziebuhr, Snijder et al. 2000). The LDV and PRRSV nonstructural polyproteins contain active PLP1α and PLP1β enzymes that each cleave at a single downstream site in *in vitro* reactions (den Boon, Faaberg et al. 1995; Sun, Xue et al. 2009; Xue, Sun et al.

2010). Although most coronavirus PLPs cleave at a canonical -Leu-X-Gly↓Gly site, the sequences of the cleavage sites used by arterivirus PLPs are not well conserved (Snijder, Wassenaar et al. 1992; den Boon, Faaberg et al. 1995; Sun, Xue et al. 2009; Xue, Sun et al. 2010; Mielech, Chen et al. 2014).

The EAV nsp1 and PRRSV nsp1α and nsp1β proteins function as interferon pathway antagonists (Beura, Sarkar et al. 2010; Kim, Sun et al. 2010; Song, Krell et al. 2010; Beura, Subramaniam et al. 2012). An N-terminal zinc-finger domain in EAV nsp1 and PRRSV nsp1α regulates sg mRNA and genome transcription abundance (van Dinten, den Boon et al. 1997; Kroese, Zevenhoven-Dobbe et al. 2008; Nedialkova, Gorbalenya et al. 2010; Kikkert 2013) and is also important for mediating the IFN-β and NF-κB immunosuppressive activities of PRRSV nsp1α (Fang and Snijder 2010).

In the present study, the catalytic residues and cleavage sites of the three SHFV PLP1 domains (PLP1α, PLP1β and PLP1γ) were predicted by sequence alignment with each other as well as with other arterivirus PLP1s and the predicted catalytic residues were shown to be proximal by homology modeling on PRRSV nsp1α and nsp1β structures (Sun, Xue et al. 2009; Xue, Sun et al. 2010). The results of analyses of the autocleavage products produced by wildtype and mutant *in vitro* translated SHFV nsp1 region polyproteins indicated that each of the SHFV PLP1s is active and identified the catalytic Cys residues and cleavage sites for each SHFV PLP. In the *in vitro* reactions, expected products as well as precursor peptides and alternative products were detected and PLPγ was shown to be able to cleave at both downstream and upstream nsp1 junction sites. However, only mature nsp1α, nsp1β and nsp1γ were detected in infected cells.

2.4 Materials and methods

2.4.1 Sequence and bioinformatics analyses.

Sequence analysis was performed on the ORF1a protein sequence that was conceptually translated from an updated SHFV genome sequence (Accession no. AF180391.2). The SHFV PLP1α, PLP1β and PLP1γ boundaries were estimated using the putative cleavage sites and catalytic domains of the PLP1s of LDV and PRRSV. The putative SHFV PLP1 sequences were aligned with each other and also with the PLP1s of other arteriviruses. All sequences were manipulated in the Geneious software suite version 5.6.3 (BoiMatters, LTD; www.geneious.com) and multiple alignments of the PLP1 amino acid sequences were performed using ClustalX in Geneious, with similarity determined using the PAM250 substitution matrix. Once the putative boundaries were predicted, homology models were separately generated for SHFV nsp1α, SHFV nsp1β and SHFV nsp1γ using the crystal structure of a PRRSV nsp1α protein (PDB code: 3IFU) (Sun, Xue et al. 2009) and also the crystal structure of a PRRSV nsp1β protein (PDB code: 3MTV) (Xue, Sun et al. 2010) as the template. Homology models were generated using Modeler (Sali and Blundell 1993) (http://salilab.org/modeller/) as implemented in the Max Planck Institute's Bioinformatics Toolkit (http://toolkit.tuebingen.mpg.de/), and the structures were rendered and compared using MacPymol (Delano Scientific, LLC; http://www.pymol.org).

2.4.2 Cells and virus.

The MA104 cell line was a gift from O. Nainan, Centers for Disease Control and Prevention. These cells were grown in Minimal Essential Medium supplemented with 10% fetal bovine serum and 10 μg/ml gentamicin at 37˚C in a 5% CO_2 atmosphere. An aliquot of SHFV,

strain LVR 42-0/M6941, passage 2, was obtained from the American Type Culture Collection, sequentially plaque-purified three times and then amplified once on MA104 cell monolayers. To make experimental virus pools, confluent MA104 monolayers were infected with the stock virus at an MOI of 0.2 and culture media was harvested at 32 h after infection and clarified by centrifugation. Virus pools contained titers of ~10^7 PFU/ml and aliquots were stored at -80 ℃.

2.4.3 Construction of SHFV nsp1 clones.

A region consisting of the 5' 1934 nts of the SHFV genome, which included the 5' NCR (209 nts) and the 5' 1725 nts of ORF1a, was amplified by RT-PCR from purified SHFV-LVR genomic RNA using the EcoRI and TaqI tailed forward primer 5'-AGgaattctcgaGATTAAAATAAAAGTGTGAAG-3' (restriction sites are indicated by lower case and two 5′ nts were added at the end to enhance restriction enzyme cleavage) and the XbaI tailed reverse primer (5'-GCtctagaACCGGCAGTACAGCATGGGT-3'). The cloned N-terminal region of ORF1a contained the entire nsp1 region as well as the first 273 nts of nsp2. An in-frame N-terminal Flag tag and an in-frame C-terminal C-myc tag were added by subcloning nts 210 through 1934 of the SHFV genome into the pFlag/C-myc expression vector (Sigma) generating pFlag-SHFV-C-myc. To construct plasmid DNA for use as template in coupled *in vitro* transcription/translation reactions, the Flag-SHFV-c-Myc region of the pFlag-SHFV-C-myc clone was subcloned into the pTNT expression vector (Promega) generating pTNT-FSM-wt. Mutagenesis of the pTNT-FSM-wt construct was performed using a QuikChange site-directed mutagenesis kit (Stratagene) according to the manufacturer's protocol. The primers used to change the putative PLP1 catalytic Cys residues to Ala or to substitute the -2 and -1 residues of

each predicted cleavage site with either Val or Ala are listed in Table 2-1. All wild type and mutant sequences generated were confirmed by sequencing.

2.4.4 In vitro, coupled transcription/translation reactions.

Wild type and mutant pTNT-FSM cDNA constructs were used as templates for *in vitro* coupled transcription/translation reactions performed using the TNT Coupled Wheat Germ Extract System (Promega) according to the manufacturer's protocol. Briefly, 1 µg of the plasmid DNA was mixed with 25 µl of wheat germ extract, 2 µl of TNT reaction buffer, 1 µl of amino acid mixture without Cys, 1 µl of TNT SP6 RNA polymerase and 20 µCi of [^{35}S]-Cys (1 mCi/mmol, Perkin Elmer) in a total volume of 50 µl. Reactions were incubated at 30 ℃ for 2 h prior to use for immunoprecipitation.

2.4.5 Immunoprecipitation.

In vitro translation reactions (50 µl) were divided into four aliquots. One aliquot (~5 µl) was stored at -20 ℃ until use as the lysate control for gel electrophoresis. One 15 µl aliquot was incubated with control murine anti-IgG antibody and another with anti-c-Myc monoclonal antibody (Sigma) for 1 h with rotation at 4 ℃, after which Protein G agarose (Roche) was added and the samples were next rotated at 4 ℃ overnight and then pelleted by centrifugation. The third 15 µl aliquot was incubated with anti-Flag® M2 affinity agarose beads (Sigma) with rotation at 4 ℃ overnight. The beads were next washed 3 times in lysis buffer (1% Triton X-100, 0.1% SDS, 150 mM NaCl, and 50 mM Tris HCl pH 7.4) containing HALT protease and phosphatase inhibitor cocktail (Pierce Scientific), and then pelleted by centrifugation. After addition of 30 µl of 2X sample loading buffer (20% SDS, 25% glycerol, 0.5 M Tris-HCl pH 6.8, 0.5%

bromophenol blue and 5% β-mercaptoethanol) to the IP pellets and the lysate aliquot, the samples were boiled for 5 min. The peptides in half of each sample were separated by SDS-PAGE on a 13% polyacrylamide gel and the remainder of the sample was frozen for use in a repeat analysis. The gels were fixed in 10% acetic acid and 30% methanol, incubated in Autofluor (National Diagnostics) and then in anti-cracking buffer (7% acetic acid, 7% methanol and 1% glycerol), dried and autoradiographed.

2.4.6 Mass spectroscopy (MS) sequencing of autoprocessed peptides.

Large volume (200 µl) *in vitro* translation reactions were performed as described above except that an amino acid mixture containing Cys was used. The wild type nsp1 construct and the PLP1β catalytic Cys246Ala mutant construct were separately transcribed and translated and the peptides immunoprecipitated with anti-c-Myc monoclonal antibody or anti-Flag® M2 affinity agarose beads. The precipitated peptides were separated on different lanes of a 12% NuPAGE bis-tris gel using NuPAGE MOPS SDS buffer (Life Technology). Precise Plus protein standards (Bio-rad) were run on a separate lane. The proteins were stained with a colloidal blue staining kit (Life Technology) and the gel was sent to the Wistar Institute proteomics facility, Philadelphia, PA. Selected bands were excised and after trypsin digestion, the peptides were analyzed by liquid chromatography-tandem mass spectrometry (LC-MS/MS) on an LTQ-Orbitrap XL mass spectrometer. The MS/MS spectra generated were searched against a custom database containing the SHFV nsp1+N-terminal nsp2 (1-575 aa) construct sequence, the Triticum proteome and common contaminants using SEQUEST.

2.4.7 Western blotting.

Infected MA104 monolayers were lysed by addition of RIPA buffer (1X phosphate-buffered saline, 1% Nonidet P-40, 0.5% sodium deoxycholate and 0.1% SDS) containing Halt protease inhibitor cocktail (Thermo Scientific). Following separation by SDS-PAGE, cell proteins were electrophoretically transferred to a nitrocellulose membrane. Membranes were blocked with 1 X TBS (0.01 M Tris and 0.15 M NaCl, pH 8) containing 5% non-fat dry milk and 0.1% Tween 20 before incubation in the presence of blocking buffer with one of the SHFV nsp1 monospecific, polyclonal antibodies made by Abgent. The peptides used to make these antibodies were: nsp1α (GDLTRPEETPLPGGC), nsp1β N-terminal (FAQKVITAFPEGVLC), nsp1β C-terminal **(**DESVPPDCQIIARF), and nsp1γ (FPPLSRKSEAQRAIL). Actin was used as a loading control and was detected with antibody C-11 (Santa Cruz Biotechnology). Blots were washed with 1X TBS and then incubated with either anti-rabbit-horseradish peroxidase (Santa Cruz Biotechnology) or anti-mouse-horseradish-peroxidase (Santa Cruz Biotechnology). After washing, the blots were processed for chemiluminescence using a Super-Signal West Pico detection kit (Pierce Scientific) according to the manufacturer's protocol.

2.5 Results:

2.5.1 Analysis of SHFV nsp1s by sequence alignment and homology modeling.

In all previously studied active arterivirus PLP1s, a Trp follows the catalytic Cys (den Boon, Faaberg et al. 1995; Sun, Xue et al. 2009; Xue, Sun et al. 2010) but this is not the case for all of the coronavirus PLP1s or for arterivirus PLP2s (Snijder, Wassenaar et al. 1995; Mielech, Chen et al. 2014). An initial inspection of the sequences of the three SHFV PLP1 catalytic domains revealed a single Cys-Trp upstream of a His for each PLP1: Cys115/Trp116/His130

(indicated by a blue diamond in Fig. 2-1B) in PLPα, Cys246/Trp247/His309 in PLP1β and Cys378/Trp379/His443 in PLP1γ (indicated by an asterisk in Fig. 2-1B). However, the PLP1α Cys115/Trp116 was located only 14 aa upstream from the catalytic His130 residue (indicated by an asterisk in Fig. 2-1B) while the Cys-Trp of PLP1β and PLP1γ were located 62 and 64 aa, respectively, upstream of their His residues (indicated by an asterisk in Fig. 2-1B). A previous study predicted that the SHFV PLP1α catalytic Cys was Cys79 **(**indicated by a blue triangle in Fig. 2-1B**)** based on an alignment of the nsp1α sequences of PRRSV, EAV, LDV and SHFV (Sun, Xue et al. 2009). The sequence context of this SHFV Cys is -X-Cys-Lys-Pro- instead of the conserved -X-Cys-Trp-Lys- context of the PRRSV and LDV nsp1α proteins (Fig. 2-1A). However, the catalytic function of SHFV Cys79 was not functionally tested in the previous study. In the present study, the putative SHFV PLP1α domain boundaries were first estimated by comparison with the previously reported cleavage sites and catalytic residues of the LDV and PRRSV PLP1α domains (Fig. 2-1A). The EAV PLP1α was not included in the alignment since it is inactive and shorter than the others. This alignment predicted Cys63 as the SHFV PLP1α catalytic Cys. An additional alignment of the three SHFV PLP1 regions with each other identified a conserved GV[Q/T]G motif approximately halfway between the catalytic Cys and His residues in the putative SHFV PLP1β and PLP1γ domains (blue underline, Fig. 2-1B). This motif GV[X]G was also conserved in the PLP1β regions of PRRSV and LDV (Xue, Sun et al. 2010). The position of this motif in SHFV PLP1α (aa positions 93 to 96) was then used as an anchor to align the three SHFV PLP1s with each other. Cys63 aligned with the predicted catalytic Cys residues of both SHFV PLP1β and PLP1γ and was located 67 aa from the PLP1α catalytic His130 (indicated by an asterisk in Fig. 2-1A and B). However, Cys63 is followed by

an Ala instead of by the Trp present in all of the other active arterivirus PLP1s (Fig. 2-1A) (Sun, Xue et al. 2009; Xue, Sun et al. 2010).

Homology models of the three SHFV nsp1 proteins were next generated as described in Materials and Methods to determine whether the predicted catalytic residues were proximal to one another in a manner that would facilitate activity. The homology model of SHFV nsp1α generated using the crystal structure of the Type II PRRSV (strain XH-GD) nsp1α (Sun, Xue et al. 2009) and the homology model of SHFV nsp1β generated using the crystal structure of the Type II PRRSV (strain XH-GD) nsp1β are shown in Fig. 2-1C. The homology models suggested that the sequence and structure of SHFV nsp1α were more similar to those of PRRSV nsp1α while the sequence and structure of SHFV nsp1β were more similar to those of PRRSV nsp1β. Because the homology modeling did not clearly suggest which of the PRRSV nsp1 structures was the most similar to that of SHFV nsp1γ, the SHFV nsp1γ homology models generated on the structures of both PRRSV nsp1α and PRRSV nsp1β are shown (Fig. 2-1C). The SHFV nsp1γ homology model made on the PRRSV nsp1β structure appeared to be less structured in the lower half than the model made on the PRRSV nsp1α structure. However, homology modeling is limited in that it models a non-identical but similar aa sequence on a fixed structural template. The actual structures of the SHFV nsp1 proteins will not be known until crystal structures for each have been solved. In the modeled structure of SHFV nsp1α, Cys63 was located close to His130 in spatially similar positions to the catalytic Cys and His residues of PRRSV nsp1α. The alignment and modeling data supported the hypothesis that the SHFV PLPα catalytic residues are Cys63 and His130.

Most coronavirus PLPs recognize and cleave canonical -Lys-X-Gly↓Gly- sites but some can cleave at -Lys-X-Gly↓Ala- sites (Mielech, Chen et al. 2014). In contrast, the arterivirus PLP1

cleavage site sequences are not well conserved. Type II PRRSV nsp1α cleaves at -Cys-Ala-Met↓Ala-Asp- and Type II PRRSV nsp1β at -Trp-Tyr-Gly↓Ala-Gly- (Sun, Xue et al. 2009; Xue, Sun et al. 2010). Both a previous alignment (Sun, Xue et al. 2009) and the new arterivirus PLP1α alignment shown in Fig. 2-1A predicted the SHFV nsp1α cleavage site to be –Thr-Thr-Gly↓Gly-Asp-. However, the alignment of the three SHFV PLP1s predicted that cleavage would occur at a downstream –Gln-Leu-Gly↓Thr-Cys- site (Fig. 2-1B, red arrow). Both the previous alignment (Sun, Xue et al. 2009) and the three SHFV PLP1 alignment predicted that SHFV nsp1β cleavage would occur at a –Thr-Phe-Gly↓Gly- site and the alignment of the three SHFV PLP1s predicted that SHFV nsp1γ cleavage occurs at a –Arg-Arg-Gly↓Gly-Gly- site (Fig. 2-1B, red arrow).

2.5.2 Analysis of the peptides produced in vitro by autoprocessing of wildtype and mutant SHFV N-terminal ORF1a polyprotein fragments.

An expression plasmid containing the N-terminal region of ORF1a (genome nts 210 through 1934) fused to an N-terminal Flag tag and a C-terminal c-Myc tag was constructed as described in Materials and Methods. The expressed polyprotein included the complete SHFV nsp1 (nsp1α, nsp1β and nsp1γ) region plus the N-terminal region (91 aa) region of nsp2. Wild type cDNA was first used as the template in a coupled *in vitro* transcription/translation reaction done in the presence of [^{35}S]-Cys. The peptides produced by PLP1 autoprocessing were immunoprecipitated (IP) with anti-Flag, anti-c-Myc or a control IgG antibody. The peptides in the IP samples and also an aliqouot of the unprecipitated reaction (lysate) were separated by SDS-PAGE (Fig. 2-2A). The wild type construct produced 9 bands (~63, ~54, ~39, ~30, ~28, ~27, ~23, ~18, and ~15 kDa) and these peptides were identified by their predicted sizes and by the presence or absence of terminal tags (Fig. 2-2B). The full length peptide (63 kDa), the

intermediate precursors nsp1α+β+γ (54 kDa) and nsp1α+β (39 kDa), and nsp1α (18 kDa) were pulled down by anti-Flag antibody, while the full length peptide (63 kDa) and the intermediate precursor nsp1γ+nsp2 (23 kDa) were pulled down by the anti-c-Myc antibody (Fig. 2-2A). The 15 kDa band present only in the lysate lane is most likely nsp1γ since neither the anti-Flag nor anti-c-Myc antibody pulled down this peptide (see Fig. 2-2B). The nsp1γ band was not as efficiently detected as the other peptides due to the presence of fewer Cys residues in this peptide (Fig. 2-2B). The 22.1 kDa nsp1β peptide would be expected to migrate to a similar position on the gel as the 24.5 kDa nsp1γ+nsp2 precursor. Because nsp1β is an internal peptide, it would not be tagged and would only be detected in the lysate lane while the c-Myc tagged nsp1γ+nsp2 band would be detected in both the lysate and c-Myc IP lanes. The detection of both nsp1α and nsp1γ indicated that efficient cleavage occurred at both the nsp1α-nsp1β and the nsp1β-nsp1γ junctions.

In addition to the expected precursor and product bands, a ~30 kDa band was present in the lysate, Flag and c-Myc IP lanes, but not in the IgG IP lane, suggesting that this lysate band contained both N-terminal and C-terminal peptides (Fig. 2-2A). Consistent with this hypothesis, the intensities of the 30 kDa bands in each of the IP lanes were lower than that of the band in the lysate lane. It is likely that the 30 kDa lysate band contained two precursor peptides, one consisting of nsp1α plus an N-terminal portion of nsp1β and a second that consists of a C-terminal portion of nsp1 plus nsp2. The N- and C-terminal 30 kDa bands were thus designated nsp1α+tβ and tβ+nsp1γ+nsp2. Three additional, unexpected bands of 23, 27 and 28 kDa were present in both the lysate and the Flag IP lanes (Fig. 2-2A). Based on the presence of the Flag tag and the estimated sizes of these peptides, they likely are nsp1α+various lengths of the N-terminal region of nsp1β (see Fig. 2-2B). These bands were designated nsp1α+ttβ due to the shorter nsp1β

sequences they contained. A ~39 kDa band was also detected in the c-Myc IP lane that likely is ttβ+nsp1γ+nsp2. The detection of bands containing the C-terminal portions of each of the alternative N-terminal bands indicated that these peptides were generated by a single cleavage of the polyprotein and not by a mechanism such as premature translation termination.

2.5.3 Functional analysis of the predicted PLP1 catalytic Cys residues.

To determine whether the predicted nsp1α catalytic Cys residue was functional, Cys63 was substituted with Ala. The resulting construct was transcribed and translated *in vitro* and the products were either analyzed directly or after IP with anti-Flag or anti-c-Myc antibody (Fig. 2-3B). The nsp1α (18 kDa) band produced by autoprocessing of the wild type polyprotein was absent from both the lysate and Flag IP lanes indicating that Cys63 is required for PLP1α cleavage of nsp1α from the polyprotein. Interestingly, the nsp1α+ttβ (27 and 28 kDa) bands were present in both the lysate and Flag lanes and the intensities of these bands were increased compared to those produced by the wild type construct, suggesting that cleavage at sites within the N-terminal region of nsp1β increased when the nsp1α PLP1 was not active. A construct with the predicted nsp1β catalytic Cys246 substituted with Ala produced full-length polyprotein (63 kDa) and nsp1α+β+γ polyprotein precursor (54 kDa) bands with increased intensity compared to the wildtype construct indicating that processing at the nsp1β-nsp1γ junction was decreased when the predicted nsp1β catalytic Cys was mutated (Fig. 2-3C). However, detection of nsp1γ (15 kDa) in the lysate lane and nsp1γ + nsp2 (23 kDa) in the lysate and c-Myc lanes indicated that cleavage at the nsp1β/nsp1γ junction still occurred when PLPβ was inactive and suggested that cleavage at this junction was not exclusively dependent on PLPβ. For the construct with the predicted nsp1γ catalytic Cys378 substituted with an Ala, the absence of nsp1γ (15 kDa) in the

lysate lane and also the absence of nsp1α+β+γ (54 kDa) in the lysate and Flag lanes indicated that cleavage at the nsp1γ-nsp2 junction did not occur (Fig. 2-3D). The detection of the ttβ/tβ+nsp1γ+nsp2 peptides (39 and 30 kDa bands in the lysate and c-Myc lanes) provided additional evidence that nsp1γ was not cleaved from nsp2 when the PLPγ catalytic Cys was mutated. Cleavage occurred efficiently at both the nsp1α-nsp1β and nsp1β-nsp1γ junctions when the predicted nsp1γ catalytic Cys378 was substituted.

Mutation of PLP1α Cys63 inhibited cleavage at the nsp1α-nsp1β junction and mutation of PLP1γ Cys378 inhibited cleavage of nsp1γ-nsp2 junction. However, cleavage at the nsp1β-nsp1γ junction was still observed when PLP1β Cys246 was mutated suggesting that either PLP1α or PLP1γ can also cleave at the nsp1β/nsp1γ junction. To analyze the activity of each PLP1 individually, double mutant constructs that produced polyproteins with only one active PLP were made. The construct with both PLP1α Cys63 and PLP1β Cys246 substituted with Ala produced an increased amount of nsp1α+β+γ (54 kDa) in the Flag IP lane compared to the wildtype construct indicating that the nsp1γ-nsp2 junction was efficiently cleaved (Fig. 2-3E). The detection of the nsp1γ (15 kDa) in the lysate lane and nsp1α+β (39 kDa) in the Flag IP lane indicated that the nsp1β/nsp1γ junction was also cleaved when only PLPγ was active. Interestingly, the nsp1α+tβ and nsp1α+ttβ (30, 28, and 27 kDa) bands were also detected in the Flag IP lanes suggesting that PLP1γ can cleave at multiple sites within nsp1β. The absence of an nsp1α (18 kDa) band in both the lysate and Flag IP lanes as well as the observed increase in the intensity of the nsp1α+β (39 kDa) band in the lysate and Flag IP lanes compared to the bands produced by the wild type construct indicated that the nsp1α/nsp1β junction was not cleaved by PLPγ.

A double mutant with PLP1β Cys246 and PLPγ Cys378 substituted with Ala produced a high intensity full length polypeptide (63 kDa) band in the lysate, c-Myc IP, and Flag IP lanes as well as nsp1α+tβ (30 kDa) and α (18 kDa) bands in the lysate and Flag IP lanes (Fig. 2-3F). Although an nsp1α band was detected when only PLPα was active, the efficiency of cleavage at the nsp1α/nsp1β junction was decreased compared to the wild type polyprotein. The detection of the 30 kDa band in both the lysate and Flag IP lanes indicated that PLP1α is also able to cleave within the catalytic region of nsp1β but this cleavage was also inefficient when PLP1α was the only protease. The lack of detection of nsp1γ+nsp2 (23 kDa) in the lysate and c-Myc lanes and of nsp1α+β (39 kDa) in the lysate and Flag lanes indicated that cleavage did not occur at the nsp1β-nsp1γ junction. Likewise, the lack of detection of nsp1γ (15 kDa) in the lysate lane and of nsp1α+β+γ (54 kDa) in the lysate and Flag lanes indicated that cleavage also did not occur at the nsp1γ-nsp2 junction. Multiple weak background bands were detected in the lysate, Flag IP and c-Myc IP lanes suggesting that either some premature termination or polyprotein breakdown had occurred.

A mutant construct with both PLP1α Cys63 and PLP1γ Cys378 substituted with Ala was next tested (Fig. 2-3G). Detection of nsp1α+β (39 kDa) in the Flag lane indicated that cleavage occurred at the nsp1β-nsp1γ junction. The detection of nsp1α+ttβ/tβ (30, 28 and 27 kDa) and ttβ/tβ+nsp1γ+nsp2 (39 and 30 kDa), unexpectedly indicated that PLP1β can also cleave within its own catalytic region as well as within the N-terminal region of nsp1β. The lack of detection of nsp1γ (15 kDa) in the lysate lane and of nsp1α (18 kDa) in the lysate and Flag lanes indicated that PLP1β is not able to cleave at either the nsp1α-nsp1β or nsp1γ-nsp2 junctions. Because self-cleavage by a protease within its own catalytic region seemed unlikely, additional bioinformatics analyses were done to determine the possible existence of an additional protease with a catalytic

domain located between Cys115/Trp116 and His181 or His185. However, neither sequence alignment nor homology modeling supported the existence of an additional protease between aa 100 and 203 (data not shown)**.**

2.5.4 Functional analysis of the cleavage sites utilized by the SHFV PLP1s.

Previous studies showed that substitutions of the -1 and -2 amino acids of PLP cleavage sites prevented their recognition and cleavage (Bonilla, Pinon et al. 1995; van Dinten, Wassenaar et al. 1996). The -1 and -2 positions of each of the predicted SHFV PLP1 cleavage sites were substituted with Ala or Val (Table 2-1). The PLP1α cleavage site, -Thr-Thr-Gly↓Gly- (aa 162-165), was predicted by the multiple arterivirus PLP1α alignment shown in Fig. 2-1A and also by a previously published alignment (Sun, Xue et al. 2009). When the peptides produced by a construct with the -1 and -2 positions of this site mutated were analyzed (Fig. 2-4A), the nsp1α (18 kDa) band was not detected in the lysate or Flag IP lanes consistent with this being the preferred nsp1α cleavage site. When the PLP1α cleavage site, -Gln-Leu-Gly↓Thr- (aa 169-172), predicted by the alignment of the three SHFV PLP1s and homology modeling (Fig. 2-1B and C) was mutated, the nsp1α (18 kDa) band was still detected in the lysate and Flag IP lanes but the intensity of these bands was lower than in the wildtype reactions (Fig. 2-4B). The intensities of the full length (63 kDa) and nsp1α+β+γ (54 kDa) precursor protein bands were increased relative to the intensities of these bands in wildtype autoprocessing reactions suggesting that the overall cleavage efficiency of this mutant polypeptide was reduced. Both the alignment of the three SHFV PLP1s (Fig. 2-1B) and a previous arterivirus PLP1β alignment predicted -Tyr-Phe-Gly↓Gly- (aa 348-351) to be the nsp1β cleavage site. A construct with the -2 and -1 aa residues of this site substituted produced nsp1α+β+γ (54 kDa) bands with increased intensity and

nsp1α+β (39 kDa) bands with decreased intensity in the lysate and Flag IP lanes (Fig. 2-4D) indicating that the -Tyr-Phe-Gly↓Gly- (aa 348-351) site is the main cleavage site at the nsp1β-nsp1γ junction**.** The alignment of the SHFV PLP1s predicted –Arg-Arg-Gly↓Gly-Gly- (aa 482-486) as the PLPγ cleavage site (Fig. 2-1B). To determine whether cleavage occurs between G484 and G485, the -2 and -1 aa residues of this site were substituted. This construct produced an nsp1γ + nsp2 (23 kDa) band that was detected in the c-Myc lane but no nsp1γ (15 kDa) band was detected in the lysate lane indicating that cleavage did not occur at the nsp1γ-nsp2 junction (Fig. 2-4E). To determine whether cleavage can also occur between G485 and G486, the -2 and -1 aa residues of this site were substituted. The bands produced by this construct were the same as those produced by the G484/G485 mutant construct (Fig. 2-4F). Because G484 was mutated in both constructs (-2 aa or -1 aa), it was not possible to distinguish experimentally which of these sites is used.

Production of the 30 kDa nsp1α+tβ and tβ+nsp1γ+nsp2 bands by cleavage within the catalytic region of PLP1β was observed with the wildtype and many of the mutant polyproteins. Each of the SHFV PLP1s was predicted to cleave between two Gly residues at each of the nsp1 junctions. Analysis of the PLP1β sequence did not reveal any -Gly-Gly- tandems. However, an -Asp-Cys-Gly-Val- (269-272 aa) sequence was found within the predicted nsp1β catalytic region (Fig. 2-1B). A construct with the -2 and -1 aa residues of this site substituted with Val produced no 30 kDa band in the lysate, Flag IP or c-Myc IP lanes suggesting that the N- and C-terminal 30 kDa bands were produced by cleavage at the Asp-Cys-Gly↓Val- (269-272 aa) site (Fig. 2-4C). The increased intensity of the full length polyprotein (63 kDa) and nsp1α+β+γ (54 kDa) bands indicated that mutation of this site decreased overall cleavage efficiency of the polyprotein.

The wildtype construct produced two other unexpected nsp1α+ttβ (28 and 27 kDa) bands that were precipitated by anti-Flag but not by anti-c-Myc antibody indicating that these two proteins contained nsp1α and part of nsp1β. Based on their sizes, these two proteins are likely to be produced by cleavages within the N-terminal region of nsp1β**.** However, the sites at which these cleavages occur are not currently known.

2.5.5 MS sequencing of selected in vitro autoprocessing cleavage products.

As an additional means of identifying the cleavage sites for each SHFV nsp1, selected autoprocessing products were subjected to sequence analysis by MS. The positions of the SHFV peptide bands were estimated by comparison to the positions of the protein standards and also to the positions of radiolabeled SHFV peptide bands on gels obtained in previous experiments because the small amounts of immunoprecipated peptides could not be detected by protein staining. The estimated nsp1α (18 kDa) and the nsp1α+β (39 kDa) precursor regions were separately excised from the Flag IP lane of a wildtype construct autoprocessing reaction while the nsp1a+β+γ (54 kDa) region was excised from the Flag IP lane of a Cys246Ala mutant construct autoprocessing reaction. Trypsin peptides aligning to the expected regions in each product was detected (Fig. 2-5). For nsp1α, ten trypsin generated peptides ending at aa 164 were detected indicating that the –Thr-Thr-Gly↓Gly- (aa 162-165) site was the preferred site for PLP1α cleavage (Fig. 2-5A). One peptide ending at aa 171 was also detected suggesting that this alternative site can also be used by PLP1α *in vitro*. No peptides were detected matching to the sequence downstream of the alternative cleavage site.

Peptide coverage for the nsp1α+β (39 kDa) precursor was lower than that for nsp1α (Fig. 2-5B). Seven peptides were detected that extended to aa 342. Trypsin primarily cleaves on the C-

terminal side of Lys or Arg residues except when a Pro follows one of these aa. Two Lys residues are located on the N-terminal side of the predicted -Tyr-Phe-Gly↓Gly- (aa 348-351) nsp1β cleavage site (-5 and -9 positions) which resulted in peptide coverage ceasing just before the predicted cleavage site because peptides shorter than 6 aa acids were not aligned. Peptides The positions of upstream trypsin cleavage sites and the lack of detection of peptides aligning to downstream sequences indicated that cleavage occurred at or very near the predicted nsp1β cleavage site. Three Arg residues are located immediately upstream of the predicted nsp1γ cleavage site -Arg–Arg-Arg-Gly↓Gly-Gly- (aa 481-486). The detection of fourteen peptides ending at Arg478 and no peptides aligning to downstream sequences strongly suggested that cleavage occurred at the predicted nsp1γ cleavage site (Fig. 2-5C).

2.5.6 Analysis of nsp1 processing in infected cells.

Nsp1α, nsp1β and nsp1γ processing was next analyzed in infected cells. MA104 cells were infected with SHFV-LVR at an MOI of 1. Cell lysates were collected at various times after infection and analyzed by Western blotting with anti-nsp1α, anti-nsp1β or anti-nsp1γ antibody. The anti-nsp1α antibody detected a single band at 22 kDa (Fig. 2-6A). A single band at 22 kDa was detected with an anti-N-terminal nsp1β antibody (Fig. 2-6B) and also with an anti-C-terminal nsp1β antibody (data not shown). Anti-nsp1γ antibody detected a single 18 kDa band (Fig. 2-6C). The estimated sizes for both nsp1α (22 versus 18 kDa) and nsp1γ (18 versus 15 kDa) were higher in cell lysates than in the *in vitro* translation assays suggesting that these proteins are post-translationally modified in infected cells. The lack of detection of any precursors suggested that rapid processing of the nsp1 proteins from the polyprotein likely occurs in infected cells. Also, no truncated forms of nsp1β or longer forms of nsp1α were detected in cell lysates during

the course of the infection suggesting that cleavage at sites within nsp1β may not occur in infected cells or that these alternative proteins are in low abundance or unstable and rapidly degraded.

2.6 Discussion

The EAV, LDV and PRRSV genomes each encode two PLP1 domains that rapidly cleave one (the EAV PLP1α is inactive) or two nsp1 proteins from the N-terminus of the 1a/1ab nonstructural polyproteins (Snijder and Kikkert 2013). The SHFV genome is unique in that it encodes three PLP1 domains PLP1α, PLP1β and PLP1γ (Snijder, Wassenaar et al. 1992). The present study showed that each of the SHFV PLP1s is an active protease and that three mature nsp1 proteins are produced in *in vitro* reactions and in infected cells. In the PLP1 domains of other arteriviruses, the catalytic Cys is adjacent to an aromatic Trp. Consistent with this conservation, the identified catalytic Cys of SHFV PLP1β (Cys246) and of SHFV PLP1γ (Cys378) are both adjacent to a Trp residue. The only Cys within the SHFV nsp1α region with an adjacent Trp was located only 14 aa from the predicted catalytic His. However, Cys63, which is adjacent to the non-aromatic residue Ala, was predicted by both alignment and homology modeling to be the nsp1α catalytic Cys and substitution of Cys63 resulted in the loss of the nsp1α cleavage product confirming that Cys63 is the PLP1α catalytic Cys. While the distance between the catalytic Cys and His residues of SHFV PLP1β and SHFV PLP1γ is 62 and 65 aa, respectively, similar to the distances in other arterivirus PLP1s, Cys63 is 67 aa upstream of its catalytic His residue. The crystal structures obtained for PRRSV nsp1α and nsp1β show that these proteins have similar but distinct structures (Sun, Xue et al. 2009; Xue, Sun et al. 2010). Both of the PRRSV nsp1 structures were used as templates for modeling each SHFV nsp1. The

sequence analysis and homology modeling data suggested that SHFV nsp1α is more similar to PRRSV nsp1α, SHFV nsp1β is more similar to PRRSV nsp1β and SHFV nsp1γ is more similar to PRRSV nsp1α than to PRRSV nsp1β. However, the actual structures of the SHFV nsp1 proteins will not be known until they are solved by X-ray crystallography.

Although arterivirus PLP1α cleavage site sequences are not conserved, arterivirus PLP1β cleavage sites show some conservation (-Trp/Tyr/Asn-Tyr/Phe-Gly↓Gly-) (Sun, Xue et al. 2009; Xue, Sun et al. 2010). Each of the three SHFV PLP1s cleaves preferentially between two Gly residues, but the amino acids upstream of these cleavage sites are not conserved. Inefficient PLP1α cleavage was also observed at an alternative downstream -Gln-Leu-Gly↓Thr- (aa 169-172) site *in vitro*. Only *cis* cleavage at a single site downstream of the catalytic His residue was observed in *in vitro* autocleavage reactions previously done with EAV, PRRSV and LDV nsp1 polyprotein fragments (Snijder, Wassenaar et al. 1992; den Boon, Faaberg et al. 1995). Because cleavage of the single expected EAV nsp1 protein occurs rapidly in *in vitro* polyprotein autoprocessing reactions, it was concluded that cleavage occurred cotranslationally as the growing peptide folded into the appropriate conformation with the cleavage site located in the active site binding pocket (Snijder, Wassenaar et al. 1992; Xue, Sun et al. 2010). The detection of "free" SHFV nsp1γ when the catalytic Cys residues of both PLP1α and PLP1β were substituted indicated that nsp1γ can cleave at both downsteam and upstream junction sites. The PLP1s encoded by the coronaviruses, murine hepatitis virus A59 and human coronavirus 229E, were previously reported to *cis* cleave at a downstream site and *trans* cleave at a site upstream of the catalytic Cys residue in *in vitro* cleavage assays (Herold, Gorbalenya et al. 1998; Teng, Pinon et al. 1999). The arterivirus PLP2 in nsp2 was also shown to possess *trans* cleavage activity (Snijder, Wassenaar et al. 1995; Han, Rutherford et al. 2009).

The 575 aa SHFV N-terminal polyprotein (599 aa with the tags) expressed in the *in vitro* transcription/translation assays in the present study is the longest arterivirus nsp1 polyprotein to be tested to date and contains three instead of two tandem PLPs. The sites cleaved by each of the SHFV PLP1s in the *in vitro* autocleavage reactions are summarized in Fig. 2-2C. PLP1α cleaves preferentially at Gly164↓Gly165. PLP1α cleavage at Gly271↓Val272 located within the catalytic region of PLP1β may occur when the polyprotein is misfolded. The abundance of the 30 kDa bands generated by cleavage at the Gly271↓Val272 site was typically low suggesting that only a small proportion of the *in vitro* synthesized polyproteins were misfolded. However, the intensity of the 30 kDa bands increased when PLP1α was mutated suggesting that retention of nsp1α may facilitate misfolding of nsp1β. The detection of cleavage at multiple sites within nsp1β when only PLP1β was active was surprising. Misfolding of this region may expose alternative sites for PLP1β *cis* or *trans* cleavage. PLP1γ cleaves at the downstream nsp1γ-nsp2 junction (Gly484/Gly485/Gly486) but also at the upstream nsp1β-nsp1γ junction (Gly350/Gly351) and at sites within nsp1β. Cleavage by PLP1γ at the upstream sites may occur *in trans* similar to upstream cleavage by coronavirus PLPs (Mielech, Chen et al. 2014).

Precursor proteins as well as alternative peptides were produced in the SHFV *in vitro* autoprocessing reactions but were not detected in infected cells. Each of the SHFV nsp1 proteins is predicted to contain multiple post-translational modification motifs, including myristolyation, phosphorylation and glycosylation sites (data not shown). That modifications at some of these predicted sites occur *in vivo* was suggested by the observed increase in the sizes of the nsp1α and nsp1γ proteins produced in infected cells. These post-translational modifications may facilitate the folding of the SHFV polyprotein into the optimal conformation for autoprocessing *in vivo*. The failure to visualize the 30-kDa proteins in infected cells could be due to low abundance or

the inability of the individual anti-nsp1 antibodies to detect these proteins. However, the possibility of the rapid degradation of uncleaved intermediate precursor peptides and alternative cleavage products in infected cells cannot be ruled out. The additional nsp1γ region in the SHFV genome is expected to have been generated by a recombination event that resulted in gene duplication. The sequence analysis and homology modeling data suggested that nsp1γ is more similar to nsp1α than to nsp1β. Only nsp1γ was shown to be able to cleave *in vitro* at both downstream and upstream junction sites. However, there was no junction site upstream of PLP1α. Preliminary data suggest that the individual SHFV nsp1 proteins may differ in their cellular localizations and functions.

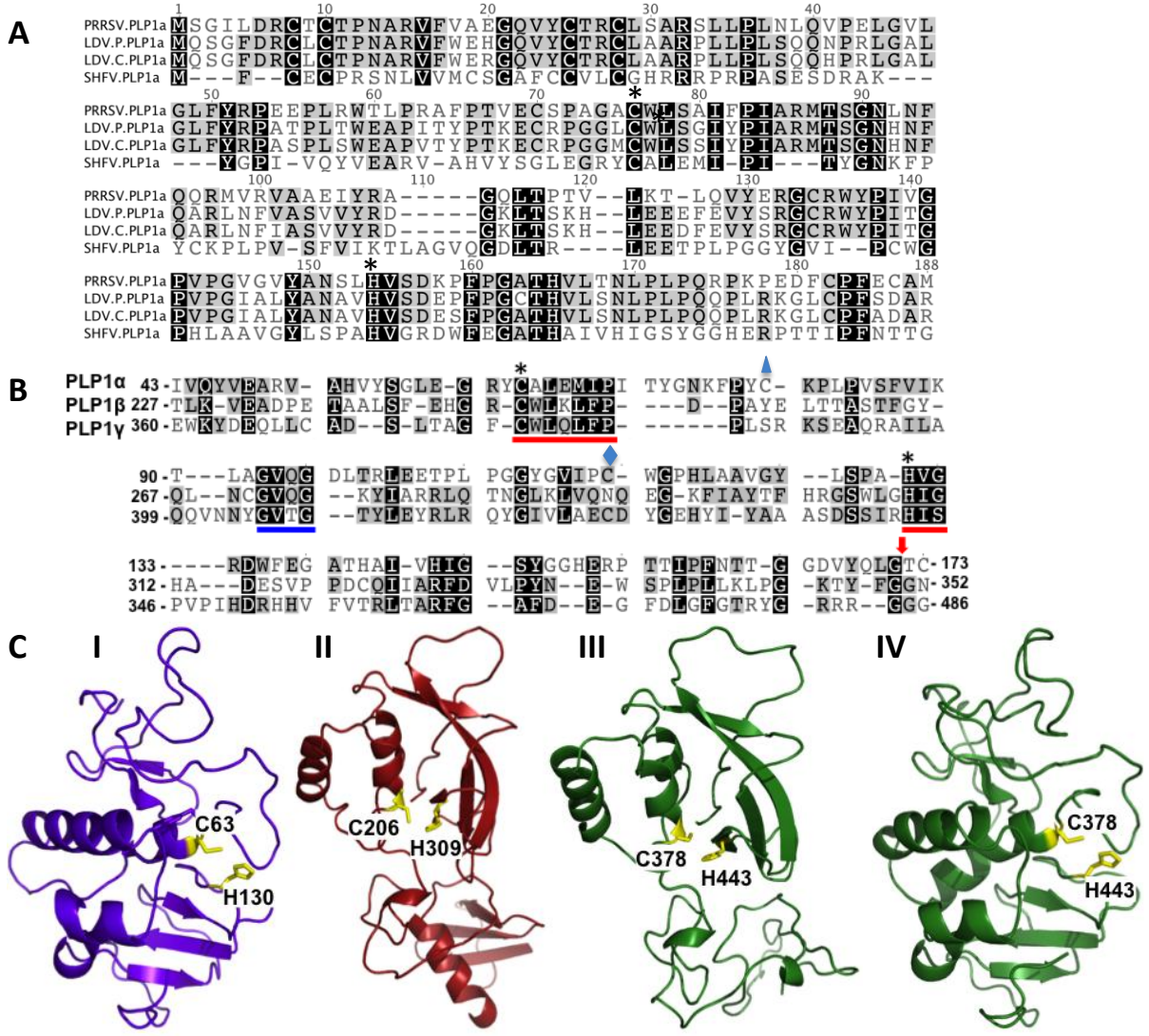


Figure 2.1. Prediction of the three SHFV PLP domains.
(A) Alignment of SHFV PLP1α with LDV and PRRSV PLP1α domains. (B) Alignment of the three SHFV PLP1 (α, β, and γ) domains. Asterisks indicate the predicted catalytic residues of each protease. A blue diamond indicates the only Cys adjacent to a Trp in SHFV PLP1α. A blue triangle indicates the Cys previously predicted to be the SHFV PLP1α catalytic residue. A red arrow indicates the position of the predicted SHFV PLP1 cleavage sites. Red lines indicate the conserved catalytic residue motifs. A blue line indicates a conserved motif used to anchor the alignment. (C) Homology models of the three SHFV PLP1 proteins generated using the crystal structure of the PRRSV nsp1α protein for PLP1α (I) and the PRRSV nsp1β structure for SHFV PLP1β (II) and PLP1γ (III). For comparison, SHFV nsp1γ was also modeled on the PRRSV nsp1α structure (IV). The predicted catalytic Cys and His residues for each SHFV PLP1 are indicated in yellow.

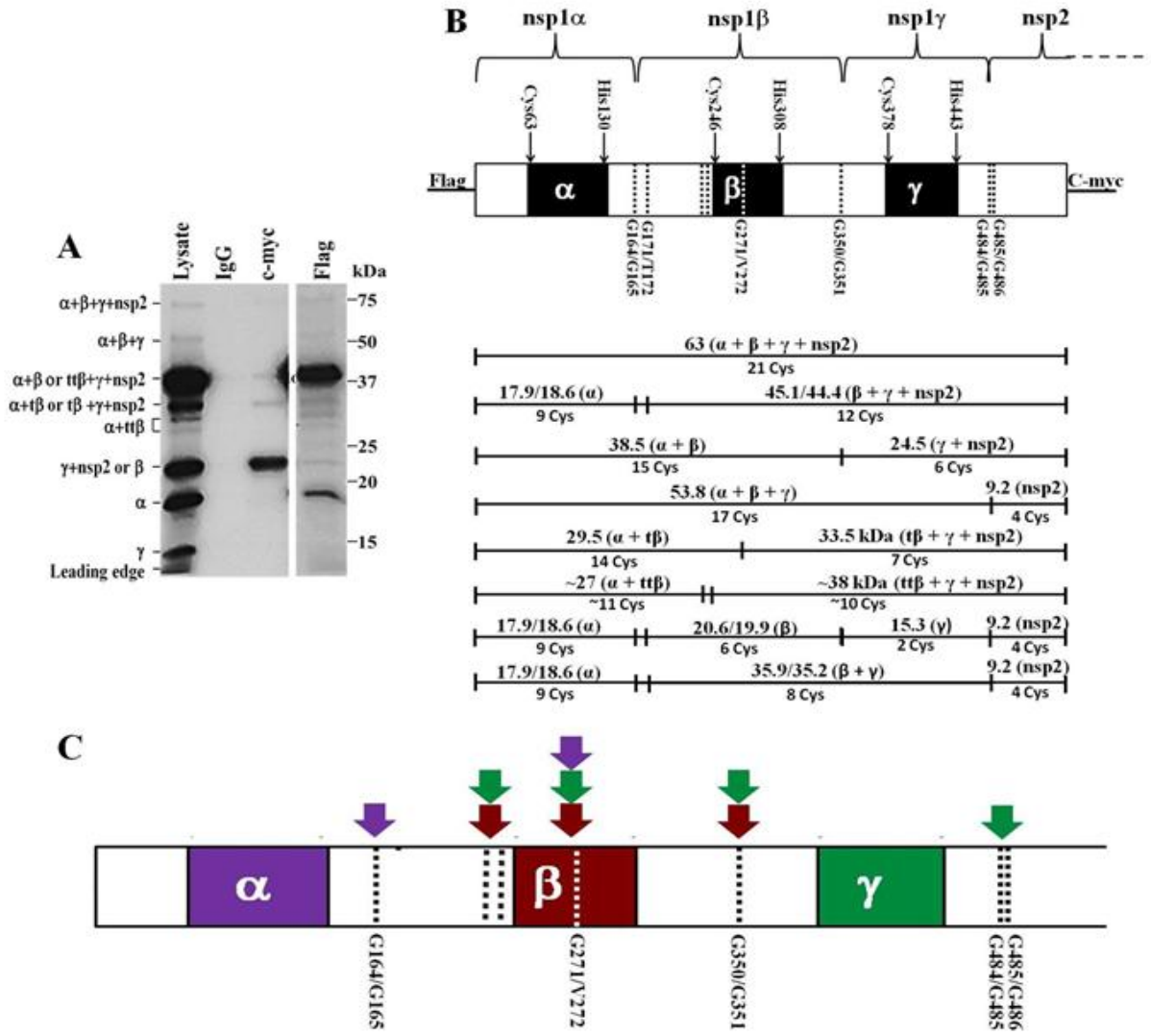


Figure 2.2. SHFV nsp1 polyprotein cleavage map.
The N-terminal 1725 nts of SHFV ORF1a were cloned into the pTnT vector. The polyprotein expressed had an N-terminal Flag tag and a C-terminal c-Myc tag. (A) Products of the wildtype SHFV nsp1 polyprotein autoprocessing. The wildtype cDNA was *in vitro* transcribed and translated in a coupled TnT reaction (Promega) as described in Materials and Methods. Protein products labeled by incorporation of [^{35}S]-Cys were immunoprecipitated with murine IgG, anti-c-Myc or anti-Flag antibody prior to separation by 13% SDS-PAGE. The positions of protein standards are indicated on the right and the identities of the SHFV peptides generated are indicated on the left. (B) Diagram showing the relative locations of the predicted PLP1 domains and cleavage sites in the SHFV nsp1 polyprotein and the predicted cleavage products. Top panel-Black boxes indicate the predicted PLP1 domains and dotted lines indicate the cleavage sites. Bottom panel-The sizes and identities of the predicted cleavage products are indicated above the lines and the number of Cys residues in each peptide is indicated below the lines. (C) The sites shown by the data to be cleaved by each SHFV PLP1 in *in vitro* autoprocessing reactions are indicated by color coordinated arrows.

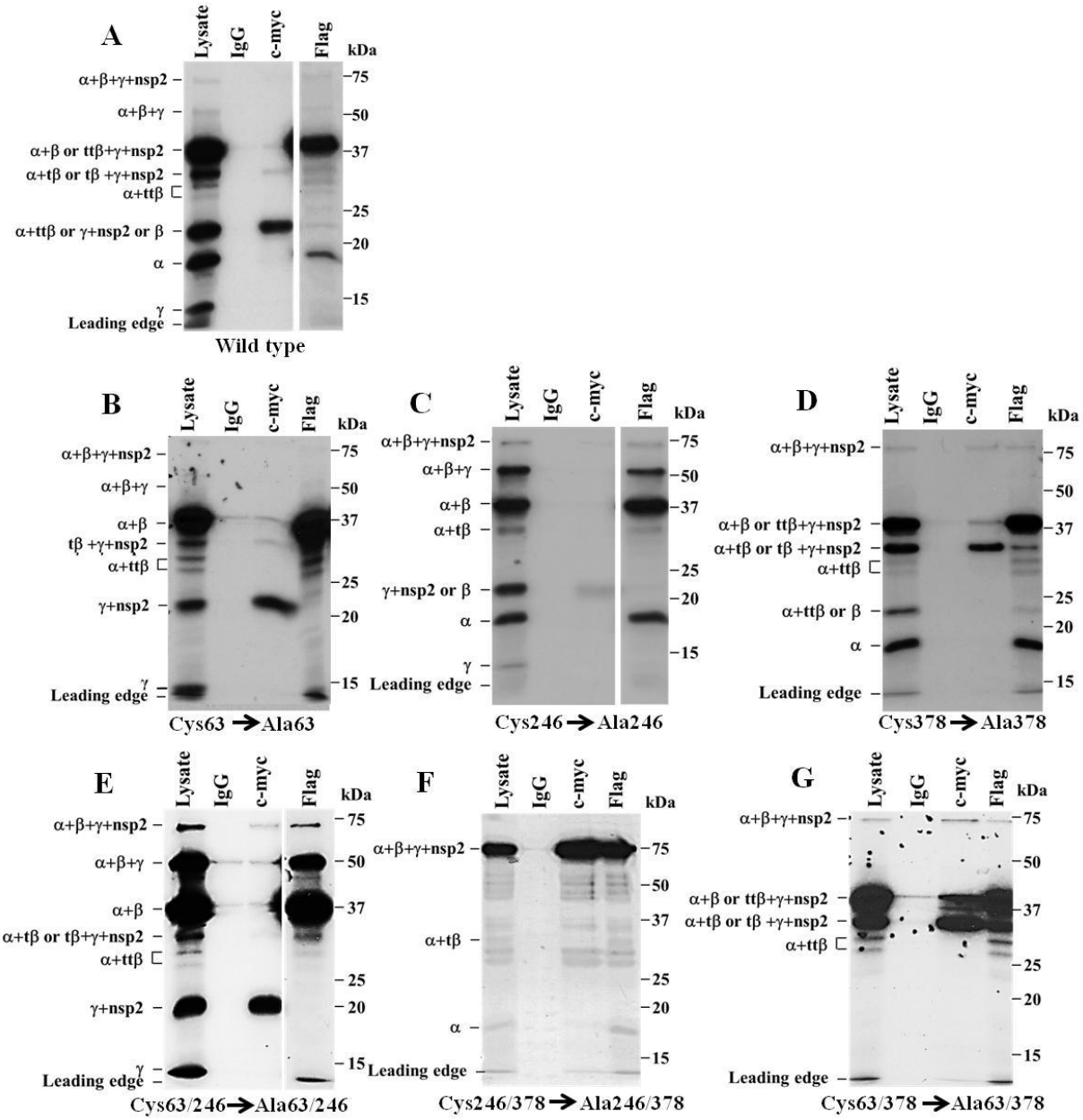


Figure 2.3. Cleavage products produced by nsp1 polyproteins containing mutations in one or two catalytic Cys residues.

(A) Wild type polyprotein. (B) PLP1α Cys 63 mutant. (C) PLP1β Cys 246 mutant. (D) PLP1γ Cys 378 mutant. (E) PLP1α Cys 63 and PLP1β Cys 246 mutant. (F) PLP1β Cys 246 and PLP1γ Cys 378 mutant. (G) PLP1α Cys 63 and PLP1γ Cys 378 mutant. Wildtype and mutant cDNAs were *in vitro* transcribed and translated and the cleavage products immunoprecipitated and analyzed as described in the legend of Fig. 2-2A. A shorter exposure is shown for the Flag IP lanes in panels A, C and E.

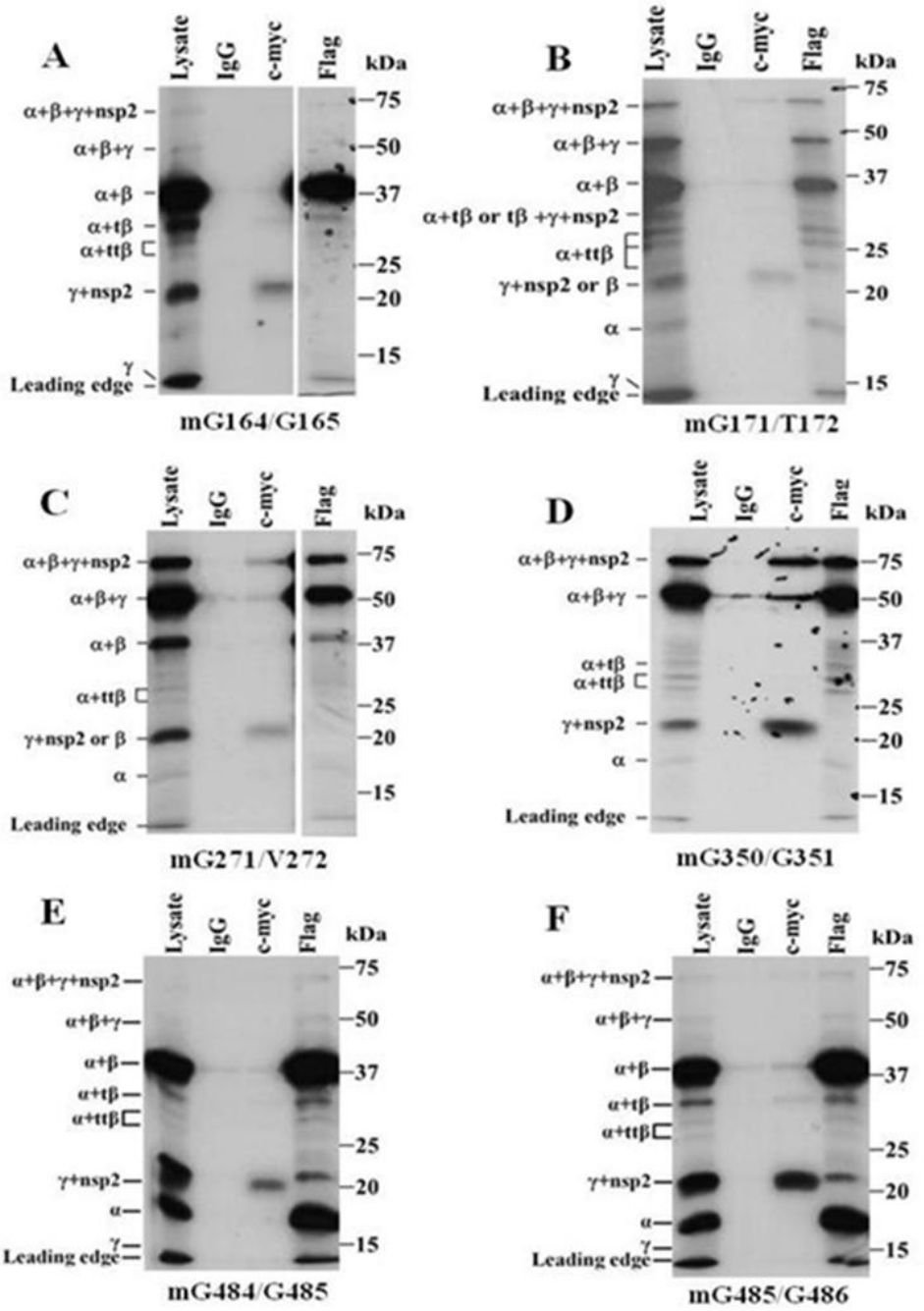


Figure 2.4. Cleavage products produced by nsp1 polyproteins with a single predicted PLP cleavage site mutated.

(A) mG164/G165, (B) mG171/T172, (C) mG271/V272, (D) mG350/G351, (E) mG484/G485 and (F) mG485/G486 cDNAs were *in vitro* transcribed and translated in a coupled TnT reaction in the presence of [^{35}S]-Cys, immunoprecipitated and analyzed as described in the legend of Fig. 2-2A. A shorter exposure was used for the Flag IP lane in panels A and C.

Figure 2.5. MassSpec analyses of selected polyprotein cleavage products.
(A) nsp1α, (B) nsp1α+β and (C) nsp1α+β+γ. Products from *in vitro* transcription/translation reactions were separated on a 12% NuPAGE bis-tris gel and stained with colloidal blue. The nsp1α (18 kDa) band and the nsp1α+β (39 kDa) precursor band were separately excised from the Flag IP lane of a wildtype construct reaction while the nsp1a+β+γ (54 kDa) band was excised from the Flag IP lane of a Cys246Ala mutant construct reaction. Proteins in the excised bands were subjected to in-gel trypsin digestion and the peptides analyzed by LC-MS/MS on an LTQ-Orbitrap XL mass spectrometer. The MS/MS spectra generated were searched against a custom database containing the SHFV nsp1+N-terminal nsp2 construct sequence, the Triticum proteome and common contaminants using SEQUEST. Regions of the nsp1 sequences for which peptides were detected are highlighted in yellow. Oxidized amino acids are indicated in green and methylated aa are indicated in red. The predicted PLP1 cleavage sites are indicated with red arrows. The N-terminal Flag tag is boxed in red and the first amino acid of the nsp1 polyprotein is indicated by a red asterisk. The entire nsp1α sequence is shown. For nsp1α+β and nsp1α+β+γ, only the C-terminal portion of the peptide is shown.

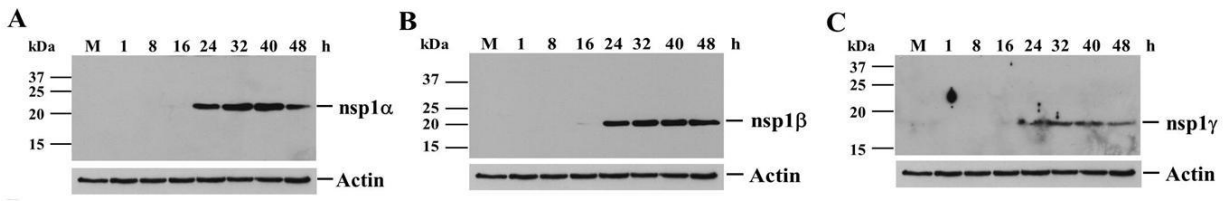


Figure 2.6. Autoprocessing of SHFV nsp1 proteins in infected cells.
MA104 monolayers were infected with SHFV, strain LVR, at an MOI of 1. Cell lysates were harvested in RIPA buffer at the indicated times after infection and the individual nsp1 viral proteins were detected by Western blotting using anti-SHFV (A) nsp1α-, (B) nsp1β-, or (C) nsp1γ-specific antibody. Anti-actin antibody was used to detect actin on stripped blots. The positions of the protein markers are indicated on the left and of the viral protein bands are indicated on the right.

Table 2.1. Primers

Construct	Mutation(s)		Primer Sequence (5' - 3')
mα	Cys63 → Ala	F	GGGCGCTATgctGCTCTTGAGATGATA[a]
		R	TATCATCTCAAGAGCagcATAGCGCCC
mβ	Cys246 → Ala	F	CTTTGAGCATGGCCGCgcCTGGCTGAAGTTGTTC
		R	GAACAACTTCAGCCAGgcGCGGCCATGCTCAAAG
mγ	Cys378 → Ala	F	CCTCACTGCTGGGTTCgcTTGGTTGCAGCTATTTCC
		R	GGAAATAGCTGCAACCAAgcGAACCCAGCAGTGAGG
G164/G165	Thr163 → Val / Gly164 → Val	F	CCATTCAACACGgtTGtAGGTGACGTTTATC
		R	GATAAACGTCACCTaCAacCGTGTTGAATGG
G171/T172	Leu170 → Ala / Gly171 → Ala	F	GGAGGTGACGTTTATCAGgcAGcAACCTGCACTATCGTTGAG
		R	CTCAACGATAGTGCAGGTTgCTgcCTGATAAACGTCACCTCC
G271/V272	Cys270 → Val / Gly271 → Val	F	CGGTTATCAATTGAACgtTGtCGTTCAAGGAAAATACATTGC
		R	GCAATGTATTTTCCTTGAACGaCAacGTTCAATTGATAACCG
G350/G351	Phe349 → Ala / Gly350 → Ala	F	TTAAGTTGCCAGGGAAGACTTATgcCGcTGGAAATGCCAGTTCGG
		R	CCGAACTGGCATTTCCAgCGgcATAAGTCTTCCCTGGCAACTTAA
G484/G485	Arg483 → Ala / Gly484 → Ala	F	CGTTACGGTCGCCGTgcgGcAGGCGGCAAGAAGTCAGGCCAATCG
		R	CGATTGGCCTGACTTCTTGCCGCCTgCcgcACGGCGACCGTAACG
G485/G486	Gly484 → Ala / Gly485 → Ala	F	CGTTACGGTCGCCGTCGAGcAGcCGGCAAGAAGTCAGGCCAATCG
		R	CGATTGGCCTGACTTCTTGCCGgCTgCTCGACGGCGACCGTAACG

[a]The substituted nts are indicated by lower case letters.

3 CHARACTERIZATION OF TRANSCRIPTION REGULATING SEQUENCES (TRS) AND PROTEINS ENCODED BY THE SHFV GENOME

3.1 Introduction

Simian hemorrhagic fever virus (SHFV) is an enveloped, single stranded, positive sense RNA virus that infects monkeys. Macrophages (MΦs) and dendritic cells are the main target cells for this virus (Brinton, Di et al. 2015). SHFV infections in African monkeys such as baboon are persistent and asymptomatic (London 1977; Lauck, Sibley et al. 2013). In contrast, SHFV infections in macaque monkeys trigger an acute, fatal hemorrhagic fever disease that causes death 7-13 days after infection. The clinical symptoms induced by SHFV infection in macaques include but are not limited to fever, facial edema, cyanosis, skin petechiae and subcutaneous hematomas (Allen, Palmer et al. 1968; Palmer, Allen et al. 1968). Because SHFV only infects monkeys and the symptoms in macaques induced by SHFV are similar to those caused by other hemorrhagic fever viruses such as Ebola virus (Bray 2005; Bray and Geisbert 2005), SHFV infected monkeys can serve as a Biosafety Level II animal model for the study of viral hemorrhagic fever disease (Palmer, Allen et al. 1968).

SHFV belongs to the family *Arteriviridae* together with porcine reproductive and respiratory syndrome virus (PRRSV), equine arteritis virus (EAV), lactate dehydrogenase-elevating virus (LDV) and the recently identified wobbly possum disease virus (WPDV) (Dunowska, Biggs et al. 2012). The arterivirus genome has a 5′-cap and a 3′-poly(A) tail. The 5' two-thirds of the genome encode two polyproteins, ORF1a and ORF1ab, that are auto-cleaved into 13-15 non-structural proteins. These non-structural proteins are required for virus replication and transcription (Snijder and Meulenberg 1998; Snijder, Kikkert et al. 2013). The 3′ one-third

of the arterivirus genome encodes five to nine minor structural proteins and three major structural proteins that are required for virus particle assembly. Among the major structural proteins, GP5 and the M form heterodimers that were shown to be essential for EAV assembly (de Vries, Post et al. 1995; Snijder, Dobbe et al. 2003; Wieringa, de Vries et al. 2004). The nucleocapsid protein (N) forms homodimers and the homodimers interact with each other as well as with the genomic RNA to form a spherical nucleocapsid containing the virus genome (Dea, Gagnon et al. 2000; Doan and Dokland 2003; Dokland 2010). Among the minor structural proteins, the minor glycoproteins GP2, GP3 and GP4 form heterotrimers and function in cell receptor recognition and cell tropism (Das, Dinh et al. 2010; Tian, Wei et al. 2012). The E protein forms oligomers and functions as an ion channel in the virus envelop that could be important for virus entry (Lee and Yoo 2006). The function of the recently discovered ORF5a is unclear but knocking out the expression of this protein in EAV reduced the virus yield (Firth, Zevenhoven-Dobbe et al. 2011; Johnson, Griggs et al. 2011). Among arteriviruses, SHFV has the largest genome at 15.7 kb that encodes an extra sets of minor structural proteins GP2′, GP3′, GP4′ and E′ proteins. The functions of this extra set of minor structural proteins are unknown but each of them is functionally important and is required for the production of infectious extracellular virus (Vatter, Di et al. 2014).

The genome organization and the transcription strategies of arteriviruses are similar to those of coronaviruses and they are both classified within the order *Nidovirales* (Gorbalenya, Enjuanes et al. 2006). Both arteriviruses and coronaviruses use discontinuous RNA synthesis to generate the templates for a nested set of 3′ and 5′ co-terminal subgenomic mRNAs (sg mRNAs) (Pasternak, Spaan et al. 2006; Sawicki, Sawicki et al. 2007). Each sg mRNA expresses at least one structural protein from the first 5′ start codon. The discontinuous RNA synthesis is regulated

by the transcription regulatory sites (TRSs) in the genomes of arteriviruses and coronaviruses (Pasternak, Spaan et al. 2006; Sawicki, Sawicki et al. 2007). The TRSs are usually 30-40 nt long with a 7-10 nt core sequence and 10-15 nt flanking sequences on both ends. Due to RNA folding, TRSs usually self assemble into stem-loop structures with the core sequences in the loops and the flanking sequences in the stems (Van Den Born, Gultyaev et al. 2004; van den Born, Posthuma et al. 2005). There is a single leader TRS at the 5′ end of the genome and multiple body TRSs at the 3′ end of the genome (de Vries, Chirnside et al. 1990). The core sequences of different body TRSs show different levels of sequence homology to the core sequence of the leader TRS. During the synthesis of the minus strand RNA, the viral RNA polymerase starts from the 3′ end of the genome and runs into a body TRS. At each body TRS, the polymerase either keeps going to the next body TRS, or terminates prematurely and disassociates with a partially transcribed minus strand RNA. The 3′ end of the nascent minus strand RNA contains the complement of the core sequence of the body TRS that is also complementary to the core sequence of the 5' leader TRS (van Marle, Dobbe et al. 1999). Therefore the nascent minus strand RNA can realign at the 5′ end by base-pairing between the leader TRS and the minus body TRS and then transcription continues, generating a minus strand sg RNA with a unique leader-body junction sequence (den Boon, Kleijnen et al. 1996; Pasternak, van den Born et al. 2001). These minus sg mRNAs are then transcribed into sg mRNAs. Although, each known structural protein ORF was assigned to a sg mRNA transcribed from the minus strand template generated from one body TRS in all the studied arterivirus, alternative body TRSs have been reported that generate different sg RNAs for the same structural protein (den Boon, Kleijnen et al. 1996; Meng, Paul et al. 1996; Lin, Chang et al. 2002; Hussain, Pan et al. 2005). Occasionally, two or more structural proteins have been assigned to a single sg mRNA because a separate sg mRNA

was not initially detected for each structural protein (Godeny, de Vries et al. 1998). So far, arteriviruses appear to only encode structural proteins at the 3′ end of the genome, whereas coronaviruses encode both structural proteins and accessory proteins in this region (Liu, Fung et al. 2014).

The SHFV genome was initially thought to encode six structural protein ORFs as do other arteriviruses. Sequencing the SHFV 3′ genome revealed the existence of an extra set of minor structural proteins, GP2′, GP3′ and GP4′. In SHFV infected MA104 cells, six strong sg mRNA bands were initially detected by Northern blotting (Zeng, Godeny et al. 1995). Subsequently, one leader TRS and eight body TRSs were identified in the SHFV genome (Smith, Wang et al. 1997; Godeny, de Vries et al. 1998). The leader-body junction for GP3′ was not identified, thus both GP3′ and GP2a′ were assigned to sg mRNA2' under the assumption of bicistronic expression (Godeny, de Vries et al. 1998). The later discovery of new ORFs encoding E and ORF5a proteins in other arteriviruses suggested that SHFV could potentially encode three additional structural proteins: E, ORF5a and the duplicated E′. In the present study, a more comprehensive analysis of SHFV sg mRNA expression and body TRSs usage was done using the recently constructed SHFV LVR infectious clone virus (SHFVic) (Vatter, Di et al. 2014). A total of 10 sg mRNA bands were detected in SHFVic infected MA104 cell extracts using Northern blotting assays with multiple long probes. Nine of these bands corresponded to individual sg mRNAs for known structural proteins including a separate sg mRNA band for GP3′. The other one was an unexpected new sg mRNA. Amplification and cloning of individual known leader-body junction regions followed by sequencing of multiple clones for each region revealed the use of 37 new body TRSs. NGS of SHFV sg mRNAs from infected MA104 cells or macaque MΦs confirmed the use of these new body TRSs and also suggested the use of

additional body TRSs located across the SHFV genome. The relative abundance of each sg mRNA remained constant at early and late times post infection in both cell types. Although the majority of the new body TRSs were redundant, alternative TRSs for sg mRNAs of known structural proteins, new body TRSs for GP3', E′ and E sg mRNAs were identified indicating that each of these proteins could be expressed from separate monocistronic sg mRNAs. A few of the new body TRSs were predicted to generate templates for sg mRNAs encoding new C-terminal in-frame ORFs of known structural proteins. Knocking out the expression of two of these C-terminal ORFs decreased virus yields in MA104 cells.

3.2 Materials and methods:

3.2.1 Cells and virus

The MA104 African green monkey embryonic kidney cell line was a gift from O. Nianan, Center for Disease Control and Prevention (Atlanta, GA). MA104 cells were cultured in 1 x Minimal Essential Medium (MEM) supplemented with 10% fetal bovine serum, 1% L-glutamine and 0.1% gentamicin and maintained at 37°C in a 5% CO_2 atmosphere.

Peripheral blood mononuclear cells (PBMCs) were isolated from whole blood collected from rhesus macaques (Yerkes Regional Primate Research Center, Atlanta, GA) using Ficoll® 400 (Mediatech Inc., Manassas VA) density gradient centrifugation according to standard protocols. The isolated monocytes were seeded in 24-well plates at a density of ~10^6 cells/well. After a 1h attachment period, monocytes were gently washed with RPMI-1640 media and incubated in RPMI-1640 media supplemented with 10% autologous serum and/or 10% heat-inactivated FBS, 50 U/ml of penicillin, 50 ug/ml of streptomycin and 25 ng/ml human recombinant MΦ colony stimulating factor (R&D Systems) for 8 days at 37°C in 5% CO_2 to

differentiate into MΦs. Two-thirds of the culture media was replaced with fresh growth media every three days during differentiation.

The SHFV LVR infectious clone was constructed from cDNA amplified from the virus genome RNA as described previously (Vatter, Di et al. 2014). SHFV RNA (100 ng) transcribed from the infectious clone *in vitro* was transfected into MA104 cells and culture fluid containing progeny virus was collected at 120 h after transfection. The progeny virus was passaged once by infecting MA104 cells at an MOI of 1 and the culture fluid was harvested at 24 h after infection, clarified by centrifugation, aliquoted and stored at -80°C. The SHFVic stock virus titer was 1.3 x 10^7 PFU/ml).

3.2.2 Digoxigenin (DIG)-labeled SHFV probes

High fidelity PCR (AccuPrime, Invitrogen) was used to amplify regions in the SHFV infectious clone to generate templates for *in vitro* synthesized RNA probes targeting different sg mRNAs. One set of primers was designed to amplify a region in the 5′-leader. Another eight sets of primers were designed based on the locations of each of the previously identified 3′ region body TRSs so that only regions between two neighboring TRSs were amplified (Table 3-1). A T7 promoter was added to the reverse primer. The PCR product templates were transcribed *in vitro* in the presence of DIG-labeled UTPs using a DIG Northern Starter Kit (Roche) according to the manufacturer's protocol. The concentrations of the DIG-labeled RNA probes were determined by a dot-blot assay using a DIG-labeled human actin RNA standard (Roche) according to the manufacturer's protocol. Briefly, a standard curve was generated using serial dilutions of the human actin RNA of known concentration. A 1 µl aliquot of each dilution was spotted onto an Amersham Hybond-N^+ membrane (GE Healthcare) and UV-crosslinked. The

membrane was then blocked in DIG blocking buffer (Roche), followed by incubation with anti-DIG antibody at a 1:10,000 dilution. After washing with DIG washing buffer (Roche), the membrane was developed with CDP-Star and imaged with a LAS4000 mini Luminescent Image Analyzer (GE Healthcare). The intensity of each spot on the membrane was measured using Multi Gauge V2.3 software and compared to the standard curve to estimate the concentration.

3.2.3 Northern blot hybridization

MA104 cells were grown in 6-well plates to confluence and infected with SHFVic at an MOI of 1. At different times after infection, cell lysates were harvested in TRI reagent (Molecular Research Center, Inc.) followed by extraction of total intracellular RNA. NorthernMax formaldehyde loading dye (Ambion) was added to 1 μg of extracted RNA and the samples were incubated at 80°C for 10 min. The denatured RNA was separated on a 1% formaldehyde agarose gel for 2.5 h at 100V. An RNA ladder (Millennium Markers-Formamide, Ambion) was used as a reference. After overnight capillary transfer to an Amersham Hybond-N^+ membrane (GE Healthcare) and UV-crosslinking of the transferred RNA to the membrane, the lane containing the RNA markers was cut from the membrane and stained with methylene blue. The rest of the membrane was pre-hybridized in DIG Easy Hyb buffer (Roche) at 68 ℃ for 30 min, followed by overnight hybridization with individual DIG-labeled, denatured RNA probes (100 ng/mL) at 68 ℃. The hybridized membrane was then washed with a low stringency buffer containing 2 X SSC plus 0.1% SDS at room temperature and then with a high stringency buffer containing 0.1 X SSC plus 0.1% SDS at 68 ℃. The membrane was then blocked in DIG blocking solution (Roche), incubated with anti-DIG antibody at 1:10,000 dilution (Roche), developed with

CDP-Star (Roche) and imaged with an LAS4000 mini Luminescent Image Analyzer (GE Healthcare).

3.2.4 Analysis of the leader-body junction sequences

One forward and ten reverse RT-PCR primers were designed to amplify the leader-body junctions of individual sg mRNAs generated from known and predicted body TRSs in the 3′ end of the SHFV genome (Table 3-2). The forward primer targeted a region in the 5′ leader (61 nt-79 nt in the SHFV genome). The reverse primers targeted sequences downstream of each known or predicted 3′ body TRS in the genome with the first one located 418 nt downstream of the identified TRS2′. MA104 cells were infected with SHFVic at an MOI of 1 and cell lysates were harvested in TRI reagent at 24 hpi. Total intracellular RNA was extracted from the cell lysate and subjected to high fidelity RT-PCR to amplify the leader-body junctions. The RT-PCR products (8 µl) were separated on a 2% agarose gel, stained with ethidium bromide and imaged with a BioDoc-it imaging system (UVP, LLC). The remainder of the RT-PCR product was run on a 2% agarose gel, stained with ethidium bromide and the appropriate bands were visualized under long-wavelength UV light, excised from the gel and the RNA was eluted. The extracted RT-PCR products (4 µl) were ligated to 1 µl of the pCR4-TOPO cloning vector and transformed into TOP10 chemically competent cells, followed by plating on LB plates containing 50 µg/ml kanamycin and cell growth overnight at 37°C. Forty colonies were picked randomly from the plates for the regions between TRS2′ and TRS4′ and between TRS5 and TRS6. Twenty colonies were picked randomly from the plates for each of the other eight regions. All the picked colonies were grown overnight in liquid culture. Plasmid DNAs were extracted from liquid cultures, digested with EcoRI and then separated on a 2% agarose gel and imaged with a BioDoc-it

imaging system (UVP, LLC). For each leader-body junction region, the plasmid DNAs carrying inserts of different sizes were sent for sequencing using a T3 primer. The sequencing results were analyzed using DNASTAR Lasergene software to identify different leader-body junctions. The body TRS sequences were identified after alignment of the leader-body junction sequences to the 5′ leader sequence and the 3′ genome sequence.

3.2.5 NGS of SHFV sg mRNAs

MA104 cells were grown in T-75 flasks to 100% confluence and subsequently infected with SHFVic at an MOI of 1. At 8 and 18 hpi, culture fluid was removed and cell lysates were harvested in 3.5 ml of TRI reagent and sent to University of North Carolina, Chapel Hill, for library construction using an Illumina TruSeq Stranded mRNA sample preparation kit (Illumina) following the manufacturer's protocol. Briefly, total intracellular RNA was extracted from the TRI reagent, followed by poly(A) selection of mRNA. The quality and quantity of the selected mRNA were assessed using an Agilent RNA 6000 nano kit (Agilent Technologies). The selected mRNA was then fragmented and subjected to RT-PCR using random hexamers, followed by end repair and adenylation. The 3′ end adenylated dsDNAs were ligated to Illumina adapters and amplified by PCR. The final dsDNAs library was validated with an Agilent High Sensitivity DNA kit (Agilent Technologies) and subjected to RNA-Seq using Illumina Hi-Seq with one sample per lane. Libraries from 3 biological repeats for each time point were prepared and sequenced

MΦs differentiated from the isolated PBMCs were grown in 24-well plates until 80% confluence and infected with SHFVic at an MOI of 1. At 7 h and 16 hpi, culture fluid was removed and cell lysates were harvested in 200 µl TRI reagent from each well. Cell lysates from

four infected wells were combined for each time point and total intracellular RNA was extracted following the manufacture's protocol. The purified total RNA was sent to the Georgia Genomics Facility at University of Georgia for library preparation using a NGS Stranded RNA library preparation kit (KAPA Biosystems) following the manufacture's protocol. This protocol starts with poly(A) selection of mRNA. The prepared libraries were subjected to RNA-Seq using Illumina Mi-Seq with four samples pooled on one lane.

3.2.6 NGS data analysis

CLC Genomics Workbench 8.5.1 software was used to analyze the RNA-Seq data. Briefly, the sequences of the SHFV full length genome (NC_003092.2) or the genome without the leader region were imported as references. The 15 nt leader-body junction sequences for the 8 previously identified and the 36 newly identified sg mRNAs were also imported as references. Each of the leader-body junction reference consists of the leader TRS core sequence at the 5′ end fused to individual body TRS region sequence. For each sample, all the fastq reads from the Illumina Hi-Seq output were trimmed before alignment to the SHFV genome using the default setting. All of the reads that aligned to the SHFV genome were then extracted and mapped again to a 15 nt region (186 nt-200 nt in the SHFV genome) located immediately upstream of the 5′-leader TRS core sequence (201 nt-209 nt) using a higher stringency setting. All of the reads that mapped to this 15 nt region were then extracted and mapped again to another l5 nt region (201 nt-215nt) containing both the 5′-leader TRS core sequence and the beginning of the ORF1a using the highest stringency setting. All of the unmapped reads were then collected and subjected to a workflow which sequentially mapped reads to the 15 nt reference sequence of each known 3′ leader-body junction sequence using the highest stringency setting. The number of reads

mapping to each leader-body junction was designated as the transcription level for a particular sg mRNA. The relative abundance of each sg mRNA was calculated as the percentage of the reads mapping to that sg mRNA among the total number of mapped reads. To estimate the expression level and relative abundance of each ORF, the transcription levels and relative abundances of each of the alternative sg mRNAs encoding the same ORF were added together. The fold change in the transcription level and relative abundance of each sg mRNA between 8 and 18 hpi were also calculated. All of the reads that contained a leader sequence but failed to map to any of the identified leader-body junctions were collected and mapped to an SHFV genome sequence without the leader region (201 nt-15717 nt in the SHFV genome) using the default setting. The mapping results were displayed with the reads aligned to the reference genome.

3.2.7 Construction of mutant SHFV infectious clone virus

Five sets of primers were designed to mutate the start codon of each predicted C-terminal in-frame ORF while keeping the amino acid sequence translated from the overlapping ORF intact (Table 3-3). With each set of primers, a single nt substitution was introduced into the appropriate SHFV infectious clone fragment (Fragment III for ΔGP2′-C, fragment IV for ΔGP4′-C and ΔGP2-C, fragment V for ΔGP5-C and ΔGP6-C) (Vatter, Di et al. 2014) using a QuikChange Lightning Site-Directed Mutagenesis kit (Agilent Technologies) according to the manufacturer's protocol. After confirming the validity of mutant fragments by sequencing, the five SHFV genome fragments were digested with PflMI and ligated simultaneously into the full-length genome cDNA and cloned into a pACYC184 vector. Mutant SHFV genome RNA was transcribed *in vitro* from the linearized full-length clone and ~100 ng of purified mutant SHFV RNA was transfected into subconfluent MA104 cells in a 6-well plate. At 120 h after

transfection, 500 ul of culture fluid was collected and used to infect a confluent MA104 monolayer in a 10 cm dish. At 28 hpi, the culture fluid was collected from the dish and centrifuged at 1,000 rpm for 5 min at 4°C, aliquoted and stored at -80°C as a P1 mutant SHFV stock. Some of the clarified supernatant was used to extract mutant SHFV genome RNA. After RT-PCR amplification of the mutant regions, the RT-PCR products were sequenced to verify the presence of the mutation. Each of the P1 virus stocks were titered by plaque assay on MA104 cells.

3.2.8 Western Blot Assay

Infected MA104 cells were lysed in 1 x RIPA buffer (1 X phosphate-buffered saline, 1% Nonidet P-40, 0.5% sodium deoxycholate, and 0.1% SDS) containing 1 x Halt protease inhibitor cocktail (Thermo Scientific). The total amount of protein in each cell lysate was measured by BCA assay (Pierce) according to the manufacturer's protocol. The same amount of total protein for each cell lysate was loaded onto a 15% SDS gel. After separation by SDS-PAGE, the lysate proteins were transferred to a nitrocellulose membrane by electrophoresis at 100 V for 1 h. The membranes were then cut into strips based on the sizes of proteins of interest, incubated in blocking buffer (1 X Tris-buffered saline containing 5% non-fat dry milk and 0.1% Tween 20) overnight at 4°C, followed by incubation with primary antibodies against the SHFV peptides (nucleocapsid: NETHYVFAEPGDLRC or nsp1β: FAQKVITAFPEGVLC) (Abgent) or against beta-actin (C-11; Santa Cruz Biotechnology) at 4°C overnight. The blots were then washed with 1 X Tris-buffered saline containing 0.1% Tween 20 and incubated with appropriate secondary antibodies (horseradish peroxidase-conjugated anti-rabbit or anti-mouse) (Santa Cruz Biotechnology) for 1 h at room temperate. Afterward, the blots were washed again and

developed with a Super-Signal West Pico or Femto detection kit (Pierce) according to the manufacturer's protocol.

3.2.9 Plaque Assay

Culture fluid containing wild type or mutant SHFV virus was collected at different times after infection of MA104 cells, followed by centrifugation at 1,000 rpm for 5 min at 4 ℃. Clarified supernatants were subjected to serial 10 fold dilution in MA104 culture media and 100 µl per well was adsorbed onto confluent MA104 monolayers in 6-well plates. After adsorption for 1 h in 37°C incubator, the inoculum was removed and replaced with 2 ml of overlay media (2 x MEM containing 5% FCS mixed 1:1 with 1% SeaKem ME agarose). The plates were incubated at 37°C for 48 h before removal of the agarose plug. Cells were stained with 0.05% crystal violet in 10% ethanol prior to plaque counting and imaging. Three biological repeats of each sample were assayed in duplicate.

3.3 Results:

3.3.1 Discovery of additional sg mRNAs generated during an SHFV infection in MA104 cells.

Although sequencing the 3′ end of the SHFV genome indicated the presence of nine ORFs, initial Northern blotting of SHFV infected MA104 cell extracts only detected six strong sg mRNA bands. The SHFV sg mRNAs were reanalyzed in SHFVic infected MA104 cells using a Northern blotting assay with a sg mRNA7 probe incorporated with DIG-labeled UTP (392 nt). MA104 cells were infected with SHFVic at a MOI of 1 and total intracellular RNA was extracted at different times after infection. A genomic RNA band and ten sg mRNA bands were detected

by 8 hpi (Figure 3-1A). The sizes of these sg mRNA bands were 5.0, 4.0, 3.5, 2.8, 2.6, 1.9, 1.2, 0.6 kb and corresponded to the sizes of the previously identified sg mRNA2', sg mRNA4', sg mRNA2, sg mRNA3, sg mRNA4, sg mRNA5, sg mRNA6 and sg m RNA7 respectively. A ~4.4 kb band was detected between the sg mRNA2' (5 kb) and sg mRNA4' (4 kb) (Figure 3-1A). This band corresponded to the size of the previously postulated but never detected sg mRNA3′. A new ~1.7 kb band was detected between sg mRNA5 (1.9 kb) and sg mRNA6 (1.2 kb) (Figure 3-1A). Northern blotting using a 5′-leader probe incorporated with DIG-labeled UTP (194nt) confirmed the detection of 10 sg mRNA bands including the potential sg mRNA3′ band and the novel sg mRNA band (data not shown).

To confirm the identity of each RNA band detected, MA104 cells were infected with SHFVic at a MOI of 1 and at 12 hour post infection (hpi), total intracellular RNA was extracted from three biological repeats and subjected to Northern blotting using a series of probes targeting the regions between each set of predicted neighboring body TRSs (Table 3-1). The probe targeting the TRS2′ to TRS4′ region (sg mRNA2′ probe) detected the strong 5 kb band consistent with the expected size of sg mRNA2′ (Figure 3-1B). This probe also detected a faint band of ~4.4 kb which was also detected by the probe targeting the TRS4′ to TRS2 region (sg mRNA4′ probe) (Figure 3-1B). This new sg mRNA contained sequence in the TRS2′ to TRS4′ region as well as in the TRS4′ to TRS2 region, which is consistent with what would be expected for a sg mRNA3′ generated from a body TRS located between the TRS2′ and TRS4′. The sg mRNA4′ probe also detected one additional smaller band at around 3.5 kb (Figure 3-1B). A known sg mRNA of that size containing sequence from the TRS4′ to TRS2 region would be expected to be sg mRNA4′. Northern blotting using individual probes (sg mRNA2 probe, sg mRNA3 probe and sg mRNA4 probe) each specifically detected one additional smaller band representing sg

mRNA2, sg mRNA3 and sg mRNA4, respectively (Figure 3-1B). Northern blotting using the probe targeting the TRS5 to TRS6 region (sg mRNA5 probe) specifically detected sg mRNA5 as well as a faint ~1.7 kb band. This band was also detected by the probe targeting the TRS6 to TRS7 region (sg mRNA6 probe) (Figure 3-1B). This novel sg mRNA band would be generated from a body TRS located between TRS5 and TRS6. The sg mRNA6 probe also specifically detected a strong 1.2 kb band of the size expected for sg mRNA6 (Figure 3-1B).

3.3.2 Additional body TRSs in the SHFV genome are used to initiate sg RNAs

The two unassigned RNA bands detected by Northern blotting in Figure 3-1 strongly suggested the existence of the predicted sg RNA3′ generated from a body TRS located between the TRS2′ and TRS4′, and also of a novel sg RNA generated from a body TRS located between the TRS5 and TRS6. Although the leader-body junction of the predicted sg mRNA3′ was not previously identified, the sequence of body TRS3′ was predicted to be 5′-CTTAAAACC-3′ located between the TRS2′ and TRS4′ (Smith, Wang et al. 1997; Godeny, de Vries et al. 1998). To test whether the predicted body TRS3′ is functional, a set of primers was designed with the forward primer targeting the leader sequence and the reverse primer (SHFV-TRS3′-R) targeting a region between the predicted TRS3′ and TRS4′ (Table 3-2 and Figure 3-2A). RT-PCR amplification of RNA extracted from 24 h SHFVic infected MA104 cells using this set of primers would be expected to amplify the leader-body junction of the predicted sg mRNA3′. But because the reverse primer would also anneal to a region downstream of TRS2′, the leader-body junction of sg mRNA2′ would also be amplified (Figure 3-2A). Separation of the amplified leader-body junction DNAs on a gel revealed a predominant ~1072 bp DNA band and a ~604 bp DNA band consistent with the predicted sizes for the leader-body junction regions of the known

sg mRNA2′ and the predicted sg mRNA3′, respectively. Multiple unexpected bands of various sizes were also detected. No bands were detected in the mock samples (Figure 3-2B). Another reverse primer targeting a different region between the predicted TRS3′ and TRS4′ was used to repeat this experiment and the same pattern of multiple bands was observed (data not shown), suggesting different leader-body junction DNAs generated from additional body TRSs were amplified as well. To identify all of the functional body TRSs located near the predicted TRS3′, all the bands near the predicted 604 bp band were excised from the DNA gel and the extracted leader-body junction DNAs were cloned into a TA vector. Forty clones were randomly selected and the insert leader-body junction DNAs were cut out by restriction digestion and separated by gel electrophoresis (Figure 3-2C). Cloned DNAs containing different sized leader-body junction DNAs were sequenced. Alignment of the resulting leader-body junction sequences with both the 5′-leader and the 3′-SHFV genome revealed eight sg mRNAs transcribed from minus strand templates generated from body TRSs located between TRS2′ and TRS4′ including the predicted TRS3′ (Figure 3-2D). Four of these sg mRNAs encoded GP3′, while the other four encoded the in-frame C-terminus of GP2′.

A novel sg mRNA band was detected between sg mRNA5 and sg mRNA6 by Northern blotting (Figure 3-1). To discover all of the body TRSs located between TRS5 and TRS6 that could generate sg RNAs, a set of primers was designed to target the leader region and a region close to TRS6 (Table 3-2 and Figure 3-3A). Separation of the amplified leader-body junction DNAs on a gel indicated the presence of multiple bands near the predicted size (852 bp) of the leader-body junction for sg mRNA5 as well as some faster migrating bands (Figure 3-3B). These bands were excised from the gel and the DNAs were cloned into a TA vector. Individual clones were picked and the inserts were cut out by restriction digestion and analyzed on gels (Figure 3-

3C). Clones containing different sized inserts were sequenced. Alignment of the leader-body junction sequences with both the 5′-leader and the 3′-SHFV genome revealed seven sg mRNAs transcribed from minus strand templates generated from functional body TRSs located between TRS5 and TRS6. All seven sg mRNAs encoded the in-frame C-terminus of GP5 (Figure 3-3D).

The discovery of multiple new body TRSs between the TRS2′ and TRS4′, and also between TRS5 and TRS6 suggested that the two extra bands detected on the Northern blots (Figure 3-1) each represented a group of sg RNAs generated from nearby body TRSs. Multiple bands were also detected near the leader-body junction bands of sg mRNA2′ (Figure 3-2B) and sg mRNA5 (Figure 3-3B), strongly suggesting that additional functional TRSs also exist near TRS2′ and TRS5. To identify additional functional TRSs within the 3′ end of the genome, eight reverse primers were designed to target downstream of each identified body TRS and amplify the leader-body junctions of the sg mRNAs generated from each of these regions (Table 3-2). After RT-PCR amplification and TA-cloning, twenty clones were randomly selected for each of the eight regions and sequenced. Each of the previously published eight body TRSs were detected and a total of 37 additional functional body TRSs were discovered (Table 3-4). Among the 37 new body TRSs, more than half functioned as redundant alternative body TRSs generating templates for sg mRNAs encoding the same structural proteins. The remainder of the new body TRSs identified generated templates for sg mRNAs encoding new ORFs. Interestingly, all of the new ORFs discovered encoded a C-terminal region of a known structural protein (GP2a′-C, GP4′-C, GP2a-C, GP5-C and GP6-C). None of these C-terminal SHFV protein ORFs have been previously reported.

The SHFV sg mRNA2′, sg mRNA2 and sg mRNA5 have been thought to bicistronically express GP2′/E′, E/GP2 and GP5/ORF5a, respectively. Two sg mRNAs generated from

additional body TRSs were identified that encode E′ as their 5′ proximal ORF. In SHFV sg mRNA2, the start codon of the E protein is located upstream of that of GP2. A separate sg mRNA generated from an additional body TRS was identified that encodes GP2 as its 5′ proximal ORF. The data indicated that GP2′ and E′, as well as E and GP2, can be expressed monocistronically from separate sg mRNAs. However, the possibility of bicistronic expression of these proteins was not ruled out. A body TRS that generates a separate sg mRNA encoding ORF5a as its 5′ proximal ORF was not identified after TA-cloning of the amplified leader-body junctions.

3.3.3 NGS analysis of SHFV sg mRNAs in MA104 cells

Although amplification and cloning of leader-body junctions in each region of the SHFV genome provided an effective means of analyzing functional body TRSs in that region and facilitated the discovery of a number of new body TRSs, the depth of this analysis was limited by the total number of clones screened in each region as well as by the relative abundance of each sg mRNA. To increase the depth of the analysis, MA104 cells were infected with SHFVic at an MOI of 1 and at 8 and 18 hpi, total intracellular RNA were extracted with TRI reagent and mRNAs were isolated and subjected to NGS using Illumina Hi-Seq. Three biological repeats for each time point were sequenced and analyzed separately using CLC genomics workbench software. A workflow was designed to first map reads to the SHFV full-length genome reference sequence (NC_003092.2) which separated the virus reads from the host reads. All the virus reads were then mapped to a 15 nt sequence (186 nt-200 nt in the SHFV genome) immediately upstream of the core sequence in the leader TRS (201 nt-209 nt) to collect all the reads that contain the leader sequence. The leader reads were then mapped against a 15 nt sequence (201

nt-215 nt in the SHFV genome) containing the leader TRS core sequence (201 nt-209 nt) and the beginning of ORF1a (210 nt-215 nt) to screen out all the leader reads that also mapped to the ORF1a region. The remaining reads were expected to contain the leader-body junctions of all of the sg mRNAs generated in infected MA104 cells (referred to as sg mRNA reads). The sg mRNA reads were then sequentially mapped to each of the 45 identified leader-body junction sequences (15 nt) and the number of reads mapped to each junction sequence was quantified to indicate the transcription level of the sg mRNA from templates generated from that body TRS. The results obtained from the three biological repeats were similar for each sg mRNA (data not shown). Data from one representative repeat are shown (Table 3-5). Reads were mapped to each of the 45 identified leader-body junction in both the 8 and 18 h infected samples, confirming that each of the identified body TRSs in the SHFV genome was used to generate sg RNAs at early and late times post infection. To calculate the relative abundance of each sg mRNA compared to the others analyzed, the number of reads for each sg mRNA was divided by the total number of reads for all the sg mRNAs analyzed. As expected, the transcription level differed for each sg mRNA. The relative abundance was the highest for sg mRNA7 which is generated from the published TRS7 (~50,000-71,000 mapped reads, ~22% - 25% abundance) and the lowest abundance was for a sg mRNA generated from a new body TRS located near the published TRS5 (3-4 mapped reads, ~0.001% -0.002% abundance) (Table 3-5). The second most abundant sg mRNA was sg mRNA6 generated from the published TRS6. Interestingly, instead of sg mRNA5, sg mRNA2′ and sg mRNA2 generated from the published TRS2′ and TRS2, respectively, located at the 5′ end of the 3′ region were the third and fourth most abundant sg mRNAs (Table 3-5).

In general, for each SHFV structural protein, the sg RNA generated from the published TRS was in higher abundance than the sg RNAs from newly identified alternative TRSs, indicating it was the major sg RNA. Exceptions were found for GP4′, GP4 and GP3′. The relative abundance of one of the newly identified GP4′ sg mRNAs was 4.44% compared to 1.42% for the previously published sg mRNA. Likewise, the relative abundance of a newly identified GP4 sg mRNA was 2.7% compared to 2.0% for that of the previously published sg mRNA. The previous publication likely did not identify the major sg mRNA for GP4 or GP4′ due to limited number of clones tested (Godeny, de Vries et al. 1998). The TRS3′ for sg mRNA3′ was previously predicted based on limited knowledge of SHFV body TRSs (Smith, Wang et al. 1997). Three alternative GP3′ sg mRNAs generated from additional body TRSs had a higher abundance (0.43% to 1.51%) than the predicted GP3′ sg mRNA (0.32%), indicating that the previously predicted sg mRNA3′ was not the major sg mRNA3′.

To compare the differences in viral gene transcription between 8 and 18 hpi, the fold change in transcription level and in relative abundance between the two times tested were calculated for each sg mRNA. The transcription level of each sg mRNA increased slightly from 8 to18 hpi, but the relative abundance of each sg mRNA remained constant at early and late times post infection (Table 3-5). The only exception was a sg mRNA transcribed from the template generated from body TRS 5′-TTCTTCGCC-3′, but there were only 4 mapped reads at 8 h and 3 mapped reads at 18 h for this sg mRNA (Table 3-5).

3.3.4 NGS analysis of SHFV sg mRNAs in macaque MΦs

MΦs and dendritic cells are targeted by SHFV during a natural infection in monkeys. To confirm the usage of all the identified body TRSs in an experimental setting that more closely

resembles a natural infection, peripheral blood mononuclear cells (PBMCs) were isolated from macaque blood by Ficoll density gradient centrifugation and differentiated into MΦs by incubation with human recombinant MΦ colony stimulating factor. The resulting MΦs were infected with SHFVic at an MOI of 1 and mRNAs were isolated from extracted total RNA at 7 and 16 hpi and subjected to NGS using Illumina Mi-Seq. The sequencing data were analyzed using CLC Genomics Workbench software with the same work flow as described above. Although fewer reads were obtained due to the lower capacity of Illumina Mi-Seq, each of the 45 body TRSs identified in MA104 cells was also functional in MΦs at both early and late times post infection with the exception for one (5′-TTCTTCGCC-3′) (Table 3-6). No reads were mapped to that body TRS in the 7 h infected MΦ sample but 2 reads were mapped to it in the 18 h infected MΦ sample. The same body TRS had only 4 mapped reads in the 8 h MA104 sample and 3 mapped reads in the18 h MA104 sample. In general, similar sg mRNA transcription patterns were observed in the infected MΦ and MA104 samples (Tables 3-5 and 3-6). Although the relative abundance of a few sg mRNAs differed slightly between the MΦ and MA104 samples, the relative abundance of each sg mRNA remained constant at early and late times post infection in MΦs as seen in MA104 cells (Tables 3-5 and 3-6). The data indicated that the regulation of SHFV sg mRNA synthesis is not affected by the stage of infection.

3.3.5 Analysis of the expression level of SHFV 3' ORFs

Each identified SHFV sg mRNA is expected to translate its 5′ proximal ORF. For ORFs that are known to be translated from only one identified sg mRNA, the transcription level and relative abundance of the sg mRNA were used to estimate the expression level and relative abundance of that ORF. For ORFs that could be translated from multiple identified sg mRNAs,

the transcription level and relative abundance of each of these sg mRNAs were added. At both early and late times post infection in MA104 cells, the three major structural protein ORFs, ORF5, ORF6 and ORF7, had higher expression levels than the minor structural protein ORFs with ORF7 being the highest (Table 3-7). Because new body TRSs were identified that generate separate sg mRNAs encoding ORF3′ (GP3′), ORF2b′ (E′) and ORF2b (E), the expression level of each of these proteins was calculated assuming only monocistronic expression. Interestingly, at both times tested, the members of the two sets of minor structural proteins differed in their expression patterns. ORF2a′ (GP2′) was expressed at much higher level than ORF2b′ (E′) but ORF2a (GP2) was expressed at much lower level than ORF2b (E). Notably, the body TRSs of ORF2a′ were located upstream of those of ORF2b′ whereas the body TRSs of ORF2a were located downstream of those of ORF2b. The expression level of ORF3′ was much higher than that of ORF3. In contrast, the expression levels of ORF4′ and ORF4 were comparable (Table 3-7). The data indicated that the expression of the members of the two sets of minor structural proteins was uniquely regulated by their body TRSs and location in the genome. Differential expression of the two sets of minor structural proteins provided further support for the previously published hypothesis that the two sets of minor structural proteins were not generated by a simple duplication event (Vatter, Di et al. 2014). Because a sg mRNA was not identified for ORF5a by the amplification and cloning protocol, the expression level of ORF5a was not analyzed. Notably, among the five new ORFs discovered that encode C-terminal in-frame proteins of known structural proteins, ORF2a′-C, ORF5-C and ORF6-C had the highest levels of expression at both times tested with ORF5-C being the highest (~4000-6000 mapped reads, 2%-2.2% relative abundance) (Table 3-7). The expression levels of ORF2a′-C and ORF6-C were higher than those of ORF2a and ORF3 which are thought to encode structural proteins. A similar

expression pattern was observed for the 3′ ORFs in MΦs as in MA104 cells, but the relative abundances of ORF2a′, ORF2b, ORF6 and ORF7 differed slightly between the MΦ and MA104 samples (Table 3-8).

To compare the differences in viral gene expression between 8 and 18 hpi in both cell types, the fold change in expression level and in relative abundance were calculated for each 3′ ORF in MA104 cells and MΦs. Although the expression level of each ORF increased at 18 compared to 8 hpi, the relative abundance of each ORF stayed constant at early and late times post infection in both MA104 cells and MΦs (Tables 3-7 and 3-8). The only exception was the new ORF4′-C, which had the lowest amount of mapped reads.

3.3.6 Many additional body TRSs exist in both the 3' end of the genome and in the ORF1a/1b region

After mapping the sg mRNA reads to all the identified leader-body junction sequences, ~60-70% of the total reads remained unmapped in samples collected at both early and late times post infection from MA104 cells and MΦs. Each of the remaining sg mRNA reads contained a leader-body junction sequence that was not 100% identical to any of the identified ones. This suggested the existence of many additional sg mRNAs. To locate the positions of additional new body TRSs in the SHFV genome, all the remaining sg mRNA reads from the MA104 samples (Figure 3-4) or MΦ samples (Figure 3-5) were mapped to a SHFV genome sequence from which the 5′ leader region had been deleted (201 nt-15717 nt). The majority of the remaining reads mapped to the 3′ end of the genome especially in the region of ORF7 as indicated by the peak in read coverage (Figures 3-4 and 3-5).

In the SHFV sg mRNA5, the distance between the start codons of GP5 and the downstream ORF5a is only 5 nt. A separate sg mRNA generated from this region encoding ORF5a from the 5′ proximal ORF was not identified after TA cloning followed by sequencing 20 clones. However, many of the remaining reads mapped to the region where GP5 and ORF5a overlap. Analysis of the junction sequences showed that some had a fusion of leader and body TRSs located within the 5 nt region between the start codons of GP5 and ORF5a (Figure 3-6). These data suggest the existence of a new functional body TRS (5′-TACTTATGT-3′, 14005 nt-14013 nt in the SHFV genome) that generates the template for a separate sg mRNA expressing ORF5a from its 5′ proximal ORF.

Unexpectedly, some of the remaining reads mapped to different locations in the ORF1a/1b region but these were in low abundance compared to the reads from the 3′ end of the genome (Figures 3-4 and 3-5). Compared to the samples collected at early times, the samples collected at later times contained more reads that mapped to the ORF1a/1b region (Figures 3-4B and 3-5B). The data showed that discontinuous RNA synthesis may also occur within the ORF1a/1b region of the genome generating long sg mRNAs that could potentially encode new proteins. A previous study on LDV identified a long sg mRNA generated from a body TRS located in ORF1b that encodes the C-terminal 200 amino acids of ORF1b (Chen, Kuo et al. 1993). A detailed analysis of all the reads mapped to the ORF1a/1b region in SHFV is in process to identify the function of body TRSs in this region and also the coding capacity of the resulting long sg mRNAs. Investigation of the reads that mapped to the 5′ end of ORF1a (Figures 3-4 and 3-5) showed they do not represent new sg mRNAs but instead are full length SHFV genome RNAs that contain mutations or sequencing errors preventing them from being mapped to the 5′-leader/ORF1a region under the stringent setting used.

3.3.7 Knocking out the expression of ORF5-C or ORF6-C reduced SHFV yield in MA104 cells

Five new ORFs, ORF2a′-C, ORF4′-C, ORF2a-C, ORF5C and ORF6C, were found to be encoded by additional sg mRNAs discovered. Each of these new ORFs encodes an in-frame C-terminal peptide from a known structural protein. The TMpred program was used to predict transmembrane domains in each of the C-terminal peptides. ORF2a′-C, ORF5C and ORF6C were not predicted to contain a transmembrane domain. The N-terminus of each of these peptides was located just after the last transmembrane domain in the respective full length protein, suggesting that these peptides would be soluble cytoplasmic proteins. ORF4′-C and ORF2a-C each contain one transmembrane domain and therefore might be located in the viral envelop (Table 3-9).

To analyze the functional relevance of the C-terminal in-frame peptides, the start codon for each peptide was mutated in the appropriate SHFV infectious clone fragment. In each case, the substitution preserved the amino acid in the overlapping ORF (Table 3-3). The mutant SHFV fragment was then ligated to the other four fragments to construct the full-length mutant SHFV clones, SHFVic-ΔORF2a′-C, SHFVic-ΔORF4′-C, SHFVic-ΔORF2a-C, SHFVic-ΔORF5-C and SHFVic-ΔORF6-C. Mutant SHFV genome RNA was then transcribed from each linearized mutant full-length clone and transfected into subconfluent MA104 monolayers. RNA from the wild type SHFV clone was used as a positive control. At 120 h after transfection, cytopathic effect was observed in wells transfected with the wild type SHFV RNA, as well as in wells transfected with each mutant SHFV RNA (data not shown). Culture fluid (500 μL) was collected at 120 hpi and passaged onto fresh MA104 cells. At 24 h after passage, cytopathic effect was

again observed for both the passage 1 (P1) wild type and each of the P1 mutant SHFV viruses. Western blotting analysis of cell lysates harvested at 24 hpi showed that each of the mutant SHFV viruses produced the viral non-structural protein nsp1β (Figure 3-7A). These data indicated that the expression of the individual C-terminal in-frame proteins was not required for the production of infectious virus or for virus entry into cells. To determine whether the loss of individual C-terminal peptides affects virus replication, MA104 cells were infected with one of the mutant P1 viruses at an MOI of 0.5 and culture fluids harvested at different times after infection were titered by plaque assay. MA104 cells infected by wild type SHFV was used as a positive control. Knocking out the expression of ORF5-C or ORF6-C decreased the virus yield starting at 24 hpi (Figure 3-7B), whereas knocking out the expression of ORF2a′-C , ORF4′-C or ORF2a-C had no effect on virus yield (Figure 3-7C). Also, the plaques formed in MA104 cells by SHFVic-ΔORF5-C and SHFVic-ΔORF6-C were much smaller than those formed by wild type SHFV (Figure 3-7D). These data suggest that ORF5-C and ORF6-C are functionally important during the viral replication cycle.

3.4 Discussion

SHFV was first thought to encode six structural protein ORFs at the 3′ end of the genome like all the other arteriviruses. A previous study using ^{32}P-labeled RNA probes detected six strong sg mRNA bands at 7 hpi in extracts from MA104 cells infected with SHFV LVR strain in a Northern blotting assay (Zeng, Godeny et al. 1995). Subsequent sequencing of the 3' region of SHFV genome revealed that nine ORFs, including a second set of GP2′, GP3′ and GP4′, were encoded in the genome (Smith, Wang et al. 1997; Godeny, de Vries et al. 1998). In the present study, MA104 cells were infected with SHFVic and at different times after infection, total

intracellular RNA was extracted and subjected to Northern blotting assay with multiple DIG-labeled long probes. A total of ten sg mRNA bands were detected, including the six bands detected previously and four additional weaker bands (Figure 3-1). Based on the estimated size of each band, nine of the bands were expected to represent the corresponding SHFV sg mRNAs for the nine ORFs encoded in the genome (sg mRNA2′ - sg mRNA7) and one band represented an unexpected novel sg mRNA.

According to the discontinuous RNA synthesis model, each sg mRNA contains a unique leader-body junction sequence generated from the base-paring between the body and the leader TRSs. A previous study identified the leader-body junction sequences for eight predicted SHFV sg mRNAs but not one for the sg mRNA3′ (Godeny, de Vries et al. 1998). Although a body TRS was predicted for the sg mRNA3′ (Smith, Wang et al. 1997), when these authors used a set of sg mRNA3′-specific primers, the only RT-PCR product detected on the gel corresponded to the size of the leader-body junction sequence from the predicted sg mRNA2′ rather than from the predicted sg mRNA3′ (Godeny, de Vries et al. 1998). In the present study, amplification of the leader-body junction sequence of the predicted sg mRNA3′ using two different sets of primers generated RT-PCR products with the expected sizes for both the sg mRNA2′ and sg mRNA3′ junction sequences when detected by long exposure under UV light (Figure 3-2B). TA-cloning and sequencing of the RT-PCR products confirmed the existence of the previously predicted body TRS for sg mRNA3′ (Table 3-4). The amount of the amplified leader-body junction sequence for the predicted sg mRNA3′ was much less than that for the known sg mRNA2′ (Figure 3-2B), suggesting a much lower abundance of sg mRNA3′ generated during infection, which could explain why the leader-body junction sequence of the predicted sg mRNA3′ was not

previously identified. Inefficient primer binding to a structured viral RNA region could also explain the reduced sg mRNA3′ leader-body junction amplification.

TA-cloning and sequencing of the amplified leader-body junction sequences from all the previously published or predicted sg mRNAs identified a total of 45 functional body TRSs including the 8 previously identified ones. Analysis of the amplified leader-body junction sequences from the predicted sg mRNA3′ not only confirmed the previously predicted body TRS3′, but also identified five additional body TRSs that generated templates for alternative sg mRNAs each encoding GP3′ from the 5′ proximal ORF. Additional body TRSs generating templates for alternative sg mRNAs were also identified for GP2′, GP4′, GP3, GP4 and GP5. These additional body TRSs were not identified in the previous study (Godeny, de Vries et al. 1998) probably because of the limited number of TA clones containing the leader-body junction inserts analyzed. In the present study, 20-40 TA clones were analyzed for each amplified leader-body junction region so that leader-body junction sequences from less abundant sg mRNAs were identified. The number of alternative sg mRNAs identified for each structural protein was determined by the total number of clones screened and also by the relative abundance of each sg mRNA. No alternative sg mRNA was identified for GP7 after TA cloning and sequencing. Because of the high abundance of the published sg mRNA7, all of the clones screened contained the same leader-body junction sequence. One or two additional body TRSs generating templates for alternative sg mRNAs encoding the structural proteins, GP3, GP4 and GP5, were previously identified by amplification and cloning for both PRRSV and EAV (Pasternak, Gultyaev et al. 2000; Lin, Chang et al. 2002). These additional body TRSs may function as back-ups when the major body TRSs are inactivated. Supportive evidence for this hypothesis was obtained for EAV (Pasternak, Gultyaev et al. 2000). Although the alternative sg mRNAs generated from the

alternative body TRSs are less abundant, they are produced in sufficient amount to generate infectious progeny virus when the major body TRSs were inactivated, but the virus yield was reduced.

In addition to generating alternative sg mRNAs encoding the known structural proteins, additional body TRSs were also identified for generating new sg mRNAs encoding the in-frame, C-terminal peptides of five structural proteins: ORF2a′-C, ORF4′-C, ORF2-C, ORF5-C and ORF6-C. The sg mRNAs encoding the ORF5-C corresponded to the unexpected novel sg mRNA band detected in Northern blotting assays (Figure 3-1). These C-terminal peptides of structural proteins have not been previously predicted or reported in arteriviruses. However, a study of the coronavirus SARS-CoV identified a new sg mRNA that encodes a C-terminal peptide of the spike protein (Hussain, Pan et al. 2005). Also, although never discussed, an additional sg mRNA band migrating between the sg mRNA5 and sg mRNA6 bands was present in multiple previously published EAV Northern blots, suggesting that EAV may also produce one additional sg mRNA (van Marle, Dobbe et al. 1999; Pasternak, Gultyaev et al. 2000; Pasternak, van den Born et al. 2001). The functions of the C-terminus of GP5 and GP6 are not known. Neither ORF5-C nor ORF6-C are predicted to contain a transmembrane domain and so are expected to be soluble proteins. Knocking out the expression of each of the SHFV C-terminal peptides did not affect the production of the progeny virus but the virus yield and the plaque size were decreased when the expression of ORF5-C or ORF6-C was inactivated (Figure 3-7). However, because the C-terminal peptide was in-frame with the full length structural protein, single nucleotide mutation of the start codon in the peptide would mutate a Met in the full-length protein. Although a conservative amino acid substitution was made for both ORF5-C and ORF6-C (Met to Leu), the possibility that these mutations affected the function of the full-length structural protein cannot

be ruled out. The NGS analyses estimated the relative abundance of these peptides to be fairly high, ORF5-C (~2%) and ORF6-C (~1%), during the infection in MA104 cells and macaque MΦs. Verification of the production of these peptides in infected cells is underway using peptide-specific antibodies. The stability and locations of these peptides will be investigated as well. Although knocking out GP2′-C, GP4′-C or GP2-C individually had no effect on the virus yield in MA104 cell, their functional relevance in infected animals were not ruled out and need to be investigated.

NGS is a powerful tool for obtaining a more comprehensive and in-depth analysis of viral sg mRNAs and the present study is the first to apply NGS for the analysis of arterivirus sg mRNA production in infected cells. The expression of all of the 45 sg mRNAs identified by TA cloning and sequencing were detected by NGS analysis in MA104 cells and macaque MΦs. A previous study showed that generally, the abundance of a subgenomic mRNA correlated with the stability of the corresponding leader-body TRS duplex (Pasternak, van den Born et al. 2003). Also, shorter sg mRNAs are usually more abundant then the larger ones, probably due to their TRSs being located closer to the 3′ end of the genome where discontinuous RNA synthesis initiates and also to their shorter minus-strand templates allowing more rapid sg mRNA transcription. The leader TRS core sequence has a higher homology to the previously published TRS5 core sequence than to the previously published TRS2′ and TRS2 core sequences. TRS5 is also located closer to the 3′ end of the genome. However, analysis of SHFV sg mRNA abundance showed that both sg mRNA2′ and sg mRNA2 are produced in higher abundance than the sg mRNA5 in MA104 cells. Previous studies have also shown that if multiple body TRSs are present in a local region, a downstream TRS may suppress the activity of an upstream TRS especially when they are close to each other (Joo and Makino 1995; Pasternak, Spaan et al.

2004). In SHFV genome, an alternative TRS2′ was identified 80 nt upstream of the previously published TRS2′ and the alternative sg mRNA2′ produced from this TRS was much less abundant (2.76%) than the previously identified sg mRNA2′ (9.49%). This observation is consistent with the previous finding. In contrast, the previously published TRS2 is located 41 nt upstream of an alternative TRS2 but the resulting sg mRNA2 was much more abundant (7.21%) than the alternative sg mRNA2 (0.63%), which is contradictory to the prediction. Together the data support a previously published hypothesis that additional factors, such as TRS flanking sequences and local RNA secondary structure, are involved in regulating sg mRNA synthesis (den Boon, Kleijnen et al. 1996; Pasternak, Gultyaev et al. 2000). Although individual SHFV sg mRNAs showed diverse transcription levels and the overall transcription level for all the sg mRNAs increased at later times after infection, the relative abundance of each sg mRNA remained constant at early and late times after infection in both MA104 cells and macaque MΦs, suggesting that the synthesis of each SHFV sg mRNA was tightly regulated and not affected by the stage of the infection cycle. However, there were some differences between MA104 cells and macaque MΦs in the relative abundance of a few sg mRNAs, namely sg mRNA2′, sg mRNA2 and sg mRNA6. This cell type specific difference suggested that host factors may also play a role in the regulation of sg mRNA abundance.

A common method for calculating cellular gene expression levels from NGS data is counting the number of reads mapped to each particular gene region. This method cannot be applied to SHFV gene expression analysis because SHFV sg mRNAs form a nested set of transcripts. Each SHFV sg mRNA contains all the downstream gene ORFs but only the 5′ proximal ORF is typically expressed (Snijder, Kikkert et al. 2013). Sg mRNA2′, sg mRNA2 and sg mRNA5 were previously thought to be bicistronic, but identification of separate sg mRNA(s)

in the present study that express E′, GP3′ and GP2 from their 5′ proximal ORF indicates the possibility of monocistronic expression of each of these proteins. Therefore, in the present study, the expression level of each SHFV 3′ gene was estimated based on the total of the transcription levels of all of the alternative sg mRNAs encoding that gene as the 5′ proximal ORF. Because of that, although the major sg mRNA5 (6.90%) is less abundant than the major sg mRNA2′ (9.49%) in the infected MA104 cells, there are 7 alternative sg mRNA5s but only 1 alternative sg mRNA2′, thereby the expression of GP5 from all of its sg mRNAs (13.91%) is estimated to be at a higher level than GP2′ (12.25%). Because not all of the alternative TRSs are yet known, the method used for analyzing SHFV gene expression based on NGS data gives a low estimation of viral transcription that will need to be recalculated when additional alternative sg mRNAs are identified in the future. Also, the method used was based on the assumption that only monocistronic expression of each 3′ gene occurs and the existence of some bicistronic expression cannot be ruled out.

No additional sg mRNA with a different leader-body junction sequence was identified in the region of ORF7 using a TA cloning protocol. NGS analysis revealed that sg mRNA7 is the most abundant sg mRNA generated during infection. A low probability of finding additional sg mRNAs produced from that region by a clone screening protocol would be expected due to the high abundance of the major sg mRNA. However, the depth of NGS expands the likelihood of finding sg mRNAs present in low abundance. After mapping all the remaining reads that contain leader-body junction sequences that differ from the identified ones to the SHFV genome, the majority of these reads were found to map to the ORF7 region (Figures 3-4 and 3-5), indicating the existence of many additional sg mRNAs generated from this region. The NGS analyses also led to the discovery of a separate sg mRNA encoding ORF5a from its 5′ proximal ORF (Figure

3-6). This sg mRNA was not found by TA cloning followed by sequencing of 20 clones. The possibility that discontinuous RNA synthesis could also occur within the ORF1a/1b region, which was not expected, was also demonstrated by the NGS analyses (Figure 3-4 and 3-5). These findings strongly suggest that NGS is the best current tool for discovering additional sg mRNAs and thereby new body TRSs in the genomes of arteriviruses and coronaviruses. Notably, the existence of an additional sg mRNA with a unique leader-body junction sequence does not necessarily indicate the use of a new body TRS in the genome because sg mRNAs generated from the same body TRSs could have heterogeneous leader-body junction sequences due to mutations caused by polymerase mistakes occurring in either the leader or body TRS regions, or more likely due to different positions within the TRSs where the fusion of the leader and body TRS sequences occurred (Chen, Kuo et al. 1993; Meulenberg, de Meijer et al. 1993). Both the complexity of leader-body junction heterogeneity and the massive amount of reads generated by NGS indicate that the development of a systematic, programmable method is needed to facilitate efficient and comprehensive discovery of all of the sg mRNAs generated and the functional body TRSs present in arterivirus and coronavirus genomes.

Table 3.1. Primers used to generate probes specific for each SHFV sg mRNA by targeting the regions between each set of adjacent predicted body TRSs

Name	Sequence (5′ to 3′)[a]	Targeted region	Probe length (nt position[b])
sg mRNA2′-F	TTGTATCTGCCTACAACAATTGGGTTCGGCGCG	TRS2′ to TRS4′	506nt (11155-11660)
sg mRNA2′-R	cggcggtaatacgactcactatagggTACATACAACGGCTGTGAGG		
sg mRNA4′-F	CCGCAACTACACTCGTTCCCTAACAATCAAGC	TRS4′ to TRS2	520nt (11902-12421)
sg mRNA4′-R	cggcggtaatacgactcactatagggACGATATGGCCATCCTGTTGC		
sg mRNA2-F	CCATCATTTGCGCATGCTTGGCGATCTTTAGAG	TRS2 to TRS3	529nt (12509-13037)
sg mRNA2-R	cggcggataatacgactcactatagggTGCGATGTCTTCGCAGTGG		
sg mRNA3-F	GATGTCCGTGGTCCTCAAGTCACACAGGCCTGC	TRS3 to TRS4	168nt (13092-13259)
sg mRNA3-R	cggcggtaatacgactcactatagggTTGCCACGGTACAGGCAAGC		
sg mRNA4-F	TTCCCAGTTCTTTTCAACGCCACTTCTGTCCAG	TRS4 to TRS5	515nt (13401-13915)
sg mRNA4-R	cggcggtaatacgactcactatagggTGATAGGCTGAAATATACCTG		
sg mRNA5-F	TCTATTACATTCAGCAGCACCGGCGCATCC	TRS5 to TRS6	541nt (14098-14638)
sg mRNA5-R	cggcggataatacgactcactatagggCCATACAATTTACCACTAAC		
sg mRNA6-F	ATTTCCGACCCACAGGGACTGCGGGTTGGACCTCATAAG	TRS6 to TRS7	508nt (14719-15226)
sg mRNA6-R	cggcggataatacgactcactatagggCATGAGACTACCATTGACTGC		
sg mRNA7-F	GCTGGCAAACCAAAAACAAATAACAAGGG	TRS7 to 3′-UTR	392nt (15298-15689)
sg mRNA7-R	cggcggtaatacgactcactatagggTTAGTCCTTAGCCTAGGGAAG		
5′-leader-F	AATAAAAGTGTGAAGCTCCCTGTGCTTTCATGCCAGG	5′-leader	194nt (7-200)
5′-leader-R	cggtagtaatacgactcactatagggTCTGCAAATCCCAAGCCAC		

[a]T7 promoter sequence is shown in lowercase.
[b]Nt numbering according to the SHFV infectious clone genome RNA.

Table 3.2. Primers used to amplify the leader-body junctions of individual sg mRNAs generated from known/predicted body TRSs in the 3′ end of the SHFV genome

Name	Sequence (5′ to 3′)
SHFV-5'-Leader -F	TAGCCCGGATTGGATAAGC
SHFV-TRS2′-R	AGTACCTGCGTGTACTGTGG
SHFV-TRS3′-R	CCAGATGCTAAGATCTGCC
SHFV-TRS4′-R	CAGTCAGCAAAGTCAAGAGC
SHFV-TRS2-R	TTGACATGAAGGGTGGCATGG
SHFV-TRS3-R	AGTGAGTGCAGAAGAAGAGCC
SHFV-TRS4-R	TTATGATGGAAGCCCACCG
SHFV-TRS5-R	GAAAGGAGGCAGTTGTAGC
SHFV-TRS5-C-R	CCTCGGTCGTCAATGATG
SHFV-TRS6-R	CATCCGCACACGTAGAATGG
SHFV-TRS7-R	ACCAGTTAGTCCTTAGCC

Table 3.3. Primers used to knock out the expression of the new C-terminal in-frame ORFs of known structural proteins in the SHFV genome.

Mutant ORFs	Primer sequences (5′ to 3′)[a]	Effects of the mutations on the overlapping ORFs
ΔORF2a′ - C	F: gcgcctcaacgcaCgagtgcacctcg	in GP2′-C: ATG to ACG (Start Codon knock out)
	R: cgaggtgcactcGtgcgttgaggcgc	in GP2′: ATG to ACG (Methionine to Threonine)
		in GP3′: CAT to CAC (Histidine to Histidine)
ΔORF4′ - C	F: gagacaattacaggcaacCtgactggtatcaaagaagc	in GP4′-C: ATG to CTG (Start Codon knock out)
	R: gcttctttgataccagtcaGgttgcctgtaattgtctc	in GP4′: ATG to CTG (Methionine to Leucine)
ΔORF2a - C	F: gtcccgaggcgcttacaaaaaCgcttttgcgcc	in GP2-C: ATG to ACG (Start Codon knock out)
	R: ggcgcaaaagcGtttttgtaagcgcctcgggac	in GP2: ATG to ACG (Methionine to Threonine)
		in E: AAT to AAC (Asparagine to Asparagine)
ΔORF5 - C	F: caagaagttagtggtaaattgtCtggccctccgc	in GP5-C: ATG to CTG (Start Codon knock out)
	R: gcggagggccaGacaatttaccactaacttcttg	in GP5: ATG to CTG (Methionine to Leucine)
ΔORF6 - C	F: ccattctacgtgtgcggCtgtgttggctcggc	in GP6-C: ATG to CTG (Start Codon knock out)
	R: gccgagccaacacaGccgcacacgtagaatgg	in GP6: ATG to CTG (Methionine to Leucine)

[a] Upper case letters indicate the mutated nucleotide in the primer sequence

Table 3.4. Body TRSs identified in the SHFV genome by amplification, cloning and sequencing of known or predicted leader-body junctions.

Body TRSs (5′ to 3′)	New or published	First ORF encoded (Gene)	Body TRSs (5′ to 3′)	New or published	First ORF encoded (Gene)
ttctgaacc	new	ORF2a′ (GP2′)	ttctaaata	new	ORF3 (GP3)
tctttaact	TRS2′ [a]		tcactaacc	TRS3 [a]	
ggtttaatc	new	ORF2b′ (E′)	acttcaaca	new	ORF4 (GP4)
tccttaaac	new		atccaaacc	new	
ttatttccc	new	ORF3′(GP3′)	tcattgacc	TRS4 [a]	
gggctaaac	new		acaacaacc	new	ORF5 (GP5)
cttaaaacc	TRS3′ [b]		cccataacc	new	
gatttcaac	new		ttcttcgcc	new	
accttcacc	new		gtgataatc	new	
ttctgtacc	new		ttgttatca	new	
cccttaact	new	ORF2a′-C (New)	atcataacc	new	
aactttact	new		tccttaact	TRS5 [a]	
tcgcaaacc	new		taattatgt	new	
caatcaacc	new		aaaataaca	new	ORF5-C (New)
gcctcaacg	new		agtcaaacc	new	
gccttaacc	new	ORF4′(GP4′)	tgttcaacg	new	
cctttcacc	TRS4′ [a]		ttcttcaca	new	
ggcgaaacg	new	ORF4′-C (New)	tcaataaca	new	
ctattaacc	TRS2 [a]	ORF2b (E)	ttactaacc	new	
tgggcaacc	new		tgtttcacc	new	
tactcaccc	new	ORF2a (GP2)	ttgtcaacc	TRS6 [a]	ORF6 (GP6)
caattaaat	new	ORF2a-C (New)	tacttgacc	new	ORF6-C (New)
			ttgttaacc	TRS7 [a]	ORF7 (GP7)

[a] Previously Published SHFV body TRS
[b] Previously Predicted SHFV body TRS

Table 3.5. NGS analysis of the transcription level and relative abundance of sg mRNAs generated from each identified body TRSs at 8 and 18 hpi in MA104 cells.

First ORF encoded (gene)	New or published	Body TRSs (5′ to 3′)	8 hpi		18 hpi		Fold change	
			Transcription level	Relative abundance	Transcription level	Relative abundance	Transcription level	Relative abundance
ORF2a′ (GP2′)	new	ttctgaacc	6151	2.76%	7458	2.61%	1.21	0.95
	TRS2′ [a]	tctttaact	21188	9.49%	24863	8.69%	1.17	0.91
ORF2b′ (E′)	new	ggtttaatc	2491	1.12%	3326	1.16%	1.34	1.04
	new	tccttaaac	3004	1.35%	3600	1.26%	1.20	0.93
ORF3′ (GP3′)	new	ttatttccc	3371	1.51%	3965	1.39%	1.18	0.92
	new	gggctaaac	187	0.08%	198	0.07%	1.06	0.83
	TRS3′ [b]	cttaaaacc	703	0.32%	935	0.33%	1.33	1.04
	new	gatttcaac	53	0.02%	74	0.03%	1.40	1.09
	new	accttcacc	964	0.43%	1118	0.39%	1.16	0.90
	new	ttctgtacc	1610	0.72%	2322	0.81%	1.44	1.12
ORF2a′-C (New)	new	cccttaact	1435	0.64%	1884	0.66%	1.31	1.02
	new	aactttact	33	0.01%	59	0.02%	1.79	1.39
	new	tcgcaaacc	267	0.12%	299	0.10%	1.12	0.87
	new	caatcaacc	81	0.04%	121	0.04%	1.49	1.16
	new	gcctcaacg	98	0.04%	125	0.04%	1.28	0.99
ORF4′(GP4′)	new	gccttaacc	9905	4.44%	11342	3.96%	1.15	0.89
	TRS4′ [a]	cctttcacc	3171	1.42%	3516	1.23%	1.11	0.86
ORF4′-C (New)	new	ggcgaaacg	16	0.01%	40	0.01%	2.50	1.95
ORF2b (E)	TRS2 [a]	ctattaacc	16083	7.21%	20073	7.01%	1.25	0.97
	new	tgggcaacc	1416	0.63%	1919	0.67%	1.36	1.06
ORF2a (GP2)	new	tactcaccc	327	0.15%	449	0.16%	1.37	1.07
ORF2a-C (New)	new	caattaaat	119	0.05%	120	0.04%	1.01	0.79
ORF3 (GP3)	new	ttctaaata	16	0.01%	33	0.01%	2.06	1.61
	TRS3 [a]	tcactaacc	859	0.38%	1474	0.51%	1.72	1.34
ORF4 (GP4)	new	acttcaaca	6029	2.70%	9112	3.18%	1.51	1.18
	new	atccaaacc	2533	1.14%	3520	1.23%	1.39	1.08
	TRS4 [a]	tcattgacc	4470	2.00%	6251	2.18%	1.40	1.09

ORF5 (GP5)	new	acaacaacc	154	0.07%	231	0.08%	1.50	1.17
	new	cccataacc	6089	2.73%	6756	2.36%	1.11	0.86
	new	ttcttcgcc	4	0.00%	3	0.00%	0.75	0.58
	new	gtgataatc	75	0.03%	130	0.05%	1.73	1.35
	new	ttgttatca	103	0.05%	178	0.06%	1.73	1.35
	new	atcataacc	9188	4.12%	10882	3.80%	1.18	0.92
	TRS5 [a]	tccttaact	15407	6.90%	17610	6.15%	1.14	0.89
	new	taattatgt	21	0.01%	31	0.01%	1.48	1.15
ORF5-C (New)	new	aaaataaca	1465	0.66%	1976	0.69%	1.35	1.05
	new	agtcaaacc	1322	0.59%	1729	0.60%	1.31	1.02
	new	tgttcaacg	226	0.10%	295	0.10%	1.31	1.02
	new	ttcttcaca	694	0.31%	971	0.34%	1.40	1.09
	new	tcaataaca	245	0.11%	424	0.15%	1.73	1.35
	new	ttactaacc	291	0.13%	352	0.12%	1.21	0.94
	new	tgtttcacc	271	0.12%	439	0.15%	1.62	1.26
ORF6 (GP6)	TRS6 [a]	ttgtcaacc	48523	21.74%	61906	21.63%	1.28	0.99
ORF6-C (New)	new	tacttgacc	1951	0.87%	3289	1.15%	1.69	1.31
ORF7 (GP7)	TRS7 [a]	ttgttaacc	50561	22.66%	70872	24.76%	1.40	1.09

[a] Previously published SHFV body TRS
[b] Previously predicted SHFV body TRS

Table 3.6. NGS analysis of the transcription level and relative abundance of sg mRNAs generated from each identified body TRSs at 7 and 16 hpi in macaque MΦs.

First ORF encoded (gene)	New or published	Body TRSs (5′ to 3′)	7 hpi		16 hpi		Fold change	
			Transcription level	Relative abundance	Transcription level	Relative abundance	Transcription level	Relative abundance
ORF2a′ (GP2′)	new	ttctgaacc	1298	2.00%	1283	2.01%	0.99	1.00
	TRS2′ [a]	tctttaact	3438	5.30%	3403	5.33%	0.99	1.01
ORF2b′ (E′)	new	ggtttaatc	599	0.92%	476	0.75%	0.79	0.81
	new	tccttaaac	611	0.94%	587	0.92%	0.96	0.98
ORF3′ (GP3′)	new	ttatttccc	1176	1.81%	1021	1.60%	0.87	0.88
	new	gggctaaac	62	0.10%	92	0.14%	1.48	1.51
	TRS3′ [b]	cttaaaacc	353	0.54%	436	0.68%	1.24	1.25
	new	gatttcaac	6	0.01%	15	0.02%	2.50	2.54
	new	accttcacc	189	0.29%	163	0.26%	0.86	0.88
	new	ttctgtacc	259	0.40%	222	0.35%	0.86	0.87
ORF2a′-C (New)	new	cccttaact	346	0.53%	429	0.67%	1.24	1.26
	new	aactttact	10	0.02%	22	0.03%	2.20	2.23
	new	tcgcaaacc	46	0.07%	54	0.08%	1.17	1.19
	new	caatcaacc	44	0.07%	29	0.05%	0.66	0.67
	new	gcctcaacg	23	0.04%	21	0.03%	0.91	0.93
ORF4′(GP4′)	new	gccttaacc	2032	3.13%	1906	2.99%	0.94	0.95
	TRS4′ [a]	cctttcacc	818	1.26%	707	1.11%	0.86	0.88
ORF4′-C (New)	new	ggcgaaacg	8	0.01%	12	0.02%	1.50	1.52
ORF2b (E)	TRS2 [a]	ctattaacc	6562	10.12%	7333	11.49%	1.12	1.14
	new	tgggcaacc	641	0.99%	637	1.00%	0.99	1.01
ORF2a (GP2)	new	tactcaccc	65	0.10%	69	0.11%	1.06	1.08
ORF2a-C (New)	new	caattaaat	36	0.06%	25	0.04%	0.69	0.71
ORF3 (GP3)	new	ttctaaata	8	0.01%	5	0.01%	0.63	0.63
	TRS3 [a]	tcactaacc	437	0.67%	405	0.63%	0.93	0.94
ORF4 (GP4)	new	acttcaaca	1392	2.15%	1409	2.21%	1.01	1.03
	new	atccaaacc	669	1.03%	714	1.12%	1.07	1.08
	TRS4 [a]	tcattgacc	1033	1.59%	1008	1.58%	0.98	0.99

ORF5 (GP5)	new	acaacaacc	34	0.05%	26	0.04%	0.76	0.78
	new	cccataacc	1499	2.31%	1409	2.21%	0.94	0.95
	new	ttcttcgcc	0	0.00%	2	0.00%	N/A	N/A
	new	gtgataatc	48	0.07%	68	0.11%	1.42	1.44
	new	ttgttatca	57	0.09%	53	0.08%	0.93	0.94
	new	atcataacc	3269	5.04%	3330	5.22%	1.02	1.03
	TRS5 [a]	tccttaact	4403	6.79%	4616	7.23%	1.05	1.06
	new	taattatgt	11	0.02%	8	0.01%	0.73	0.74
ORF5-C (New)	new	aaaataaca	711	1.10%	485	0.76%	0.68	0.69
	new	agtcaaacc	281	0.43%	182	0.29%	0.65	0.66
	new	tgttcaacg	73	0.11%	53	0.08%	0.73	0.74
	new	ttcttcaca	142	0.22%	109	0.17%	0.77	0.78
	new	tcaataaca	105	0.16%	107	0.17%	1.02	1.04
	new	ttactaacc	64	0.10%	49	0.08%	0.77	0.78
	new	tgtttcacc	75	0.12%	86	0.13%	1.15	1.16
ORF6 (GP6)	TRS6 [a]	ttgtcaacc	17989	27.74%	19779	30.99%	1.10	1.12
ORF6-C (New)	new	tacttgacc	533	0.82%	628	0.98%	1.18	1.20
ORF7 (GP7)	TRS7 [a]	ttgttaacc	13382	20.64%	10353	16.22%	0.77	0.79

[a] Previously published SHFV body TRS
[b] Previously predicted SHFV body TRS

Table 3.7. NGS analysis of the expression level and relative abundance of SHFV 3′ ORFs encoded by sg mRNAs generated from identified body TRSs at 8 and 18 hpi in MA104 cells.

Body TRSs (5′ to 3′)	New or published	First ORF encoded (gene)	8 hpi		18 hpi		Fold change	
			Expression level	Relative abundance	Expression level	Relative abundance	Expression level	Relative abundance
ttctgaacc	new	ORF2a′ (GP2′)	27339	12.25%	32321	11.29%	1.18	0.92
tctttaact	TRS2′ [a]							
ggtttaatc	new	ORF2b′ (E′)	5495	2.46%	6926	2.42%	1.26	0.98
tccttaaac	new							
ttatttccc	new	ORF3′ (GP3′)	6888	3.09%	8612	3.01%	1.25	0.97
gggctaaac	new							
cttaaaacc	TRS3′ [b]							
gatttcaac	new							
accttcacc	new							
ttctgtacc	new							
cccttaact	new	ORF2a′-C (New)	1914	0.86%	2488	0.87%	1.30	1.01
aactttact	new							
tcgcaaacc	new							
caatcaacc	new							
gcctcaacg	new							
gccttaacc	new	ORF4′(GP4′)	13076	5.86%	14858	5.19%	1.14	0.89
cctttcacc	TRS4′ [a]							
ggcgaaacg	new	ORF4′-C (New)	16	0.01%	40	0.01%	2.50	1.95
ctattaacc	TRS2 [a]	ORF2b (E)	17499	7.84%	21992	7.68%	1.26	0.98
tgggcaacc	new							
tactcaccc	new	ORF2a (GP2)	327	0.15%	449	0.16%	1.37	1.07
caattaaat	new	ORF2a-C (New)	119	0.05%	120	0.04%	1.01	0.79
ttctaaata	new	ORF3 (GP3)	875	0.39%	1507	0.53%	1.72	1.34
tcactaacc	TRS3 [a]							
acttcaaca	new	ORF4 (GP4)	13032	5.84%	18883	6.60%	1.45	1.13
atccaaacc	new							
tcattgacc	TRS4 [a]							

acaacaacc	new	ORF5 (GP5)	31041	13.91%	35821	12.51%	1.15	0.90
cccataacc	new							
ttcttcgcc	new							
gtgataatc	new							
ttgttatca	new							
atcataacc	new							
tccttaact	TRS5 [a]							
taattatgt	new							
aaaataaca	new	ORF5-C (New)	4514	2.02%	6186	2.16%	1.37	1.07
agtcaaacc	new							
tgttcaacg	new							
ttcttcaca	new							
tcaataaca	new							
ttactaacc	new							
tgtttcacc	new							
ttgtcaacc	TRS6 [a]	ORF6 (GP6)	48523	21.74%	61906	21.63%	1.28	0.99
tacttgacc	new	ORF6-C (New)	1951	0.87%	3289	1.15%	1.69	1.31
ttgttaacc	TRS7 [a]	ORF7 (GP7)	50561	22.66%	70872	24.76%	1.40	1.09

[a] Previously published SHFV body TRS
[b] Previously predicted SHFV body TRS

Table 3.8. NGS analysis of the expression level and relative abundance of SHFV 3′ ORFs encoded by sg mRNAs generated from identified body TRSs at 7 and 16 hpi in macaque M Φs.

Body TRSs (5′ to 3′)	New or published	First ORF encoded (gene)	7 hpi		16 hpi		Fold change	
			Expression level	Relative abundance	Expression level	Relative abundance	Expression level	Relative abundance
ttctgaacc	new	ORF2a′ (GP2′)	4736	7.30%	4686	7.34%	0.99	1.01
tctttaact	TRS2′ [a]							
ggtttaatc	new	ORF2b′ (E′)	1210	1.87%	1063	1.67%	0.88	0.89
tccttaaac	new							
ttatttccc	new	ORF3′ (GP3′)	2045	3.15%	1949	3.05%	0.95	0.97
gggctaaac	new							
cttaaaacc	TRS3′ [b]							
gatttcaac	new							
accttcacc	new							
ttctgtacc	new							
cccttaact	new	ORF2a′-C (New)	469	0.72%	555	0.87%	1.18	1.20
aactttact	new							
tcgcaaacc	new							
caatcaacc	new							
gcctcaacg	new							
gccttaacc	new	ORF4′(GP4′)	2850	4.40%	2613	4.09%	0.92	0.93
cctttcacc	TRS4′ [a]							
ggcgaaacg	new	ORF4′-C (New)	8	0.01%	12	0.02%	1.50	1.52
ctattaacc	TRS2 [a]	ORF2b (E)	7203	11.11%	7970	12.49%	1.11	1.12
tgggcaacc	new							
tactcaccc	new	ORF2a (GP2)	65	0.10%	69	0.11%	1.06	1.08
caattaaat	new	ORF2a-C (New)	36	0.06%	25	0.04%	0.69	0.71
ttctaaata	new	ORF3 (GP3)	445	0.69%	410	0.64%	0.92	0.94
tcactaacc	TRS3 [a]							
acttcaaca	new	ORF4 (GP4)	3094	4.77%	3131	4.91%	1.01	1.03
atccaaacc	new							
tcattgacc	TRS4 [a]							

acaacaacc	new	ORF5 (GP5)	9321	14.38%	9512	14.90%	1.02	1.04
cccataacc	new							
ttcttcgcc	new							
gtgataatc	new							
ttgttatca	new							
atcataacc	new							
tccttaact	TRS5 [a]							
taattatgt	new							
aaaataaca	new	ORF5-C (New)	1451	2.24%	1071	1.68%	0.74	0.75
agtcaaacc	new							
tgttcaacg	new							
ttcttcaca	new							
tcaataaca	new							
ttactaacc	new							
tgtttcacc	new							
ttgtcaacc	TRS6 [a]	ORF6 (GP6)	17989	27.74%	19779	30.99%	1.10	1.12
tacttgacc	new	ORF6-C (New)	533	0.82%	628	0.98%	1.18	1.20
ttgttaacc	TRS7 [a]	ORF7 (GP7)	13382	20.64%	10353	16.22%	0.77	0.79

[a] Previously published SHFV body TRS
[b] Previously predicted SHFV body TRS

Table 3.9. Prediction of transmembrane domains in the C-terminal in-frame ORFs of known structural proteins.

New ORFs (size)	Transmembrane domain	Protein sequence
ORF2a′-C (25aa)	Does not contain a transmembrane domain	MSAPRFVRLPVGSGQILASGKGQGL
ORF4′-C (84aa)	Contains 1 transmembrane domain	MTGIKEAFIALDFADCLLTGILYREHNVTAVFQQQDGHIVLCWNGT DPRISINQIPTPWFISPGALRWATIICACLAIFRAFYS
ORF2a-C (161aa)	Contains 1 transmembrane domain	MLLRRCHQSIIPYPNHPLGITTHAMVNALAAFSLQKAEDQAHATLH VKGITSAEYTYNVSCEPSSFTLDVTGLSKYLTSKNRALERLRHCEDI APIVGYLLSNRTHSYLTNPWMSVVLKSHRPAVVFCYYVCCFLLVLQ IKHIFAFVTYKLRSSCTSTPQS
ORF5-C (68aa)	Does not contain a transmembrane domain	MALRFAWTRHTNFIIDDRGRLFVNHDDVLISDPQGLRVGPHKVRAA KVILGGREANLLRQAHVEEWSW
ORF6-C (79aa)	Does not contain a transmembrane domain	MCWLGRQYITAPSSMVESSLGRLAINATGSTAVVTRRSGMTAVNGS LMPDVKRIILNGRVAAKRGLVNLRKYGWQTKNK

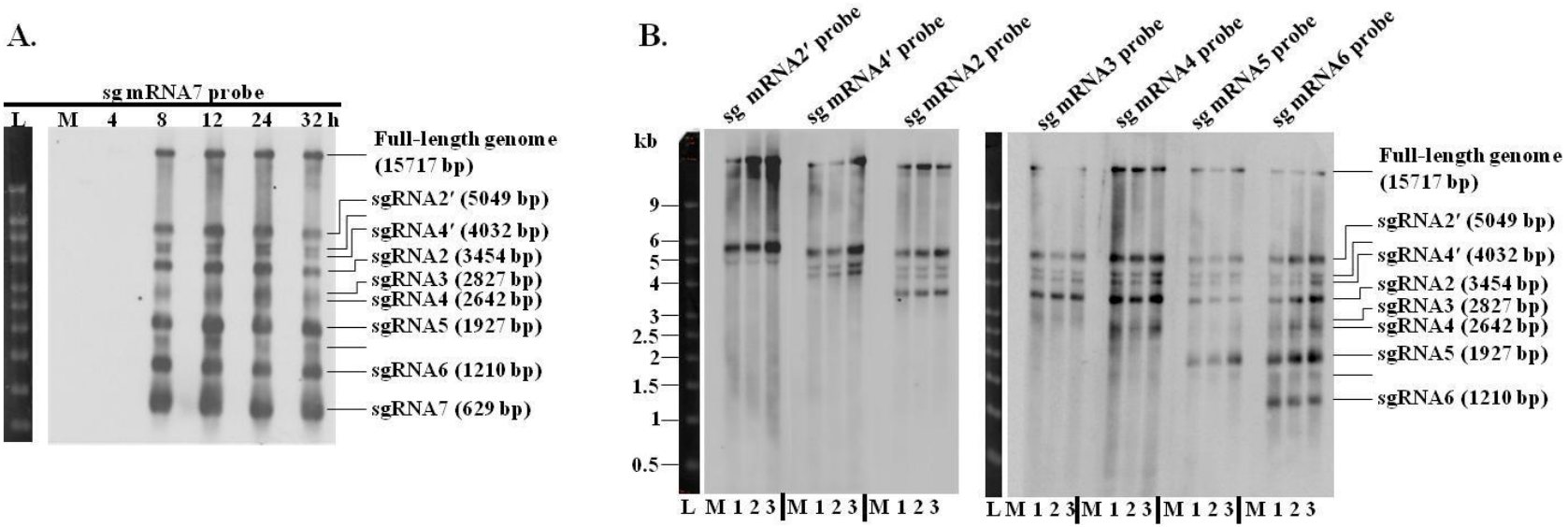


Figure 3.1. Northern blotting analysis of sg mRNAs produced in SHFVic infected MA104 cells.
(A) MA104 cells were mock infected or infected with SHFVic at an MOI of 1. At different times post infection, total intracellular RNA was extracted and 1 µg RNA was separated on a 1% denaturing agarose gel followed by transfer to an Hybond-N^+ membrane. After UV-crosslinking, the membrane was hybridized with a DIG-labeled RNA probe specific for sg mRNA7. (B) Total intracellular RNA collected at 12 hours after infection from three biological repeats was hybridized separately to DIG-labeled RNA probes specific for each known SHFV sg mRNA (sg mRNA2′ - sg mRNA6). The SHFV full-length genome and each known sg mRNA was assigned to the corresponding bands on the membrane based on the estimated mRNA sizes and the probe used. Additional bands detected are indicated with lines. The RNA ladder was cut from the membrane, stained with methylene blue and imaged. L, ladder; M, mock infected; 1, biological repeat 1; 2, biological repeat 2; 3, biological repeat 3.

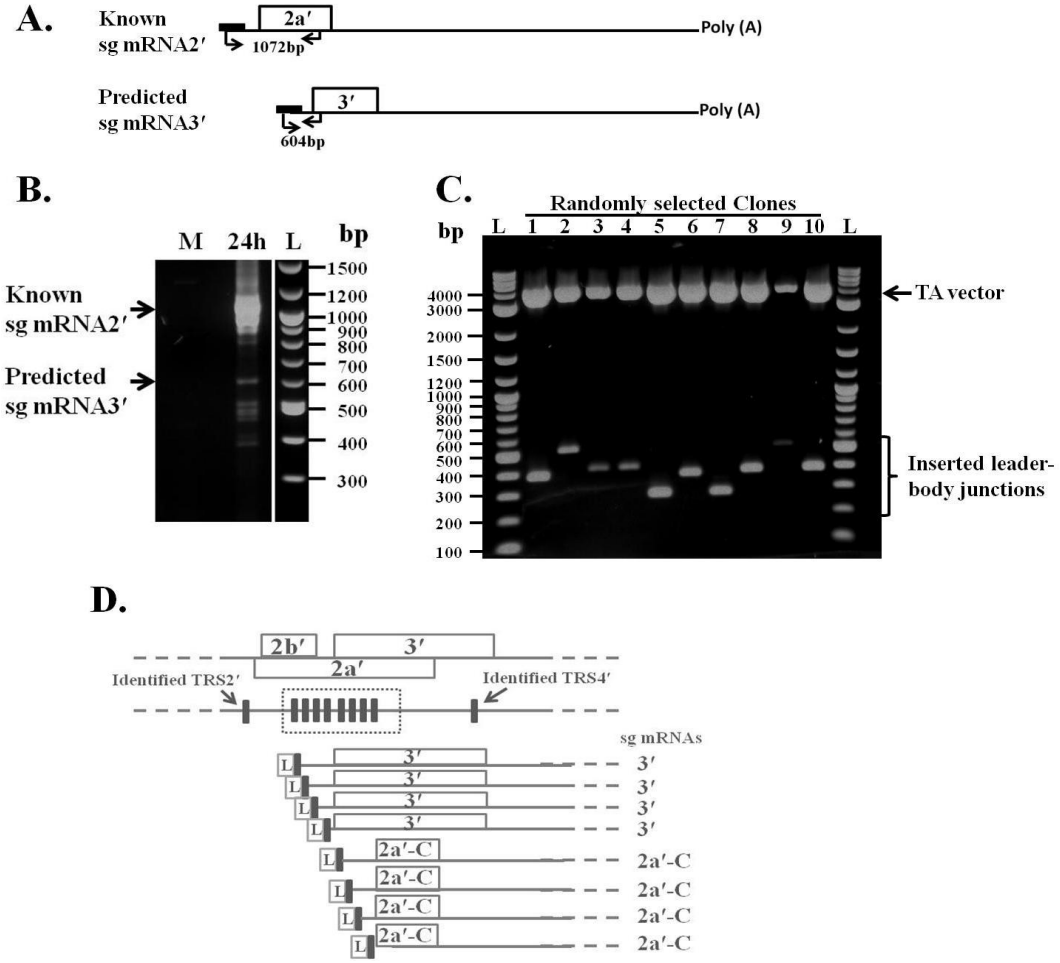


Figure 3.2. Amplification and cloning of sg mRNA leader-body junctions generated from the predicted TRS3′ region.

(A) Diagram indicating the positions of the primers used and the estimated sizes for the amplified leader-body junctions. A white open box indicates an ORF and the small black box indicates the leader sequence. (B) Electrophoresis separation of the amplified leader-body junction DNAs. MA104 cells were either mock infected (M) or infected with SHFVic at an MOI of 1. At 24 hpi, total intracellular RNA was extracted, subjected to RT-PCR and the products were separated on a DNA gel. The bands with the estimated sizes of the leader-body junctions of the known sg mRNA2′ and the predicted sg mRNA3′ are indicated by arrows. (C) All the bands that are around and below 500 bp were excised from the gel and the extracted leader-body junctions were cloned into a TA-vector. Forty clones were randomly selected, subjected to restriction digestion followed by gel electrophoresis. The results for 10 clones are shown. L, ladder. (D) Diagram of new body TRSs discovered between the preciously identified TRS2′ and TRS4′. The known ORFs are shown on the top line. The previously identified TRSs are indicated by black vertical bars on the second line. The black vertical bars within the dashed line box indicate new functional TRSs found. The ORFs encoded by the individual sg mRNAs are indicated by white open boxes. L, leader region.

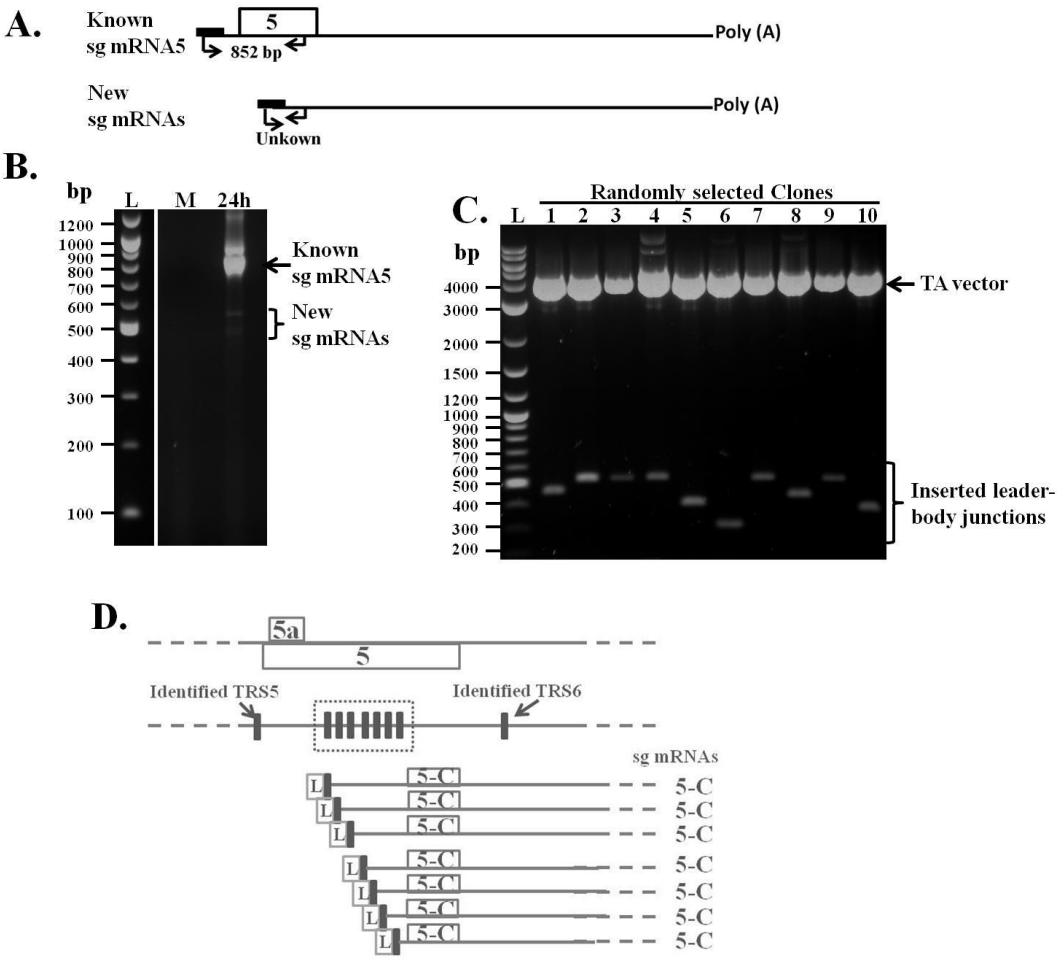


Figure 3.3. Amplification and cloning of sg mRNA leader-body junctions generated from the region between the identified TRS5 and TRS6.

(A) Diagram indicating the positions of the primers used and the estimated size for the amplified leader-body junctions. A white open box indicates an ORF and a small black box indicates the leader sequence. (B) Electrophoresis separation of the amplified leader-body junction DNAs. MA104 cells were either mock infected (M) or infected with SHFVic at an MOI of 1. At 24 hpi, total intracellular RNA was extracted, subjected to RT-PCR and the products were separated on a DNA gel. The band with the estimated size for the leader-body junction of the known sg mRNA5 is indicated by an arrow. All the bands represented the leader-body junctions of the new sg mRNAs were indicated by parenthesis and excised from the gel. (C) The extracted leader-body junctions were cloned into a TA-vector. Forty clones were randomly selected, subjected to restriction digestion followed by gel electrophoresis. The results for 10 clones are shown. L, ladder. (D) Diagram of new body TRSs discovered between the preciously identified TRS5 and TRS6. The known ORFs are shown on the top line. The previously identified TRSs are indicated by black vertical bars on the second line. The black vertical bars within the dashed line box indicate new functional TRSs found. The ORFs encoded by the individual sg mRNAs are indicated by white open boxes. L, leader region.

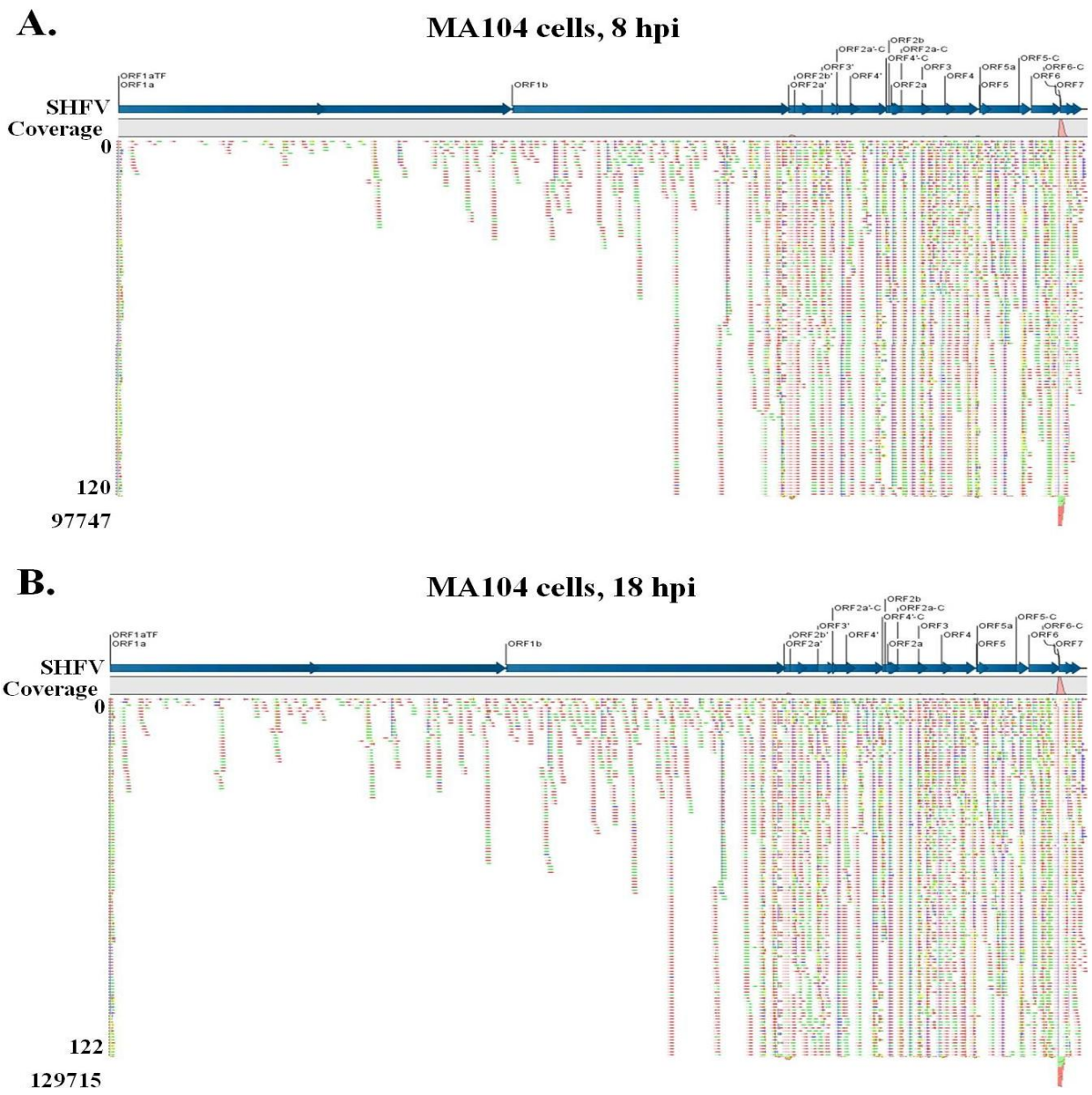


Figure 3.4. Mapping the NGS reads not containing identified leader-body junction sequences to the SHFV genome.

MA104 cells were infected with SHFVic at an MOI of 1. At 8 and 18 hpi, total intracellular RNA was extracted and mRNA was isolated and subjected to library preparation and RNA-Seq using Illumina Hi-Seq. The resulting data was trimmed and mapped to the SHFV genome sequence (NC_003092.2) and also to each of the identified leader-body junctions. All the remaining reads were mapped to the SHFV genome sequence without the leader region. The mapping results were displayed with read depth 0-120 in detail and the rest in compact view. The SHFV genome is indicated as a thick blue line with vertical lines indicating the beginning of each ORF. The pink peak indicates high coverage of reads. Underneath that, each green line represents a single forward read, each red line represents a single reverse read and each blue line represents a paired read. The numbers on the left side indicate the read depth.

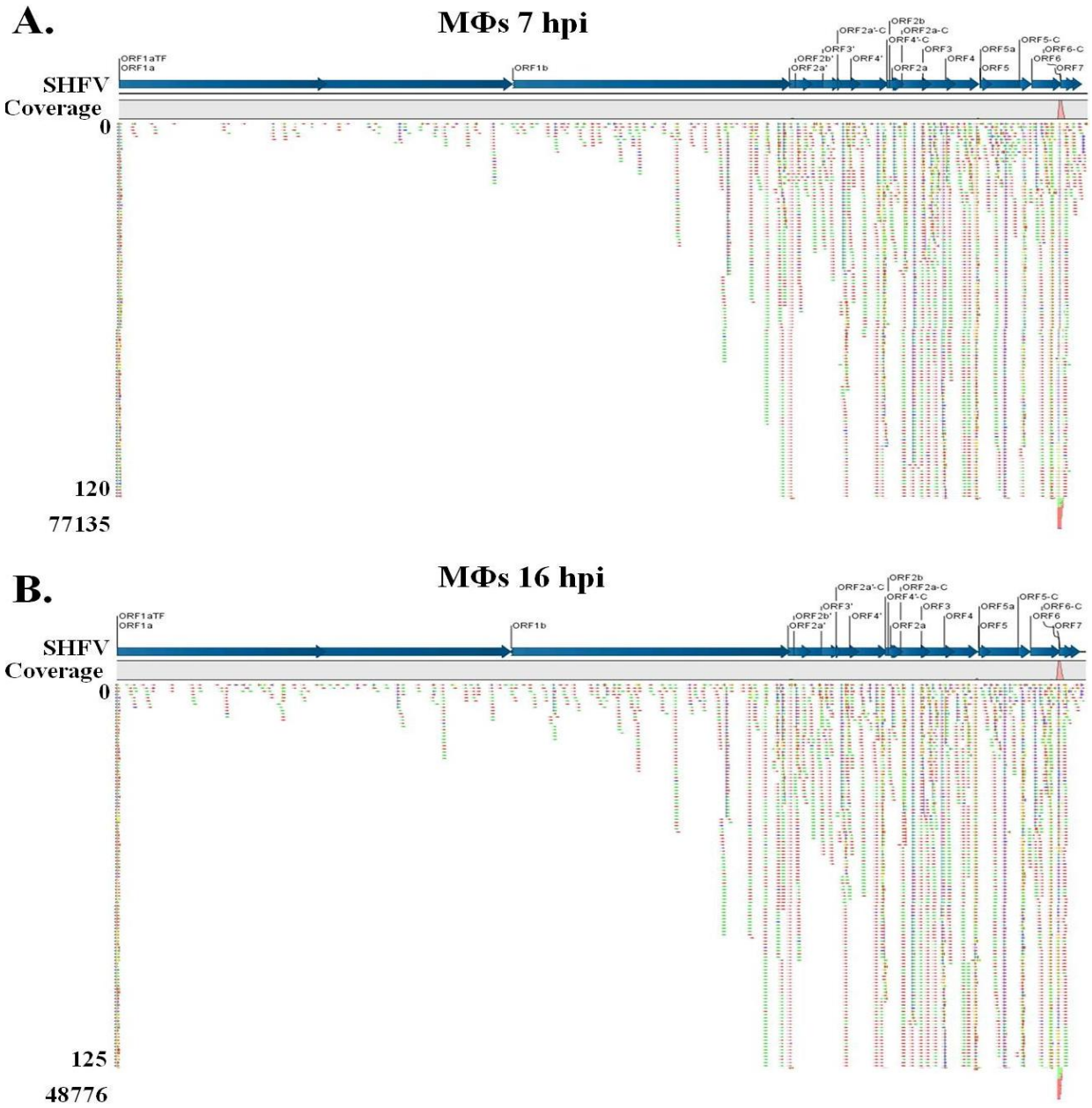


Figure 3.5. Mapping the NGS reads not containing identified leader-body junction sequences to the SHFV genome.

Macaque MΦs were infected with SHFVic at an MOI of 1. At 7 h and 16 hpi, total intracellular RNA was extracted and mRNA was isolated and subjected to library preparation and RNA-Seq using Illumina Mi-Seq. The resulting data was trimmed and mapped to the SHFV genome sequence (NC_003092.2) and also to each of the identified leader-body junctions. All the remaining reads were mapped to the SHFV genome sequence without the leader region. The mapping results were displayed with read depth 0-120 in detail and the rest in compact view. The SHFV genome is indicated as a thick blue line with vertical lines indicating the beginning of each ORF. The pink peak indicates high coverage of reads. Underneath that, each green line represents a single forward read, each red line represents a single reverse read and each blue line represents a paired read. The numbers on the left side indicate the read depth.

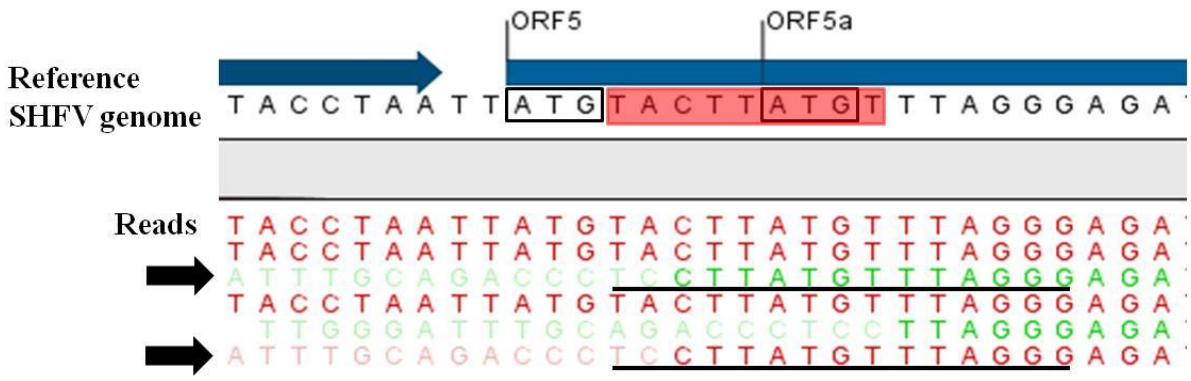


Figure 3.6. Discovery of a separate sg mRNA encoding ORF5a as its 5′ proximal ORF. MA104 cells were infected with SHFVic at an MOI of 1. At 8 hpi, total intracellular RNA was extracted and mRNA was isolated and subjected to library preparation and RNA-Seq using Illumina Hi-Seq. The resulting data was trimmed and mapped to the SHFV genome reference sequence (NC_003092.2) and then to each of the identified leader-body junctions. All the remaining reads were mapped to the SHFV genome reference without the leader region. The SHFV genome reference sequence is displayed and the vertical lines indicate the beginning of ORF5 and ORF5a. The start codons of ORF5 and ORF5a are boxed. The reads in green represent single forward reads, the reads in red represent single reverse reads. Faint nts represent the region that does not map to the reference. The two reads indicating the existence of a separate sg mRNA encoding ORF5a as its 5′ proximal ORF are indicated by black arrows. The unique 15 nt leader-body junction sequence of these reads is underlined in black. The body TRS for generating a sg mRNA encoding ORF5a is highlighted in red in the reference SHFV genome.

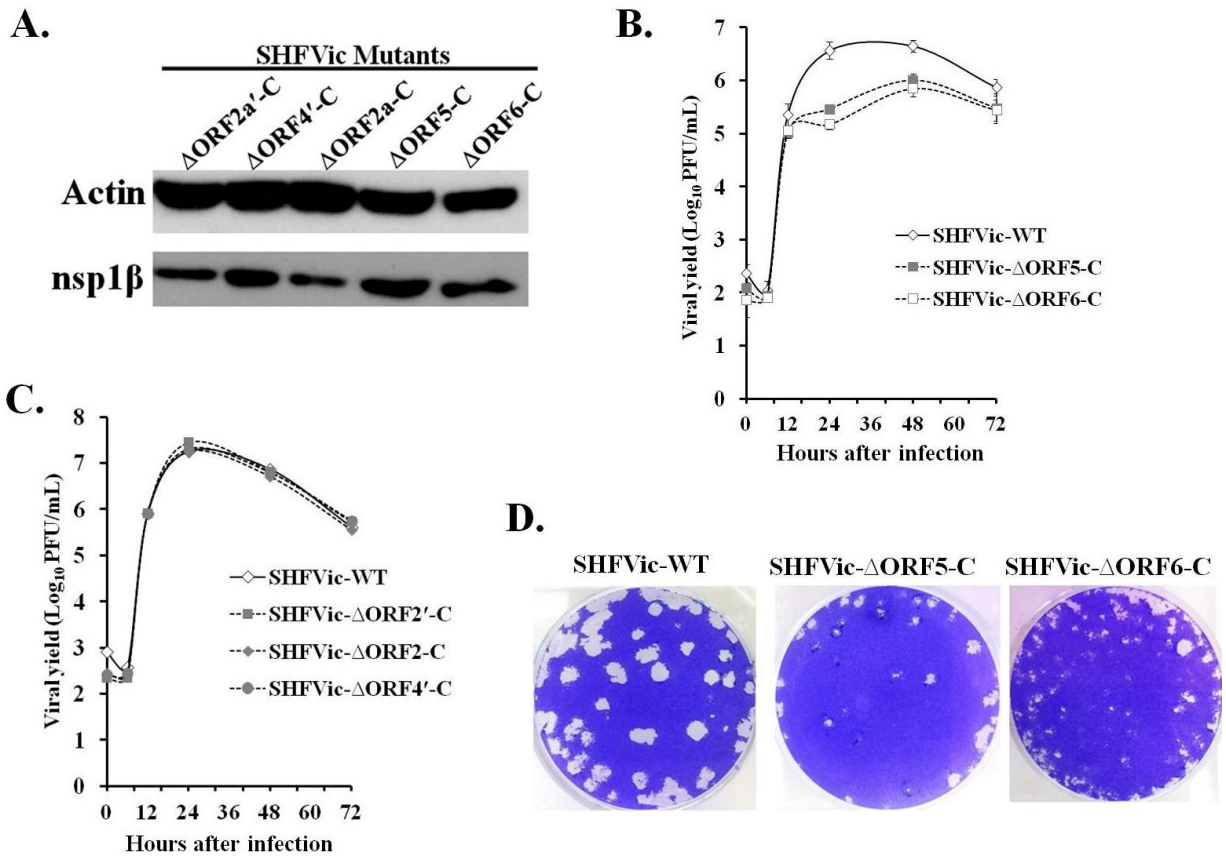


Figure 3.7. Analysis of the effect of knocking out the expression of each new C-terminal ORF on virus replication in MA104 cells.

(A) MA104 cells were infected with one of the five mutant viruses lacking the expression of a C-terminal in-frame ORF. At 24 hpi, cell lysates were harvested in RIPA buffer and subjected to Western blotting using anti-β-actin and anti-nsp1β antibodies. (B and C) MA104 cells were infected with wild type or the indicated mutant SHFVic virus. At different times after infection, the culture fluid was collected and virus was titered by plaque assay on MA104 cells. (D) MA104 cells were infected with wild type SHFVic or mutant virus lacking the expression of ORF5-C or ORF6-C. At 24 hpi, the culture fluid was collected and virus was titered by plaque assay and the plaque sizes are shown.

4 DEVELOPMENT OF STRATEGIES FOR EXPRESSING AN EXOGENOUS GENE FROM SHFV GENOME

4.1 Introduction

SHFV is a simian arterivirus that belongs to the family *Arteriviridae*. Besides other simian arteriviruses, The family *Arteriviridae* also includes PRRSV, LDV, EAV, WPDV and a novel arterivirus from the forest giant pouched rat (Dunowska, Biggs et al. 2012; Snijder, Kikkert et al. 2013; Kuhn, Lauck et al. 2016). All the arteriviruses have a similar genome arrangement and SHFV has the largest genome of ~15.7 kb. The 5′ two-thirds of the SHFV genome encode two polyproteins that self-cleave into 14 non-structural proteins. These non-structural proteins function in viral replication and transcription (Snijder, Kikkert et al. 2013; Brinton, Di et al. 2015). Compared to other arteriviruses that encode two PLP1 domains at the 5′ end of the genome, SHFV encodes three PLP1 domains in that region and they are all functionally active (Vatter, Di et al. 2014). The 3′ one-third of the SHFV genome encodes structural proteins that are translated from a nested set of sg mRNAs. The transcription of the minus strand template for each sg mRNA is regulated by a unique body TRS motif in the genome that includes a 9 nt core sequence and two ~15-20 nt adjacent flanking sequences (Godeny, de Vries et al. 1998; Sola, Moreno et al. 2005). Compared to other arteriviruses, SHFV encodes an extra set of minor structural proteins E′, GP2′, GP3′ and GP4′. Both sets of minor structural proteins are functionally important for generating infectious progeny virus (Vatter, Di et al. 2014). Recently, the full-length cDNA of the SHFV LVR strain was successfully cloned into a pACYC184 vector. The virus generated from the full-length clone has the same characteristics as the parental virus.

The potential of arterivirus as a viral vector has been explored by inserting a foreign gene into the infectious cDNA clones of EAV and PRRSV using different strategies. Fusion of the HA epitope of influenza virus directly to the PRRSV nucleocapsid protein was tried first. However, it was shown that direct fusion at either the N terminus or the C terminus of nucleocapsid protein decreased the viability and the genetic stability of PRRSV (Bramel-Verheije, Rottier et al. 2000). Insertion of a fluorescent protein gene fused to the foot-and-mouth disease virus (FMDV) 2A self-cleaving oligopeptide into the EAV polyprotein pp1a generated viable recombinant virus only when it was inserted between the nsp1 and nsp2 genes (van den Born, Posthuma et al. 2007; Mondal, Cook et al. 2015). Another successful strategy has been insertion of a foreign gene at the 3′ end of the arterivirus genome to express it from an additional sg mRNA utilizing the unique TRS regulation system (Pei, Hodgins et al. 2009).

The potential of SHFV as a viral expression vector has not previously been explored. SHFV encodes an extra nsp1 protein at the 5′ end of the genome. Insertion of a foreign gene in that region may affect the overall folding of the polyprotein and hence the efficiency of autocleavage. In addition, SHFV has the largest genome (15.7 kb) among arteriviruses and due to proposed icosahedral nucleocapsid packaging constraints, a gene smaller than the common fluorescent protein GFP would be expected to be better tolerated. M013 protein (~14 kDa) is a small myxoma virus protein that antagonizes inflammation (Rahman, Mohamed et al. 2009; Rahman and McFadden 2011). SHFV infection in African monkeys such as baboons induces an asymptomatic persistent infection whereas SHFV infection in macaque monkeys triggers acute hemorrhagic fever disease which is characterized by an elevated inflammatory response (Johnson, Dodd et al. 2011; Vatter and Brinton 2014; Vatter, Donaldson et al. 2015). In the present study, the first recombinant SHFV infectious clone expressing a foreign gene M013 from

an additional sg mRNA was successfully constructed. The M013 gene with a C-terminal V5 tag and a body TRS motif was inserted into the SHFV genome either between ORF4′ and ORF2b or between ORF4 and ORF5. Only insertion between ORF4′ and ORF2b generated viable progeny virus. The inserted M013 gene used the TRS2 for its transcription in each recombinant clone constructed but one of three different synthetic body TRSs, TRS2′, TRS4′ and TRS7 was added after the M013 gene to rescue the transcription of ORF2b. Although all three TRSs rescued the transcription of ORF2b and generated viable virus, only the recombinant genome with the synthetic TRS2′ maintained the M013 gene during serial virus passage.

4.2 Materials and methods:

4.2.1 Cells

The MA104 cell line was a gift from O. Nianan, Center for Disease Control and Prevention (Atlanta, GA). MA104 cells are embryonic kidney cells from African green monkeys. They were cultured at 37°C in a 5% CO_2 atmosphere using 1 x Minimal Essential Medium (MEM) supplemented with 1% L-glutamine, 10% fetal bovine serum and 0.1% gentamicin.

4.2.2 Construction of an M013 mammalian expression plasmid pEF6-M013-V5

The pcDNA-DEST40-M013 construct was a gift from Dr. Grant McFadden, University of Florida. A set of primers was designed to amplify the M013 gene from this construct and a V5 tag (GKPIPNPLLGLDST) was introduced in the reserve primer fused at the C-terminus of the M013 gene (Table 4-1). AccuPrime Tag DNA polymerase (Thermo Fisher Scientific) was used for PCR and the parameters were set as 35 cycles at 95°C for 15 sec, 60°C for 30 sec and 68°C

for 35 sec. The amplified PCR product was verified on and extracted from a 1% agarose gel, digested with restriction enzymes and ligated into the mammalian expression vector pEF6-V5/His TOPO. After transformation into TOP10 competent cells, pEF6-M013-V5 plasmids were extracted from selected clones and sequenced.

4.2.3 Insertion of the M013 gene into the full-length SHFVic to construct SHFVic-M013

The SHFV full length cDNA is divided into five fragments and each fragment is maintained in a pCR-XL-TOPO vector as described previously (Vatter, Di et al. 2014). One set of primers was designed to separate the 1 nt overlapping ORF4′ from ORF2b and insert AflII and BglII restriction sites between these two ORFs in fragment IV. Another set of primers was designed to insert AflII and BglII restriction sites between ORF4 and ORF5 in fragment V (Table 4-1). Fragment IV and V plasmids were subjected to mutagenesis using these two sets of primers and the QuikChange lighting kit following the manufacturer's protocol (Agilent Genomics). The mutated fragment IV and V plasmids were transformed into XL10-Gold Ultracompetent cells (Agilent Genomics), purified from selected clones and sequenced.

One forward primer and three reverse primers, that introduced the core and flanking sequences of either SHFV body TRS2′, TRS4′ or TRS7, were designed to amplify the M013-V5 gene from the pEF6-M013-V5 plasmid (Table 4-1). The PCR reactions were done using AccuPrime Tag DNA polymerase (Thermo Fisher Scientific) under the same conditions described above and the three PCR products, M013-V5-TRS2′, M013-V5-TRS4′ and M013-V5-TRS7, were verified on and excised from a 1% agarose gel and extracted. The extracted PCR products and the mutated fragment IV were digested with AflII and BglII and ligated together to produce Fragment IV-M013-TRS2′, Fragment IV-M013-TRS4′ and Fragment IV-M013-TRS7

constructs, respectively. Each of the constructs was transformed into TOP10 competent cells, purified from selected clones and sequenced. To construct the full-length SHFVic-M013, each of the three M013 inserted fragment IV constructs was digested with PflMI and ligated simultaneously with the other four PflMI digested wild type SHFV fragments. The ligated full-length mutant SHFV sequences were each cloned into a pACYC184 vector to generate the SHFVic-4′/2b-M013-TRS2′, SHFVic-4′/2b -M013-TRS4′ and SHFVic-4′/2b -M013-TRS7 constructs. The same strategy was used to insert the M013 gene into the mutated fragment V to generate the full-length constructs, SHFVic-4/5-M013-TRS2′, SHFVic-4/5-M013-TRS4′ and SHFVic-4/5-M013-TRS7.

4.2.4 SHFVic-M013 virus cell passage protocol

Each SHFVic-M013 construct plasmid DNA was linearized at the introduced PvuI restriction site downstream of the SHFV genome on the pACYC184 vector (Vatter, Di et al. 2014). Viral genome RNA was *in vitro* transcribed from the linearized plasmid DNA and purified using the mMESSAGE mMACHINE SP6 Transcription kit following the manufacturer's protocol (Thermo Fisher Scientific). MA104 cells were grown in a 6-well plate until ~50% confluence and transfected with ~100 ng purified SHFV genome RNA in DMRIE following the manufacturer's protocol (Thermo Fisher Scientific). Five days after transfection, culture fluid (2 ml) was collected from wells when cytopathic effect (CPE) was observed, centrifuged at 4°C for 5 min at 1,000 rpm, aliquoted and designated as passage 0 virus (P0). P0 virus (500 µl) was then used to infect an MA104 monolayer in a 10 cm dish. Culture fluid (10 ml) was collected from the dish at 24 hpi, centrifuged, aliquoted and designated as passage 1 virus (P1). Each P1 virus stock was titered by plaque assay on MA104 cells. For additional

passages, MA104 monolayers in 6-well plates were infected with P1 virus at an MOI of 1. At 24 hpi, culture fluid (2ml) was collected, centrifuged, aliquoted, titered and designated as passage 2 virus (P2). The same protocol was repeated to generate a passage 3 virus (P3) from the P2 virus.

4.2.5 Reverse transcription and amplification of SHFVic-M013 genome RNA

Viral genome RNA was extracted from 250 µl culture fluid at each passage using TRI-LS reagent following the manufacturer's protocol (Molecular Research Center, Inc.). The extracted viral genome RNA (~50-100 ng) was subjected to RT-PCR using primers designed to the regions flanking the inserted M013 gene (Table 4-1). A SuperScript III one-step RT-PCR kit was used and the manufacturer's protocol was followed. The RT-PCR product was divided into two portions, separated on a 1% agarose gel, stained with ethidium bromide and visualized under long-wavelength UV light. The appropriate band(s) were excised from the gel and the extracted RT-PCR products were either sequenced directly or subjected to TA cloning.

4.2.6 TA-cloning of RT-PCR products

The extracted RT-PCR products were ligated into a pCR4-TOPO cloning vector and transformed into TOP10 chemically competent cells using the manufacturer's protocol (Thermo Fisher Scientific). Ten colonies were picked randomly from the plates and grown overnight in liquid culture (LB media plus 50 µg/ml kanamycin). Plasmid DNA was extracted from the liquid cultures and digested with FastDigest EcoRI (Thermo Fisher Scientific). The digested plasmids were then separated on a 1% agarose gel, stained with ethidium bromide and imaged with a BioDoc-it imaging system (UVP, LLC). The plasmids carrying different sized inserts were sequenced.

4.2.7 Western Blot Assay

Transfected or infected MA104 cell lysates were harvested in 1 x RIPA buffer (1 X phosphate-buffered saline, 1% Nonidet P-40, 0.5% sodium deoxycholate, and 0.1% SDS) containing 1 x Halt protease inhibitor cocktail (Thermo Scientific). Total protein in cell lysates was measured by a BCA assay following the manufacturer's protocol (Pierce). The same amount of total protein was separated in each well of a 15% SDS-PAGE gel and transferred to a nitrocellulose membrane at 100 V for 1 h. The membrane was incubated in blocking buffer (1 X Tris-buffered saline containing 5% non-fat dry milk and 0.1% Tween 20) at 4°C overnight and cut into strips. The strips were incubated at 4°C overnight with a primary antibody targeting the SHFV nsp1β peptide (FAQKVITAFPEGVLC) (Abgent custom antibody) or the V5 tag (R960-25, Thermo Fisher Scientific) or beta-actin (C-11; Santa Cruz Biotechnology). The membrane strips were then washed three times for 10 min with 1 X Tris-buffered saline containing 0.1% Tween 20, followed by incubation with a secondary antibody (horseradish peroxidase-conjugated anti-rabbit or anti-mouse antibody) (Santa Cruz Biotechnology) for 1 h at room temperature. After washing, the membrane strips were developed with a Super-Signal West Pico detection kit (Pierce) following the manufacturer's protocol.

4.2.8 Immunofluorescent Assay (IFA)

MA104 cells were seeded on cover slips in a 24-well plate and grown until ~70-80% confluent. Cells were either infected with SHFVic virus at an MOI of 1 for 24 h or transfected with 0.5 µg pEF6-M013-V5 plasmid DNA using lipofectamine LTX/PLUS reagent following the manufacturer's protocol (Thermo Fisher Scientific) for 24 h before infection with SHFVic

virus (MOI of 1) for another 24 h. The cells were then fixed with 4% paraformaldehyde (PFA) for 10 min, followed by permeabilization for 10 min with 0.1% Triton-X and then blocking with 5% horse serum at room temperature for 1 h. After incubation with rabbit anti-V5 antibody (Abcam, 1:500) and mouse anti ds-RNA antibody (English & Scientific Consulting Kft. 1:1000) overnight at 4°C, cells were incubated with Alexa Fluor 488-donkey anti-rabbit antibody (Thermo Fisher Scientific, 1:400), Alexa Fluor 594 donkey anti-mouse antibody (Thermo Fisher Scientific, 1:400) and Hoechst 33342 (Thermo Fisher Scientific, 0.05%) for 1 h at room temperature. The cells on the cover slips were then mounted on a glass slide and imaged with a Zeiss Axio Observer 1 microscope using a 40x or 63x oil emersion objective.

.

4.2.9 Plaque Assay

MA104 cells were seeded in 6-well plates, grown to confluence, and infected with 100 µl of a serial 10 fold dilution of a virus sample for 1 h at 37°C. The inoculum was removed and 2 ml of overlay media (2 x MEM containing 5% FCS mixed 1:1 with 1% SeaKem ME agarose) was added. After incubation at 37°C for 48 h, the overlay media was removed and the cells were stained with 0.05% crystal violet in 10% ethanol prior to plaque counting and imaging. Three biological repeats of each sample were assayed in duplicate wells.

4.3 Results:

4.3.1 Overexpression of M013 does not negatively affect SHFV replication in MA104 cells

Overexpression of the myxoma virus M013 protein was previously shown to inhibit the NF-κB pathway in Hela cells (Rahman, Mohamed et al. 2009) and NF-κB pathway activation was reported to be essential for efficient replication of EAV (Mottahedin, Paidikondala et al. 2013). To analyze the effect of overexpressing M013 protein on SHFV replication, a mammalian expression plasmid pEF6-M013-V5 was constructed with a V5 tag fused to the C-terminus of M013 protein. MA104 cells were seeded on cover slips and transfected with the pEF6-M013-V5 plasmid DNA 24 h before infection with wild type SHFVic (SHFVic-WT) virus at an MOI of 1. At 24 hpi, MA104 cells were fixed, permeabilized and processed for IFA using anti-V5 and anti-dsRNA antibodies. The intensity of the viral ds RNA signal (red) was comparable between cells with and without overexpression of M013 (green) (Figure 4-1), suggesting that M013 overexpression does not negatively affect the level of SHFV RNA replication in MA104 cells and supporting the feasibility of constructing a SHFVic virus that expresses the M013 gene.

4.3.2 Generation of SHFVic virus expressing M013 protein from an additional sg mRNA

The 3′ ORFs in the SHFV genome overlap each other except for ORF4 and ORF5, which are separated by two nucleotides. ORF4′ and ORF2b only overlap by 1 nucleotide. To accommodate the insertion of the M013 gene while minimizing disruption of the genome organization, two sets of primers (Table 4-1) were designed to introduce two unique restriction sites (AflII and BglII) between either ORF4 and ORF5 or between ORF4′ and ORF2b using QuikChange Lighting kit following the manufacturer's protocol (Agilent Genomics). Because the body TRS5 and body TRS2, which would normally be used for ORF5 and ORF2b expression,

are now used for M013 expression when M013 was inserted between ORF4 and ORF5 or between ORF4′ and ORF2b, a separate body TRS was needed after the M013 gene to restore the transcription regulation of the downstream ORF. If a duplicate body TRS5 or body TRS2 was to be inserted after the M013 gene, intramolecular homologous recombination between the two close copies of body TRS would be expected resulting in excision of the inserted M013 gene. Therefore, a different body TRS was inserted to regulate the transcription of the downstream ORF. Three body TRSs, TRS2′, TRS4′ and TRS7, were selected to be inserted as the new body TRS for the downstream ORF. The M013 mammalian expression vector pEF6-M013-V5 was used as the template to amplify the M013-V5 DNA. Three reverse primers were designed to introduce the core and flanking sequences (~40-55 nt) of each of the three selected body TRSs immediately downstream of the V5 tag sequence (Table 4-1). Each of the purified M013-V5-TRS cassettes was then inserted between ORF4 and ORF5 or between ORF4′ and ORF2b using the introduced AflII and BglII restriction sites to make each of the full-length SHFVic-4/5-M013-TRS or SHFVic-4′/2b-M013-TRS constructs.

To generate SHFVic-M013 virus, mutant viral genome RNA was transcribed *in vitro* from each of the full-length SHFVic-4/5-M013-TRS or SHFVic-4′/2b-M013-TRS constructs and purified. MA104 cells were mock transfected, transfected with SHFVic-WT RNA or mutant viral RNA. At 5 days after transfection, CPE was observed in wells transfected with SHFVic-WT RNA and with RNA generated from each of the three SHFVic-4′/2b-M013-TRS constructs (SHFVic-4′/2b-M013-TRS2′, SHFVic-4′/2b-M013-TRS4′ and SHFVic-4′/2b-M013-TRS7) (data not shown). Western blot analysis detected viral nsp1β protein in cell lysates harvested from each of the wells showing CPE, but only wells transfected with viral RNA generated from the three SHFVic-4′/2b-M013-TRS constructs showed detectable M013-V5 expression (Figure 4-2).

The data indicate that infectious progeny virus expressing the M013 protein was produced from each of the three SHFVic-4′/2b-M013-TRS genome RNAs. Progeny virus was collected and designated as P0 virus. No CPE was observed in any of the wells transfected with RNA transcribed from the SHFVic-4/5-M013-TRS constructs even after two additional repeated attempts and Western blot analyses detected no viral protein in the transfected cell lysates (data not shown), indicating that infectious progeny virus was not produced from any of the SHFVic-4/5-M013-TRS constructs.

4.3.3 Analysis of M013 expression from the SHFVic-M013 viruses during serial virus passage

The P0 virus generated from each of the three SHFVic-4′/2b-M013-TRS constructs was passaged once on MA104 cells. At 24 hpi, the culture fluid was harvested and designated as P1 virus. Cell lysates was also harvested and subjected to Western blot analysis for M013 protein expression. Only the SHFVic-4′/2b-M013-TRS2′ virus maintained M013 expression during passage 1 whereas both the SHFVic-4′/2b-M013-TRS4′ and SHFVic-4′/2b-M013-TRS7 viruses lost M013 expression after one passage (Figure 4-3A). To check whether the M013 gene sequence was maintained, viral genome RNA was extracted from the P0 and P1 stocks of each SHFVic-4′/2b-M013-TRS virus and subjected to RT-PCR using primers flanking the genome region where M013 was inserted. RT-PCR amplification of the wild type SHFV genome RNA was used as a negative control. The RT-PCR products were separated on a 1% agarose gel and imaged (Figure 4-3B). The full-length M013 gene was present in all of the P0 SHFVic-4′/2b-M013-TRS viral genomes but was only maintained in the genome of the P1 SHFVic-4′/2b-M013-TRS2′ virus consistent with the Western blot data (Figure 4-3B).

The P1 SHFVic-4′/2b-M013-TRS2′ virus was serially passaged on MA104 cells at an MOI of 1 to generate the P2 and P3 virus. Cell lysates harvested from MA104 cells infected with P2 or P3 SHFVic-4′/2b-M013-TRS2′ virus were subjected to Western blot analysis. M013 expression was maintained through each of these passages (Figure 4-3A). RT-PCR amplification of viral genome RNA extracted from P2 and P3 SHFVic-4′/2b-M013-TRS2′ virus using the same sets of primers confirmed that the full-length M013 gene was maintained in the viral genome through these passages (Figure 4-3B). The data indicate that although infectious progeny virus expressing M013 protein was produced from each of the SHFVic-4′/2b-M013-TRS constructs, only the SHFVic-4′/2b-M013-TRS2′ virus stably expressed M013 gene for at least three passages.

4.3.4 Analysis of the M013 sequence in the genomes of SHFVic-4′/2b-M013-TRS viruses

Although the band representing the full-length M013 RT-PCR product was not detected on the gel for the P1 SHFVic-4′/2b-M013-TRS4′ and SHFVic-4′/2b-M013-TRS7 viruses (Figure 4-3B), three smaller sized bands and one smaller sized band were detected for the P1 SHFVic-4′/2b-M013-TRS4′ and SHFVic-4′/2b-M013-TRS7 viruses, respectively, suggesting the existence of a truncated M013 gene (Figure 4-4A). To analyze the M013 sequence in the genomes of the P1 SHFVic-4′/2b-M013-TRS viruses, the smaller RT-PCR products generated from the P1 SHFVic-4′/2b-M013-TRS4′ or SHFVic-4′/2b-M013-TRS7 virus were subjected to TA-Cloning. The cloned plasmids were digested with EcoRI and separated on a 1% agarose gel. Plasmids with three different sized inserts were observed for SHFVic-4′/2b-M013-TRS4′ (Figure 4-4B), whereas all the plasmids from SHFVic-4′/2b-M013-TRS7 contained the same size insert (Figure 4-4C). Selected plasmids were sequenced and the sequences were aligned to that of the

full-length M013 gene. The three smaller RT-PCR products generated from the SHFVic-4′/2b-M013-TRS4′ genome contained M013 gene internal deletions of different lengths (Figure 4-5A). The smaller RT-PCR product generated from the SHFVic-4′/2b-M013-TRS7 genome also contained an internally deleted M013 gene (Figure 4-5B). In most cases, both the V5 tag and the inserted body TRS were maintained in the viral genome, but the majority of the M013 gene was deleted.

4.3.5 Analysis of the M013 gene expression pattern in infected cells

Among the three SHFVic-4′/2b-M013-TRS viruses, SHFVic-4′/2b-M013-TRS2′ was the only virus that stably expressed the M013 gene through at least three passages. To analyze the M013 expression pattern during infection, MA104 cells were infected with P1 SHFVic-4′/2b-M013-TRS2′ virus at an MOI of 1. At different times after infection, cell lysates were harvested and subjected to Western blot analysis of M013 and viral nsp1β expression. M013 expression was detected as early as 12 hpi whereas nsp1β expression could be detected as early as 2 hpi (Figure 4-6A). This was expected because nsp1β is translated and autocleaved from the 5′ end of the polyprotein starting immediately after genome uncoating. Whereas M013 is expressed from a sg mRNA and the production of sg mRNA requires the functions of viral non-structural proteins. At 48 hpi, the expression level of M013 peaked and did not decrease by 72 hpi. MA104 cells seeded on cover slips were infected with either P1 SHFVic-WT virus or P1 SHFVic-4′/2b-M013-TRS2′ virus at an MOI of 1. At 24 hpi, cells were fixed, permeabilized and processed for IFA using anti-dsRNA antibody and anti-V5 antibody. The dsRNA signal was detected in all of the infected cells, but only cells infected with SHFVic-4′/2b-M013-TRS2′ virus also had an

M013-V5 signal (Figure 4-6B). However, not all of the SHFVic-4′/2b-M013-TRS2′ infected cells had detectable levels of M013, which may due to non-synchronous virus infection cycles.

4.3.6 Comparison of the growth kinetics of SHFVic-WT virus and SHFVic-M013 virus

Insertion of a foreign gene into the genome of an arterivirus can generate a highly attenuated virus (Wang, Hou et al. 2014). To analyze the effect of M013 insertion on virus yields, MA104 cells were infected with either P1 SHFVic-WT virus or P1 SHFVic-4′/2b-M013-TRS2′ virus at an MOI of 1. Virus titers in culture fluids collected at different times after infection were determined by plaque assay. The growth kinetics of the two viruses were similar but the SHFVic-4′/2b-M013-TRS2′ virus produced slightly lower yields after 12 hpi (Figure 4-7). The data indicate that insertion of M013 in the SHFVic-4′/2b-M013-TRS2′ virus had only a slightly negative effect on virus production.

4.4 Discussion

The potential for expressing a foreign gene from the viral genome has been extensively studied in the arteriviruses EAV and PRRSV. The successful generation of a recombinant EAV and PRRSV expressing a fluorescent protein would facilitate the tracking of virus entry and spread in cells and serve as a valuable tool for screening neutralizing antibodies and antiviral inhibitors *(van den Born, Posthuma et al. 2007; Pei, Hodgins et al. 2009; Wang, Huang et al. 2013; Mondal, Cook et al. 2015).* The generation of recombinant PRRSV expressing the antigens of another swine virus, porcine circovirus type 2, suggested the potential use of PRRSV as a vaccine vector (Pei, Hodgins et al. 2009; Zhang, Liu et al. 2015). SHFV has the longest genome

among the arteriviruses. In the present study, a recombinant SHFV that expresses a foreign gene from an additional sg mRNA was generated for the first time.

In previous studies, a foreign gene was inserted between the 5′ ORF1b and the 3′ ORF2a in the PRRSV genome or between the 3′ ORF5 and ORF6 in the EAV genome and expressed from an additional sg mRNA (de Vries, Glaser et al. 2001; Pei, Hodgins et al. 2009). There is no overlap between the PRRSV ORF1b and ORF2a and a 12 nt overlap between the EAV ORF5 and ORF6. The previous study with EAV showed that separation of overlapping ORFs and insertion of an AflII restriction site did not negatively affect virus viability or yield (de Vries, Glaser et al. 2000). In the SHFV genome, the only two 3′ ORFs that do not overlap are ORF4 and ORF5. ORF4′ and ORF2b overlap but with only 1 nt. Therefore, in this study, the M013 gene was separately inserted at each of these locations using AflII and BglII restriction sites. Interestingly, although ORF4 and ORF5 do not overlap, insertion of M013 between them abolished virus viability, suggesting this region in SHFV is very sensitive to mutations and does not tolerate insertion of a foreign gene. In contrast, insertion of M013 between ORF4′ and ORF2b generated viable virus (Figure 4-2). This was not totally unexpected because the SHFV genome acquired four extra minor structural proteins in this region during its evolution. The M013 gene was not inserted between ORF7 and the 3′ untranslated region (3′-UTR) because the interaction between a stem-loop structure located in the 3′-UTR and a hairpin structure in the ORF7 was shown to be important for arterivirus RNA synthesis (Verheije, Olsthoorn et al. 2002; Beerens and Snijder 2007). However, recent publications have reported successful generation of recombinant PRRSV expressing a foreign gene inserted between ORF7 and 3′-UTR, indicating the mutagenesis potential of this location (Wang, Huang et al. 2013; Zhang, Liu et al. 2015).

When the M013 gene is inserted between ORF4′ and ORF2b in the SHFV genome, the published TRS2 (Godeny, de Vries et al. 1998), which is normally used to regulate ORF2b expression, is instead used for regulating the M013 expression. A different body TRS is then needed downstream of M013 gene to rescue the expression of ORF2b. Different body TRSs regulate the transcription at various efficiencies which are determined by the primary sequence of the body TRS, the flanking sequences and the local RNA secondary structure (Pasternak, Gultyaev et al. 2000; Pasternak, van den Born et al. 2003; Sola, Moreno et al. 2005). In its normal location, TRS2 is an efficient TRS that generates sg mRNA with high abundance (~7%, unpublished data). Among all the other TRSs, TRS2′ is most similar to TRS2 in the abundance of sg mRNA it generates (~9%, Chapter III, Table 3-5). However, when moved to a new location in the genome, the expression efficiency from a body TRS may change. A less efficient TRS4′ (~1%, Chapter III, Table 3-5) and a more efficient TRS7 (~23%, Chapter III, Table 3-5) were also selected to test for insertion into the new location between M013 and ORF2b to restore the expression of ORF2b to an appropriate level. Interestingly, insertion of each of the three body TRSs generated viable progeny virus (P0) after transfection of viral genome RNA into MA104 cells (Figure 4-2). However, only insertion of TRS2′ generated virus that stably maintained M013 gene expression during serial viral passage (Figure 4-3). The observed genome instability may be due to an inappropriate amount of ORF2b being produced from the inserted TRS4′ or TRS7.

Insertion of M013-V5-TRS4′ or M013-V5-TRS7 in the SHFV genome resulted in internal deletions in the M013 gene after one passage (Figure 4-5), which could be due to polymerase jumping across an RNA structure. Although SHFVic virus inserted with M013-V5-TRS2′ maintained M013 expression for 3 passages, the expression level started to decrease at

passage 4 and was undetectable by Western blot at passage 7 (data not shown). Similar genetic instability resulting from the expression of a foreign gene from an additional sg mRNA has been previously reported (de Vries, Glaser et al. 2001; Wang, Hou et al. 2014). Insertion of an additional transcription unit may promote foreign gene deletion by affecting genome RNA folding and/or altering optimal regulation of 3′ sg RNA transcription. An alternative strategy for expressing a foreign gene from the viral genome would be to fuse the foreign protein to a viral protein using the FMDV 2A peptide as a linker. The conformation of the 2A peptide can trigger a ribosomal "skip" in the linker so that the Gly-Pro peptide bond is not synthesized. Therefore, the foreign protein can be released from the fusion protein (Ryan, Donnelly et al. 1999; Donnelly, Luke et al. 2001). A previous study showed that fusion of an HA epitope to the PRRSV nucleocapsid protein using the 2A linker generated viable virus that maintained the HA epitope in the genome after passage (Bramel-Verheije, Rottier et al. 2000). A similar strategy could be used to fuse M013 gene to ORF2b in the future. Two additional studies have successfully generated stable recombinant EAV expressing fluorescent protein and the 2A linker fused between nsp1 and nsp2 in the 5′ ORF1a region (van den Born, Posthuma et al. 2007; Mondal, Cook et al. 2015). SHFV encodes three nsp1 proteins, nsp1α, nsp1β and nsp1γ, in that region (Vatter, Di et al. 2014). Fusion of M013-FMDV-2A between nsp1γ and nsp2 is worth trying. The other functions of the three nsp1 proteins are not known. If one of the nsp1 proteins is functionally redundant, replacement of that protein with M013 can be another strategy for generating a stable recombinant SHFV.

Table 4.1. Primers

Primer name	**Sequence (5′ to 3′)**
BamHI-M013-V5-XbaI-F	ATAAGGATCCACCATGGAGCACCGAG [a]
BamHI-M013-V5-XbaI-R	GGCGTCTAGATCA**CGTAGAATCGAGACCGAGGAGAGGGTTAGGGATAGGCTTACC**AA ATAACAATTTGCGAC [b]
Fragment IV-AflII+BglII-QC-F	GGCGATCTTTAGAGCATTCTATTCTTAACTTAAGAGATCTATGGGTTCTATACTCACCC
Fragment IV-AflII+BglII-QC-R	GGGTGAGTATAGAACCCATAGATCTCTTAAGTTAAGAATAGAATGCTCTAAAGATCGCC
Fragment V-AflII+BglII-QC-F	CCGCTCCTTAACTACCTAACTTAAGAGATCTATGTACTTATGTTTAGGG
Fragment V-AflII+BglII-QC-R	CCCTAAACATAAGTACATAGATCTCTTAAGTTAGGTAGTTAAGGAGCGG
AflII-M013-V5-BglII-F	GGGCTTAAGATGGAGCACCGAGGCGTCATTATAACCGTGTTGG
AflII-M013-V5-TRS2′-BglII-R	GCAGATCTcaggctgtaggtgtgaaacagaagttaaagaccttggctagTCACGTAGAATCGAGACC [c]
AflII-M013-V5-TRS4′-BglII-R	TGAGATCTtgcggctgagaccaaagagttcagggtgaaaggttaaggcagctaTCACGTAGAATCGAGACC
AflII-M013-V5-TRS9-BglII-R	TGAGATCTtggtttgccagccacacttcctcaggttaacaagacccctttggcggcaaccctTCACGTAGAATCGAGACC
Fragment IV-RTPCR-F	TATCGTCCTCTGCTGGAACGGG
Fragment IV-RTPCR-R	CAGGTCAAAGAACGAGACGAGG

[a] Introduced restriction sites are underlined

[b] The V5 tag sequence is in bold letters

[c] Lowercase letters indicate inserted TRS2′, TRS4′ or TRS7 core and flanking sequences

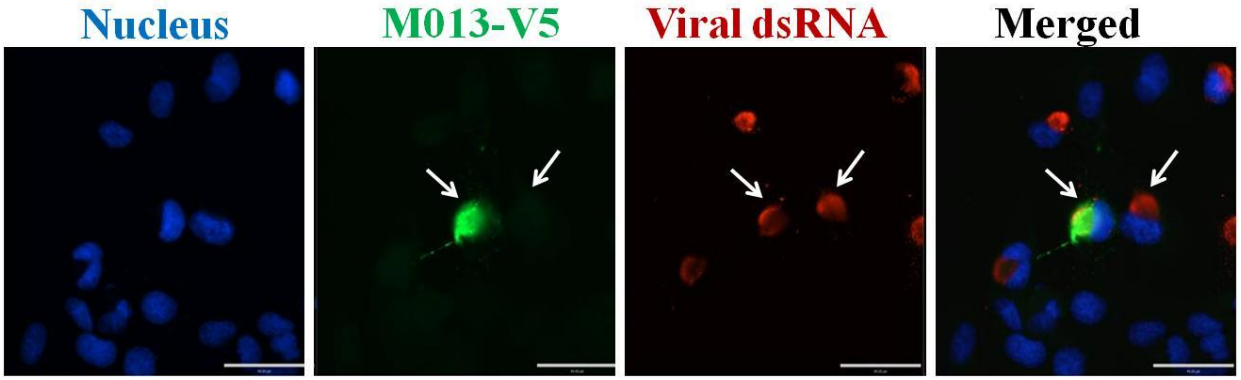


Figure 4.1. Overexpression of M013 did not negatively affect SHFV RNA replication in MA104 cells.
MA104 cells were seeded on cover slips in a 24-well plate and transfected with 0.5 µg pEF6-M013-V5 plasmid DNA using Lipofectamine LTX/PLUS. At 24 h post transfection, cells were infected with SHFVic-WT virus at an MOI of 1 for 24 h and then fixed, permeabilized and incubated with anti-V5 antibody (green), anti-dsRNA antibody (red) and Hoechst 33342 (blue). Cells were imaged with Zeiss Axio Observer 1 microscope using a 40x objective. White arrows indicate two neighboring cells, one expressing M013 protein and the other not. Scale bar, 82 µm.

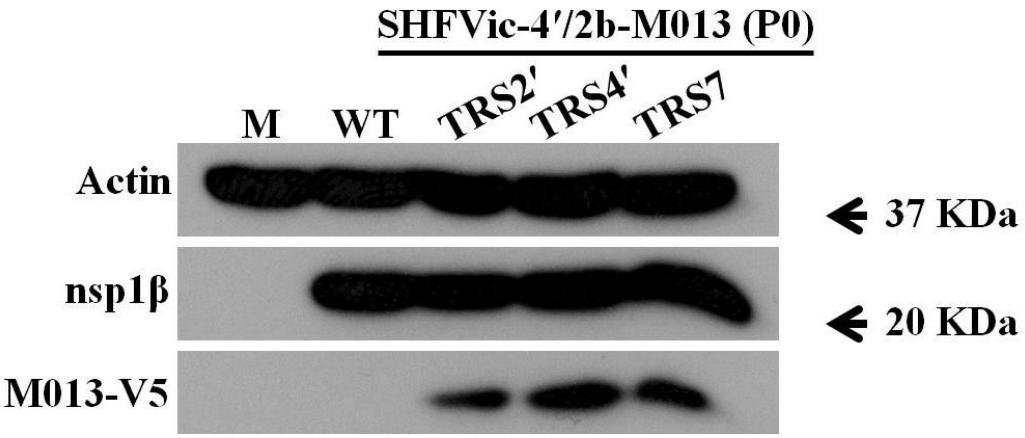


Figure 4.2 Western blot analysis of protein expression in cells transfected with one of the three SHFVic-4′/2b-M013-TRS virus RNAs.
MA104 cells were either mock transfected (M), transfected with SHFVic-WT RNA (WT) or transfected with one of the three SHFVic-4′/2b-M013-TRS viral RNAs. At 5 days after transfection, cell lysates were harvested in RIPA buffer and subjected to Western blotting using anti-beta actin, anti-nsp1β and anti-V5 antibodies.

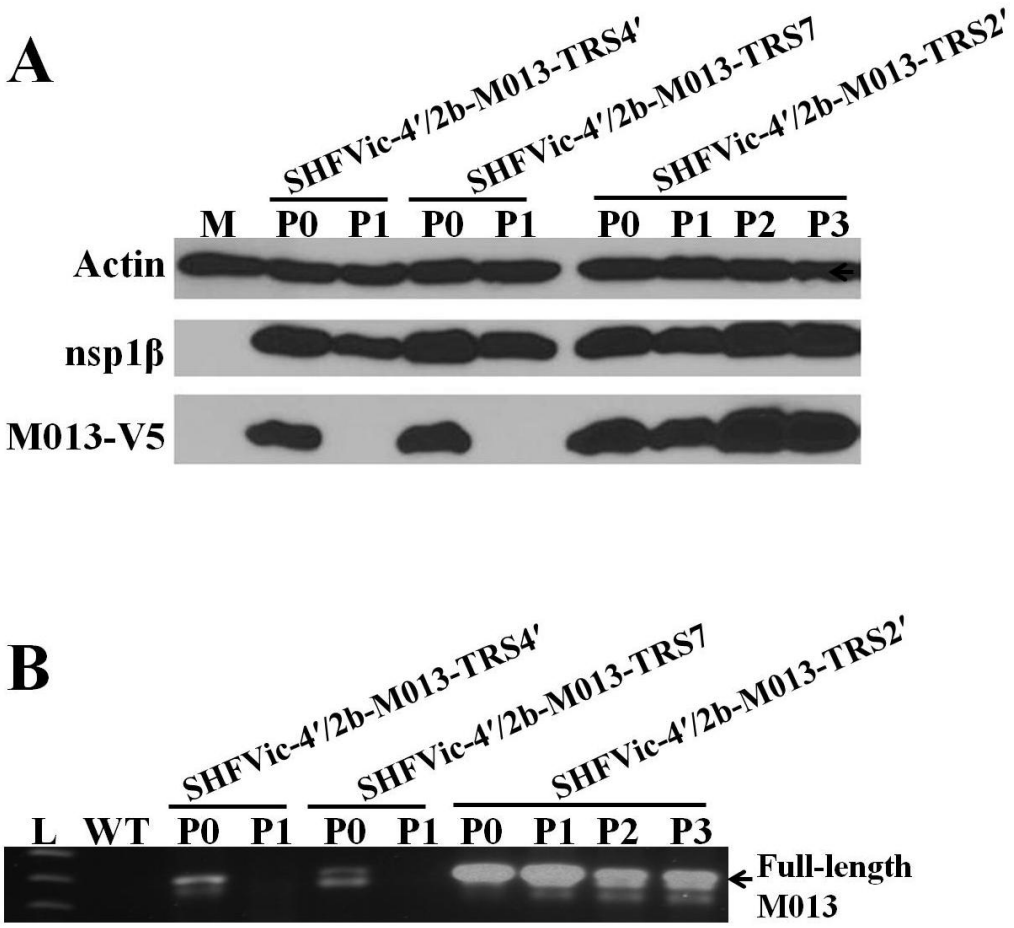


Figure 4.3. Effect of virus passage on M013 expression from the SHFVic-M013 viruses. For each SHFVic-4′/2b-M013-TRS virus, MA104 cells were either transfected with *in vitro* transcribed viral genome RNA (P0) or infected with virus from different passages (P1 to P3). Mock transfection (M) and MA104 cells transfected with SHFVic-WT RNA (WT) were used as controls. (A) At 24 hpi, cell lysates were harvested and subjected to Western blot analysis for the expression of viral nsp1β and M013 protein using anti-nsp1β and anti-V5 antibodies, respectively. Beta actin was used as a loading control. (B) Culture fluid was also collected at 24 hpi and the viral genome RNA was extracted and subjected to RT-PCR to amplify the M013 inserted region. The RT-PCR products were separated on a 1% agarose gel and imaged under UV light. L, DNA ladder.

.

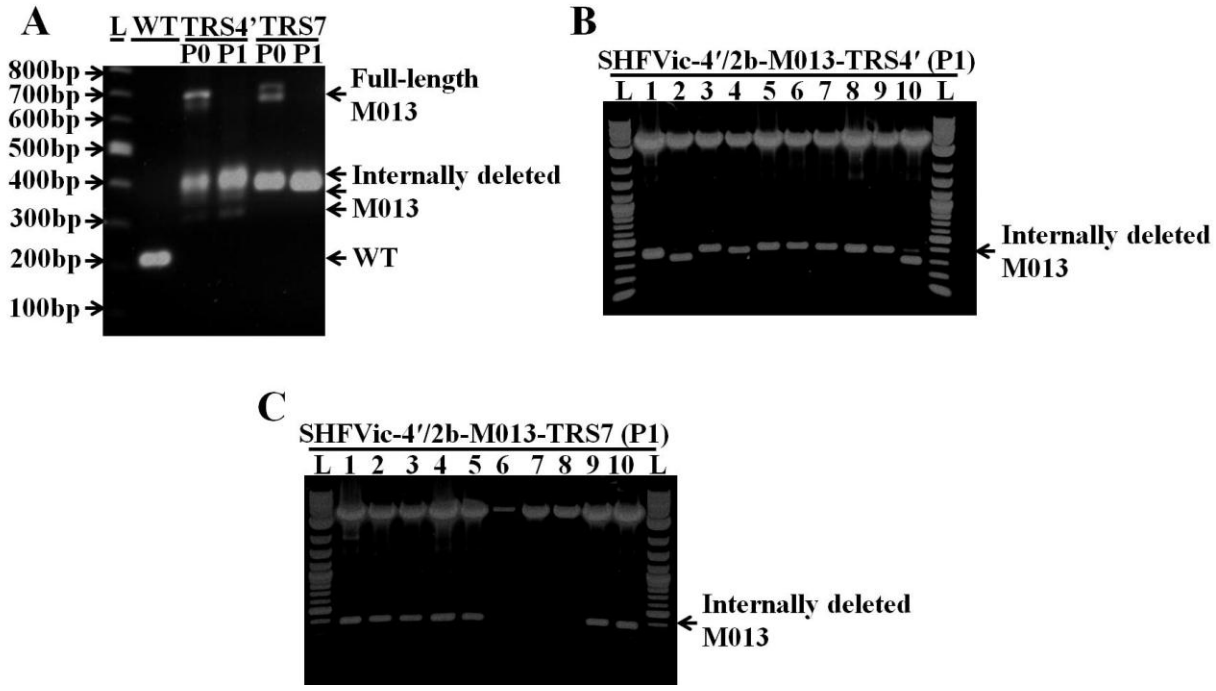


Figure 4.4. Analysis of M013 inserts in the genomes of SHFVic-M013 viruses after passage. (A) For each SHFVic-4′/2b-M013-TRS virus, MA104 cells were either transfected with *in vitro* transcribed viral genome RNA (P0) or infected with P1 virus for 24 h. MA104 cells transfected with SHFVic-WT RNA (WT) were used as a negative control. Viral genome RNA was extracted from the collected culture fluid and subjected to RT-PCR to amplify the region containing the inserted M013 gene. The RT-PCR products were separated on a 1% agarose gel and imaged under UV light. The RT-PCR products amplified from the WT genome, full-length M013 insert and internally deleted M013 inserts are indicated by black arrows. The extracted RT-PCR products were subjected to TA cloning and 10 clones were digested with EcoRI and the inserts separated on a 1% agarose gel for P1 SHFVic-4′/2b-M013-TRS4′ virus (B) and P1 SHFVic-4′/2b-M013-TRS7 virus (C) The inserts representing the internally deleted M013 gene are indicated by black arrows. L, ladder

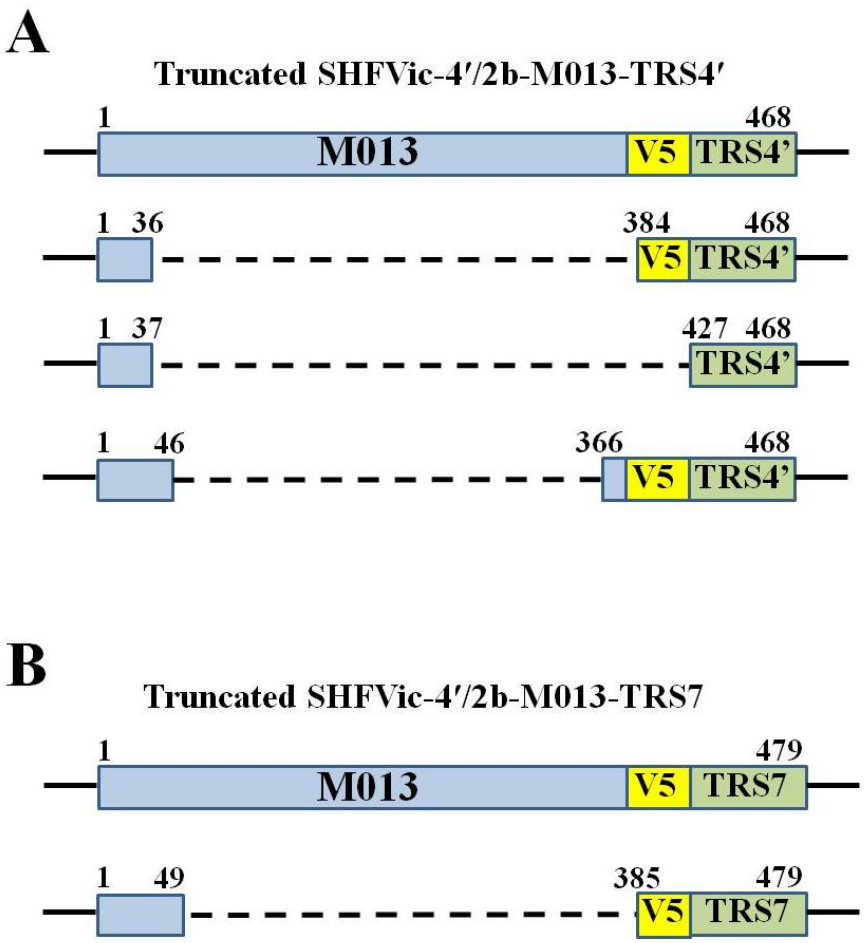


Figure 4.5. Diagrams of internal deletions in the M013-V5-TRS sequence in recombinant SHFVic genomes.
(A) Three different internal deletions were found in P1 SHFVic-4′/2b-M013-TRS4′ genomes. (B) One internal deletion was found in the genome of P1 SHFVic-4′/2b-M013-TRS4′. The M013 coding sequence is indicated by a blue box. The V5 tag is indicated by a yellow box and the body TRS is indicated by a green box. The numbers above the boxes indicate the nt boundaries of the inserts and internal deletions.

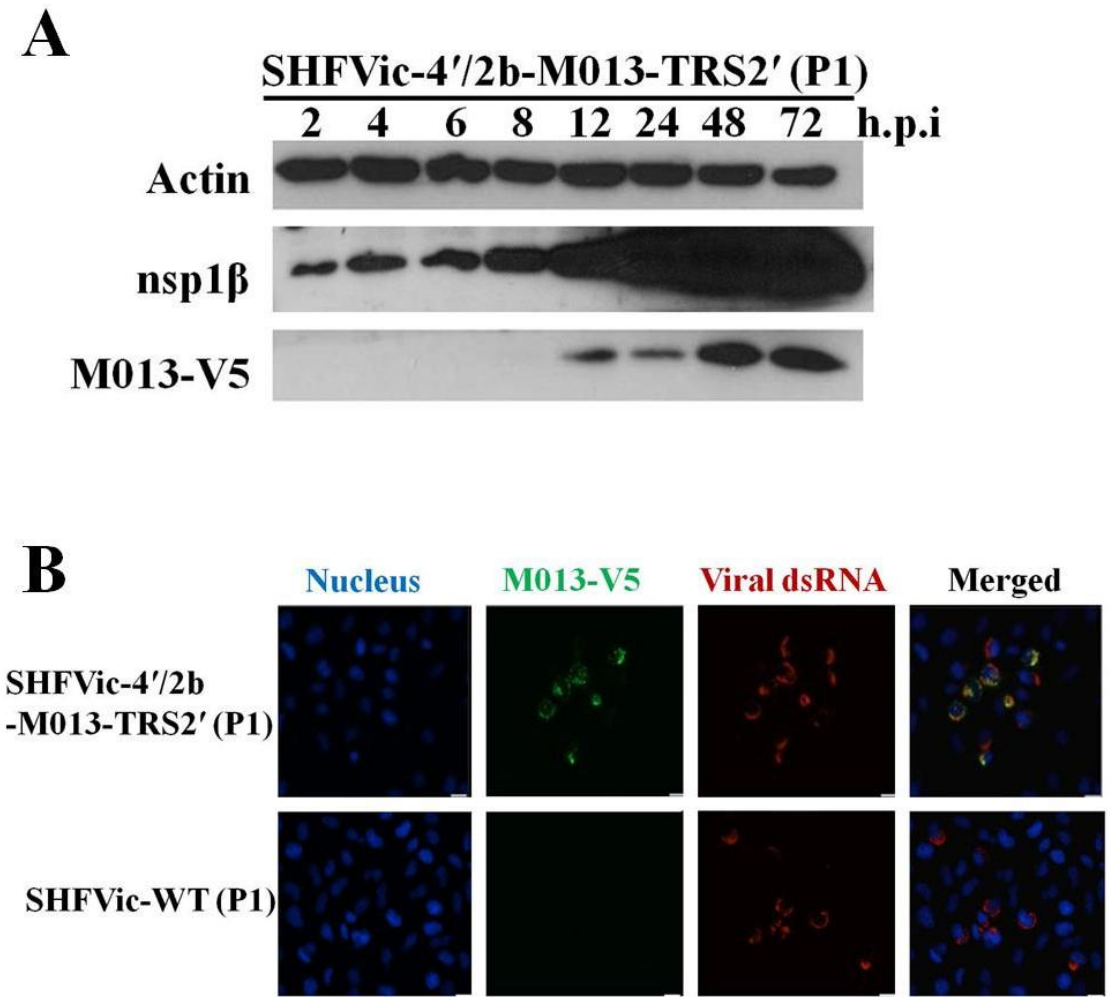


Figure 4.6. Analysis of the M013 expression pattern from the SHFVic-4′/2b-M013-TRS2′ virus. (A) MA104 cells were infected with P1 SHFVic-4′/2b-M013-TRS2′ virus at an MOI of 1. At different times after infection, cell lysates were harvested in RIPA buffer and subjected to Western blot analysis for M013 expression using anti-V5 antibody. The expression of viral nsp1β protein and beta actin were also detected with protein-specific antibodies. (B) MA104 cells seeded on cover slips were infected with SHFVic-WT virus or P1 SHFVic-4′/2b-M013-TRS2′ virus at an MOI of 1. At 24 hpi, cells were fixed, permeabilized, blocked and processed for IFA using anti-V5 antibody (green), anti-dsRNA antibody (red) and Hoechst 33342 (blue). The cells were imaged with Zeiss Axio Observer 1 microscope using a 63x oil emersion objective. Scale bar, 25μm.

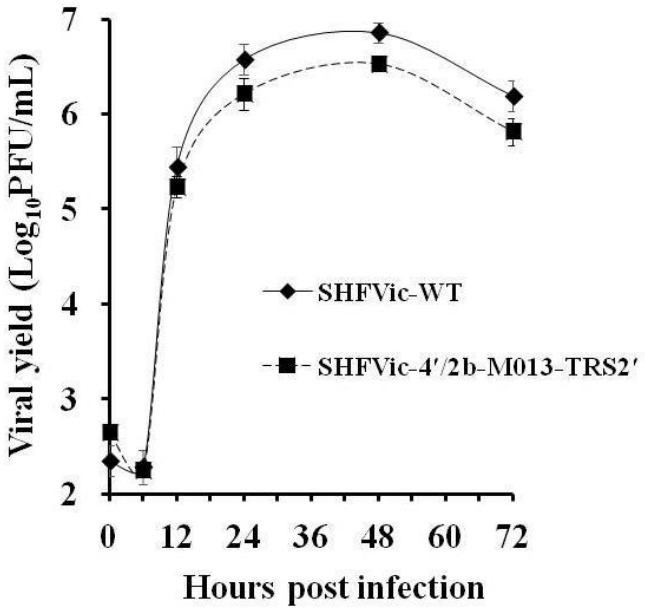


Figure 4.7. Growth curves of SHFVic-WT virus and SHFVic-4′/2b-M013-TRS2′ virus in MA104 cells.

MA104 cells were infected with SHFVic-WT virus or P1 SHFVic-4′/2b-M013-TRS2′ virus. At different times after infection, culture fluid was collected and subjected to plaque assay on MA104 cells. Average values are shown for triplicate biological samples that were each titered in duplicate.

5 CONLUSIONS AND FUTURE DIRECTIONS

SHFV is a simian arterivirus and has the largest genome among arteriviruses. The 5′ end of the SHFV genome encodes two polyproteins that are autocleaved into non-structural proteins by the viral proteases PLP1, PLP2 and SP (Snijder, Kikkert et al. 2013). The 3′ end of the genome encodes structural proteins. Compared to other arteriviruses that encode one or two PLP1 domains at the 5′ end of the genome, SHFV encodes three functional PLP1 domains, PLP1α, PLP1β and PLP1γ. The catalytic residues and cleavage sites were determined for each of the three PLP1s by members of the Brinton lab. The PLP1s of other arteriviruses have been shown to only *cis* cleave at a single site downstream of the catalytic domain (Snijder, Wassenaar et al. 1992; den Boon, Faaberg et al. 1995). Interestingly, SHFV PLP1γ was the first arterivirus PLP1 shown to cleave at sites located both downstream and upstream of its own catalytic domain in *in vitro* cleavage assays. The observation that SHFV PLP1β could cleave within its catalytic domain *in vitro* was also surprising. These findings suggest that SHFV PLP1β and PLP1γ may have *trans* cleavage activity. The PLP1s of coronaviruses have been shown to cleave in *trans* at a site upstream of their catalytic domains *in vitro* (Herold, Gorbalenya et al. 1998; Teng, Pinon et al. 1999). The possibility of *trans* cleavage by SHFV PLP1β and PLP1γ could be tested by separately expressing PLP1β or PLP1γ and a substrate peptide in an *in vitro* cleavage assay. Three SHFV nsp1 proteins are produced after autocleavage by PLP1s. The intracellular location of each nsp1 protein is not clear. IFA analysis of each overexpressed SHFV nsp1 protein in MA104 cells is underway. The nsp1 proteins of other arteriviruses have been shown to antagonize IFN induction and signaling pathways (Beura, Sarkar et al. 2010; Chen, Lawson et al. 2010). A preliminary study showed that SHFV infection inhibits poly(I:C) induced IFN-β gene upregulation (Han and Brinton, unpublished data). The function of each SHFV nsp1 protein in

IFN induction and signaling pathways needs to be analyzed. Crystal structures for PRRSV nsp1α and nsp1β have been obtained (Sun, Xue et al. 2009; Xue, Sun et al. 2010). Homology modeling of each SHFV nsp1 protein on PRRSV nsp1α and nsp1β structures suggested that SHFV nsp1γ is more similar to PRRSV nsp1α than to PRRSV nsp1β. However, crystal structures of each of the SHFV nsp1 proteins are needed to obtain accurate information about the SHFV PLP1s.

SHFV was predicted to encode 12 structural proteins within the 3′ end of the genome (Snijder, Kikkert et al. 2013). However, only eight SHFV sg mRNAs had been identified to express these structural proteins (Godeny, de Vries et al. 1998). In the present study, a total of 37 additional sg mRNAs transcribed from minus strand templates generated from newly identified functional body TRSs in the SHFV genome were discovered after an in-depth analysis of the sg mRNA leader-body junction sequences. The majority of these additional sg mRNAs are alternative sg mRNAs that encode the same structural proteins as the previously identified eight sg mRNAs. A previous study on EAV suggested that alternative sg mRNAs function as back-ups to maintain the production of infectious virus when the major sg mRNA is not produced (Pasternak, Gultyaev et al. 2000). To test this hypothesis in SHFV, viral production will need to be measured after the transcription of each major sg mRNA is abolished by mutating the corresponding TRSs.

The arterivirus sg mRNAs are a nested set of 3′ and 5′ co-terminal transcripts. Although each SHFV sg mRNA contains all the downstream ORFs, only the 5′ proximal ORF is typically expressed based on a monocistronic expression model. However, bicistronical expression had not been ruled out because separate sg mRNAs encoding some of the structural proteins as the 5′ proximal ORF had not been found (Godeny, de Vries et al. 1998). SHFV sg mRNA2′, sg mRNA2 and sg mRNA5 were thought to express GP2′/GP3′/E′, GP2/E and GP5/ORF5a,

respectively (Godeny, de Vries et al. 1998; Snijder, Kikkert et al. 2013). In the present study, at least one sg mRNA was found that encodes each SHFV structural protein as the 5′ proximal ORF. These findings suggest that all of the SHFV structural proteins can be monocistronically expressed. Previous data showed that both E and E′ expression were required for infectious progeny virus production (Vatter, Di et al. 2014). To test the possibility of bicistronic expression, the transcription of the sg mRNAs monocistronically expressing E′, E or ORF5a from the 5′ proximal ORF will be abolished by mutating the corresponding TRSs, if the lack of expression of the corresponding protein occurs, production of infectious progeny virus would be inhibited and this observation would rule out bicistronic expression.

Although the previous study showed that both sets of minor structural proteins are involved in infectious progeny virus production (Vatter, Di et al. 2014), the exact function(s) of each minor structural protein is unknown. Another previous study showed that the EAV GP2, GP3 and GP4 form heterotrimers in the viral envelope (Wieringa, de Vries et al. 2003). The components and structural morphology of the SHFV minor structural protein complexes are not yet known. The present study showed that the expression patterns of the members of the two minor structural protein sets were quite different, suggesting that the members of each set of minor structural proteins may not form separate complexes and function as two independent units. A few of the additional sg mRNAs encode a C-terminal peptide of a structural protein. Two of these C-terminal peptides, ORF5-C and ORF6-C, may be functionally relevant because mutating the start codon of ORF5-C or ORF6-C, individually, reduced the SHFV yield in MA104 cells. Although unlikely, the possibility that this reduction was due to the Met to Leu change in the full-length ORF5 or ORF6 protein was not ruled out. To confirm the production of the ORF5-C and ORF6-C proteins during natural infection, two peptide-specific antibodies were

recently produced and purified by peptide-affinity purification. SHFV infected cell lysates will be subjected to Western blot analysis with these antibodies to access the production of ORF5-C and ORF6-C in infected cells. To confirm that the reduction in virus yield is due to the lack of function of ORF5-C, a compensation experiment could be done by expressing the wild type ORF5-C from the SHFVic-ΔORF5-C mutant genome to see whether this restores the virus yield to wild type levels. The same strategy could be used to confirm the functional relevance of ORF6-C. Future studies are also needed to identify the functions of ORF5-C and ORF6-C in progeny virus production and host-virus interactions. Verification of their production and functional studies of these C-terminal peptides will expand knowledge about the arterivirus coding capacity.

NGS has been widely used to analyze the cellular transcriptome response against arterivirus infection (Miller, Fleming et al. 2012; Badaoui, Rutigliano et al. 2014; Sang, Brichalli et al. 2014). However, it has never been used to study the arterivirus transcriptome probably due to the unique arterivirus nested set of 3′ transcripts. In the present study, the first NGS workflow to analyze the SHFV transcriptome in infected cells was designed and consisted of sequentially mapping screened reads to the unique leader-body junction sequence of each known sg mRNA. Exactly how arteriviruses regulate their sg mRNA production is still unclear. This regulation can be divided into two steps: syntheses of minus-strand sg RNA templates, which are discontinuous and guided by the TRS, and transcription of sg mRNAs from the minus-strand templates. The SHFV transcriptome results confirmed the published hypothesis that in addition to sequence homology between leader and body TRSs, TRS location in the genome and the length of the minus-strand sg RNA templates, other factors also play a role in regulating the abundance of the sg mRNAs (van Marle, Dobbe et al. 1999; Pasternak, Gultyaev et al. 2000; Pasternak, Spaan et

al. 2004). Also, the observation that the relative abundance of each sg mRNA stays constant at early and late times after infection indicates that the regulation is consistent during the infection cycle. Interestingly, the relative abundance of particular sg mRNAs differed between infected MA104 cells and macaque MΦs, suggesting that cell type-specific host factors may also play a role in regulation. In the present NGS study, only the SHFV genome RNA and sg mRNAs were sequenced due to a poly(A) selection step before library preparation. To further analyze how the syntheses of minus-strand sg RNA templates are regulated, the abundance of each minus-strand sg RNA could be analyzed by NGS of total RNA extracted from the infected cells.

The current workflow for SHFV transcriptome analysis is based on the known sg mRNAs generated during infection. Mapping of the remaining reads (60% of the total reads) suggested the existence of a large amount of additional sg mRNAs generated from different locations in the 3′ end of the genome, especially in the ORF7 region. Also, although much rarer, sg mRNAs could be generated from the 5′ ORF1a/1b region as well, especially from the 3′ end of ORF1b. The TRSs used for generating these unknown sg mRNAs need to be analyzed and their abundance calculated to expand current knowledge of the SHFV transcriptome. The large number of unknown SHFV sg mRNAs indicates that the capacity and diversity of arterivirus sg mRNA production has been greatly underestimated.

The potential of SHFV, the largest arterivirus, as an expression vector was explored and the first recombinant SHFVic virus was made expressing the foreign gene M013 from an additional sg mRNA. Two possible locations were tested for insertion of this foreign gene in the SHFV genome. Insertion between the two sets of minor structural protein ORFs, rather than between the major and minor structural protein ORFs, resulted in generation of infectious progeny virus. The data indicated that the location where natural recombination had occurred

during evolution to acquire a second set of minor structural proteins was also tolerant for artificial insertion of a foreign gene. However, although recombinant SHFV progeny virus was produced from each of three different infectious clone constructs, only progeny virus produced from one of these constructs maintained M013 expression through several passages. This may be due to the additional transcription cassette disrupting the optimal folding of the viral RNA genome. An alternative strategy would be fusion of the foreign gene to a SHFV structural or non-structural protein using the 2A peptide as a linker to avoid introduction of an additional transcription cassette. Also, a small fluorescent Light, Oxygen, or Voltage-sensing (LOV) flavoprotein domain, which is only half the size of green fluorescent protein, could be inserted into the SHFV genome to generate the first auto-fluorescent SHFV (Chapman, Faulkner et al. 2008; Christie, Hitomi et al. 2012).

6 ADDITIONAL WORK: CHARACTERIZATION OF HUMAN OAS1 ISOFORMS

6.1 Introduction

Oligoadenylate synthetase (OAS) is a double-stranded RNA (dsRNA) sensor that functions as part of the innate immune response against virus infection. After binding to viral dsRNA or dsRNA regions in the viral single-stranded RNA (ssRNA), activated OAS polymerizes ATP into short 2′-5′ linked oligoadenylates (2′-5′A). The 2′-5′A trimers and higher order 2′-5′A can bind to endoribonuclease L (RNase L), inducing dimerization and activation. Activated RNase L cleaves viral and cellular ssRNAs including 28S and 18S ribosomal RNA (rRNA) (Dong, Xu et al. 1994; Player and Torrence 1998; Kristiansen, Gad et al. 2011). OAS genes are evolutionally conserved in vertebrates (Kjaer, Poulsen et al. 2009; Kristiansen, Gad et al. 2011). In human genome, the OAS family consists of one copy each of the OAS1, OAS2, OAS3 and OASL genes. The OAS1, OAS2 and OAS3 proteins differ in the number of OAS domains they contain. Human OASL contains an inactive OAS domain plus two domains of ubiquitin-like sequences (Marie and Hovanessian 1992; Hartmann, Olsen et al. 1998; Hovanessian and Justesen 2007). The human OAS1 gene can produce five isoforms, p42, p44, p46, p48 and p52 by alternative splicing. The OAS1 p42 and p44 isoforms are predicted to be expressed in humans of all genotypes (Justesen, Hartmann et al. 2000; Bonnevie-Nielsen, Field et al. 2005). The generation of either p46 or p48/p52 depends on the allele of the single nucleotide polymorphism (SNP) rs10774671 located at the splicing acceptor site of the exon 7 in the OAS1 gene (Bonnevie-Nielsen, Field et al. 2005). Preliminary data obtained by the Brinton lab suggested that additional human OAS1 splicing isoforms may exist (Perelygin and Brinton, unpublished data). The exon1 to exon5 sequence of all of the predicted human OAS1 isoforms are the same and contain an intact catalytic domain, but the C-terminus sequence varies. Both the

crystal structures of porcine OAS1 alone and of dsRNA bound human OAS1 p46 suggested a bilobal structure in which the C-terminus folds toward the N-terminus (Hartmann, Justesen et al. 2003; Donovan, Dufner et al. 2013). The distinct C-terminus of each OAS1 isoform therefore could affect the overall folding of the protein which could result in different synthetase activity or binding partners for the different isoforms.

Preliminary data from the Brinton lab showed that in response to poly(I:C) activation, recombinant human OAS1 isoforms, p42, p44, p46 and p48, can synthesize 2′-5′A trimers and higher order 2′-5′A *in vitro* (Elbahesh and Brinton, unpublished data). The synthetase activity of the human OAS1 p52 isoform was not tested *in vitro* in this previous study. However, a previous publication reported that in response to dengue virus infection in cells, only the overexpressed OAS1 p42 and p46 isoforms, but not p44, p48 and p52 isoforms, activated RNase L cleavage activity (Lin, Yu et al. 2009). The discrepancy in the OAS1 activities observed *in vitro* and in dengue virus infected cells suggested that either the p44 and p48 isoforms are not active when expressed in mammalian cells or they are active but cannot be activated by dengue virus RNA.

Dengue virus belongs to the family *Flaviviridae,* genus *Flavivirus.* Mice display genetically controlled differential susceptibility to flavivirus-induced disease. Although there are 8 *Oas1* genes (*Oas1a-Oas1h*) in the mouse genome, only Oas1a and Oas1g are known to be active synthetase. Although murine Oas1b has a four amino acid deletion in its catalytic domain and is inactive, it was identified as the flavivirus resistance gene and the resistance mechanism was shown not to be dependent on the Oas-RNase L pathway (Perelygin, Scherbik et al. 2002; Scherbik, Paranjape et al. 2006; Elbahesh, Jha et al. 2011). Previous studies in Brinton lab identified murine oxysterol-binding protein homologue (Orp1L) and ATP-binding cassette, sub-family F, member 3 (Abcf3) as binding partners for murine Oas1b and Abcf3 was shown to play

a role in flavivirus-specific resistance in mouse cells (Courtney, Di et al. 2012). Among human OAS1 isoforms, p48 isoform was reported to have proapoptotic activity by interacting with the anti-apoptotic proteins Bcl-2 and Bcl_{XL} (Ghosh, Sarkar et al. 2001).

In the present study, the human p52 isoform was expressed, purified and shown to be an active synthetase *in vitro.* Human OAS1 p42, p44, p46, p48 and p52 isoforms were also overexpressed individually in mammalian cells and each of the overexpressed OAS1 isoform was shown to activate RNase L cleavage in response to poly(I:C) transfection, suggesting they are all functionally active in cells. In addition, using a yeast two hybrid assay and an *in vitro* pull-down assay, a novel binding partner supervillin (SVIL) was identified for human OAS1 p44 isoform and another novel binding partner fibrillin-1 (FBN1) was identified for the p42 isoform. The interaction between overexpressed p44 and endogenous full-length SVIL was confirmed by a co-immunoprecipitation assay in mammalian cells

6.2 Materials and methods:

6.2.1 Cells

Human embryonic kidney HEK293 cells were cultured at 37°C in a 5% CO2 atmosphere using 1 x Dulbecco's Modified Eagle Medium (DMEM) supplemented with 1% L-glutamine, 5% fetal bovine serum and 0.1% gentamicin. Human lung carcinoma A549 cells were cultured at 37°C in a 5% CO_2atmosphere using 1 x F-12K nutrient media supplemented with 1% L-glutamine, 10% fetal bovine serum and 0.1% gentamicin.

6.2.2 Protein purification

Human OAS1 isoform p52 cDNA was inserted in a pET151-TOPO expression vector (Thermo Fisher Scientific) with a 6X His tag and a V5 epitope fused at the N-terminus and transformed into One Shot BL21(DE3)-pLysS chemically competent cells (Thermo Fisher Scientific) for expression. The transformed cells were grown in LB media supplemented with 0.05% glucose and 100 μg/ml of carbenicillin (CRB) until the OD_{600} reached ~0.8-0.9. The cells were then induced with IPTG (1mM) overnight at 16 ℃. The induced cells were centrifuged at 4 ℃ for 10 min at 6,000 x g and the pellets were resuspended in 1 X Equilibration buffer [NaCl (300 mM), sodium phosphate (50 mM) and 1 X Complete Mini EDTA-Free Protease inhibitor cocktail (Roche)]. To lyse the cells, the resuspended pellets were mixed with cellytic express lysis powder (Sigma) and shaked at 37 ℃ for 30 min, followed by centrifugation at 4 ℃ for 10 min at 15,000 x g. The supernatant was transferred to a column containing TALON metal affinity resin (Clontech) for protein binding. After washing the column with 1 X washing buffer [NaCl (300 mM), sodium phosphate (50 mM) and 5 mM imidazole], the bound proteins were eluted with 1 X Elution buffer [NaCl (300 mM), sodium phosphate (50 mM) and 150 mM imidazole] and the buffer was exchanged with 1 X Storage buffer [KCl (50 mM), $Mg(OAc)_2$ (25 mM), HEPES-KOH (20 mM), β-mercaptoethanol (7 mM), EDTA (0.03 mM), glycerol (0.25%) and 1 X Complete Mini EDTA-Free Protease inhibitor cocktail (Roche)]. After concentration using a Microcon-10 kDa Centrifugal Filter Unit (Millipore), the purified protein was aliquoted and stored at -80 ℃.

6.2.3 In vitro OAS1 synthesis assay

Purified OAS protein was mixed with α^{32}p-ATP (15 µCi) and poly(I:C) (50 ng/µl) in 1 X Assay buffer [KCl (50 mM), $Mg(OAc)_2$ (25 mM), HEPES-KOH (20 mM), β-mercaptoethanol (7 mM), ATP (5 mM), creatine phosphate (10 mM) and creatine kinase (1U/µl)]. The reaction was incubated at 30°C for 18 h before adding an equal amount of Gel Loading Buffer II (Amibion). An aliquot of the reaction was separated on a 20% polyacrylamide/Urea gel at 800V for 3.5 h and the production of 2′-5′A was detected by autoradiography.

6.2.4 Functional analysis of human OAS1 isoforms in mammalian cells

Human OAS1 p42, p44, p46, p48 and p52 isoform cDNAs were inserted into the p3xFlag-CMV mammalian expression vector with the 3X Flag tag fused at the N-terminus. HEK293 cells were seeded in a 6-well plate and grown to ~70% confluence before transfection with either empty vector or with one of the human OAS1 isoform plasmid DNAs using Lipofectamine LTX/PLUS reagent (Thermo Fisher Scientific). At different times after transfection, cells were transfected with poly(I:C) for 6 h using Lipofectamine 2000 (Thermo Fisher Scientific) before harvesting the cell lysates in TRI reagent (Molecular Research Center. Inc.). Total intracellular RNA was extracted and purified following the manufacturer's protocol and separated on a denaturing formaldehyde/MOPS agarose gels. The RNA gel was stained in ethidium bromide and imaged under UV light.

6.2.5 Yeast two hybrid assay

A yeast two hybrid assay was done using the Matchmaker Gold Yeast Two-Hybrid System following the manufacturer's protocol (Clontech). Briefly, full length human OAS1-p42

or OAS1-p44 isoform was fused with the GAL4-BD domain in the pGBKT7 vector and transformed into the Y2HGold yeast strain as the bait. A Y187 yeast strain containing a universal human cDNA library fused to the GAL4-AD domain in a pGADT7-RecAB vector was purchased from Clontech and used as the prey. The Y2HGold yeast strain was mated with the Y187 yeast strain at 30 ℃ for 24 h and plated on double dropout (DDO, minus Trp and Leu) plates for selection of mated diploid cells. The diploid cells were then plated on triple dropout plates supplemented with Aureobasidin A (TDO/A, minus His, Trp and Leu) and quadruple dropout plates supplemented with Aureobasidin A (QDO/A, minus Ade, His, Trp and Leu) for selection of positive clones that contained interacting prey and bait. The prey and bait plasmid constructs were extracted from the yeast colonies grown on QDO/A plates and sequenced. A BLAST search was then performed to determine the identity of the putative binding peptide.

6.2.6 Yeast co-transformation

The plasmid expressing the putative binding peptide was transformed into the Y187 yeast strain. The transformed Y187 strain was then mated with the Y2HGold strain containing the human OAS1 bait at 30 ℃ for 24 h. The mated diploid cells were next plated on DDO, TDO/A and QDO/A plates to confirm the co-transformation as well as the positive interaction between the bait and the putative binding peptide.

6.2.7 In vitro transcription/translation and pull-down assay

The full-length human OAS1 bait and the putative binding peptide were individually cloned into a pTNT expression vector (Promega) with a c-myc tag or a HA tag fused at the N-terminus, respectively. The resulting bait and prey constructs were expressed *in vitro* in the

presence of [^{35}S]-methionine using a TnT coupled transcription/translation system (Promega) following the manufacture's protocol. The *in vitro* translated bait and prey peptide were incubated together for 1 h at room temperature and divided into three portions. One portion was saved at -20 ℃ as a lysate sample, one portion was incubated with control agarose beads conjugated with IgG antibody and one portion was incubated with agarose beads conjugated with either anti-c-myc or anti-HA antibody. After an overnight incubation at 4 ℃ with rotation, the beads were washed with 1 X Lysis buffer [Triton X-100 (1%), SDS (0.1%), NaCl (150 mM), Tris-HCl (50 mM) and HALT protease and phosphatase inhibitor cocktail (Pierce Scientific)]. The washed beads were suspended in 2 X sample buffer [(SDS (20%), glycerol (25%), Tris-HCl (0.5 M), bromophenol blue (0.5%) and β-mercaptoethanol (5%)] and boiled for 5 min. The pulled-down protein complexes and the proteins in the original lysate sample were separated by 10% SDS-PAGE. The gels were then incubated in a fixing solution (10% acetic acid and 30% methanol), followed by incubation in Autofluor (National Diagnostics) and then in anti-cracking buffer (7% acetic acid, 7% methanol and 1% glycerol). The gel was dried onto a 3 mm chromatography paper (Whatman) and autoradiographed.

6.2.8 Mammalian cell co-immunoprecipitation

Human lung carcinoma A549 cells were seeded in a 10-cm plate and transiently transfected with the p3xFlag-CMV-OAS1-p44 construct. At 48 h after transfection, cell lysates were harvested in 1 X Lysis buffer [Triton X-100 (1%), SDS (0.1%), NaCl (150 mM), Tris-HCl (50 mM) and HALT protease and phosphatase inhibitor cocktail (Pierce Scientific)]. The cell lysates were divided into three portions. One portion was saved at -20 ℃ as a lysate sample, one

portion was incubated with control agarose beads conjugated with IgG antibody and one portion was incubated with agarose beads conjugated with anti-Flag antibody. After an overnight incubation at 4 ℃ with rotation, the beads were washed with 1 X Lysis buffer and suspended in the 2 X Sample buffer before denaturing at 95 ℃ for 5 min. The immunoprecipitated proteins were separated by 8% SDS-PAGE, transferred to a nitrocellulose membrane and blocked with 5% non-fat dry milk at room temperature for 1 h. The blocked membrane was then cut into strips and the strips were incubated with either anti-SVIL (Sigmal, 1:1000 dilution) or anti-Flag (Sigma, 1:1000 dilution) antibody overnight at 4 ℃. After washing with 1 X Tris-buffered saline containing 0.1% Tween 20, the membrane was incubated with anti-mouse or anti-rabbit secondary antibody (Cell signaling, 1:2000 dilution) for 1 h at room temperature, followed by washing and development with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

6.2.9 Immunofluorescent assay (IFA)

HEK293 cells were seeded on cover slips in a 24-well plate and grown until ~70-80% confluent. Cells were either transfected with 1 µg of empty vector or with 1 µg of one of the human OAS1 isoforms cloned into p3XFlag-CMV vector using lipofectamine LTX/PLUS reagent following the manufacturer's protocol (Thermo Fisher Scientific). At 24 h after transfection, the cells were fixed with 4% PFA for 10 min, permeabilized with 0.1% Triton-X for 10 min and then blocked with 5% horse serum for 1 h at room temperature. After incubation with mouse anti-Flag antibody (Sigma, 1:500) overnight at 4°C, cells were incubated with Alexa Fluor 488-donkey anti-mouse antibody (Thermo Fisher Scientific, 1:400) and Hoechst 33342 (Thermo Fisher Scientific, 0.05%) for 1 h at room temperature. The cells on the cover slips were

then mounted on a glass slide and imaged with a Zeiss Axio Observer 1 microscope using a 25X objective.

6.2.10 Western Blot Assay

Transfected HEK293 cell lysates were harvested in 1 X RIPA buffer [1 X phosphate-buffered saline, 1% Nonidet P-40, 0.5% sodium deoxycholate, and 0.1% SDS and 1 X Halt protease inhibitor cocktail (Thermo Scientific)] and same amount of total protein was separated on a 10% SDS-PAGE gel and transferred to a nitrocellulose membrane at 100 V for 1 h. The membrane was incubated in blocking buffer (1 X Tris-buffered saline containing 5% non-fat dry milk and 0.1% Tween 20) at room temperature for 1 h and cut into strips. The strips were incubated at 4^{o}C overnight with mouse anti-Flag antibody (Sigma, 1:1000) or mouse anti-β actin antibody (Cell signaling, 1: 10,000), followed by washing three times, 10 min each with 1 X Tris-buffered saline containing 0.1% Tween 20. The washed membrane strips were incubated with horseradish peroxidase-conjugated anti-mouse antibody (Santa Cruz Biotechnology) for 1 h at room temperature. After washing, the membrane strips were then developed with a Super-Signal West Pico detection kit (Pierce) following the manufacturer's protocol.

6.3 Results:

6.3.1 Human OAS1 p52 is functionally active in vitro

Previous data obtained in the Brinton lab showed that the human OAS1 isoforms p42, p44, p46 and p48 are functionally active *in vitro* in response to poly(I:C). However, human OAS1 p52 isoform was not tested at that time. In the present study, cloned p52 isoform was

expressed in bacteria, partially purified *in vitro* and incubated with poly(I:C) and α-^{32}P-ATP for 18 h at 30 ℃. The reaction samples were then separated on a 20% PAGE/8M urea gel, visualized by autoradiography to assess the pattern of the synthesized 2′-5′A. The human OAS1 p46 isoform and the LacZ protein were purified using the same protocol used for the p52 isoform and used as the positive control and negative control, respectively. The substrate α-^{32}P-ATP was run alone to indicate the position of free ATP on the gel. As expected, no 2′-5′A synthesis was detected for LacZ or the free ATP controls. The human p52 isoform synthesized higher order 2′-5′A (2′-5′A trimer and above) similar to the p46 isoform (Figure 6-1), indicating that the human OAS1 p52 isoform is an active synthetase.

6.3.2 Overexpressed human OAS1 isoforms activate RNase L cleavage after transfection of cells with poly (I:C)

Although the human OAS1 isoforms p42, p44, p46, p48 and p52 can synthesize higher order 2′-5′A *in vitro*, their functionalities in cells are not well characterized. A previous study in human A549 cells claimed that overexpression of the human OAS1 p42 or p46 isoform activated RNase L cleavage in response to Dengue virus infection whereas the p44, p48 and p52 isoforms did not (Lin, Yu et al. 2009). This observation suggested that either the p44, p48 and p52 isoforms are inactive synthetases in cells or they are functionally active but cannot be activated by dengue virus RNA. To test whether each of the five human OAS1 isoforms is functionally active in cells, the p42, p44, p46, p48 and p52 isoforms were cloned individually into a p3xFlag-CMV mammalian expression vector and 1 μg of each clone plasmid DNA was separately transfected into HEK293 cells. HEK293 cells were selected instead of A549 cells for this study because HEK293 cells have a much lower background endogenous OAS1 activity compared to

A549 cells in response to poly(I:C) (data not shown). At 24 h after transfection, the expression levels of each OAS1 isoform were analyzed by Western blotting and IFA assay using anti-Flag antibody. Interestingly, although the same transfection protocol was used, the p42 and p46 isoforms expression levels were much higher than those of the p44, p48 and p52 isoforms (Figure 6-2A and 6-2B). To analyze the activity of the OAS1 isoforms in cells, HEK293 cells overexpressing each OAS1 isoform for 24 h were either mock transfected or transfected with poly(I:C) (10 μg/μl) for 6 h and the total intracellular RNA was extracted, purified and separated on a denaturing agarose gel. Cellular rRNA cleavage was assayed as a measure of RNase L activation. Surprisingly, among the five OAS1 isoforms, the overexpressed p42 and p46, but not p44, p48 or p52, activated RNase L cleavage in the absence of poly(I:C) transfection (Figure 6-2C). After poly(I:C) transfection, the overexpressed p44, p48 and p52 isoforms activated RNase L cleavage as indicated by the appearance of cleavage bands underneath the 18S and 28S rRNA bands, which were not observed with the vector control (Figure 6-2D). The data indicate that the human OAS1 p44, p48 and p52 isoforms are functionally active in HEK293 cells. Half the amount of the p42 or p46 plasmid DNA (0.5 μg) was transfected into HEK293 cells for a shorter time (9 h), followed by mock transfection or transfection with poly(I:C) (10 μg/μl) for 6 h. Total intracellular RNA was extracted, purified and separated on a denaturing agarose gel. The levels of p42 and p46 expressed from 0.5 μg of plasmid DNA for 9 h were lower (data not shown) and did not activate RNase L cleavage in the absence of poly(I:C). However, rRNA cleavage by activated RNase L was observed in cells expressing either of these proteins after poly(I:C) transfection, indicating that both the p42 and p46 isoforms are functionally active in mammalian cells (Figure 6-2E)

6.3.3 Yeast two hybrid assay identified novel binding partners for human OAS1 p42 and p44 isoforms

Binding partners have been reported for the mouse Oas1b that function in Oas1b mediated flavivirus resistance (Courtney, Di et al. 2012) and also for the human OAS1 p48 isoform that function in apoptosis (Ghosh, Sarkar et al. 2001). The human OAS1 p42 and p44 isoforms are thought to be expressed by humans of all genotypes.. Binding partners for the p42 and p44 isoforms were searched for using yeast two-hybrid screening of a universal human cDNA library. Full-length human OAS1-p42 or OAS1-p44 was used as the bait. After screening approximately 2 x 10^6 clones, five and three unique peptide candidates were identified as binding partners for p42 and p44, respectively (Table 6-1). A yeast co-transformation assay was then used to confirm binding between the human OAS1 isoform and each of the peptide candidates. The transformed yeast cells were plated under three different levels of stringency: DDO, TDO/A and QDO/A. Growth of yeast colonies on DDO plates indicated successful transformation of both OAS1 isoform and a peptide candidate. Growth of yeast colonies on TDO/A and QDO/A plates indicated that the transformed OAS1 isoform and peptide candidate bind each other in yeast. As a control, an empty vector, instead of an OAS1 isoform vector, was co-transformed with each of the screened peptide candidates into yeast. SV40 large T-antigen and murine p53 are known to interact and were co-transformed as a positive control. SV40 large T-antigen and lamin have not been reported to interact and were co-transformed as a negative control. The peptide candidates, fibrillin1 (FBN1) and supervillin (SVIL) were confirmed as binding partners in yeast for p42 and p44, respectively (Table 6-2). Yeast co-transformed with p42 and the phosphoglucomutase 1 (PGM1) peptide grew on both TDO/A and QDO/A plates but so did yeast co-transformed with empty vector and phosphoglucomutase 1 (PGM1) peptide, indicating a false

positive (Figure 6-2). Yeast co-transformed with p42 and voltage-dependent anion channel2 (VDAC2) peptides grew on TDO/A but not on QDO/A plates, suggesting possible binding between these proteins in yeast.

6.3.4 Human OAS1 p42 and p44 isoforms interact with their binding partners in vitro

Yeast co-transformation confirmed binding between FBN1 and p42 as well as between SVIL and p44 in yeast. It also suggested potential binding between VDAC2 and p42 in yeast. However, this binding could be indirect and mediated by another protein in yeast. To test whether p42 and p44 directly interact with their binding partners, wheat germ extract was used for *in vitro* transcription and translation of c-myc tagged full-length human OAS1 isoforms and HA-tagged peptide partners in the presence of [^{35}S]-labeled methionine. After incubation of each OAS1 isoform with their peptide partners individually, *in vitro* pull-down of p42 and p44 was done. Increased amounts of FBN1 and SVIL peptide, respectively, were detected compared to the non-specific IgG control (Figure 6-3A and 6-3C). After pull-down of FBN1 or SVIL peptide, increased amounts of p42 and p44, respectively, were detected compared to the control IgG (Figure 6-3B and 6-3D). The data indicate that p42 and p44 interact directly *in vitro* with FBN1 and SVIL peptide, respectively. The same amount of p42 was pulled-down by the VDAC2 peptide as by the non-specific control IgG (data not shown), indicating that either p42 does not directly interact with the VDAC2 peptide *in vitro*, or the interaction between these two proteins require some post-translational modifications that are not provided in the wheat germ extract expression system.

6.3.5 Human OAS1 p44 isoform interacts with its binding partner in cells

Although p42 was shown to directly interact with a FBN1 peptide and p44 to interact with a SVIL peptide *in vitro*, their binding capacities for the full-length FBN1 and SVIL proteins in cells was still unknown. The mammalian expression vector p3xFlag-CMV-p44 was transfected into A549 cells. At 48 h post transfection, cell lysate was collected and subjected to co-immunoprecipitation with anti-Flag antibody to detect the interaction between overexpressed full-length p44 and the endogenous full-length SVIL protein. The immunoprecipitated complex was separated by SDS-PAGE and anti-Flag and anti-SVIL antibodies were used to detect the proteins by Western blotting. The endogenous full-length SVIL was enriched in the complex pulled down by p44 compared to the non-specific IgG control and the lysate input, indicating that human OAS1 p44 interacts with full-length SVIL in cells (Figure 6-4).

6.4 Discussion

The human OAS family consists of OAS1, OAS2, OAS3 and OASL (Hovanessian and Justesen 2007). Unlike the mouse genome, the human genome only encodes one OAS1 gene that can generate at least 5 isoforms, p42, p44, p46, p48 and p52 (Bonnevie-Nielsen, Field et al. 2005). A previous study in Brinton lab showed that in response to poly(I:C), the purified recombinant p42, p44, p46 and p52 isoforms potently synthesized 2′-5′A trimers and higher order 2′-5′A *in vitro* which are activators of RNase L (Elbahesh and Brinton, unpublished data). In the present study, the recombinant OAS1 p52 isoform was expressed, purified and shown to potently synthesize 2′-5′A trimers and higher order 2′-5′A in response to poly(I:C) *in vitro.* In addition, overexpression of the human OAS1 isoforms individually in HEK293 cells showed that in response to poly(I:C) transfection in cells, human OAS1 isoforms p42, p44, p46, p48 and p52

each activated RNase L cleavage of cellular 18S/28S rRNA, indicating that all five human OAS1 isoforms are active synthetases in mammalian cells. A quantitative FRET assay triggered by RNase L cleavage of RNA could be used to compare the relative activation levels of RNase L by the different OAS1 isoforms (Thakur, Xu et al. 2005). A previous publication showed that in response to dengue virus infection in A549 cells, only the overexpressed human p42 and p46 isoforms activated RNase L cleavage of 18S/28S rRNA (Lin, Yu et al. 2009). The possibility that only the overexpressed p42 and p46 isoforms can recognize and be activated by dengue virus RNA in infected cells will be analyzed next. In the present study, we also observed that overexpression of either the p42 or p46 isoform at high levels activated RNase L cleavage of 18S/28S rRNA in the absence of poly(I:C) treatment (Figure 6-2C), which is unexpected and requires further investigation.

Novel binding partners were previously identified for the murine Oas1b protein and for the human OAS1 p48 isoform (Courtney, Di et al. 2012). In the present study, FBN1 was identified as a binding partner for human OAS1 p42 and SVIL was identified as a binding partner for human OAS1 p44. FBN1 is an extracellular matrix glycoprotein that can down-regulate the TGF-β signaling pathway (Hubmacher and Reinhardt 2009; Wipff, Allanore et al. 2009). Previous publications have reported OAS activity in the serum and showed that extracellular OAS1 could enter cells and exert an antiviral activity independent of RNase L (Giannelli, Antonelli et al. 1993; Kristiansen, Scherer et al. 2010; Thavachelvam, Gad et al. 2015). Antibody against endogenous FBN1 protein is now available and could be used to confirm the interaction between the p42 isoform and the full length FBN1 protein. Whether FBN1 functions as a scaffold protein to tether the extracellular OAS1 proteins on the cell surface forming an initial immune response barrier is worth further investigation. RNase L was reported

recently to bind to an actin-binding protein filamin A and to function as innate immune component maintaining a cellular barrier to viral entry (Ezelle, Malathi et al. 2016). In the present study, SVIL, an actin binding protein from the villin family was identified as the binding partner for human OAS1 p44 isoform and its C-terminal headpiece domain was shown to interact with p44 (Pope, Pestonjamasp et al. 1998). The C-terminal headpiece domain is conserved among many members in the villin family (Khurana and George 2008). Whether the p44 isoform can interact with other members of the villin family and whether these interactions bring p44 close to RNase L on the cytoskeleton where RNase L is also located is worth further study.

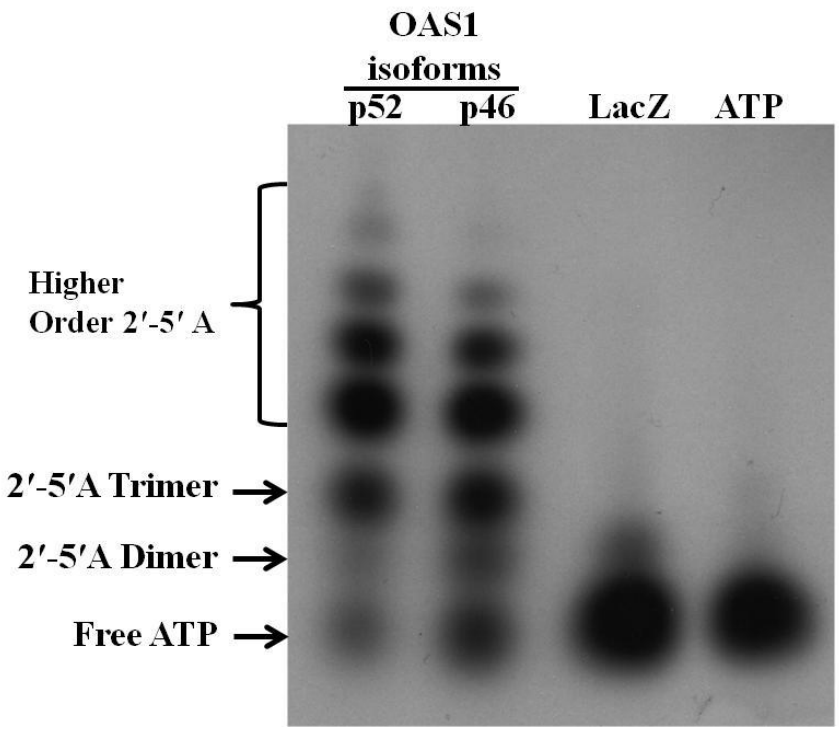


Figure 6.1. Analysis of the synthetase activity of the human OAS1 p52 isoform in vitro
In vitro expressed and purified LacZ protein, human OAS1 p46 isoform and p52 isoform were incubated individually with α^{32}p-ATP (15 µCi) and poly(I:C) (50 ng/µl) in a 1x Assay buffer at 30°C for 18 h. Each of the reactions was then mixed with Gel Loading Buffer II and separated on a 20% polyacrylamide/8M urea gel at 800V for 3.5 h. An aliquot of the free α^{32}p-ATP (15 µCi) was also separated on the gel as a control. The production of the 2′-5′A was detected by autoradiography. The free ATP, 2′-5′A dimer and 2′-5′A trimer are indicated by arrows, the higher order 2′-5′As are indicated by a bracket.

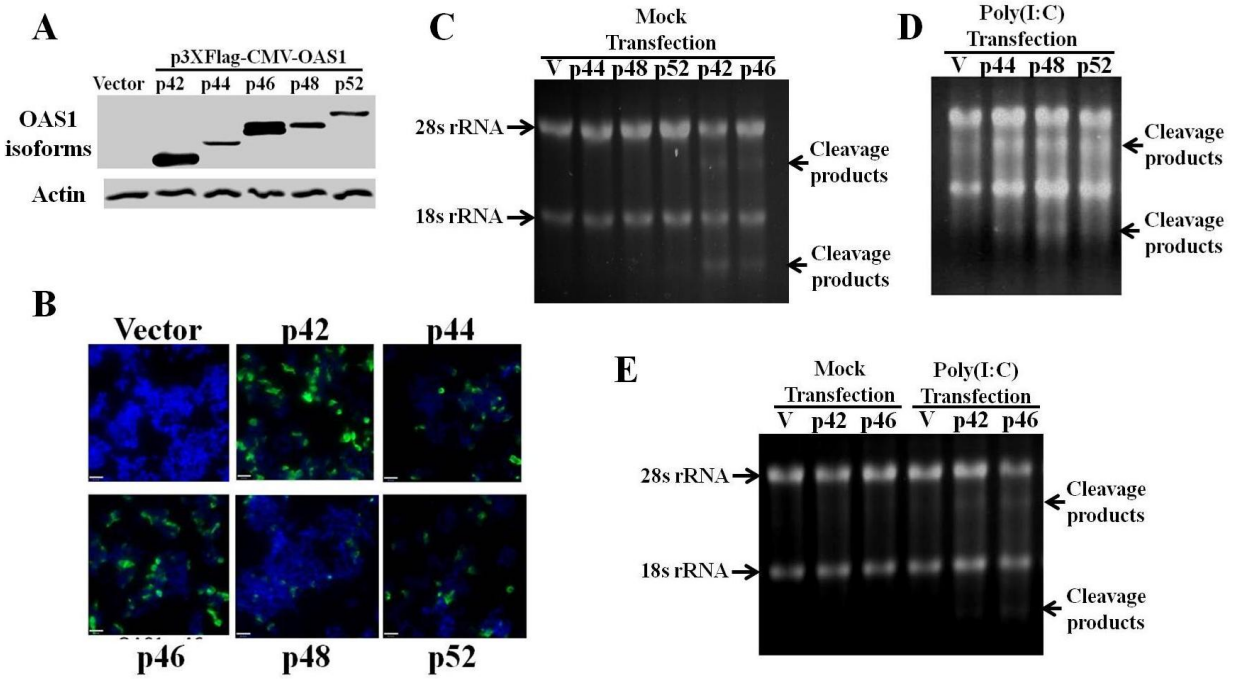


Figure 6.2. Analysis of RNase L cleavage in cells overexpressing individual human OAS1 isoforms in response to poly (I:C)

HEK293 cells were transfected with 1 µg of empty vector DNA or with one of the human OAS1 isoforms cloned into the p3xFlag-CMV plasmid for 24 h. (**A**) Western blotting analysis of the expression levels of each OAS1 isoform using anti-Flag and anti-β actin antibodies. (**B**) IFA analysis of the expression levels of each OAS1 isoform using anti-Flag antibody (green) and Hoechst 33342 for nuclei staining (Blue). HEK293 cells overexpressing each OAS1 isoform (1 µg) for 24 h were either mock transfected **(C)** or transfected with poly(I:C) for 6 h (**D**). Total intracellular RNA was extracted, purified and separated on a denaturing agarose gel and stained with ethidium bromide. **(E)** HEK293 cells were transfected with 0.5 µg p42 or p46 plasmid for 9 h, followed by either mock transfection or transfection with poly(I:C) for 6 h. Total intracellular RNA was extracted, purified and separated on a denaturing agarose gel and stained with ethidium bromide. The 18S/28S rRNA and their cleavage products are indicated by arrows.

Table 6.1. Identification of peptide candidate binding partners for the full-length human OAS1 p42 and p44 isoforms using yeast two hybrid analysis

Human OAS1 bait	Peptide candidate size (aa)	Putative protein containing the peptide
p42	188, 225	fibrillin 1 (FBN1)
	95	phosphoglucomutase 1 (PGM1)
	127	voltage-dependent anion channel 2 (VDAC2)
	120	trafficking protein particle complex 8 (TRAPPC8)
	115	lysyl Oxidase-Like 3 (LOXL3)
p44	178	supervillin (SVIL)
	206	G elongation factor, mitochrondria 2
	139	COX11 cytochrome c oxidase assembly homolog

Table 6.2. Confirmation of the peptide candidates as binding partners for the full-length human OAS1 p42 and p44 isoforms using yeast co-transformation

Peptide candidate	Co-transformed bait	DDO[a]	TDO/A[a]	QDO/A[a]
FBN1	P42	>100[b]	7	1
	Empty vector	>100		
PGM1	p42	>100	8	1
	Empty vectorr	>100	2	2
VDAC2	P42	>100	17	
	Empty vector	>100		
KIAA1012	P42	>100		
	Empty vector	>100		
LOXL3	P44	>100		
	Empty vector	>100		
supervillin	P44	>100	6	1
	Empty vector	>100		
G elongation factor	P44	>100		
	Empty vector	>100		
COX11	P44	>100		
	Empty vector	>100		
SV40 large T antigen	p53[c]	>100	43	14
SV40 large T antigen	Lamin[d]	28		

[a] Growth conditions with different stringencies
[b] Numbers of yeast colonies grown on the plates
[c] Positive control
[d] Negative control

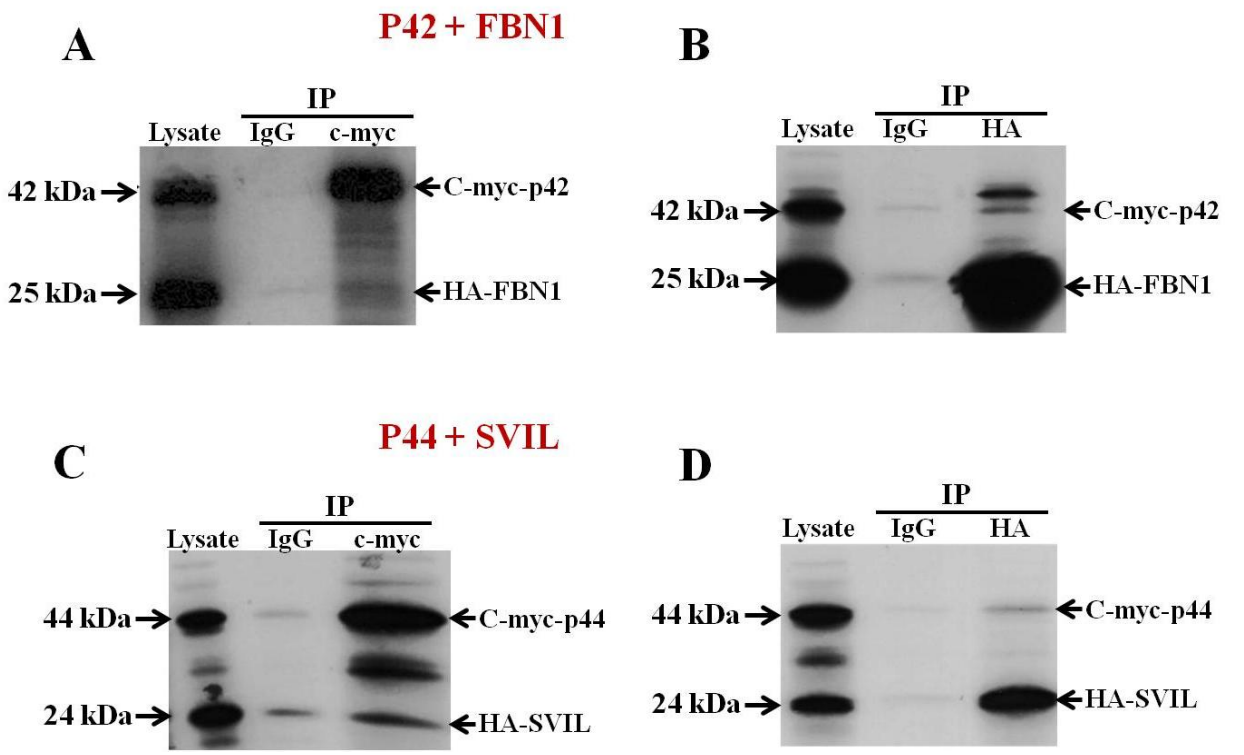


Figure 6.3. Confirmation of direct interaction between human OAS1 isoforms p42 and p44 and their respective binding partners

Wheat germ extract was used for *in vitro* transcription and translation of c-myc tagged full length human OAS1 isoforms and HA-tagged FBN1 or SVIL peptides in the presence of [^{35}S]-methionine. Reciprocal pull-down assays were done with anti-HA and anti-c-myc antibodies. IgG was used as a control. The protein complexes were resolved by SDS-PAGE and detected by autoradiography. **(A)** HA tagged FBN1 peptide was pulled down by c-myc tagged full-length OAS1 p42 *in vitro*. **(B)** C-myc tagged full-length human OAS1 p42 was pulled down by HA-tagged FBN1 peptide *in vitro*. **(C)** HA-tagged SVIL peptide was pulled down by c-myc tagged OAS1 p44 *in vitro*. **(D)** C-myc tagged full-length human OAS1 p44 was pulled down by HA-tagged SVIL peptide *in vitro*. The sizes of the human OAS1 isoforms and their binding partners are indicated by arrows.

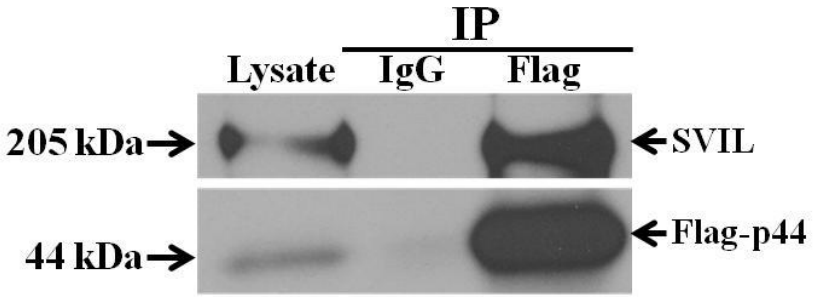


Figure 6.4. The interaction between human OAS1 p44 and the endogenous full-length SVIL in cells was confirmed by co-immunoprecipitation.
Human A549 cells were transiently transfected with 3xFlag-tagged full-length human OAS1 p44. At 48 hours after transfection, cell lysates were collected and incubated with anti-Flag antibody conjugated agarose beads. IgG conjugated agarose beads were used as the control. After co-immunoprecipitation, protein complexes were resolved by SDS-PAGE and detected by Western blotting. The OAS1 p44 and endogenous SVIL protein bands are indicated by arrows.

REFERENCES

Allen, A. M., A. E. Palmer, et al. (1968). "Simian hemorrhagic fever. II. Studies in pathology." Am J Trop Med Hyg **17**(3): 413-421.

An, S. and S. Makino (1998). "Characterizations of coronavirus cis-acting RNA elements and the transcription step affecting its transcription efficiency." Virology **243**(1): 198-207.

Badaoui, B., T. Rutigliano, et al. (2014). "RNA-sequence analysis of primary alveolar macrophages after in vitro infection with porcine reproductive and respiratory syndrome virus strains of differing virulence." PLoS One **9**(3): e91918.

Bailey, A. L., M. Lauck, et al. (2014). "Two novel simian arteriviruses in captive and wild baboons (Papio spp.)." J Virol **88**(22): 13231-13239.

Bailey, A. L., M. Lauck, et al. (2014). "High Genetic Diversity and Adaptive Potential of Two Simian Hemorrhagic Fever Viruses in a Wild Primate Population (vol 9, e90714, 2014)." Plos One **9**(7).

Balasuriya, U. B., J. C. Dobbe, et al. (2004). "Characterization of the neutralization determinants of equine arteritis virus using recombinant chimeric viruses and site-specific mutagenesis of an infectious cDNA clone." Virology **321**(2): 235-246.

Baric, R. S., S. A. Stohlman, et al. (1983). "Characterization of replicative intermediate RNA of mouse hepatitis virus: presence of leader RNA sequences on nascent chains." J Virol **48**(3): 633-640.

Bautista, E. M., K. S. Faaberg, et al. (2002). "Functional properties of the predicted helicase of porcine reproductive and respiratory syndrome virus." Virology **298**(2): 258-270.

Beerens, N., B. Selisko, et al. (2007). "De novo initiation of RNA synthesis by the arterivirus RNA-dependent RNA polymerase." J Virol **81**(16): 8384-8395.

Beerens, N. and E. J. Snijder (2007). "An RNA pseudoknot in the 3' end of the arterivirus genome has a critical role in regulating viral RNA synthesis." J Virol **81**(17): 9426-9436.

Beura, L. K., S. N. Sarkar, et al. (2010). "Porcine reproductive and respiratory syndrome virus nonstructural protein 1beta modulates host innate immune response by antagonizing IRF3 activation." J Virol **84**(3): 1574-1584.

Beura, L. K., S. Subramaniam, et al. (2012). "Identification of amino acid residues important for anti-IFN activity of porcine reproductive and respiratory syndrome virus non-structural protein 1." Virology **433**(2): 431-439.

Bonilla, P. J., J. L. Pinon, et al. (1995). "Characterization of the leader papain-like protease of MHV-A59." Adv Exp Med Biol **380**: 423-430.

Bonnevie-Nielsen, V., L. L. Field, et al. (2005). "Variation in antiviral 2',5'-oligoadenylate synthetase (2'5'AS) enzyme activity is controlled by a single-nucleotide polymorphism at a splice-acceptor site in the OAS1 gene." Am J Hum Genet **76**(4): 623-633.

Bramel-Verheije, M. H. G., P. J. M. Rottier, et al. (2000). "Expression of a foreign epitope by porcine reproductive and respiratory syndrome virus." Virology **278**(2): 380-389.

Bray, M. (2005). "Pathogenesis of viral hemorrhagic fever." Curr Opin Immunol **17**(4): 399-403.

Bray, M. and T. W. Geisbert (2005). "Ebola virus: the role of macrophages and dendritic cells in the pathogenesis of Ebola hemorrhagic fever." Int J Biochem Cell Biol **37**(8): 1560-1566.

Brinton, M. A., H. Di, et al. (2015). "Simian hemorrhagic fever virus: Recent advances." Virus Res **202**: 112-119.

Cai, Y., E. N. Postnikova, et al. (2015). "Simian hemorrhagic fever virus cell entry is dependent on CD163 and uses a clathrin-mediated endocytosis-like pathway." J Virol **89**(1): 844-856.

Calvert, J. G., D. E. Slade, et al. (2007). "CD163 expression confers susceptibility to porcine reproductive and respiratory syndrome viruses." J Virol **81**(14): 7371-7379.

Chapman, S., C. Faulkner, et al. (2008). "The photoreversible fluorescent protein iLOV outperforms GFP as a reporter of plant virus infection." Proc Natl Acad Sci U S A **105**(50): 20038-20043.

Chen, Z., L. Kuo, et al. (1993). "Sequences of 3' end of genome and of 5' end of open reading frame 1a of lactate dehydrogenase-elevating virus and common junction motifs between 5' leader and bodies of seven subgenomic mRNAs." Journal of General Virology **74 (Pt 4)**: 643-659.

Chen, Z., S. Lawson, et al. (2010). "Identification of two auto-cleavage products of nonstructural protein 1 (nsp1) in porcine reproductive and respiratory syndrome virus infected cells: nsp1 function as interferon antagonist." Virology **398**(1): 87-97.

Cheng, V. C., S. K. Lau, et al. (2007). "Severe acute respiratory syndrome coronavirus as an agent of emerging and reemerging infection." Clin Microbiol Rev **20**(4): 660-694.

Christie, J. M., K. Hitomi, et al. (2012). "Structural tuning of the fluorescent protein iLOV for improved photostability." J Biol Chem **287**(26): 22295-22304.

Conzelmann, K. K., N. Visser, et al. (1993). "Molecular characterization of porcine reproductive and respiratory syndrome virus, a member of the arterivirus group." Virology **193**(1): 329-339.

Courtney, S. C., H. Di, et al. (2012). "Identification of novel host cell binding partners of Oas1b, the protein conferring resistance to flavivirus-induced disease in mice." J Virol **86**(15): 7953-7963.

Cowley, J. A., C. M. Dimmock, et al. (2002). "Gill-associated nidovirus of Penaeus monodon prawns transcribes 3'-coterminal subgenomic mRNAs that do not possess 5'-leader sequences." Journal of General Virology **83**(Pt 4): 927-935.

Das, P. B., P. X. Dinh, et al. (2010). "The minor envelope glycoproteins GP2a and GP4 of porcine reproductive and respiratory syndrome virus interact with the receptor CD163." J Virol **84**(4): 1731-1740.

de Vries, A. A., E. D. Chirnside, et al. (1990). "All subgenomic mRNAs of equine arteritis virus contain a common leader sequence." Nucleic Acids Research **18**(11): 3241-3247.

de Vries, A. A., S. M. Post, et al. (1995). "The two major envelope proteins of equine arteritis virus associate into disulfide-linked heterodimers." J Virol **69**(8): 4668-4674.

de Vries, A. A. F., A. L. Glaser, et al. (2000). "Genetic manipulation of equine arteritis virus using full-length cDNA clones: Separation of overlapping genes and expression of a foreign epitope." Virology **270**(1): 84-97.

de Vries, A. A. F., A. L. Glaser, et al. (2001). "Recombinant equine arteritis virus as an expression vector." Virology **284**(2): 259-276.

Dea, S., C. A. Gagnon, et al. (2000). "Current knowledge on the structural proteins of porcine reproductive and respiratory syndrome (PRRS) virus: comparison of the North American and European isolates." Archives of Virology **145**(4): 659-688.

Delputte, P. L., N. Vanderheijden, et al. (2002). "Involvement of the Matrix Protein in Attachment of Porcine Reproductive and Respiratory Syndrome Virus to a Heparinlike Receptor on Porcine Alveolar Macrophages." J Virol **76**(9): 4312-4320.

den Boon, J. A., K. S. Faaberg, et al. (1995). "Processing and evolution of the N-terminal region of the arterivirus replicase ORF1a protein: identification of two papainlike cysteine proteases." J Virol **69**(7): 4500-4505.

den Boon, J. A., M. F. Kleijnen, et al. (1996). "Equine arteritis virus subgenomic mRNA synthesis: analysis of leader-body junctions and replicative-form RNAs." J Virol **70**(7): 4291-4298.

deVries, A. A., M. C. Horzinek, et al. (1997). "The Genome Organization of the Nidovirales: Similarities and Differences between Arteri-, Toro-, and Coronaviruses." Sem Virol **8**: 33-47.

Doan, D. N. and T. Dokland (2003). "Structure of the nucleocapsid protein of porcine reproductive and respiratory syndrome virus." Structure **11**(11): 1445-1451.

Dokland, T. (2010). "The structural biology of PRRSV." Virus Res **154**(1-2): 86-97.

Dong, B., L. Xu, et al. (1994). "Intrinsic molecular activities of the interferon-induced 2-5A-dependent RNase." J Biol Chem **269**(19): 14153-14158.

Donnelly, M. L., G. Luke, et al. (2001). "Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'." J Gen Virol **82**(Pt 5): 1013-1025.

Donovan, J., M. Dufner, et al. (2013). "Structural basis for cytosolic double-stranded RNA surveillance by human oligoadenylate synthetase 1." Proc Natl Acad Sci U S A **110**(5): 1652-1657.

Dunowska, M., P. J. Biggs, et al. (2012). "Identification of a novel nidovirus associated with a neurological disease of the Australian brushtail possum (Trichosurus vulpecula)." Vet Microbiol **156**(3-4): 418-424.

Elbahesh, H., B. K. Jha, et al. (2011). "The Flvr-encoded murine oligoadenylate synthetase 1b (Oas1b) suppresses 2-5A synthesis in intact cells." Virology **409**(2): 262-270.

Ezelle, H. J., K. Malathi, et al. (2016). "The Roles of RNase-L in Antimicrobial Immunity and the Cytoskeleton-Associated Innate Response." Int J Mol Sci **17**(1).

Fang, Y. and E. J. Snijder (2010). "The PRRSV replicase: exploring the multifunctionality of an intriguing set of nonstructural proteins." Virus Res **154**(1-2): 61-76.

Fehr, A. R. and S. Perlman (2015). "Coronaviruses: an overview of their replication and pathogenesis." Methods Mol Biol **1282**: 1-23.

Firth, A. E., J. C. Zevenhoven-Dobbe, et al. (2011). "Discovery of a small arterivirus gene that overlaps the GP5 coding sequence and is important for virus production." The Journal of General Virology **92**(Pt 5): 1097-1106.

Ghosh, A., S. N. Sarkar, et al. (2001). "A specific isozyme of 2'-5' oligoadenylate synthetase is a dual function proapoptotic protein of the Bcl-2 family." J Biol Chem **276**(27): 25447-25455.

Giannelli, G., G. Antonelli, et al. (1993). "2',5'-Oligoadenylate synthetase activity as a responsive marker during interferon therapy for chronic hepatitis C." J Interferon Res **13**(1): 57-60.

Godeny, E. K., A. A. de Vries, et al. (1998). "Identification of the leader-body junctions for the viral subgenomic mRNAs and organization of the simian hemorrhagic fever virus genome: evidence for gene duplication during arterivirus evolution." J Virol **72**(1): 862-867.

Godeny, E. K., A. A. F. deVries, et al. (1998). "Indentification of the Leader-Body Junctions for the Viral Subgenomic mRNAs and Organization of the Simian Hemorrhagic Fever Virus Genome: Evidence for Gene Duplication during Arterivirus Evolution." J Virol **72**(1): 862-867.

Godeny, E. K., L. Zeng, et al. (1995). "Molecular characterization of the 3' terminus of the simian hemorrhagic fever virus genome." Journal of Virology **69**(4): 2679-2683.

Gorbalenya, A. E., L. Enjuanes, et al. (2006). "Nidovirales: evolving the largest RNA virus genome." Virus Res **117**(1): 17-37.

Han, J., M. S. Rutherford, et al. (2009). "The porcine reproductive and respiratory syndrome virus nsp2 cysteine protease domain possesses both trans- and cis-cleavage activities." J Virol **83**(18): 9449-9463.

Han, M., Y. Du, et al. (2013). "Degradation of CREB-binding protein and modulation of type I interferon induction by the zinc finger motif of the porcine reproductive and respiratory syndrome virus nsp1alpha subunit." Virus Res **172**(1-2): 54-65.

Hartmann, R., J. Justesen, et al. (2003). "Crystal structure of the 2'-specific and double-stranded RNA-activated interferon-induced antiviral protein 2'-5'-oligoadenylate synthetase." Mol Cell **12**(5): 1173-1185.

Hartmann, R., H. S. Olsen, et al. (1998). "p59OASL, a 2'-5' oligoadenylate synthetase like protein: a novel human gene related to the 2'-5' oligoadenylate synthetase family." Nucleic Acids Res **26**(18): 4121-4128.

Herold, J., A. E. Gorbalenya, et al. (1998). "Proteolytic processing at the amino terminus of human coronavirus 229E gene 1-encoded polyproteins: identification of a papain-like proteinase and its substrate." J Virol **72**(2): 910-918.

Holtkamp, D. J., J. B. Kliebenstein, et al. (2013). "Assessment of the economic impact of porcine reproductive and respiratory syndrome virus on United States pork producers." Journal of Swine Health and Production **21**(2): 72-84.

Hovanessian, A. G. and J. Justesen (2007). "The human 2'-5'oligoadenylate synthetase family: unique interferon-inducible enzymes catalyzing 2'-5' instead of 3'-5' phosphodiester bond formation." Biochimie **89**(6-7): 779-788.

Huang, C., Q. Zhang, et al. (2014). "Porcine reproductive and respiratory syndrome virus nonstructural protein 4 antagonizes beta interferon expression by targeting the NF-kappaB essential modulator." J Virol **88**(18): 10934-10945.

Hubmacher, D. and D. P. Reinhardt (2009). "One more piece in the fibrillin puzzle." Structure **17**(5): 635-636.

Hussain, S., J. Pan, et al. (2005). "Identification of novel subgenomic RNAs and noncanonical transcription initiation signals of severe acute respiratory syndrome coronavirus." Journal of Virology **79**(9): 5288-5295.

Johnson, C. R., T. F. Griggs, et al. (2011). "Novel structural protein in porcine reproductive and respiratory syndrome virus encoded by an alternative ORF5 present in all arteriviruses." The Journal of General Virology **92**(Pt 5): 1107-1116.

Johnson, R. F., L. E. Dodd, et al. (2011). "Simian hemorrhagic fever virus infection of rhesus macaques as a model of viral hemorrhagic fever: Clinical characterization and risk factors for severe disease." Virology **421**(2): 129-140.

Joo, M. and S. Makino (1995). "The effect of two closely inserted transcription consensus sequences on coronavirus transcription." J Virol **69**(1): 272-280.

Justesen, J., R. Hartmann, et al. (2000). "Gene structure and function of the 2'-5'-oligoadenylate synthetase family." Cell Mol Life Sci **57**(11): 1593-1612.

Khurana, S. and S. P. George (2008). "Regulation of cell structure and function by actin-binding proteins: villin's perspective." FEBS Lett **582**(14): 2128-2139.

Kikkert, E. J. S. a. M. (2013). Arteriviruses Fields Virology, Sixth Edition. D. M. K. a. P. M. Howley. Philadelphia Wolters Kluwer/ Lippincott Williams & Wilkins 859-879.

Kim, O., Y. Sun, et al. (2010). "Modulation of type I interferon induction by porcine reproductive and respiratory syndrome virus and degradation of CREB-binding protein by non-structural protein 1 in MARC-145 and HeLa cells." Virology **402**(2): 315-326.

Kjaer, K. H., J. B. Poulsen, et al. (2009). "Evolution of the 2'-5'-oligoadenylate synthetase family in eukaryotes and bacteria." J Mol Evol **69**(6): 612-624.

Knoops, K., M. Barcena, et al. (2012). "Ultrastructural characterization of arterivirus replication structures: reshaping the endoplasmic reticulum to accommodate viral RNA synthesis." J Virol **86**(5): 2474-2487.

Kreutz, L. C. and M. R. Ackermann (1996). "Porcine reproductive and respiratory syndrome virus enters cells through a low pH-dependent endocytic pathway." Virus Res **42**(1-2): 137-147.

Kristiansen, H., H. H. Gad, et al. (2011). "The oligoadenylate synthetase family: an ancient protein family with multiple antiviral activities." J Interferon Cytokine Res **31**(1): 41-47.

Kristiansen, H., C. A. Scherer, et al. (2010). "Extracellular 2'-5' oligoadenylate synthetase stimulates RNase L-independent antiviral activity: a novel mechanism of virus-induced innate immunity." J Virol **84**(22): 11898-11904.

Kroese, M. V., J. C. Zevenhoven-Dobbe, et al. (2008). "The nsp1alpha and nsp1 papain-like autoproteinases are essential for porcine reproductive and respiratory syndrome virus RNA synthesis." J Gen Virol **89**(Pt 2): 494-499.

Kroese, M. V., J. C. Zevenhoven-Dobbe, et al. (2008). "The nsp1alpha and nsp1 papain-like autoproteinases are essential for porcine reproductive and respiratory syndrome virus RNA synthesis." Journal of General Virology **89**(Pt 2): 494-499.

Kuhn, J. H., M. Lauck, et al. (2016). "Reorganization and expansion of the nidoviral family Arteriviridae." Archives of Virology **161**(3): 755-768.

Kuo, L., Z. Chen, et al. (1992). "Lactate dehydrogenase-elevating virus (LDV): subgenomic mRNAs, mRNA leader and comparison of 3'-terminal sequences of two LDV isolates." Virus Res **23**(1-2): 55-72.

Kuo, L. L., J. T. Harty, et al. (1991). "A nested set of eight RNAs is formed in macrophages infected with lactate dehydrogenase-elevating virus." J Virol **65**(9): 5118-5123.

Lauck, M., S. V. Alkhovsky, et al. (2015). "Historical Outbreaks of Simian Hemorrhagic Fever in Captive Macaques Were Caused by Distinct Arteriviruses." J Virol **89**(15): 8082-8087.

Lauck, M., D. Hyeroba, et al. (2011). "Novel, Divergent Simian Hemorrhagic Fever Viruses in a Wild Ugandan Red Colobus Monkey Discovered Using Direct Pyrosequencing." Plos One **6**(4).

Lauck, M., S. D. Sibley, et al. (2013). "Exceptional simian hemorrhagic fever virus diversity in a wild African primate community." J Virol **87**(1): 688-691.

Lee, C. and D. Yoo (2006). "The small envelope protein of porcine reproductive and respiratory syndrome virus possesses ion channel protein-like properties." Virology **355**(1): 30-43.

Lin, R. J., H. P. Yu, et al. (2009). "Distinct antiviral roles for human 2',5'-oligoadenylate synthetase family members against dengue virus infection." J Immunol **183**(12): 8035-8043.

Lin, Y. C., R. Y. Chang, et al. (2002). "Leader-body junction sequence of the viral subgenomic mRNAs of porcine reproductive and respiratory syndrome virus isolated in Taiwan." Journal of Veterinary Medical Science **64**(11): 961-965.

Liu, D. X., T. S. Fung, et al. (2014). "Accessory proteins of SARS-CoV and other coronaviruses." Antiviral Res **109**: 97-109.

London, W. T. (1977). "Epizootiology, transmission and approach to prevention of fatal simian haemorrhagic fever in rhesus monkeys." Nature **268**(5618): 344-345.

Marie, I. and A. G. Hovanessian (1992). "The 69-kDa 2-5A synthetase is composed of two homologous and adjacent functional domains." J Biol Chem **267**(14): 9933-9939.

Meng, X. J., P. S. Paul, et al. (1996). "A nested set of six or seven subgenomic mRNAs is formed in cells infected with different isolates of porcine reproductive and respiratory syndrome virus." Journal of General Virology **77**: 1265-1270.

Meng, X. J., P. S. Paul, et al. (1996). "A nested set of six or seven subgenomic mRNAs is formed in cells infected with different isolates of porcine reproductive and respiratory syndrome virus." Journal of General Virology **77 (Pt 6)**: 1265-1270.

Meulenberg, J. J., E. J. de Meijer, et al. (1993). "Subgenomic RNAs of Lelystad virus contain a conserved leader-body junction sequence." Journal of General Virology **74 (Pt 8)**: 1697-1701.

Mielech, A. M., Y. Chen, et al. (2014). "Nidovirus papain-like proteases: Multifunctional enzymes with protease, deubiquitinating and deISGylating activities." Virus Res.

Miller, L. C., D. Fleming, et al. (2012). "Analysis of the swine tracheobronchial lymph node transcriptomic response to infection with a Chinese highly pathogenic strain of porcine reproductive and respiratory syndrome virus." BMC Vet Res **8**: 208.

Mondal, S. P., R. F. Cook, et al. (2015). "Development and characterization of a synthetic infectious cDNA clone of the virulent Bucyrus strain of equine arteritis virus expressing mCherry (red fluorescent protein)." Archives of Virology.

Mottahedin, A., M. Paidikondala, et al. (2013). "NF-kappaB activation by equine arteritis virus is MyD88 dependent and promotes viral replication." Archives of Virology **158**(3): 701-705.

Nedialkova, D. D., A. E. Gorbalenya, et al. (2010). "Arterivirus Nsp1 modulates the accumulation of minus-strand templates to control the relative abundance of viral mRNAs." PLoS Pathog **6**(2): e1000772.

Nedialkova, D. D., R. Ulferts, et al. (2009). "Biochemical characterization of arterivirus nonstructural protein 11 reveals the nidovirus-wide conservation of a replicative endoribonuclease." J Virol **83**(11): 5671-5682.

Nitschke, M., T. Korte, et al. (2008). "Equine arteritis virus is delivered to an acidic compartment of host cells via clathrin-dependent endocytosis." Virology **377**(2): 248-254.

Palmer, A. E., A. M. Allen, et al. (1968). "Simian hemorrhagic fever. I. Clinical and epizootiologic aspects of an outbreak among quarantined monkeys." Am J Trop Med Hyg **17**(3): 404-412.

Pasternak, A. O., A. P. Gultyaev, et al. (2000). "Genetic manipulation of arterivirus alternative mRNA leader-body junction sites reveals tight regulation of structural protein expression." Journal of Virology **74**(24): 11642-11653.

Pasternak, A. O., W. J. Spaan, et al. (2004). "Regulation of relative abundance of arterivirus subgenomic mRNAs." J Virol **78**(15): 8102-8113.

Pasternak, A. O., W. J. Spaan, et al. (2006). "Nidovirus transcription: how to make sense...?" Journal of General Virology **87**(Pt 6): 1403-1421.

Pasternak, A. O., E. van den Born, et al. (2001). "Sequence requirements for RNA strand transfer during nidovirus discontinuous subgenomic RNA synthesis." EMBO J **20**(24): 7220-7228.

Pasternak, A. O., E. van den Born, et al. (2003). "The stability of the duplex between sense and antisense transcription-regulating sequences is a crucial factor in arterivirus subgenomic mRNA synthesis." Journal of Virology **77**(2): 1175-1183.

Pedersen, K. W., Y. van der Meer, et al. (1999). "Open reading frame 1a-encoded subunits of the arterivirus replicase induce endoplasmic reticulum-derived double-membrane vesicles which carry the viral replication complex." J Virol **73**(3): 2016-2026.

Pei, Y. L., D. C. Hodgins, et al. (2009). "Porcine reproductive and respiratory syndrome virus as a vector: Immunogenicity of green fluorescent protein and porcine circovirus type 2 capsid expressed from dedicated subgenomic RNAs." Virology **389**(1-2): 91-99.

Perelygin, A. A., S. V. Scherbik, et al. (2002). "Positional cloning of the murine flavivirus resistance gene." Proc Natl Acad Sci U S A **99**(14): 9322-9327.

Plagemann, P. G. (2004). "GP5 ectodomain epitope of porcine reproductive and respiratory syndrome virus, strain Lelystad virus." Virus Res **102**(2): 225-230.

Plagemann, P. G., R. R. Rowland, et al. (2002). "The primary neutralization epitope of porcine respiratory and reproductive syndrome virus strain VR-2332 is located in the middle of the GP5 ectodomain." Archives of Virology **147**(12): 2327-2347.

Player, M. R. and P. F. Torrence (1998). "The 2-5A system: modulation of viral and cellular processes through acceleration of RNA degradation." Pharmacol Ther **78**(2): 55-113.

Pope, R. K., K. N. Pestonjamasp, et al. (1998). "Cloning, characterization, and chromosomal localization of human superillin (SVIL)." Genomics **52**(3): 342-351.

Posthuma, C. C., D. D. Nedialkova, et al. (2006). "Site-directed mutagenesis of the Nidovirus replicative endoribonuclease NendoU exerts pleiotropic effects on the arterivirus life cycle." J Virol **80**(4): 1653-1661.

Posthuma, C. C., K. W. Pedersen, et al. (2008). "Formation of the arterivirus replication/transcription complex: a key role for nonstructural protein 3 in the remodeling of intracellular membranes." J Virol **82**(9): 4480-4491.

Rahman, M. M. and G. McFadden (2011). "Myxoma virus lacking the pyrin-like protein M013 is sensed in human myeloid cells by both NLRP3 and multiple Toll-like receptors, which independently activate the inflammasome and NF-kappaB innate response pathways." J Virol **85**(23): 12505-12517.

Rahman, M. M., M. R. Mohamed, et al. (2009). "Co-regulation of NF-kappaB and inflammasome-mediated inflammatory responses by myxoma virus pyrin domain-containing protein M013." PLoS Pathog **5**(10): e1000635.

Rascon-Castelo, E., A. Burgara-Estrella, et al. (2015). "Immunological features of the non-structural proteins of porcine reproductive and respiratory syndrome virus." Viruses **7**(3): 873-886.

Ryan, M. D., M. Donnelly, et al. (1999). "A Model for Nonstoichiometric, Cotranslational Protein Scission in Eukaryotic Ribosomes." Bioorganic Chemistry **27**(1): 55-79.

Sagripanti, J. L. (1985). "Polyadenylic acid sequences in the genomic RNA of the togavirus of simian hemorrhagic fever." Virology **145**(2): 350-355.

Sagripanti, J. L., R. O. Zandomeni, et al. (1986). "The cap structure of simian hemorrhagic fever virion RNA." Virology **151**(1): 146-150.

Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.

Sang, Y., W. Brichalli, et al. (2014). "Genome-wide analysis of antiviral signature genes in porcine macrophages at different activation statuses." PLoS One **9**(2): e87613.

Sawicki, S. G. and D. L. Sawicki (1995). "Coronaviruses use discontinuous extension for synthesis of subgenome-length negative strands." Coronaviruses and Arteriviruses **380**: 499-506.

Sawicki, S. G., D. L. Sawicki, et al. (2007). "A contemporary view of coronavirus transcription." J Virol **81**(1): 20-29.

Scherbik, S. V., J. M. Paranjape, et al. (2006). "RNase L plays a role in the antiviral response to West Nile virus." J Virol **80**(6): 2987-2999.

Sethna, P. B., S. L. Hung, et al. (1989). "Coronavirus subgenomic minus-strand RNAs and the potential for mRNA replicons." Proc Natl Acad Sci U S A **86**(14): 5626-5630.

Smith, S. L., X. Wang, et al. (1997). "Sequence of the 3' end of the simian hemorrhagic fever virus genome." Gene **191**(2): 205-210.

Snijder, E. J., J. C. Dobbe, et al. (2003). "Heterodimerization of the two major envelope proteins is essential for arterivirus infectivity." J Virol **77**(1): 97-104.

Snijder, E. J. and M. Kikkert (2013). Arteriviruses. Fields Virology. D. M. Knipe and P. M. Howley. Philadelphia, PA, Lippincott, Williams and Wilkins. **1:** 859-879.

Snijder, E. J., M. Kikkert, et al. (2013). "Arterivirus molecular biology and pathogenesis." Journal of General Virology **94**: 2141-2163.

Snijder, E. J. and J. J. M. Meulenberg (1998). "The molecular biology of arterivirus." J Gen Virol **79**: 961-979.

Snijder, E. J., H. van Tol, et al. (1999). "Identification of a novel structural protein of arteriviruses." J Virol **73**(8): 6335-6345.

Snijder, E. J., H. van Tol, et al. (2001). "Non-structural proteins 2 and 3 interact to modify host cell membranes during the formation of the arterivirus replication complex." Journal of General Virology **82**(Pt 5): 985-994.

Snijder, E. J., A. L. Wassenaar, et al. (1992). "The 5' end of the equine arteritis virus replicase gene encodes a papainlike cysteine protease." J Virol **66**(12): 7040-7048.

Snijder, E. J., A. L. Wassenaar, et al. (1995). "The arterivirus Nsp2 protease. An unusual cysteine protease with primary structure similarities to both papain-like and chymotrypsin-like proteases." J Biol Chem **270**(28): 16671-16676.

Snijder, E. J., A. L. Wassenaar, et al. (1996). "The arterivirus nsp4 protease is the prototype of a novel group of chymotrypsin-like enzymes, the 3C-like serine proteases." J Biol Chem **271**(9): 4864-4871.

Sola, I., J. L. Moreno, et al. (2005). "Role of nucleotides immediately flanking the transcription-regulating sequence core in coronavirus subgenomic mRNA synthesis." J Virol **79**(4): 2506-2516.

Song, C., P. Krell, et al. (2010). "Nonstructural protein 1alpha subunit-based inhibition of NF-kappaB activation and suppression of interferon-beta production by porcine reproductive and respiratory syndrome virus." Virology **407**(2): 268-280.

Sun, L., Y. Li, et al. (2013). "Porcine reproductive and respiratory syndrome virus ORF5a protein is essential for virus viability." Virus Res **171**(1): 178-185.

Sun, Y., F. Xue, et al. (2009). "Crystal structure of porcine reproductive and respiratory syndrome virus leader protease Nsp1alpha." J Virol **83**(21): 10931-10940.

Sun, Z., Z. Chen, et al. (2010). "The cysteine protease domain of porcine reproductive and respiratory syndrome virus nonstructural protein 2 possesses deubiquitinating and interferon antagonism functions." J Virol **84**(15): 7832-7846.

Sun, Z., Y. Li, et al. (2012). "Nonstructural protein 2 of porcine reproductive and respiratory syndrome virus inhibits the antiviral function of interferon-stimulated gene 15." J Virol **86**(7): 3839-3850.

Tauraso, N. M., A. Shelokov, et al. (1968). "Simian hemorrhagic fever. 3. Isolation and characterization of a viral agent." Am J Trop Med Hyg **17**(3): 422-431.

Teng, H., J. D. Pinon, et al. (1999). "Expression of murine coronavirus recombinant papain-like proteinase: efficient cleavage is dependent on the lengths of both the substrate and the proteinase polypeptides." J Virol **73**(4): 2658-2666.

Thakur, C. S., Z. Xu, et al. (2005). "A convenient and sensitive fluorescence resonance energy transfer assay for RNase L and 2',5' oligoadenylates." Methods Mol Med **116**: 103-113.

Thavachelvam, K., H. H. Gad, et al. (2015). "Rapid Uptake and Inhibition of Viral Propagation by Extracellular OAS1." J Interferon Cytokine Res **35**(5): 359-366.

Tian, D., Z. Wei, et al. (2012). "Arterivirus minor envelope proteins are a major determinant of viral tropism in cell culture." J Virol **86**(7): 3701-3712.

Tian, X., G. Lu, et al. (2009). "Structure and cleavage specificity of the chymotrypsin-like serine protease (3CLSP/nsp4) of Porcine Reproductive and Respiratory Syndrome Virus (PRRSV)." J Mol Biol **392**(4): 977-993.

Tijms, M. A., D. D. Nedialkova, et al. (2007). "Arterivirus subgenomic mRNA synthesis and virion biogenesis depend on the multifunctional nsp1 autoprotease." J Virol **81**(19): 10496-10505.

Tijms, M. A., L. C. van Dinten, et al. (2001). "A zinc finger-containing papain-like protease couples subgenomic mRNA synthesis to genome translation in a positive-stranded RNA virus." Proc Natl Acad Sci U S A **98**(4): 1889-1894.

Van Breedam, W., H. Van Gorp, et al. (2010). "The M/GP(5) glycoprotein complex of porcine reproductive and respiratory syndrome virus binds the sialoadhesin receptor in a sialic acid-dependent manner." PLoS Pathog **6**(1): e1000730.

Van Den Born, E., A. P. Gultyaev, et al. (2004). "Secondary structure and function of the 5'-proximal region of the equine arteritis virus RNA genome." Rna-a Publication of the Rna Society **10**(3): 424-437.

van den Born, E., C. C. Posthuma, et al. (2005). "Discontinuous subgenomic RNA synthesis in arteriviruses is guided by an RNA hairpin structure located in the genomic leader region." J Virol **79**(10): 6312-6324.

van den Born, E., C. C. Posthuma, et al. (2007). "An infectious recombinant equine arteritis virus expressing green fluorescent protein from its replicase gene." J Gen Virol **88**(Pt 4): 1196-1205.

van der Meer, Y., H. van Tol, et al. (1998). "ORF1a-encoded replicase subunits are involved in the membrane association of the arterivirus replication complex." J Virol **72**(8): 6689-6698.

van Dinten, L. C., J. A. den Boon, et al. (1997). "An infectious arterivirus cDNA clone: identification of a replicase point mutation that abolishes discontinuous mRNA transcription." Proc Natl Acad Sci U S A **94**(3): 991-996.

van Dinten, L. C., S. Rensen, et al. (1999). "Proteolytic processing of the open reading frame 1b-encoded part of arterivirus replicase is mediated by nsp4 serine protease and Is essential for virus replication." J Virol **73**(3): 2027-2037.

van Dinten, L. C., H. van Tol, et al. (2000). "The predicted metal-binding region of the arterivirus helicase protein is involved in subgenomic mRNA synthesis, genome replication, and virion biogenesis." J Virol **74**(11): 5213-5223.

van Dinten, L. C., A. L. Wassenaar, et al. (1996). "Processing of the equine arteritis virus replicase ORF1b protein: identification of cleavage products containing the putative viral polymerase and helicase domains." J Virol **70**(10): 6625-6633.

van Kasteren, P. B., B. A. Bailey-Elkin, et al. (2013). "Deubiquitinase function of arterivirus papain-like protease 2 suppresses the innate immune response in infected host cells." Proc Natl Acad Sci U S A **110**(9): E838-847.

van Kasteren, P. B., C. Beugeling, et al. (2012). "Arterivirus and nairovirus ovarian tumor domain-containing Deubiquitinases target activated RIG-I to control innate immune signaling." J Virol **86**(2): 773-785.

van Marle, G., J. C. Dobbe, et al. (1999). "Arterivirus discontinuous mRNA transcription is guided by base pairing between sense and antisense transcription-regulating sequences." Proc Natl Acad Sci U S A **96**(21): 12056-12061.

van Vliet, A. L., S. L. Smits, et al. (2002). "Discontinuous and non-discontinuous subgenomic RNA transcription in a nidovirus." EMBO J **21**(23): 6571-6580.

Vanderheijden, N., P. L. Delputte, et al. (2003). "Involvement of sialoadhesin in entry of porcine reproductive and respiratory syndrome virus into porcine alveolar macrophages." J Virol **77**(15): 8207-8215.

Vatter, H. A. and M. A. Brinton (2014). "Differential responses of disease-resistant and disease-susceptible primate macrophages and myeloid dendritic cells to simian hemorrhagic fever virus infection." J Virol **88**(4): 2095-2106.

Vatter, H. A., H. Di, et al. (2014). "Each of the eight simian hemorrhagic fever virus minor structural proteins is functionally important." Virology **462-463C**: 351-362.

Vatter, H. A., H. Di, et al. (2014). "Each of the eight simian hemorrhagic fever virus minor structural proteins is functionally important." Virology **462-463**: 351-362.

Vatter, H. A., H. Di, et al. (2014). "Functional analyses of the three simian hemorrhagic fever virus nonstructural protein 1 papain-like proteases." J Virol **88**(16): 9129-9140.

Vatter, H. A., H. Di, et al. (2014). "Functional analyses of the three simian hemorrhagic fever virus nonstructural protein 1 papain-like proteases." J Virol.

Vatter, H. A., E. F. Donaldson, et al. (2015). "A simian hemorrhagic fever virus isolate from persistently infected baboons efficiently induces hemorrhagic fever disease in Japanese macaques." Virology **474**: 186-198.

Veit, M., A. K. Matczuk, et al. (2014). "Membrane proteins of arterivirus particles: structure, topology, processing and function." Virus Res **194**: 16-36.

Verheije, M. H., R. C. Olsthoorn, et al. (2002). "Kissing interaction between 3' noncoding and coding sequences is essential for porcine arterivirus RNA replication." J Virol **76**(3): 1521-1526.

Wahl-Jensen, V., J. C. Johnson, et al. (2015). "Divergent Simian Arteriviruses Cause Simian Hemorrhagic Fever of Differing Severities in Macaques." MBio **7**(1).

Wang, C., B. Huang, et al. (2013). "A novel porcine reproductive and respiratory syndrome virus vector system that stably expresses enhanced green fluorescent protein as a separate transcription unit." Vet Res **44**: 104.

Wang, L., J. Hou, et al. (2014). "Attenuation of highly pathogenic porcine reproductive and respiratory syndrome virus by inserting an additional transcription unit." Vaccine **32**(43): 5740-5748.

Welch, S. K. and J. G. Calvert (2010). "A brief review of CD163 and its role in PRRSV infection." Virus Res **154**(1-2): 98-103.

Wieringa, R., A. A. de Vries, et al. (2003). "Formation of disulfide-linked complexes between the three minor envelope glycoproteins (GP2b, GP3, and GP4) of equine arteritis virus." J Virol **77**(11): 6216-6226.

Wieringa, R., A. A. de Vries, et al. (2004). "Structural protein requirements in equine arteritis virus assembly." J Virol **78**(23): 13019-13027.

Wipff, J., Y. Allanore, et al. (2009). "[Interactions between fibrillin-1 and tgf-beta: consequences and human pathology]." Med Sci (Paris) **25**(2): 161-167.

Wissink, E. H., M. V. Kroese, et al. (2005). "Envelope protein requirements for the assembly of infectious virions of porcine reproductive and respiratory syndrome virus." J Virol **79**(19): 12495-12506.

Wood, O., N. Tauraso, et al. (1970). "Electron microscopic study of tissue cultures infected with simian haemorrhagic fever virus." Journal of General Virology **7**(2): 129-136.

Xue, F., Y. Sun, et al. (2010). "The crystal structure of porcine reproductive and respiratory syndrome virus nonstructural protein Nsp1beta reveals a novel metal-dependent nuclease." J Virol **84**(13): 6461-6471.

Zeng, L., E. K. Godeny, et al. (1995). "Analysis of simian hemorrhagic fever virus (SHFV) subgenomic RNAs, junction sequences, and 5' leader." Virology **207**(2): 543-548.

Zhang, T., X. Liu, et al. (2015). "[Construction and identification of a recombinant PRRSV expressing ORF2 of porcine circovirus type 2]." Bing Du Xue Bao **31**(1): 65-73.

Ziebuhr, J., E. J. Snijder, et al. (2000). "Virus-encoded proteinases and proteolytic processing in the Nidovirales." J Gen Virol **81**(Pt 4): 853-879.