11-29-2007

Functional Analyses of West Nile Virus (WNV) Bicistronic Replicons Containing Different Sequence Elements and of Simian Hemorrhagic Fever Virus (SHFV) Polyprotein Processing

Gertrud Ulrike Radu
Georgia State University

Follow this and additional works at: http://scholarworks.gsu.edu/biology_diss

Recommended Citation

This Dissertation is brought to you for free and open access by the Department of Biology at ScholarWorks @ Georgia State University. It has been accepted for inclusion in Biology Dissertations by an authorized administrator of ScholarWorks @ Georgia State University. For more information, please contact scholarworks@gsu.edu.
Functional analyses of West Nile virus (WNV) bicistronic replicons containing different sequence elements and of simian hemorrhagic fever virus (SHFV) polyprotein processing

by

Gertrud U. Radu

Under the Direction of Margo A. Brinton

ABSTRACT

The flavivirus West Nile virus (WNV) encodes a single polyprotein that is processed into three structural and seven nonstructural proteins. Various WNV bicistronic replicons that direct cap-dependent translation of an N-terminal viral capsid or capsid/Renilla luciferase fusion protein as well as IRES-dependent translation of the nonstructural proteins were constructed. An original replicon consisting of the WNV 5' NCR, the 5' 198 nts of the capsid coding sequence, which included the 5' cyclization sequence (Cyc), and an EMCV IRES followed by the WNV nonstructural genes and 3' NCR was generated. Real time qRT-PCR analysis of intracellular levels of this replicon RNA showed a 4 fold increase by 96 hr after transfection of BHK cells. Increasing the distance between the 5' Cyc and IRES by insertion of a 5' IRES flanking sequence alone or together with a Renilla luciferase reporter did not increase RNA replication. Addition of only a reporter decreased RNA replication. The insertion of an extended capsid coding sequence also did not enhance RNA replication, but did enhance both cap- and IRES-dependent translation of replicon RNA, as indicated by immunofluorescence and
Western blot analysis. These results suggest the presence of a translation enhancer in the 3' portion of the capsid coding region.

Simian hemorrhagic fever virus (SHFV) is a member of the family *Arteriviridae*, order *Nidovirales*. SHFV is unique among Nidoviruses in having three instead of two papain-like cysteine protease (PCP) motifs designated α, β, and γ, within the N-terminal region of its ORF1a. Mutations of putative PCP cleavage sites showed that the most efficient cleavage was by PCPβ at its downstream cleavage site. A large deletion located between the two catalytic residues of PCPα was hypothesized to render this protease inactive. However, processing was observed at the cleavage site following PCPα. Mutational analyses confirmed that PCPα is an inactive protease, and that the cleavage sites downstream of PCPα are cleaved by PCPγ. When the catalytic residues of PCPγ were mutated, PCPβ was also able to back cleave at these sites. This "back" cleavage is a previously unreported activity for an arterivirus PCP.

**INDEX WORDS:** West Nile virus, Replicon, IRES, Capsid, RNA Replication, Translation, Simian hemorrhagic fever virus, Papain-like cysteine protease.
Functional analyses of West Nile virus (WNV) bicistronic replicons containing different sequence elements and of simian hemorrhagic fever virus (SHFV) polyprotein processing

by

Gertrud U. Radu

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

2007
Functional analyses of West Nile virus (WNV) bicistronic replicons containing different sequence elements and of simian hemorrhagic fever virus (SHFV) polyprotein processing

by

Gertrud U. Radu

Major Professor: Margo A. Brinton
Committee Members: Teryl K. Frey
Irene T. Weber

Electronic Version Approved by:

Office of Graduate Studies
College of Arts and Sciences
Georgia State University
December 2007
DEDICATION

I dedicate this dissertation to my family. To my parents, Gerda and Stefan, who instilled in me the desire for a higher education from a very young age, and who taught me that with diligence, patience and perseverance, anything can be achieved. To my brother Uwe and his wife Valerie, for their emotional support and encouragement during this very long journey, and last, but not least, to my adorable niece and nephew, Katie and Max, who never fail to delight and inspire me.

"Imagination is more important than knowledge.
Knowledge is limited. Imagination encircles the world."

-- Albert Einstein
ACKNOWLEDGEMENTS

My deepest gratitude goes to my mentor, Dr. Margo Brinton. I thank her for allowing me into her lab, for introducing me to the exciting world of Virology, and for guiding me along the way. I would also like to thank my committee members, Dr. Teryl Frey and Dr. Irene Weber, for their help and encouragement. My most sincere thanks go to Dr. Svetlana Sherbik, without whom this dissertation would not exist, and Drs. William Davis and Mohamed Emara for their abundant advice and critical input. I also thank Dr. Taronna Maines for her initial help, as well as her friendship and encouragement over the years. My appreciation also goes to all my other colleagues in the laboratory, Slava Stockman, Joanna Pulit-Penaloza, Dr. Mausumi Basu, Sean Courtney, Husni Elbahesh, Dr. Natalia Astakhova, Dr. Andrey Perelygin, Heather Hardcastle and Hsuan Liu. Finally, I would like to thank members of Dr. Frey's lab Suganthi Suppiah, Heather Mousa and Jason Matthews for their friendship and support.
# TABLE OF CONTENTS

DEDICATION ................................................................................................................................................... iv

ACKNOWLEDGEMENTS .................................................................................................................................. v

Chapter I .......................................................................................................................................................... 1

Functional analyses of West Nile virus (WNV) bicistronic replicons containing different sequence elements .......................................................................................................................... 1

INTRODUCTION ........................................................................................................................................... 1

General Characteristics ................................................................................................................................. 1
Classification ............................................................................................................................................... 2
Virion ........................................................................................................................................................... 2
Genome ....................................................................................................................................................... 3
Replication cycle .......................................................................................................................................... 6
Secondary structures and *cis*-acting elements in the flaviviral noncoding regions ............................................ 8
Mechanisms of translation initiation ......................................................................................................... 12
*Cap-dependent translation of cellular mRNAs* ...................................................................................... 12
*Cap-dependent translation of some viral mRNAs* .................................................................................. 14
*IRES-dependent translation* .................................................................................................................. 15
Replicons as tools for studying viral RNA replication and translation ................................................... 18
Goals .......................................................................................................................................................... 24

MATERIALS AND METHODS ....................................................................................................................... 25

Cell culture ................................................................................................................................................ 25
Preparation of S100 cell extracts ............................................................................................................ 25
Preparation of cDNA clones and DNA templates for *in vitro* RNA transcription ................................. 25
Transformation .......................................................................................................................................... 26
*In vitro* RNA transcription ...................................................................................................................... 26
Real time qRT-PCR ..................................................................................................................................... 27
Western blotting ...................................................................................................................................... 28
Immunofluorescence analysis ................................................................................................................... 29
Luciferase assays ..................................................................................................................................... 30
Gel mobility shift assays .......................................................................................................................... 30

RESULTS .......................................................................................................................................................... 31

Construction of a WNV bicistronic replicon ......................................................................................... 31
Measurement of CNI and CNI LI RNA replication by real time qRT-PCR ........................................ 33
Mung bean nuclease treatment .................................................................................................................. 33
Addition of an IRES 5' flanking sequence and the 3' E signal sequence to the WNV bicistronic replicon ........................................................................................................................................ 35
Analysis of the efficiency of transfection and replicon function.......................... 37
Analysis of IRES-dependent translation by Western blotting.......................... 40
Extension of the capsid coding sequence.............................................................. 42
The extended capsid coding region has an enhancing effect on both IRES- and cap-dependent translation ..................................................................................... 43
Analysis of WNV RNA-cell protein binding activity........................................... 49
Detection of cap-dependent translation by analyzing luciferase activity ............. 49
Effect of mutation of the major eEF1A binding site in a bicistronic WNV replicon ............................................................................................................................... 52
DISCUSSION........................................................................................................... 57

Chapter II .......................................................................................................................... 67
Functional analysis of simian hemorrhagic fever virus (SHFV) polyprotein processing ................................................................. 67
INTRODUCTION .................................................................................................... 67
Classification......................................................................................................... 67
Virion .................................................................................................................... 68
Genome ................................................................................................................. 69
Replication cycle................................................................................................... 72
Proteases in Arteriviruses ..................................................................................... 75
Papain-like proteases ......................................................................................... 77
Goals ......................................................................................................................... 84
Aim 1: Analysis of the activity of SHFV PCPα. ................................................. 84
Aim 2: Determination of which of the predicted cleavage sites are used............ 84
Aim 3: Determination of which PCP cleaves which cleavage sites. ............... 84
MATERIALS AND METHODS .............................................................................. 85
Mutagenesis of an SHFV cDNA clone ................................................................. 85
Preparation of cDNA clones ................................................................................. 85
In vitro transcription ............................................................................................. 86
In vitro translation ................................................................................................ 86
In vitro transcription/translation ......................................................................... 86
Western blotting .................................................................................................... 87
Estimation of protein molecular mass by relative electrophoretic mobility ....... 87
RESULTS ................................................................................................................. 88
Prediction of PCP cleavage sites in the N-terminal region of SHFV ORF1a..... 88
Analysis of the activity of PCPα ........................................................................... 90
Identification of the cleavage sites used by individual SHFV PCPs ................. 93
Determination of which PCP cleaves which cleavage sites............................... 97
LIST OF FIGURES

Figure 1.1. Schematic of a flavivirus genome. ................................................................. 1
Figure 1.2. WNV replication cycle. ................................................................................. 7
Figure 1.3. West Nile virus NCRs. .................................................................................. 9
Figure 1.4. Models of cap-dependent mRNA cyclization. ............................................. 13
Figure 1.5. Two major classes of picornavirus IRES. .................................................... 17
Figure 1.6. Construction of WNV bicistronic replicons. ................................................ 32
Figure 1.7. Analysis of (A) CNI and (B) CNLI WNV replicon RNA replication by real
time qRT-PCR. ............................................................................................................. 34
Figure 1.8. Effect of treating AflII digested replicon DNA with mung bean nuclease prior
to in vitro RNA transcription. .................................................................................... 36
Figure 1.9. Analysis of (A) CNI and (B) CNLI WNV replicon RNA replication. ....... 38
Figure 1.10. Analysis of translation competence and transfection efficiency of a
nonrepliating replicon, and viral RNA decay kinetics, and replication of a
nonreplicating and replicating replicon...................................................................... 39
Figure 1.11. Comparison of Western blots of viral proteins present in extracts of cells
transfected with 1μg of WNV infectious clone RNA or CNI replicon RNA. .......... 41
Figure 1.12. Analysis of (A) CFI and (B) CLFI replicon RNA replication.................... 44
Figure 1.13. Comparison of WNV nonstructural protein levels produced by the CNI and
CFI replicon RNAs. .................................................................................................... 45
Figure 1.14. Schematics of WNV capsid RNA and replicons with truncated capsid
coding regions. .......................................................................................................... 46
Figure 1.15. Effect of capsid length on cap- and IRES-dependent translation by replicons. ..................................................................................................................... 48

Figure 1.16. Thermodynamic folding prediction of a region of the WNV capsid RNA. 50

Figure 1.17. Gel mobility shift assay done with a [α-32P]-radiolabeled capsid RNA (2000 cpm) and BHK cytoplasmic S100 extracts. .................................................................................. 51

Figure 1.18. Analysis of Renilla and Firefly luciferase activity in extracts from cells transfected with reporter constructs or WNV replicons. ................................................. 53

Figure 1.19. Analysis of C^N^I and C^N^I/GUGU RNA replication by real time qRT-PCR. 55

Figure 1.20. Sequence analysis of replicon RNA extracted from cells transfected with the C^N^I/GUGU replicon.................................................................................................. 55

Figure 1.21. Schematics of standard bicistronic replicons utilized in previous studies. . 62

Figure 2.1. Genomic organization of SHFV. ........................................................................ 71

Figure 2.2. Models for the generation of sg mRNAs in Nidoviruses. .............................. 76

Figure 2.3. Overview of the proteolytic processing cascade of the EAV, LDV, and PRRSV replicase polyproteins................................................................. 78

Figure 2.4. The structure of the SARS-CoV PCP.......................................................... 80

Figure 2.5. Alignment of the three SHFV PCPs................................................................. 83

Figure 2.6. Locations of predicted PCP cleavage sites and predicted cleavage products resulting from SHFV PCP cleavage. .............................................................. 89

Figure 2.7. Cleavage products produced by polyproteins with mutations in the catalytic cysteines of individual or combinations of PCPs. ........................................... 92
Figure 2.8. Comparison of PCP activity in the (A) absence, or (B) presence of microsomal membranes. ................................................................. 94

Figure 2.9. Cleavage products produced by polyproteins with mutations in the predicted cleavage sites. ........................................................................................................... 96

Figure 2.10. Cleavage products generated from polyproteins with mutations in five out of six predicted cleavage sites ............................................................................. 98

Figure 2.11. Effect of combination catalytic and cleavage site mutations. ............... 100

Figure 2.12. Schematic of expression vectors pFLAG/myc and pTNT. ......................... 102

Figure 2.13. Comparison of translation efficiency of pFSm with that of the wild type polyprotein. ............................................................................................................. 103

Figure 2.14. Analysis of pTNT/FSm N-terminal cleavage products by Western blotting. .............................................................................................................................. 106

Figure 2.15. Putative SHFV PCP processing cascade. ................................................ 110
CHAPTER I

Functional analyses of West Nile virus (WNV) bicistronic replicons containing different sequence elements

INTRODUCTION

General Characteristics

West Nile virus (WNV) was first isolated from a febrile woman in the West Nile region of Uganda in 1937 (Gubler, 2007; Smithburn, 1940). Until 1999 the geographic distribution of WNV was limited to the Mediterranean countries, parts of Europe, the Middle East, Africa, West Asia, and Australia. In 1999, WNV entered the Western hemisphere, where it has since spread rapidly (Lanciotti et al., 2002). WNV is maintained in a mosquito-bird cycle in nature and the spread of the virus follows the migratory patterns of infected birds. Humans and domestic animals are considered incidental hosts that do not play a role in maintaining the viral life cycle in nature because the viral titers in their blood are not high enough to infect mosquitoes (Gubler, 2007). The majority of WNV infections in humans are asymptomatic, but flu-like symptoms are observed in about 20% of infected individuals. Less than 1% of infected persons develop severe neurologic disease, such as meningitis, encephalitis, or acute flaccid paralysis, which is sometimes fatal. The incidence of neurologic disease and consequent fatality is higher among the elderly (Petersen, Marfin, and Gubler, 2003; Saad et al., 2005).
Classification

The virus family *Flaviviridae* has been subdivided into the genera *Pestivirus*, *Hepacivirus*, and *Flavivirus*. *Pestiviruses* are of major importance in the livestock industry. This genus includes bovine viral diarrhea virus (BVDV) and classical swine fever virus (CSFV). *Hepaciviruses* include hepatitis C virus and GB virus B. *Hepaciviruses* differ from *Pestiviruses* and *Flaviviruses* mainly in the organization of their structural proteins which are encoded at the 5’ end of the genome. However, *Pestiviruses* and *Hepaciviruses* are more closely related to each other in both their genome organization and mechanism of translation initiation than to *Flaviviruses*. WNV is a member of the genus *Flavivirus*. This genus also includes other mosquito-borne viruses, such as Japanese encephalitis virus, yellow fever virus (YFV), dengue virus types 1-4 (DENV), as well as tick-borne viruses, such as tick-borne encephalitis virus (TBEV) (Lindenbach, 2007).

Virion

WNV virions are small enveloped particles, about 50 nm in size that are assembled from three structural proteins. The viral nucleocapsid is composed of the capsid protein (C) and the unsegmented, positive sense (+) RNA genome. Surrounding the nucleocapsid is a host-derived membrane in which the two surface proteins membrane (M) and envelope (E) are anchored by C-terminal hydrophobic regions that traverse the membrane twice (Zhang et al., 2003a; Zhang et al., 2003b). M is derived from a larger precursor termed prM which is present in immature virions. Full length
prM fragments provide structural support for E protein trimers in the immature virion, and following cleavage of prM by the viral serine protease, the E complexes collapse onto the surface of the virion membrane, and the E trimers dissociate to form the E dimers seen in mature virions. Ultimately, E and M interact to form an icosahedral shell on the outer surface of the virus particle (Lindenbach, 2007).

**Genome**

The WNV genome is a (+) strand RNA of about 11 kb. It has a 5' type I cap but lacks a 3' poly(A) tail. The genomic RNA is an mRNA, which encodes a single polyprotein that is co-translationally and post-translationally cleaved by viral and host proteases to produce three structural proteins (E, M and C), and seven nonstructural proteins (Figure 1.1) (Lindenbach, 2007).

Capsid (C) protein is encoded at the 5' end of the genome. It is about 11 kDa in size and is very rich in basic amino acid residues (Kofler, Heinz, and Mandl, 2002). Nascent C (anchored C) contains a C-terminal hydrophobic region that serves as a signal sequence for the translocation of prM (the next protein in the polyprotein) across the membrane of the endoplasmic reticulum (ER). Cleavage of the C-terminus of the capsid protein is catalyzed by the viral NS2B/NS3 protease complex on the cytoplasmic side of the ER membrane, while the C-prM junction on the lumenal side is cleaved by host signal peptidase (Stocks and Lobigs, 1998). Released mature C folds into a dimer. Each monomer contains four alpha helices. However, the mechanism by which encapsidation occurs and the organization of C protein dimers within the nucleocapsid are still
Figure 1.1. Schematic of a flavivirus genome. The polyprotein and mature processed proteins are shown. Figure from (Brinton, 2002).
unknown (Jones et al., 2003; Lindenbach, 2007; Ma et al., 2004). Recent studies suggest
the possibility of additional roles for the C protein. It may activate cell apoptotic
pathways, and is present in the nucleus as well as the cytoplasm of infected cells (Jones et
al., 2003; Westaway et al., 1997; Yang et al., 2002).

The M protein is about 26 kDa and is cleaved from its precursor prM. The N-
terminal region of prM has one to three N-linked glycosylation sites, as well as six
conserved, disulfide-linked cysteine residues. The C-terminal region of prM contains
two transmembrane domains which help to secure prM and M in the ER membranes, and
may aid in the heterodimerization of prM and E. Besides providing structural support for
E in the immature virion, prM also protects E from undergoing conformational changes
during the transit of immature virions through acidic compartments of the trans-Golgi
network (Heinz et al., 1994). Furin-mediated cleavage of virion-associated prM to
release the pr fragment occurs in the Golgi and leads to the rearrangement of E trimers
into the homodimers seen in mature virion structures (Lindenbach, 2007).

The E protein is about 53 kDa and is located on the surface of flavivirus virions.
During infection, it is primarily responsible for facilitating entry into the cell, by binding
to an unknown receptor and mediating the fusion of viral and host membranes (Allison et
al., 1995). The native form of E is composed of three domains. Domain I is located at
the N-terminus and contains the most conserved sequence, which has been hypothesized
to be part of the flavivirus receptor-binding site (Mandl et al., 1989). Domain II contains
a putative fusion peptide, which mediates insertion into the host membrane during virus
entry, and domain III, which is located at the C-terminus of the E protein contains a
hydrophobic domain which serves as a signal sequence for the translocation of NS1 into the lumen of the ER during polyprotein translation (Allison et al., 2001; Lindenbach, 2007; Markoff, Chang, and Falgout, 1994).

Nonstructural protein 1 (NS1) is about 47 kDa and has been shown to exist as a dimer (Winkler et al., 1988). NS1 is the only flaviviral nonstructural protein known to be glycosylated (Grun and Brinton, 1986; Lindenbach, 2007; Winkler et al., 1989). The N-terminal end of NS1 is cleaved from the E protein by host signal peptidase, while the NS1/NS2a junction is cleaved by an unknown host protease (Falgout, Chanock, and Lai, 1989; Falgout and Markoff, 1995). Although the majority of NS1 is retained in infected cells, it is also detected at the cell surface where it is slowly secreted (Lindenbach and Rice, 2003). A YFV replicon with a large deletion in the NS1 gene was shown to be deficient in RNA replication. Upon transfection of this replicon into cells expressing YFV NS1 from noncytopathic Sindbis virus vectors, YFV RNA replication levels were restored, demonstrating that YFV NS1 can be supplied in trans (Lindenbach and Rice, 1997). A significant decrease in minus (-) strand synthesis was observed in the absence of complementation, suggesting that NS1 functions at an early stage in RNA replication (Lindenbach and Rice, 1997). Moreover, YFV NS1 could not be trans-complemented by NS1 from DENV, suggesting that NS1 functions in a flavivirus-specific manner (Lindenbach and Rice, 1999). Additional studies showing colocalization of NS1 with viral replication complexes also suggested a role for this protein in RNA replication, but its exact function in viral RNA replication is not known (Lee, Crooks, and Stephenson, 1989; Schlesinger, Brandriss, and Monath, 1983; Winkler et al., 1988).
NS2A and NS2B are small hydrophobic proteins of about 21 and 13 kDa, respectively. Studies on YFV have shown that a mutation in the C-terminal cleavage site of NS2A resulted in a failure to yield infectious virus in spite of normal viral RNA replication levels, suggesting that this protein may be involved in virus assembly (Kummerer and Rice, 2002). NS2B has been shown to act as an NS3 protease cofactor. Studies have revealed that it forms a stable complex with NS3, and that its absence abrogates NS3 protease stability and activity (Erbel et al., 2006). Both the NS2A/NS2B and NS2B/NS3 junctions are cleaved by the viral serine protease (Falgout and Markoff, 1995).

NS3 is a large ~70 kDa protein with both protease and helicase functions. The N-terminal region comprises the protease domain, which together with NS2B forms an active complex. This protease autocatalytically cleaves itself from the polyprotein, and also cleaves at multiple other sites in the polyprotein (Nall et al., 2004). The C-terminal region of NS3 encodes domains with homology to the superfamily II of RNA helicases and NTPases (Borowski et al., 2000; Wengler and Wengler, 1991). Moreover, the NS3 protein has been shown to act as an RNA 5' triphosphatase, dephosphorylating the 5' end of the genome prior to cap addition (Bartelma and Padmanabhan, 2002). Collectively, these functions establish NS3 as a protein that is indispensable to the replication cycle of flaviviruses.

NS4A and NS4B are small hydrophobic proteins of 16 and 27 kDa, respectively. The C-terminus of NS4A is generated by a combination of host signal peptidase and viral serine protease cleavage (Lin et al., 1993). NS4A has been shown to interact with NS1,
as they are both suspected to reside on the luminal side of the ER membrane (Lindenbach and Rice, 1999). Furthermore, NS4A has been shown to colocalize with double-stranded RNA and vesicle packets within the replication complex, suggesting a role for NS4A/NS1 in RNA replication (Mackenzie et al., 1998). Studies with DENV have detected the presence of NS4B in reticular structures, along with NS3 and viral double-stranded RNA (Miller, Sparacio, and Bartenschlager, 2006). This colocalization suggests that these proteins are associated with the membrane-bound viral replication complexes, but the exact role of NS4B has yet to be determined.

NS5 is a large, multifunctional, and very well conserved protein of about 96 kDa, that is cleaved from the polyprotein by viral serine protease (Cahour, Falgout, and Lai, 1992). The N-terminal region of this protein shows homology to methyltransferase domains (Koonin, 1993) and methyltransferase activity assays done with the N-terminal domain of DENV NS5 confirmed that it was capable of transferring methyl groups to capped substrates (Egloff et al., 2002). Recently, similar studies employing the purified methyltransferase domains of DENV, YFV, JEV and TBEV, demonstrated the ability of these enzymes to specifically methylate the cap structures of flaviviral RNA (Dong et al., 2007). The C-terminal region of NS5 harbors motifs characteristic of viral RNA-dependent RNA-polymerases (RdRps) (Koonin, 1991). Mutations made in the RdRp GDD motif of Kunjin virus, an Australian strain of WNV, completely abolished viral replication (Khromykh, Varnavski, and Westaway, 1998). In studies using purified DENV NS5 protein, de novo initiation of (-) strand viral RNA synthesis from genome RNA templates was observed (Ackermann and Padmanabhan, 2001). Several groups
have shown that NS5 is phosphorylated \textit{in vivo}, and have detected the presence of some NS5 in the nucleus of infected cells (Buckley et al., 1992; Kapoor et al., 1995; Miller, Sparacio, and Bartenschlager, 2006). However, the significance of this localization is not currently understood.

\textbf{Replication cycle}

The flaviviral replication cycle takes place in the cytoplasm of infected cells (Figure 1.2). The infection cycle is initiated when the E protein on the surface of the virion binds to an unknown receptor on a host cell, and virions enter the cell by clathrin-mediated endocytosis. Subsequently, low pH induced fusion between the viral envelope protein and the membrane of the host endocytic vesicle facilitates the release of the viral nucleocapsid into the cytoplasm (Chu and Ng, 2004). The viral genome is released from the nucleocapsid by an unknown mechanism and is translated. Genome translation is dependent on a 5' cap. Translation of the single open reading frame results in a polyprotein that is cleaved by viral and host proteases to produce the mature viral proteins (Lindenbach, 2007). In addition to its function as the only viral mRNA, the genome also serves as a template for the synthesis of (-) strand RNA, which in turn is copied into (+) strand RNA. In infected cells, (+) strand RNA is about 10 to 100 times more abundant than (-) strand RNA (Chu and Westaway, 1985). Replication of the genomic RNA occurs in invaginations of perinuclear membranes, while virion assembly occurs in vesicles of the ER (Cardiff et al., 1973; Lindenbach, 2007). The cleavage of the
Figure 1.2. WNV replication cycle. (A) Attachment and entry of virion and uncoating of the virion RNA. (B) Translation of the virion RNA. (C) Proteolytic processing of the viral polyprotein. (D) Synthesis of minus strand RNA. (E) Synthesis of genomic RNA from the minus strand RNA. (F) Encapsidation of the viral genome and assembly of immature virions. (G) Transport of the virions to the plasma membrane. (H) Exocytosis of the virions. Figure and legend from (Brinton, 2002).
viral structural protein prM to M on immature virions is executed by the Golgi-resident protease furin and leads to the production of mature virions, which are transported to the plasma membrane in vesicles (Mackenzie and Westaway, 2001; Stadler et al., 1997). Virus-containing vesicles fuse with the plasma membrane and the virions are released from the cell (Mackenzie and Westaway, 2001).

**Secondary structures and cis-acting elements in the flaviviral noncoding regions**

The WNV 5' and 3' NCRs are 96 nt and 631 nt in length, respectively (Brinton, Fernandez, and Dispoto, 1986; Markoff, 2003). Stable secondary structures are predicted to form in both the 5' and 3' NCRs (Brinton and Dispoto, 1988). The 3' terminal 95 nt of the 3'NCR are predicted to form two stem loop structures, a short SL of 16 nts and a long terminal SL of 79 nts (Figure 1.3) (Brinton, Fernandez, and Dispoto, 1986; Hahn et al., 1987). The structures of these SLs are conserved. These structures were previously shown to contain binding sites for three cellular proteins, one of which was identified as eukaryotic elongation factor 1 alpha (eEF1A) (Fig. 1.3) (Blackwell and Brinton, 1995; Blackwell and Brinton, 1997). The sequence 5' CACA 3' located on the 5' side of the 3' terminal SL was identified as the major eEF1A binding site, while a sequence 5' CACAG 3' in the top loop of this SL, known as the pentanucleotide motif was identified as a minor eEF1A binding site (Fig. 1.3). A second minor binding site was mapped to the lower part of the 3' SL, and the adjacent small SL (Blackwell and Brinton, 1997). eEF1A
Figure 1.3. West Nile virus NCRs. Cyclization sequences, the major eEF1A binding site and the pentanucleotide motif are indicated in red, and purple, respectively.
is the second most abundant protein in eukaryotic cells, after actin (Condeelis, 1995). It plays a pivotal role in the translation of cellular proteins; it is responsible for recruiting aminoacyl-tRNA to the A site of the ribosome (Condeelis, 1995; Grassi et al., 2007; Lewin, 2000). In addition to its role in translation, eEF1A has also been implicated in interactions with the cytoskeleton, in apoptosis, and in cell proliferation (Grassi et al., 2007). Recent data also suggest that the interaction between eEF1A and the WNV 3' (+) SL facilitates viral (-) strand RNA synthesis (Davis, 2007).

Mutations in the pentanucleotide motif of WNV were shown reduce eEF1A binding by 20%, and similar mutations of this loop inhibited Kunjin replicon RNA replication (Blackwell and Brinton, 1997; Khromykh et al., 2003). Subsequently, mutational analyses in a WNV infectious clone identified two nts in the pentanucleotide motif that are essential for virus replication in vivo (Elghonemy, Davis, and Brinton, 2005). Collectively, these data indicate that the pentanucleotide motif plays an important role in viral RNA replication.

The genomes of all mosquito-borne flaviviruses have been shown to contain 8 nt conserved cyclization (Cyc) sequences near their 5' and 3' termini that are complementary to each other. These sequences are located within the capsid gene (5' Cyc) and in the 3' NCR just upstream of the 3' (+) SL (3' Cyc). The sequences of the 5' and 3' Cyc are identical among the mosquito-borne flaviviruses, and consist of 5' UCAAUAUG 3' and 5' CAUAUUGA 3', respectively (Hahn et al., 1987; Markoff, 2003; Men et al., 1996). Base pairing between these sequences is hypothesized to cyclize (+) strands (Fig. 1.3) (Hahn et al., 1987). Complementarity between the 5' and 3' Cyc was subsequently shown to be
essential for flavivirus replication, but the specific role that this long distance RNA-RNA interaction plays in the viral life cycle is unknown (Khromykh et al., 2001; Men et al., 1996; You and Padmanabhan, 1999).

There are additional nucleotides flanking the 8 nt 5' and 3' Cyc that are complementary to each other, but the sequences of these regions vary among the different flaviviruses. These sequences can range in length from 13 to 21 nt, and are referred to as the 5' CS and CS1 in the 3' NCR (Hahn et al., 1987). Additional conserved sequences which are tandem repeats (CS2 and RCS2) were shown to be located upstream of CS1 (Hahn et al., 1987). Mutational analyses in a DENV infectious clone showed that although deletion of CS2 and RCS2 resulted in viruses with reduced replication competence, only deletion of CS1 resulted in a lethal phenotype (Men et al., 1996).

In a recent study, the 3'-5' cyclization of chimeric DENV RNA molecules was visualized by atomic force microscopy (Alvarez et al., 2005b). The RNA molecules used in this study consisted of the complete viral 5' NCR, followed by the first 186 nt of the capsid coding region, a luciferase coding sequence, and the complete viral 3' NCR. An antisense luciferase RNA sequence was hybridized to its complementary region in the chimeric RNA, generating a double-stranded luciferase sequence flanked by single stranded viral 5' and 3' termini. The detection of circular RNA molecules by atomic force microscopy indicated that the 5' and 3' single stranded regions could base pair under these conditions. Subsequent analyses showed that in addition to the 5' and 3' Cyc sequences, adjacent complementary 3' and 5' regions termed upstream AUG regions (UARs), of 15 to 20 nt could base pair. Studies with mutated subgenomic replicons and recombinant
viruses suggested that base pairing between these UARs might also be needed for efficient RNA replication (Alvarez et al., 2005a; Alvarez et al., 2005b).

**Mechanisms of translation initiation**

*Cap-dependent translation of cellular mRNAs*

During the generation of mRNAs in cell nuclei, the termini of the RNA are modified prior to translocation into the cytoplasm. The 5' cap and 3' poly(A) tail are hypothesized to both increase the efficiency of translation and provide stability for the mRNA (Varani, 1997). Although there is consensus about the synergistic functions of the 5' cap and poly(A) tail in translation initiation, there are various models for how assembly of the translation initiation complex occurs on the cellular 5' cap. The poly(A) binding protein (PABP) binds to the 3' poly(A) and to the PABP-interacting protein-1 (PAIP-1). The cap-binding complex eIF4F, which consists of eIF4G, eIF4E and eIF4A recognizes and binds the 5' cap structure via its 4E subunit (Craig et al., 1998; Gingras, Raught, and Sonenberg, 1999). The 4G subunit of eIF4F acts as a scaffolding protein and also interacts with PAIP-1. This interaction circularizes the mRNA by bridging its 5' and 3' termini (Figure 1.4A) (Baird et al., 2006). The binding of the 40S ribosomal subunit to the cap and the eIF4F complex leads to the formation of the larger 43S preinitiation complex. This complex scans along the mRNA in the 5' to 3' direction in search of a start
A. Protein complex involved in cyclization of cell mRNAs. The eIF4G, eIF4A and eIF4E subunits make up the cap-binding complex eIF4F, which binds the 5' cap via its 4E subunit. The interaction between eIF4F and PABP is mediated by PAIP-1.

B. Protein complex involved in rotavirus mRNA cyclization. The interaction between eIF4F and the viral 3' end is mediated by viral NSP3. Figure modified from Edgil and Harris (2006).
codon, a process that requires energy which is provided by ATP (Baird et al., 2006; Lopez-Lastra, Rivas, and Barria, 2005).

That mRNA circularization increases the efficiency of protein synthesis has been well supported by experimental evidence (Craig et al., 1998; Gray et al., 2000; Kozlov et al., 2001; Varani, 1997). However, it is not known exactly how circularization of the mRNA stimulates translation. It has been hypothesized that the 3'-5' interaction (closed-loop model) allows efficient recycling of 40S ribosomes, and rapid re-initiation of translation. In addition, cyclization may provide increased stability for the mRNA and the translation complex, and may also facilitate the translation of full length RNAs (Edgil and Harris, 2006; Lopez-Lastra, Rivas, and Barria, 2005).

**Cap-dependent translation of some viral mRNAs**

Translation of flavivirus genomes is also thought to initiate after ribosomal scanning from the 5' cap to the first AUG (Lindenbach, 2007). However, it is not known which cellular and/or viral proteins are involved in regulating translation of flavivirus genomic RNA. The genome of rotavirus, a member of the family Reoviridae, consists of eleven double-stranded RNA segments, which are transcribed into mRNAs that have a 5' cap but lack a 3' poly(A) tail. Studies of rotavirus RNAs revealed that the rotavirus-encoded protein NSP3 binds to the 3' end of the genome as well as to the cap binding complex scaffolding protein eIF4G. This protein-protein interaction facilitates circularization of the mRNA in the absence of a viral 3' poly(A) tail (Fig. 1.4B) (Groft and Burley, 2002). Studies of BVDV have suggested that interactions between the
NF90/NFAR (nuclear factor 90/nuclear factor associated with RNA) group of cellular proteins and both the viral 5' and 3' NCRs, may facilitate cyclization of Pestivirus genomes (Isken et al., 2003). This cyclization is thought to play a role in viral translation, because the 5' and 3' motifs that are bound by these cellular proteins are considered important regulators of translation (Yu et al., 2000). It is possible that cellular proteins binding to the WNV NCRs may also establish a functional bridge between the ends of the genome and thus provide a regulatory switch between RNA replication and translation template functions.

**IRES-dependent translation**

Several viruses as well as some eukaryotic mRNAs have been shown to initiate translation independently of a 5' cap, by use of an internal ribosome entry site (IRES). IRESs are nucleotide sequences that fold into distinct secondary and tertiary structures and that recruit translation initiation complex components for initiation of translation at an internal region of the mRNA (Lopez-Lastra, Rivas, and Barria, 2005). IRESs were discovered in studies of poliovirus (PV) and encephalomyocarditis virus (EMCV), members of the family *Picornaviridae*. The insertion of the PV 5' NCR between two ORFs of a bicistronic construct mediated efficient translation of the second ORF that was independent of the first ORF (Pelletier and Sonenberg, 1988). Subsequently, IRESs were identified in members of other virus families, such as the *Flaviviridae, Retroviridae*, and *Herpesviridae* (Lopez-Lastra, Rivas, and Barria, 2005).
Two main types of IRES have been identified in the Picornaviridae. Type I IRESs are found in the genomes of rhinoviruses and enteroviruses (such as PV). Type II IRESs are found in cardioviruses such as EMCV, and aphthoviruses such as foot-and-mouth disease virus (FMDV). These two types of IRES differ in their structure, as well as in their location in the viral genome relative to the initiation codon. Type I IRESs are located up to 150 bases upstream of the initiation codon, while type II IRESs are located adjacent to the initiation codon (Fig. 1.5) (Andino et al., 1999; Belsham and Sonenberg, 1996). An element that is common to both types of structures is the GNRA sequence (G, guanine; N, any nt; R, purine; A, adenine), which is located in SL IV of the type I and SL I of the type II IRESs (Fig. 1.5) (Racaniello, 2007).

IRES-mediated translation initiation has also been observed in some cellular eukaryotic mRNAs. The first cellular IRES element was discovered in the mRNA of the immunoglobulin heavy chain binding protein BiP, when it continued to be actively translated in picornavirus-infected cells in spite of virus-mediated shutdown of cellular cap-dependent translation (Sarnow, 1989). IRES elements were later found in numerous other cell mRNAs that are involved in regulating gene expression during development, differentiation, cell cycle progression, cell growth, apoptosis and stress. It is hypothesized that about 3-5% of cellular mRNAs are capable of being translated in an IRES-dependent manner (Komar and Hatzoglou, 2005).
Figure 1.5. Two major classes of picornavirus IRES. Figure modified from Racaniello (2007).
Replicons as tools for studying viral RNA replication and translation

Replicons are autonomously replicating constructs that contain the viral genes necessary for RNA amplification in host cells, but lack the viral structural genes, so no virus particles are produced. Self-replicating replicons can only be derived from viruses with (+) sense RNA genomes, because translation of the viral genes that are needed for subsequent RNA replication can occur from the input RNA. Bicistronic flavivirus replicons were constructed by inserting an IRES sequence in front of a reporter gene into the viral 3' NCR. In this type of bicistronic system, the translation of the nonstructural genes is controlled separately from that of the reporter gene, so that cap-dependent translation controls the expression of the viral nonstructural proteins, while the IRES controls the expression of the reporter gene. The expression of the viral nonstructural proteins leads to initiation of RNA replication. Genome RNA templates are thought to switch between RNA replication and translation.

The detection of reporter proteins in transfected cells within a few hours after RNA transfection is assumed to result from the translation of input RNA. Proteins detected at later time points are assumed to have been translated from replicated RNA. Reporter assays done at later time points have been used to quantify the level of replicon RNA replication. Alternatively, RNA replication can be analyzed directly by reverse transcription PCR (RT-PCR), real-time qRT-PCR or Northern blotting. An alternative strategy for uncoupling the translation of the viral nonstructural genes from that of a reporter gene is the insertion of a reporter gene downstream of the viral 5' cap, and insertion of an IRES 5' of the viral nonstructural genes. For this type of bicistronic
replicon, viral cap-dependent translation produces the reporter protein and IRES-dependent translation produces the viral nonstructural proteins.

Virus replicon systems have facilitated studies of several types of viruses, including Sindbis virus (Frolov and Schlesinger, 1994), PV (Kaplan and Racaniello, 1988), and rhinovirus (McKnight and Lemon, 1996). Uncoupled bicistronic replicons have been used to analyze the effect of viral cis-acting elements on various aspects of the viral life cycle.

The initial flaviviral replicon studies were done with constructs derived from Kunjin virus (Khromykh and Westaway, 1997; Varnavski and Khromykh, 1999). The first replicons generated were monocistronic, and were prepared from full-length cDNA clones by deleting the prM, E and capsid genes. These replicons were transfected into BHK cells by electroporation, and RNA replication was analyzed by immunofluorescence analysis (IFA) of cells using antibodies to viral nonstructural proteins. Replicons lacking the capsid gene were shown to be defective in RNA replication. Subsequent studies with replicons containing various lengths of the capsid coding regions revealed that the first 60 nt of the capsid protein were required for RNA replication (Khromykh and Westaway, 1997). The 5' Cyc sequence is located in this region of the capsid sequence.

Kunjin replicons with deletions in the 3' NCR were subsequently used to investigate the role of the 3' NCR in flavivirus RNA replication. The replication of RNA from these replicons was analyzed by IFA of cells transfected with these mutant constructs, as well as by Northern blotting of cell lysates. These analyses revealed that
deletion of 76 nt at the 5' end of the 3' NCR did not affect RNA replication, whereas a longer deletion of 352 nt resulted in a significant impairment of RNA replication (Khromykh and Westaway, 1997). This suggested that a portion of the 5' end of the flaviviral 3' NCR, which varied significantly among flaviviruses, could be removed or modified without affecting RNA replication. Replicons containing chloramphenicol acetyl transferase (CAT) reporters in both the (+) and (-) sense orientations inserted into this variable region of the 3' NCR were subsequently generated. Readily detectable expression of CAT in lysates from transfected cells was detected with replicons containing CAT in the (+) sense orientation. However, no CAT activity was detected in lysates from cells transfected with replicons containing CAT in the (-) sense orientation (Khromykh and Westaway, 1997).

Kunjin replicons carrying an IRES/neomycin cassette (Neo) in the variable region of the 3' NCR were used to select Neo-expressing cells by maintaining the cells in medium containing Geneticin (G418), an antibiotic that is toxic for normal cells. This was done to extend the period of replicon expression and to increase the proportion of replicon-expressing cells. Continued passaging of surviving cells resulted in the recovery of a cell culture that persistently expressed this replicon for as long as 9 passages or 41 days after transfection (Khromykh and Westaway, 1997). Transfection of these cell lines with replicons containing a mutation in the GDD motif allowed trans-complementation of these replicons by endogenous viral RdRp, and this resulted in efficient replication of GDD mutant replicons (Khromykh, Kenney, and Westaway, 1998). The importance of the 5' and 3' Cyc for RNA replication was analyzed using a Kunjin monocistronic
replicon. The predicted 5' or 3' Cyc sequences were mutated simultaneously at multiple positions and RNA replication was monitored by IFA and Northern blotting. The results showed that base pairing between 3' and 5' complementary sequences was essential for efficient replication of Kunjin virus RNA (Khromykh et al., 2001).

Bicistronic replicons have also been used to study WNV. A WNV replicon that contained dual reporter genes, a *Renilla* luciferase (Rluc) gene in frame with the partial capsid coding sequence at the 5' end, and an IRES/Neo cassette in the variable region of the 3' NCR, was constructed. A stable cell line persistently expressing this replicon was established by selection with G418. The reporting cells were then incubated with a series of known WNV inhibitors and the activity of Rluc was monitored. Decreased luciferase activity was observed after treatment with inhibitors, suggesting that reporting cell lines could be used for anti-WNV drug discovery screens (Lo, Tilgner, and Shi, 2003; Shi, Tilgner, and Lo, 2002).

Other studies employed bicistronic replicons expressing a Rluc/FMDV2A/Neo cassette from an IRES located in the viral 3' NCR. Replicon carrying cells were selected with G418, and treatment of these cells with the nucleoside analog ribavirin resulted in reduced luciferase activity (Rossi et al., 2005). Later it was shown that when this replicon was expressed in different cell lines, it displayed different levels of sensitivity to interferon (Scholle and Mason, 2005).

Monocistronic replicons derived from TBEV were used in studies to generate a cell line that was able to package replicons into virus-like particles. Cells that constitutively expressed the viral structural proteins E and M were transfected with
monocistronic TBEV replicons that lacked the E and M proteins. Supernatants from these transfected cells were transferred to fresh Vero cells and were able to initiate a further round of infection, demonstrating that structural proteins provided in trans produced infectious viral particles (Gehrke et al., 2003).

Additional studies with TBEV replicons examined the functional role of the second transmembrane region (TM2) of the E protein independently of its role as an internal signal sequence in the polyprotein. A bicistronic replicon was made, expressing the viral envelope proteins prM and E from a separate IRES-dependent cistron located in the viral 3’ NCR. The modified cap-dependent cistron had a complete copy of the capsid gene but lacked the entire prM and E genes, except for TM2, which contained the signal sequence for NS1 membrane translocation. In order to avoid possible homologous recombination between the two copies of TM2, which were present in both cistrons, the codon sequences in the 3’ copy were wobbled, which changed the nucleotide sequence without changing the amino acid sequence. Incremental deletions were then made in the functional domains of the 3’ copy of TM2. BHK cells were transfected with mutant replicon RNAs and the harvested supernatants were passaged several times. Only the mutant with the shortest deletion was able to infect new cells, and only at reduced efficiency, suggesting that the presence of TM2 is essential for the assembly of new viral particles (Orlinger et al., 2006).

Monocistronic flavivirus reporter replicons have also been reported. Cap-dependent translation of an Rluc reporter occurred when it was inserted in frame with the truncated viral capsid coding sequence. The proper processing of the capsid-reporter
fusion protein from the nonstructural proteins was ensured by inserting the coding sequence of the autocatalytic FMDV 2A protease between the reporter gene and the viral nonstructural genes. To evaluate the efficiency of RNA replication by these replicons, transfected BHK cells were assayed for reporter activity at 12 to 48 hr after transfection. Increased levels of activity at 48 hr were observed, and were assumed to indicate efficient RNA replication (Jones, Patkar, and Kuhn, 2005).

In bicistronic replicons derived from YFV, green fluorescent protein (GFP) or Rluc was expressed from an IRES inserted in the viral 3' NCR. However, the IRES-driven translation of reporter genes inserted in the 3' NCR was observed to be less efficient than cap-dependent translation of the same reporter gene inserted at the 5' end of the viral genome (Jones, Patkar, and Kuhn, 2005). The same effect was observed in bicistronic Kunjin and WNV replicons. A strong inhibitory effect on CAT and GFP activity was noted when these reporters were placed downstream of an IRES in the viral 3' NCR (Khromykh and Westaway, 1997; Shi, Tilgner, and Lo, 2002).
Goals

The original goal of this part of the study was to utilize a unique uncoupled WNV replicon to analyze the effects of mutation of various viral cis-acting elements on RNA replication and/or translation. In this uncoupled replicon, the translation of the viral nonstructural proteins was under the control of an EMCV IRES, while the translation of the N-terminal portion of the C protein alone or fused to Renilla luciferase was under the control of the 5' cap in the viral 5' NCR. Since the replication of the original replicon was not efficient, the goal became to improve the functioning of the uncoupled replicon. A highly sensitive real time qRT-PCR assay was proposed as a direct means of assaying viral RNA replication, and luciferase assays, IFA and Western blotting were proposed to analyze viral protein translation.
MATERIALS AND METHODS

Cell culture

Baby hamster kidney-21/W12 cells (BHK cells) were grown at 37°C with 5% CO₂ in Eagle’s minimum essential media (MEM) supplied with 4.5% heat-inactivated fetal bovine serum (FBS) in 10 μg/ml of gentamicin.

Preparation of S100 cell extracts

Confluent cells grown in T150 tissue culture flasks were washed in ice-cold PBS, harvested by scraping, and pelleted by centrifugation at 150 x g for 5 minutes at 4°C. After removal of PBS, the cell pellet was stored at -80°C or was used immediately. The cell pellet was resuspended in cytolysis buffer [10 mM HEPES (pH 7.5), 100 mM NaCl, 1% Triton X-100, 5 mM DTT, and complete Mini, EDTA-free protease inhibitor (Roche)] at 5 x 10⁷ cells/ml, vortexed for 30 seconds, and stored on ice for 30 minutes. The lysate was centrifuged at 2,200 x g for 5 minutes to pellet the nuclei, and the supernatant was subsequently centrifuged at 100,000 x g for 1 hour at 4°C. The resulting supernatant was aliquoted and stored at -80°C until use. The total protein concentration was determined on a spectrophotometer (Pharmacia Biotech), and ranged from 0.2 to 1 μg/μl.

Preparation of cDNA clones and DNA templates for in vitro RNA transcription

Viral RNA was purified from lineage I WNV, strain EG101, and fragments were reverse transcribed with a set of viral sequence specific primers. cDNAs amplified by PCR were Topo-XL cloned and sequenced. Correct fragments were assembled according to the scheme illustrated in Figure 1.6B to produce CNI. The IRES sequence was
amplified from a DNA clone of EMCV kindly provided by Dr. Ann Palmenberg. To construct \( C^{\text{NLI}} \), a \textit{Renilla} luciferase reporter gene (Promega) was inserted in frame with the N-terminal capsid coding region by double PCR. \( C^{\text{NFI}} \) was generated by inserting a longer flanking sequence on the 5' side of the IRES as well as the C-terminal 89 nts of the E gene upstream of NS1 using appropriate restriction sites. Mutation of the viral RNA-dependent RNA-polymerase motif (GDD to GAA) in this construct yielded \( C^{\text{NFI/GDD}} \), while addition of \textit{Renilla} luciferase to \( C^{\text{NFI}} \) produced \( C^{\text{NFLI}} \). Addition of 117 nts of the capsid coding region (but not the anchor sequence) to \( C^{\text{NFI}} \) created \( C^{\text{FI}} \), and addition of \textit{Renilla} luciferase to this construct produced \( C^{\text{FLI}} \). Replicons with truncated capsid coding regions \( C^{249\text{LFI}}, C^{279\text{LFI}}, \) and \( C^{297\text{LFI}} \) were generated by double PCR using a plasmid containing the entire capsid coding region. Finally, addition of \textit{Firefly} luciferase to \( C^{\text{NFI}} \) yielded \( C^{\text{NFLFI}} \).

**Transformation**

All replicons are cloned into pBR322 and were transformed into One Shot Top 10 Chemically Competent \textit{E.Coli} cells (Invitrogen). Transformed cells were plated on LB agar plates containing Ampicillin.

**\textit{In vitro} RNA transcription**

Purified plasmid DNA (Qiagen QIAprep Miniprep Kit) was linearized at 37°C for 3 hr using the AflII restriction enzyme. Linearized plasmid DNA was purified using QIA quick Gel Extraction Kit PB buffer (Qiagen) and \textit{in vitro} transcribed using the AmpliCap SP6 High Yield Message Maker Kit (Epicentre) according to the manufacturer’s protocol. Briefly, 1 \( \mu g \) of template DNA, 2 \( \mu l \) 10X transcription buffer, 5 \( \mu l \) Cap/NTP
PreMix, 2 μl 100 mM DTT, and 2 μl AmpliCap SP6 Enzyme Solution were combined at room temperature with enough RNase-free water to bring the volume to 20 μl. The mixture was incubated at 37°C for 2 hr and subsequently treated with RNase-free DNase at 37°C for 15 min. The resulting RNA was stored at -80°C until transfection.

**Real time qRT-PCR**

Cells were seeded in 6-well dishes 24 hr before transfection to achieve 50-60% confluency. Cells were transfected with 0.1 μg per well of *in vitro* transcribed replicon RNA in 1 ml of OptiMEM serum free media using 10 μl of DMRIE-C transfection reagent (Invitrogen). Transfection media were replaced with MEM after 2 hr and time points were taken every 24 hr over the following days. Cells were washed in PBS and trypsinized at 24, 48, 72, 96 and 120 hr after transfection. Cells were spun down, pellets were resuspended in TRI Reagent (Molecular Research Center, Inc.) and stored at -80°C until further use. Total RNA was extracted and purified according to the manufacturer’s instructions (Molecular Research Center, Inc.). Purified RNA was resuspended in 20 μl of RNase-free water and stored at -80°C until further use. The relative amount of intracellular viral RNA was determined on an Applied Biosystems 7500 real time RT-PCR system, using the TaqMan one-step RT-PCR kit according to the manufacturer’s instructions (Applied Biosystems). The primers used to detect WNV RNA targeted the NS1 region of the genome and were For 5’-GGC GGT CCT GGG TGA AGT TAA-3’ and Rev 5’-CTC CGA TTG TGG TTG CTT CGT-3’ and the FRET probe was 5’Fam-TGC ACT TGG CCT GAA ACG CAC ACT TTG T-3’ TAMRA. Briefly, 1 μl (0.2 – 0.5 μg) of each RNA sample was added to a 24 μl TaqMan PCR universal cocktail.
to give a final volume of 25 μl. The cocktail consisted of 12.5 μl 2X Master Mix, 0.625 μl 40X MultiScribe and RNase Inhibitor Mix, and 0.5 μl each of 100 pm/μl forward and reverse primers, as well as 1 μl of a 10 pm/μl probe. Real time RT-PCR was performed under the following conditions: 30 min at 48°C, 10 min at 95°C, 40 cycles of 15 s at 95°C, and 1 min at 60°C. Viral RNA levels were normalized to the levels of cell GAPDH mRNA and are shown relative to the levels of replicon RNA detected at 5 hr (CNFI/GDD) or 24 hr (all other replicons) after transfection. The calibrator levels were set at a value of 1. All assays were performed in triplicate for each sample using the singleplex format and a no-template control was included in each assay.

**Western blotting**

BHK cells were seeded in 6-well dishes 24 hr before transfection to achieve 50-60% confluency. Cells were transfected with 1 μg per well of *in vitro* transcribed replicon RNA in 1 ml of OptiMEM serum free media using 10 μl of DMRIE-C transfection reagent (Invitrogen). Transfection media were replaced with MEM after 2 hr and time points were taken every 24 hours over the following days. Cells were washed in PBS and trypsinized at 5, 24, 48, and 72 hr after transfection. Cells were spun down and pellets were resuspended in RIPA buffer containing protease inhibitors. Proteins were separated by 10% SDS-PAGE, transferred electrophoretically to a nitrocellulose membrane, then immunoblotted using primary antibodies to WNV nonstructural proteins or capsid protein followed by HRP-linked secondary antibodies. WNV nonstructural proteins were detected using a 1:300 dilution of mouse hyperimmune anti-WNV sera (a gift from Dr. Robert Tesh, UTMB, Galveston, TX), and a 1:3000 dilution of HRP-linked
anti-mouse IgG (Cell Signaling). Capsid protein was detected using a 1:500 dilution of anti-West Nile virus Core protein made in rabbits (ProSci, Inc.), and a 1:2000 dilution of HRP-linked anti-rabbit IgG (Cell Signaling). Protein bands were detected using SuperSignal West Pico Chemiluminescent Substrate (Pierce).

**Immunofluorescence analysis**

Cells were seeded on microscope coverslips placed in 24-well dishes 24 hr before transfection to achieve 20-30% confluency. Cells were transfected with 1 μg per well of *in vitro* transcribed replicon RNA in 1 ml of OptiMEM serum free media using 10 μl of DMRIE-C transfection reagent (Invitrogen). Transfection media were replaced with MEM after 2 hr and time points were taken at 5, 24, 48, and 72 hr after transfection. Cells were washed in PBS, fixed by incubation with 4% paraformaldehyde in PBS for 10 min at room temperature and then permeabilized with methanol at -20°C for 10 min. Coverslips were blocked overnight with 5% horse serum (Invitrogen) in PBS. Coverslips were incubated with a 1:100 dilution of mouse hyperimmune serum against WNV, or rabbit anti-West Nile virus Core protein (ProSci, Inc.) in PBS containing 5% horse serum for 1 hr at 37°C, and then washed 3 times with PBS. Coverslips were then incubated with a 1:300 dilution of chicken anti-mouse or chicken anti-rabbit IgG-TR (Santa Cruz Biotechnology) in PBS containing 5% horse serum and 0.5 μg/ml Hoechst 33258 (Molecular probes) to stain the nuclei. Coverslips were washed with PBS, mounted on glass slides with Prolong mounting media (Invitrogen) and visualized with a 100 X oil immersion objective on a LSM 510 laser confocal microscope using LSM 5 (Ver. 3.2).
software (Carl Zeiss Inc., Thornwood, NY). All images were obtained using the same experimental conditions and microscope settings.

**Luciferase assays**

BHK cells were seeded in 6-well dishes 24 hr before transfection to achieve 50-60% confluency. Cells were transfected with 1 μg per well of *in vitro* transcribed RNA in 1 ml of OptiMEM serum free media using 10 μl of DMRIE-C transfection reagent (Invitrogen). Transfection media were replaced with MEM after 2 hr and at 5, 24, 48, 72, 96, and 120 hr, cells were rinsed in PBS and lysed in 100 μl of 1X reporter lysis buffer (Promega). Luciferase assays were done by mixing 20 μl of cell lysate with 50 μl of *Renilla* luciferase substrate (Promega) and measuring luciferase activity on a luminometer. All data are expressed in terms of relative light units (RLU).

**Gel mobility shift assays**

Increasing concentrations (0.2, 0.5, and 0.8 μg) of BHK S100 cytoplasmic extracts were incubated in gel shift buffer (GS buffer) with yeast tRNA (200 ng) used as a nonspecific competitor RNA, an RNase inhibitor (Roche) (10 units), and a $^{32}$P-labeled WNV capsid region RNA probe (2000 cpm final concentration per reaction) for 30 minutes at room temperature. The RNA-protein complexes formed were resolved by 5% polyacrylamide gel electrophoresis (PAGE) in 1X TBE buffer at 100 volts at 4ºC. The gel was then dried and analyzed on a Bio-Imaging Analyzer PhosphorImager (Molecular Dynamics).
RESULTS

Construction of a WNV bicistronic replicon

Members of the family Flaviviridae, such as pestiviruses and hepatitis C virus, have a 5' IRES. The members of the genus Flavivirus do not. The effect of a 5' IRES on the replication and translation of flavivirus RNA from WNV was analyzed using a replicon system. A replicon consisting of the WNV 5' NCR, the 5' 198 nt of the capsid coding region, an EMCV IRES (I), all the WNV nonstructural genes, and the 3' NCR was constructed by another member of the lab. The 5' region of the capsid was included because it contained the 5' Cyc. Cyc motifs are widely conserved among flaviviruses, and to avoid a possible negative effect on replication, these elements were included in the replicon. To generate this bicistronic replicon, viral RNA was purified from lineage I WNV, strain EG101, and fragments were reverse transcribed with a set of viral sequence specific primers. cDNAs amplified by PCR were Topo-XL cloned (Invitrogen) and sequenced. Correct fragments were assembled according to the scheme illustrated in Figure 1.6B and produced CNI. The IRES sequence was amplified from a DNA clone of EMCV kindly provided by Dr. Ann Palmenberg. To generate a bicistronic replicon that could be used to measure cap-dependent translation, a Renilla luciferase (L) reporter gene was inserted in frame with the partial capsid coding region by double PCR. This replicon was named CNI (Fig. 1.6B). Plasmid DNA was linearized with AflIII restriction enzyme, and then in vitro transcribed using SP6 RNA polymerase. Cap-dependent translation of this transcript produces a truncated capsid-luciferase fusion protein, that
Figure 1.6. Construction of WNV bicistronic replicons. (A) Schematic of a flavivirus genome. The polyprotein and mature processed proteins are shown. (B) Strategy used to engineer the CNI and CNLI replicon cDNAs. The restriction sites used to assemble genomic fragments and their positions in the genome are indicated. Renilla luciferase was inserted by double PCR using appropriate primers.
was used to analyze translation efficiency by measurement of the activity of luciferase. IRES-dependent translation of the nonstructural region would produce viral replication complexes that would allow viral RNA replication. RNA replication was analyzed directly by measurement of genomic RNA levels by relative real time qRT-PCR.

Measurement of CNI and CNLI RNA replication by real time qRT-PCR

CNI and CNLI replicon DNA was digested with AflII restriction enzyme and subsequently in vitro transcribed using SP6 RNA polymerase. Residual DNA was removed from the reaction with DNase and the RNA was used to directly transfect 30-50% confluent BHK monolayers. Cells were harvested at 24, 48, 72, 96, and 120 hr after transfection, RNA was extracted and the amounts of intracellular viral RNA levels were determined by real time qRT-PCR (Fig. 1.7). Although RNA levels of CNI increased almost 4 fold by 96 hr (panel A), they were significantly lower than those seen after transfection of WNV infectious clone RNA. However, RNA levels produced by CNLI were even lower, never reaching those of residual input RNA remaining at 24 hr (panel B). This result suggested that the addition of a Renilla luciferase reporter suppressed the ability of the replicon to replicate viral RNA.

Mung bean nuclease treatment

In an effort to improve the quality and authenticity of WNV replicon RNA transfected into BHK cells, AflII digested replicon DNA was treated with mung bean nuclease prior to in vitro transcription. This treatment was intended to remove single-
Figure 1.7. Analysis of (A) $\text{C}^\text{NI}$ and (B) $\text{C}^\text{NL}$ WNV replicon RNA replication by real time qRT-PCR. BHK cells were transfected with 100 ng of $\text{C}^\text{NI}$ or $\text{C}^\text{NL}$ replicon RNA per well. The amount of viral genomic RNA detected at 24 hr was set at 1 and RNA levels at later time points were expressed relative to the 24 hr level. Viral RNA levels were normalized to the levels of cell GAPDH, and all assays were done in triplicate. Error bars represent the standard error of the mean (n=3) and are based on an RQMin/Max of 95% confidence level.
stranded protruding ends resulting from restriction enzyme digestion, and produce the authentic 3' end of the WNV genome. However, this treatment did not result in higher levels of replicon RNA replication in BHK cells (Fig. 1.8). Therefore, mung bean nuclease was omitted from all subsequent experiments.

**Addition of an IRES 5' flanking sequence and the 3' E signal sequence to the WNV bicistronic replicon**

It was expected that because the addition of luciferase to $C^N_I$ would increase the distance between the 5' Cyc and the IRES in the replicon $C^N_{LI}$ and that this would improve the efficiency of long distance 5'-3' RNA interactions, as well as the replication efficiency. However, decreased replication was observed after transfection of $C^N_{LI}$ (Fig. 1.7B). An alternative strategy for improving replicon RNA replication involved the insertion of the authentic IRES 5' flanking sequence. This sequence was previously shown to facilitate the proper folding of the IRES, keeping it from interfering with the folding of adjacent viral RNA structures. NS1 is the N-terminal gene in the nonstructural polyprotein. Another possible reason for the poor replication of the bicistronic replicons could be the lack of the E signal sequence which is thought to be required for NS1 translocation into the lumen of the ER, and proper orientation of the polyprotein in the ER membrane. The addition of the IRES 5' flanking sequence and the E signal sequence yielded $C^N_{FI}$, and insertion of *Renilla* luciferase into this replicon produced $C^N_{LFI}$. Analysis of viral RNA replication levels in BHK cells transfected with these replicons showed that while the RNA replication of $C^N_{FI}$ was less efficient than that of $C^N_I$, that of
Figure 1.8. Effect of treating AflII digested replicon DNA with mung bean nuclease prior to *in vitro* RNA transcription. The AflII digested replicon DNA was either (A) untreated or (B) treated with mung bean nuclease prior to *in vitro* transcription. BHK monolayers (30-50% confluency) were transfected with 100 ng of *in vitro* transcribed C^NI^ replicon RNA per well. The amount of viral genomic RNA detected at 24 hr was set at 1 and RNA levels at later time points were expressed relative to the 24 hr level. Viral RNA levels were normalized to the levels of cell GAPDH, and all assays were done in triplicate. Error bars represent the standard error of the mean (n=3) and are based on an RQMin/Max of 95% confidence level.
C\textsuperscript{NLFI} was as efficient as C\textsuperscript{NI} but was sustained at higher levels through 120 hr. However, the addition of these two elements did not significantly increase the level of viral RNA replication, as RNA levels never exceeded a 3.5 fold increase (Fig. 1.9).

**Analysis of the efficiency of transfection and replicon function**

Previous studies have shown that a mutation in the GDD motif of the NS5 RdRp rendered Kunjin virus RNAs replication defective (Khromykh, Kenney, and Westaway, 1998). A mutation was introduced into C\textsuperscript{NLFI}, which changed GDD to GAA. This mutant was designated C\textsuperscript{NLFI/GDD}. Since the protein produced from C\textsuperscript{NLFI/GDD} RNA only represented translation of input RNA, this replicon was used to analyze the efficiency of transfection. To determine the transfection efficiency, viral antigen production was detected by IFA. Cells were transfected with C\textsuperscript{NLFI/GDD} RNA, and were then analyzed by confocal microscopy using anti-WNV mouse hyperimmune ascites fluid (MHIAF) (Fig. 1.10A, and B). Enumeration of total and virus antigen positive cells indicated that the transfection efficiency was about 88% (Fig. 1.10C). These results indicate that the low level of replicon RNA replication was not the result of poor transfection efficiency.

C\textsuperscript{NLFI/GDD} was also used to measure the kinetics of decay of input RNA in the absence of replication. Comparison of the levels of RNA at various times after C\textsuperscript{NLFI/GDD} RNA transfection allowed a more accurate estimate of the levels of RNA replication by polymerase active replicons. BHK cells were transfected with C\textsuperscript{NLFI} or C\textsuperscript{NLFI/GDD} RNA, and RNA was extracted and purified from cells at various timepoints. Analyses of RNA replication by real time qRT-PCR showed that the levels of C\textsuperscript{NLFI/GDD}
Figure 1.9. Analysis of (A) C\textsuperscript{N}FI and (B) C\textsuperscript{NL}FI WNV replicon RNA replication. BHK cells were transfected with 100 ng of C\textsuperscript{N}FI or C\textsuperscript{NL}FI replicon RNA per well. The amount of viral genomic RNA detected at 24 hr was set at 1 and RNA levels at later time points were expressed relative to the 24 hr level. Viral RNA levels were normalized to the levels of cell GAPDH, and all assays were done in triplicate. Error bars represent the standard error of the mean (n=3) and are based on an RQMin/Max of 95% confidence level.
Figure 1.10. Analysis of translation competence and transfection efficiency of a nonreplicating replicon, and viral RNA decay kinetics, and replication of a nonreplicating and replicating replicon. (A) WNV nonstructural proteins translated from input RNA in transfected BHK cells detected by immunofluorescence using anti-WNV hyperimmune serum. BHK cells were transfected with 1 μg of C\textsuperscript{N}FI/GDD replicon RNA per well, and cells were fixed and permeabilized 5 hr after transfection. (B) Mock-transfected cells. (C) Transfection efficiency. The average number of cells detected by Hoechst 33258 stained nuclei as well as of viral protein positive cells were determined after counting the cells in three fields. (D) Comparison of relative RNA levels in cells transfected with a replicating replicon to those cells transfected with a nonreplicating replicon. BHK cells were transfected with 100 ng of C\textsuperscript{N}FI or C\textsuperscript{N}FI/GDD replicon RNA per well. The amount of viral genomic RNA detected at 24 hr was set at 1 and RNA levels at later time points were expressed relative to the 24 hr level. Viral RNA levels were normalized to the levels of cell GAPDH, and all assays were done in triplicate. Error bars represent the standard error of the mean (n=3) and are based on an RQMin/Max of 95% confidence level.
RNA decreased steadily over time. In contrast, the RNA levels produced by the C^NI replicon were twice as high at 72 hr as those observed at 24 hr after transfection (Fig. 1.10D). By 72 hr, very little residual input RNA was still present as indicated by the C^NI/GDD RNA levels. Replicons used in previous studies that contained an IRES/reporter cassette in the 3' NCR, exhibited similar levels of RNA replication, with RNA levels increasing two fold from 24 hr after transfection to 72 hr after transfection (Shi, Tilgner, and Lo, 2002). These results indicate that C^NI RNA replicates as efficiently as replicons that have an IRES inserted at the 3' end of the genome.

**Analysis of IRES-dependent translation by Western blotting**

As an alternative method of analyzing the translation of proteins from replicon RNA, Western blotting of cell lysates was used. BHK cells were transfected with C^NI replicon RNA and harvested, followed by lysis 5, 24, 48 or 72 hr after transfection. Proteins were separated by 10% SDS-PAGE, transferred to nitrocellulose membranes and immunoblotted with mouse hyperimmune anti-WNV sera. No bands were detected after exposure of the Western blots for 10 min, but after prolonged exposure (30 min), WNV nonstructural protein bands were visible. Lysates from cells transfected with one of two different WNV infectious clone RNAs and harvested at 24 hr after transfection, served as positive controls in these Western analyses. Strong bands were detected after just 5 sec of exposure with extracts from cells transfected with RNA from either infectious clone (Fig. 1.11, left panel, lanes 2 and 3). In contrast, the same exposure time did not reveal any bands for Western blots done with lysates from cells transfected with C^NI RNA even
Figure 1.11. Comparison of Western blots of viral proteins present in extracts of cells transfected with 1μg of WNV infectious clone RNA or CNI replicon RNA. Lane 1: lysates of mock-transfected BHK cells; Lanes 2 and 3: lysates of cells transfected with infectious clone RNA derived from either WNV strain EG101 or strain Wengler, respectively, and harvested at 24 hr after transfection; Lanes 4-7: lysates from cells transfected with replicon CNI RNA, at 5, 24, 48, and 72 hr after transfection, respectively. The three panels represent the same blot exposed for different lengths of time. The time of exposure is indicated below each panel.
though the number of replicon RNA molecules used was 1.2 times higher than that of the infectious clone RNA used for transfection (Fig. 1.11, left panel, lanes 4-7). After an exposure time of 15 min, weak replicon protein bands were detected (middle panel, lanes 4-7), and after a 30 min exposure, replicon protein bands of moderate intensity were observed (right panel, lanes 4-7). Also, the level of translation of $C^{NI}$ replicon RNA increased at 72 hr, consistent with the amplification of RNA due to RNA replication (see Fig. 1.7A). In contrast, efficient amplification of infectious clone RNA by 24 hr resulted in the production of high levels of viral protein.

**Extension of the capsid coding sequence**

One explanation for the low replication levels of $C^{NI}$, $C^{NLI}$, $C^{NFI}$ and $C^{NLFI}$ replicons could be the absence of regulatory elements in the missing capsid coding sequence. The effect of extending the capsid coding sequence in replicon $C^{NFI}$ was next investigated. This replicon was designated CFI. The addition of *Renilla* luciferase in frame with the extended capsid sequence yielded replicon CLFI. The extended capsid sequence did not include the C-terminal hydrophobic (anchor) domain. The anchor sequence serves as a signal sequence for ER translocation of the adjacent prM in the polyprotein, and its inclusion would introduce a transmembrane region at the junction between the two proteins in the fusion protein. The overexpression of a capsid protein containing this region was previously shown to induce apoptosis in infected cells (Yang et al., 2002). Analysis of RNA levels produced by CFI and CLFI replicons showed that
they replicated less efficiently than the $C^N$FI and $C^N$LFI replicons (compare Fig. 1.12 to Fig. 1.9).

Detection of intracellular viral nonstructural proteins by immunofluorescence at 5 and 24 hr after CFI RNA transfection showed that the protein levels were significantly higher (Fig. 1.13B) than those in cells transfected with replicons expressing a truncated capsid ($C^N$I) (Fig. 1.13A). Viral protein levels were also assayed by Western blotting. Western blots done with extracts from cells transfected with the $C^N$I replicon required an exposure time of 45 min to visualize viral nonstructural proteins (Fig. 1.13, panel C), while strong viral protein bands were detected in extracts from CFI replicon-transfected cells after a 10 min exposure (Panel D). These results suggest the possible presence of a translational enhancer in the 3' portion of the capsid coding region.

The extended capsid coding region has an enhancing effect on both IRES- and cap-dependent translation

In a WNV whole genome RNA fold (Sgro and Palmenberg, unpublished data), four single stranded regions were predicted within the RNA structures in the extended capsid sequence region (nt 195 to 315) (Fig. 1.14A). One or more of these predicted single stranded regions might bind to cellular proteins involved in translation initiation. To further investigate possible translational enhancer elements in this region, replicons with sequential truncations in the extended capsid coding region were generated (Fig. 1.14B and C). Reporter replicons encoding the first 249, 279 or 297 nt of the capsid coding region were designated C249LFI, C279LFI, and C297LFI, respectively. BHK
Figure 1.12. Analysis of (A) CFI and (B) CLFI replicon RNA replication. BHK cells were transfected with 100 ng of CFI or CLFI replicon RNA per well. The amount of viral genomic RNA detected at 24 hr was set at 1 and RNA levels at later time points were expressed relative to the 24 hr level. Viral RNA levels were normalized to the levels of cell GAPDH, and all assays were done in triplicate. Error bars represent the standard error of the mean (n=3) and are based on an RQMin/Max of 95% confidence level.
Figure 1.13. Comparison of WNV nonstructural protein levels produced by the CNI and CFI replicon RNAs. BHK cells were transfected with 1 µg of (A) CNI or (B) CFI replicon RNA. Viral proteins were detected by immunofluorescence at 5 or 24 hr after transfection. Cells were analyzed by confocal microscopy and all images were obtained using the same experimental conditions and microscope settings. Viral proteins were also detected by Western blotting in lysates of cells transfected with 1 µg of (C) CNI or (D) CFI replicon RNA at 5, 24, 48 or 72 hr after transfection. The time of exposure of the Western blots was either 10 or 45 min.
Figure 1.14. Schematics of WNV capsid RNA and replicons with truncated capsid coding regions. (A) Linear representation of the WNV capsid RNA, with predicted single stranded regions depicted in a whole WNV genome fold in pink, blue, green and yellow. (B) Sequential truncation of the capsid coding region to exclude the yellow, green, and blue single stranded regions, resulted in truncated capsid coding regions of 297, 279, or 249 nt, respectively. (C) Schematics of replicons containing capsid coding regions of various lengths.
cells were transfected with C^{NLFI} (encoding the first 198 nt of the capsid coding region), C249LFI, C279LFI, C297LFI, or CLFI (encoding the first 315 nt of the capsid coding region) RNA, and cells were harvested and lysed at 5, 24, 48 or 72 hr after transfection. Cell lysates were analyzed by Western blotting for both IRES- and cap-dependent translation (Fig. 1.15). The viral nonstructural proteins NS1, NS3, and an NS1 dimer (NS1_d) were detected, but no NS5 bands were detected. This may be partly due to the fact that it is the 3’ most protein in the polyprotein. However, it is also the least antigenic of the nonstructural proteins. Actin was used as a protein control. Capsid was detected using an antibody to the N-terminal region of the WNV capsid protein, which was present in all constructs.

However, due to the small size of the capsid fragments in constructs lacking luciferase, they ran off the gels under the conditions used. Capsid/Renilla luciferase fusion proteins were retained on the gels. C^{NLFI}, which contained the shortest capsid coding sequence, showed barely detectible protein bands after a 10 min exposure. Inclusion of increasing lengths of capsid sequence in the replicons resulted in an increased efficiency of both IRES- and cap-dependent translation (Fig. 1.15). Moreover, with all of the replicons, cap-dependent translation was less efficient and delayed as compared to IRES-dependent translation. This suggested that the IRES competes more efficiently for ribosomes.
Figure 1.15. Effect of capsid length on cap- and IRES-dependent translation by replicons. (A) Analysis of IRES-dependent translation of replicons C^NI, C249LFI, C279LFI, C297LFI, and CLFI by Western blotting using hyperimmune serum. The WNV proteins detected were NS1, NS3, and a dimer of NS1 (NS1\(_d\)). Actin was used as a protein control. (B) Analysis of cap-dependent translation of C^NI, C249LFI, C279LFI, C297LFI, and CLFI using an anti-WNV N-terminal capsid antibody. Exposure times for both (A) and (B) were 10 min.
Analysis of WNV RNA-cell protein binding activity

Thermodynamic folding predictions of the WNV capsid RNA sequence were done using the Mfold program. These analyses predicted a stable RNA structure that consisted of two stem loops (Fig. 1.16B). One large loop corresponded to the first predicted single stranded region in the extended capsid RNA (Fig. 1.16A) (pink), while the second, smaller loop corresponded to the second predicted single stranded region (blue). This predicted structure was used to design an RNA probe, that was used in gel mobility shift assays to analyze possible interactions between the RNA probe and cell proteins in BHK S100 cytoplasmic extracts.

The RNA probe was labeled with $\alpha^{32}$P-ATP, and incubated with increasing amounts of BHK cytoplasmic extracts for 30 min at room temperature. The RNA-protein complexes formed were resolved on a 10% nondenaturing polyacrylamide gel, and detected by phosphorimaging. Three RNA-protein complexes (RPCs) were detected in the cytoplasmic extracts (Fig. 1.17). The intensity of the RPC bands increased with increasing amounts of cell extract. The detection of two RNA probe bands suggests that the probe forms a dimer. These results suggest that this RNA region can interact with cellular proteins.

Detection of cap-dependent translation by analyzing luciferase activity

It was of interest to determine whether cap-dependent translation could be detected by analyzing luciferase activity. BHK cells were separately transfected with 1 µg of each of the replicons that encoded Renilla luciferase, and cells were harvested and
Figure 1.16. Thermodynamic folding prediction of a region of the WNV capsid RNA. (A) Linear representation of the whole WNV capsid RNA. Pink, blue, green and yellow regions represent four putative single stranded regions as predicted in the WNV whole genome fold. The top line indicates the unfolded RNA probe depicted in (B), comprising nt 194 to 279 in the capsid coding region. (B) Predicted secondary structure of a region comprising the first two single stranded regions. This structure is thermodynamically very stable, with the first predicted single stranded region forming the large loop (pink), and the second predicted single stranded region forming the small loop (blue). This structure also represents the WNV capsid RNA probe that was used in cell protein binding assays.
Figure 1.17. Gel mobility shift assay done with a $[^\alpha-^{32}\text{P}]$-radiolabeled capsid RNA (2000 cpm) and BHK cytoplasmic S100 extracts. The numbers below the image represent the amount of total cell protein added to the reaction in micrograms.
lysed at 5, 24, 48, or 72 hr after transfection. No activity was detected in cell lysates at any of the time points from cells transfected with C\textsuperscript{NLI}, C\textsuperscript{NLFI}, or CLFI RNA. Only the 24 hr time point data are shown (Fig. 1.18A). Possible explanations for the lack of detectable activity are a misfolding of the reporter protein when fused with truncated capsid, or alternatively, low levels of protein. Renilla luciferase was then replaced with Firefly luciferase in C\textsuperscript{NLFI} to yield C\textsuperscript{NFLFI} (Fig. 1.18B). Firefly luciferase is about 700 nt longer than Renilla luciferase, and it was thought that its longer sequence might produce a capsid-luciferase fusion protein that retained luciferase activity. However, no detectable luciferase activity was observed in C\textsuperscript{NFLFI} RNA-transfected cell lysates (Fig. 1.18A, lane 5). An Oas1b-RLuc construct, which encodes Renilla luciferase downstream of the promoter for the cellular gene Oas1b was used as a positive control in this set of experiments (Fig. 1.18B). A very high level of luciferase activity was detected in cells transfected with this construct (Fig. 1.18A, lane 1).

**Effect of mutation of the major eEF1A binding site in a bicistronic WNV replicon**

The major eEF1A binding site was previously mapped on the WNV 3' (+) SL using RNase footprinting and nitrocellulose filter binding assays (Blackwell and Brinton, 1997). This site consists of the 4 nt sequence 3' ACAC 5', located at nt 62 to 65 from the 3' end of the viral genome (Fig. 1.3). Subsequent studies showed that mutations in this site in a WNV infectious clone, which preserved the SL secondary structure (3' GUGU 5'), had a negative effect on viral (-) strand synthesis and reduced the efficiency of virus replication (Davis et al., 2007). Partial revertants were observed after passage and
Figure 1.18. Analysis of **Renilla** and **Firefly** luciferase activity in extracts from cells transfected with reporter constructs or WNV replicons. (A) BHK cells were transfected with 1 μg of an Oas1b-RLuc reporter construct RNA or of C^{NL}I, C^{NL}FI, CLFI, or C^{NF}FLFI WNV replicon RNA per well. Cells were harvested 24 hr after transfection, lysed and assayed for **Renilla** or **Firefly** luciferase activity. (B) Schematics of replicons assayed for the luciferase activity assay. **Renilla** luciferase reporters are depicted in blue and the **Firefly** luciferase reporter is depicted in purple.
sequencing, and revealed that the mutated U at positions 63 or 65 reverted to the parental C after three viral passages in BHK cells, resulting in the sequence 3' GUGC 5' or 3' GCGU 5'. The partial revertant sequence 3' GUGC 5' was subsequently introduced into the infectious clone. Separate plates of BHK cells were transfected with infectious clone RNA containing the original 3' GUGU 5' and the partial revertant 3' GUGC 5' mutations. Real time qRT-PCR analyses revealed that although the levels of GUGU and GUGC genomic RNA were similar at 48 hr after transfection, there was an increase in GUGC genomic RNA as compared to GUGU genomic RNA by 72 hr after transfection (Davis et al., 2007).

It was of interest to determine the effect of a mutation in the major binding site of eEF1A on the replication of a bicistronic replicon, and to determine whether reversion would occur during replicon replication. The 3' GUGU 5' mutation was introduced into the major eEF1A binding site of replicon CNI to produce CNI/GUGU. CNI/GUGU RNA was transfected into BHK cells, and cells were harvested and lysed at 5, 24, 48, 72, 96, and 120 hr after transfection. RNA was extracted and purified, and subjected to real time qRT-PCR. CNI/GUGU RNA only increased 1.5-fold by 96 hr as compared to CNI RNA, which increased 4-fold by 96 hr (Fig. 1.19). The decreased RNA replication observed is in agreement with the data obtained with the WNV infectious clone carrying the same mutation (Davis et al., 2007). To analyze whether revertants arose, CNI/GUGU RNA was extracted at 5, 96, and 120 hr after transfection, reverse transcribed, and cloned into TOPO-XL (Invitrogen). Ten clones from each time point were sequenced. Sequence analysis revealed that while all of the clones derived from the 5 hr time point retained the
Figure 1.19. Analysis of $C^\text{NI}$ and $C^\text{NI}/\text{GUGU}$ RNA replication by real time qRT-PCR. BHK cells were transfected with 100 ng of replicon RNA per well. The amount of viral genome RNA detected at 24 hr was set at 1 and RNA levels at later time points were expressed relative to the 24 hr level. Viral RNA levels were normalized to the levels of cell GAPDH, and all assays were done in triplicate. Error bars represent the standard error of the mean (n=3) and are based on an RQMin/Max of 95% confidence level.

Figure 1.20. Sequence analysis of replicon RNA extracted from cells transfected with the $C^\text{NI}/\text{GUGU}$ replicon. (A) Sequence of the region spanning the major eEF1A binding site in the plasmid DNA of replicon $C^\text{NI}/\text{GUGU}$. This mutant replicon RNA was transfected into BHK cells. The same sequence was detected in cDNA reverse transcribed from RNA extracted at 5 hr after transfection. (B) Sequence of cDNA reverse transcribed from viral RNA extracted at 96 and 120 hr after transfection. (C) Sequence of the wild type eEF1A binding site.
GUGU mutation (Fig. 1.20A), 5 out of 10 clones at 96 hr, and 6 out of 10 clones at 120 hr after transfection had a reverted C at position 65, resulting in the sequence 3’ GUGC 5’ (Fig. 1.20B). These data indicated that replicon RNAs could rapidly revert even though they replicate less efficiently than infectious clone RNA.
DISCUSSION

In the majority of previous flavivirus replicon studies, RNA replication was analyzed indirectly by IFA, by detecting the translation of viral proteins in replicon transfected cells. Most of these replicon studies analyzed the translation of viral proteins by IFA at only a single time point, either at 24 (Jones, Patkar, and Kuhn, 2005), 48 (Corver et al., 2003), or 72 hr after transfection (Gehrke et al., 2003; Kofler et al., 2006). This approach was based on the assumption that the RNA replication kinetics of replicons are similar to those observed with viral infectious clones, and that the viral proteins detected at 24, 48 and 72 hr after transfection were translated from replicated replicon RNA. However, in only one previous study, were the relative amounts of viral proteins detected by IFA compared at two different times after transfection (Khromykh and Westaway, 1997). The data described in the present study clearly indicate that the replication of a bicistronic replicon RNA can be significantly delayed, with replicon RNA replication often not reaching its peak until 96 hr after transfection. Also, viral proteins could be detected by IFA throughout this period in cells transfected with replicons with different RNA replication efficiencies (Fig. 1.7A, 1.9B, 1.13A, B, and data not shown), suggesting that the protein detected was translated primarily from the transfected input replicon RNA.

While IFA is a useful assay for the detection of the presence or location of specific proteins, it is not a quantitative assay. The present study differs from previous studies of flavivirus replicons in the use of Western blotting as the primary method of measuring viral RNA translation levels. Products from both cistrons in a flavivirus
bicistronic replicon were assayed. Initial IFA studies of the replicons with an extended capsid coding sequence suggested enhanced IRES-dependent translation. Subsequent Western blotting assays demonstrated that the efficiency of both cap-dependent and IRES-dependent translation increased as the length of the capsid sequence was increased. A previous study done with Kunjin virus monocistronic replicons focused on analyzing the effect of capsid sequences on RNA replication. This study showed that the first 60 nt of the capsid coding region, which included the 5′ Cyc sequence, were needed for efficient RNA replication. A replicon that included the capsid coding sequence up to the anchor (321nt) had an enhancing effect on RNA replication, but had no effect on translation. However, the translation assays in this study were done in vitro using rabbit reticulocyte lysates and the proteins produced were analyzed by Western blotting. It could be that the in vitro nature of these assays did not accurately reflect events that occur in vivo. Although these assays were subsequently repeated in vivo, viral protein production was only analyzed by IFA, and only at 24 hr after transfection (Khromykh and Westaway, 1997). Furthermore, the monocistronic structure of the replicons used in those studies differs from that of the bicistronic replicons used in the present study.

Although real time qRT-PCR has been used in previous flavivirus replicon studies to analyze RNA replication, there are differences between the method used in those studies and the methods used in the present study. Real time qRT-PCR allows simultaneous amplification of cDNAs transcribed from viral genomic RNA by sequence specific primers and quantification of the number of copies of the PCR product that are present. There are two types of chemistries that facilitate the detection of PCR products
by real time qRT-PCR. One type involves a dye called SYBR green. SYBR green intercalates into double stranded DNA and consequently produces a fluorescent signal. However, SYBR green binds to all double stranded DNA, including nonspecific PCR products. This can compromise accurate quantification of the intended target sequence. The other type of chemistry involves a fluorescent reporter probe (TaqMan fluorogenic 5' nuclease) that binds to a specific region in the amplified cDNA. RNA can be quantified by real time qRT-PCR by an absolute or a relative quantification method. For absolute quantification, the levels of target gene RNA are compared to a defined standard that has been accurately quantified. For relative quantification, the levels of target gene RNA in each sample are measured relative to that of a gene that is present in the cell in stable amounts. The housekeeping gene GAPDH is often used for this purpose. In the present study, qRT-PCR assays were done using the TaqMan fluorogenic 5' nuclease chemistry, and intracellular replicon RNA levels in transfected BHK cell lysates were quantified relative to cellular GAPDH mRNA and expressed relative to the level of replicon RNA present at 24 hr after transfection.

In one previous flavivirus replicon study, SYBR green chemistry was used to measure the RNA replication of a monocistronic YFV replicon. A 300 to 400 fold increase of wild type replicon RNA was reported when compared to a nonreplicating replicon, although the results were not shown (Jones, Patkar, and Kuhn, 2005). The high amount of replicon RNA detected could in part be attributed to the use of SYBR green dye in these assays, which often detects nonspecific cDNAs. Also, a monocistronic replicon was used, and monocistronic flavivirus replicons have been consistently
observed to replicate more efficiently than bicistronic replicons (Gehrke et al., 2003; Kofler et al., 2006; Orlinger et al., 2006; Shi, Tilgner, and Lo, 2002). In a second study, TaqMan fluorogenic 5' nuclease chemistry was used to measure the RNA replication efficiency of a monocistronic TBEV replicon. The levels of replicon RNA were determined by absolute quantification and were reported to increase 100 fold by 24 hr after transfection (Kofler et al., 2006). The monocistronic nature of this replicon likely contributed to its replication efficiency. In a third study, TaqMan fluorogenic 5' nuclease chemistry was used to measure the RNA replication of a bicistronic WNV replicon. In this replicon, translation of the viral nonstructural genes was cap-dependent, while a GFP reporter gene inserted in the variable region of the 3' NCR was translated from an IRES. The RNA levels produced by this replicon were determined by absolute quantification, using RNA extracted from a virus stock of a known titer as the reference for quantification. The replicon RNA levels detected in this assay increased 2 fold from 24 hr to 72 hr after transfection. RNA levels at earlier timepoints were not reported (Shi, Tilgner, and Lo, 2002). These levels of RNA replication are comparable or lower than the ones observed for the replicons analyzed in the present study (see Fig. 1.10D). Although the presence of an IRES in a flavivirus bicistronic replicon increases the translation efficiency of the cistron it controls (Fig. 1.15, compare panels A and B), it always seems to have an overall negative effect on the RNA replication of flavivirus replicons. The extent of this negative effect varies with its location in the replicon.

A third aspect that distinguishes this replicon study from previous ones is the use of a unique type of bicistronic replicon. Many of the bicistronic replicons used in
previous studies employed cap-dependent translation of the viral nonstructural genes, and IRES-dependent translation of a reporter gene located at the 5' end of the viral 3' NCR. Figure 1.21 lists several of these types of replicons. In contrast, the replicons used in the present study had cap-dependent translation of a capsid/luciferase fusion protein and IRES-dependent translation of the viral nonstructural genes. Although the construction of a YFV replicon with a structure similar to that of the replicons used in the present study was previously reported, its replication was not analyzed (Jones, Patkar, and Kuhn, 2005). Since IRESs have such an extensive secondary structure, their location in a bicistronic replicon in close proximity to either the 5' or 3' Cyc sequences might affect the ability of the Cyc sequences to form a long distance interaction, and the consequence of this would be a reduced efficiency of RNA replication. The addition of a luciferase reporter to replicon CNI increased the distance between the 5' Cyc sequence and the IRES by 936 nt, and was therefore expected to enhance the RNA replication efficiency of this replicon. However, the opposite effect was observed (Fig. 1.7, compare panels A and B). The addition of a 5' IRES flanking sequence alone or together with an Rluc reporter increased the distance between the 5' Cyc sequence and the IRES by 103 or 1039 nt, respectively. However, no appreciable increase in replicon RNA replication levels was observed with either of these replicons (Fig. 1.9). The addition of the luciferase gene to CNI decreased RNA replication but this effect was not observed in replicons with an IRES 5' flanking sequence or the E signal sequence. These results suggest that possible effects on RNA folding of inserted sequence rather than the distance between the IRES
Figure 1.21. Schematics of standard bicistronic replicons utilized in previous studies. These replicons facilitate cap-dependent translation of the nonstructural proteins and IRES-dependent translation of a reporter gene that is located in the viral 3' NCR. (A) Schematics of bicistronic replicons used by (Shi, Tilgner, and Lo, 2002). The prM and E genes are deleted, and a partial capsid coding region containing the 5' Cyc sequence was fused to the nonstructural region. Reporter genes inserted into the viral 3' NCR are translated from an IRES. (B) Schematics of bicistronic replicons used by (Jones, Patkar, and Kuhn, 2005). A partial capsid coding region containing the 5' Cyc sequence was followed by all the nonstructural genes. The black bar indicates the presence of the E signal sequence. Reporter genes inserted into the viral 3' NCR are translated from an IRES.
and the 5' Cyc sequence may be more important. However, the translation efficiency of
the nonstructural region was also not enhanced by increasing the distance between the
IRES and the 5' cap. The addition of neither the IRES 5' flanking sequence nor the E
signal sequence increased nonstructural protein synthesis. These results suggest that the
IRES is inefficiently initiated when it is located near the 5' end of a flavivirus RNA. The
addition of an extended capsid coding sequence alone or together with an Rluc reporter
increased the distance between the 5' Cyc sequence and the IRES by 120 or 1056 nt,
respectively. The marked decrease in replicon RNA replication of both of these replicons
was due to the enhancement of translation from both cistrons by the extended capsid
sequence.

The identity of the cell protein(s) that interact with the translational enhancer
element defined remains to be determined. Gel mobility shift assays done with
cytoplasmic extracts showed the formation of several RPCs, suggesting that one or more
cell proteins may interact with this RNA region (Fig. 1.17). The cell protein(s) binding
to this region may be a translation initiation complex component. Trans-acting factors
binding to this RNA element could facilitate the assembly of the translation initiation
complex at both the cap and the IRES. It is also possible that a viral protein(s) could be
involved in facilitating this translational enhancement.

Another question that remains unanswered, is the reason for the lack of detectable
luciferase activity in cell extracts from replicon-transfected cells. Several previous
studies utilized flavivirus reporter replicons that encoded Renilla luciferase and high
levels of luciferase activity were detected. In some of these replicons, the luciferase
reporters were not translated from the viral 5' cap as a capsid/luciferase fusion protein, but were translated from an IRES that was located in the viral 3' NCR and were not fused with capsid (Jones, Patkar, and Kuhn, 2005; Scholle and Mason, 2005). Replicons that did encode capsid/luciferase fusion proteins were either monocistronic, with the NS3 cleavage site between E and NS1 included or the FMDV2A protease immediately following luciferase, or were bicistronic, with a second reporter gene located in the viral 3' NCR downstream of an IRES (Jones, Patkar, and Kuhn, 2005; Puig-Basagoiti et al., 2005; Rossi et al., 2007). The presence of an IRES immediately downstream of a luciferase reporter fusion protein in the replicons used in the present study, may have interfered with full elongation of luciferase. Premature termination might lead to improper folding of luciferase, thereby inhibiting its activity, but still allow detection of the capsid gene by Western blotting. However, since only a single band of about the right size was detected (see Fig. 1.15, panel B), termination would have to occur at a consistent position near the C terminus. Alternatively, the lack of detectable luciferase activity might be due to possible nuclear localization of the capsid/luciferase fusion proteins. Studies of Kunjin virus and JEV have shown that capsid protein localizes primarily to the nucleus in the absence of viral assembly (Mori et al., 2005; Westaway et al., 1997; Westaway, Mackenzie, and Khromykh, 2003). In the present study, luciferase activity was assayed in cytoplasmic extracts and was not detected. However, capsid/luciferase fusion proteins were easily detected by Western blotting in total cell lysates (see Fig. 1.15, panel B).
For all of the replicons used in the present study, cap-dependent translation levels were lower and delayed compared to IRES-dependent translation levels, indicating that the IRES outcompetes the cap for ribosomes (see Fig. 1.15). Increased IRES-dependent translation resulted in significantly lower replicon RNA levels, while replicons that replicated more efficiently showed reduced levels of IRES-dependent translation (compare Fig. 1.15A and 1.12B; and Fig. 1.15A and 1.7A). In a recent study that used TBEV bicistronic replicons with an IRES controlling the translation of prM and E proteins inserted in the 3' NCR, second site mutations were observed to arise in the IRES sequence. These mutations decreased the efficiency of IRES-dependent translation, but improved the efficiency of replicon RNA replication (Orlinger et al., 2007). The reduced efficiency of IRES-dependent translation in the original CNI replicon due to the lack of a 5' IRES flanking sequence may explain the replication efficiency of this replicon. Also, previous studies of picornaviruses demonstrated that initiation of RNA replication requires clearance of ribosomes from the viral genome, suggesting that ribosomes always "win" over RNA polymerases copying the same template (Barton, Morasco, and Flanegan, 1999). Picornaviruses and possibly also pestiviruses and hepatitis C virus may have mechanisms that enhance ribosome clearance from the 3' end that flaviviruses do not have.

The pestivirus and hepacivirus members of the family Flaviviridae have a 5' IRES. The structure of this IRES is similar to the Type II EMCV IRES. However, neither of these viruses have been reported to have 3' and 5' Cyc sequences. A long distance RNA-RNA interaction occurs between the highly conserved 98 nt 3' X domain
located in the 3' NCR and the 5BSL3.2 conserved SL structure in the NS5B coding region. This interaction is restricted to the 3' region of the genome. Inserting an IRES into either the 3' NCR or the 5' coding region of a flavivirus RNA was shown by this and other studies to decrease but not inhibit RNA replication.
CHAPTER II

Functional analysis of simian hemorrhagic fever virus (SHFV) polyprotein processing

INTRODUCTION

Classification

The order Nidovirales comprises the families Coronaviridae, Roniviridae, and Arteriviridae. Viruses belonging to these families all produce a 5', 3' co-terminal, nested set of subgenomic mRNAs during their replication cycles (Snijder, 2007). Members of the family Coronaviridae have been divided into three serogroups. Groups I, II and III include the prototype strains of human coronavirus-229E (HCoV-229E), murine hepatitis virus (MHV), and avian infectious bronchitis virus (IBV), respectively. The recently characterized severe acute respiratory syndrome coronavirus (SARS-CoV) was initially thought to represent a fourth serogroup, but was subsequently assigned to serogroup II as a result of phylogenetic analyses (Kim, Lee, and Lee, 2006; Snijder, 2007). Coronavirus are distinguished by the crown-like appearance of their virions that results from long spikes protruding from their surface, as well as by their unusually large genomes, often as large as 32 kb (Ziebuhr, Thiel, and Gorbunova, 2001). Recently, viruses that infect invertebrate hosts, such as prawns, were classified into the family Roniviridae, which is the newest member of the order Nidovirales (Pasternak, Spaan, and Snijder, 2006).
Members of the family *Arteriviridae* include equine arteritis virus, lactate dehydrogenase-elevating virus (LDV), porcine reproductive and respiratory syndrome virus (PRRSV), and simian hemorrhagic fever virus (SHFV). EAV is the best studied arterivirus, and it is a significant horse pathogen. It is transmitted by the respiratory route and causes flu-like symptoms, abortion in pregnant mares, and can often be fatal in adult animals (Snijder, 2007). LDV causes asymptomatic infections in mice, which result in lifelong viremia. Infection is characterized by abnormally high levels of lactic dehydrogenase in the plasma (Hull, 1989). PRRSV infects domestic pigs and causes fever, anorexia, lesions in the lungs, and abortion, and is the most recently characterized Arterivirus (Snijder, 2007).

SHFV was originally isolated in 1964 from macaque colonies in Soviet and American primate facilities (Tauraso et al., 1968). It has since been concluded that African green monkeys, African patas monkeys and baboons provide a natural reservoir for SHFV, since these species develop asymptomatic persistent infections. On the other hand, infections of macaques with this virus cause hemorrhagic symptoms, with death occurring within one to two weeks. Common symptoms include bloody diarrhea, fever, as well as hemorrhages in the intestine, lung, nasal mucosa, dermis, spleen, liver, and periocular connective tissues (London, 1977).

**Virion**

Arterivirus virions are spherical, about 40 to 60 nm in diameter, with a relatively smooth surface. The nucleocapsid is about 25 to 35 nm in diameter and is composed of
the viral (+) strand RNA genome and the nucleocapsid protein N (Snijder and Meulenberg, 1998). The nucleocapsid is surrounded by a lipid bilayer obtained from the infected cell. This lipid envelope has been shown to contain as many as six envelope proteins, with the number varying between different Arteriviruses (Snijder and Meulenberg, 1998; Snijder, 2007). Studies with EAV and PRRSV have shown that there are two structures formed by the viral envelope proteins. GP5 and the M protein organize into a disulfide-linked heterodimer, while the minor structural proteins GP2, GP3 and GP4 form a heterotrimer (Gorbalenya et al., 2006).

**Genome**

The SHFV genome is a (+) sense RNA, about 15.7 kb in length, that is flanked by a 5' NCR of 209 nt and a 3' NCR of 76 nt. The 5' end contains a type I cap structure and the 3' end is polyadenylated. The viral proteins are encoded by multiple ORFs, some of which overlap. The nonstructural polyproteins are encoded at the 5' end of the genome and this region comprises about three quarters of the genome length (Snijder, 2007). The structural genes are located at the 3' end of the genome, and are expressed from a nested set of 5'-3' co-terminal subgenomic mRNAs. SHFV has been shown to encode ten ORFs for structural proteins, 2a, 2b, 3, 4a, 4b, 5, 6, 7, 8 and 9 (Maines and Brinton, unpublished data) (Maines, ; Snijder, 2007). The other arteriviruses only encode seven 3' ORFs, and studies suggest that ORFs 2a, 2b and 3 of SHFV may be the result of evolutionary gene duplication of ORFs 4a, 4b, and 5 (Godeny et al., 1998). The nonstructural proteins are translated as two polyproteins (nsp1a and nsp1ab). ORF1a is translated from the
genomic RNA. ORF1b, which is longer, is also copied from the genome but is only expressed as part of the nsp1ab polyprotein when a -1 ribosomal frameshift occurs (Figure 2.1). This frameshift occurs just before the ORF1a stop codon, and results in a –1 reading frame for ORF1b relative to ORF1a (den Boon et al., 1995; Snijder and Meulenberg, 1998; Snijder, 2007). Studies with EAV have shown that ribosomal frameshifting occurs about 15-20% of the time (den Boon et al., 1991).

Cleavage of the replicase polyproteins gives rise to 12 nonstructural proteins with various functions. ORF1a encodes nsp1 to nsp8 (Gorbalenya et al., 2006; Maines) (Maines and Brinton, unpublished data). There are three types of protease domains that are located in different nonstructural protein regions. The papain-like cysteine protease (PCP), cysteine protease (CP), and 3C-like serine protease (3CL) domains are located in nsp1, nsp2, and nsp4, respectively (Ziebuhr, Snijder, and Gorbalenya, 2000). Hydrophobic domains have been predicted in nsp2, nsp3 and nsp5 of EAV and are thought to be involved in anchoring the replication complex to intracellular membranes (Snijder and Meulenberg, 1998; Snijder, Wassenaar, and Spaan, 1994). ORF1b encodes an RNA-dependent RNA-polymerase (RdRp) in its N-terminal portion, followed by a zinc finger domain, an RNA helicase domain, and a conserved domain of unknown function, which is specific to Nidoviruses (Fig. 2.3) (Godeny et al., 1993; Gorbalenya et al., 2006).

Nsp1 is the N-terminal cleavage product of the replicase polyprotein. Analyses of the EAV nsp1 have shown that although this protein is not required for genome
Figure 2.1. Genomic organization of SHFV. The genome length is indicated by a line. The SHFV ORFs are indicated by boxes. ORF1a and 1ab are translated directly from the genomic RNA to produce the individual nonstructural proteins. ORF1b is copied from the genome as a result of a -1 ribosomal frameshift (RFS). The 3' ORFs are expressed from subgenomic mRNAs. The nonstructural ORFs are in red and the ORFs translated from subgenomic mRNAs are in blue.
replication, it is vital for subgenomic RNA synthesis (Tijms et al., 2001). The specific function of nsp1 in subgenomic RNA synthesis is still unknown. It has been hypothesized that nsp1 contributes to the viral life cycle by its involvement in replicase polyprotein processing, RNA synthesis, and virion biogenesis (Tijms et al., 2007). In addition, nsp1 has been shown to localize to the nucleus early in infection, and to interact with p100, a cellular transcriptional co-activator, which is located mainly in the cytoplasm but efficiently shuttles to and from the nucleus (Tijms and Snijder, 2003; Tijms, van der Meer, and Snijder, 2002). Current studies are aimed at identifying additional cellular proteins which interact with nsp1, at understanding the significance of these interactions, and the reason that nsp1 goes to the nucleus (Tijms et al., 2007).

**Replication cycle**

The replication cycle of arteriviruses is presumed to take place in the cytoplasm of infected cells (Snijder, 2007). Arteriviruses primarily infect cells of the monocyte/macrophage lineage, most likely because of a specific preference for receptors present on the surface of these cells. PRRSV, the only arterivirus for which attachment and entry studies have been published, has been shown to enter cells by attaching to sialoadhesin, a member of the sialic acid binding immunoglobulin-like lectin family (Vanderheijden et al., 2003).

Following viral entry by receptor-mediated endocytosis, the replication cycle begins with the translation of the viral replicase gene (Kreutz and Ackermann, 1996; Snijder, 2007). Deletion analyses in the EAV infectious clone have suggested that
despite its lengthy 5' NCR (224 nt), initiation of translation most likely occurs by the conventional "ribosome-scanning" mechanism (van den Born et al., 2005). The amount of nonstructural polyprotein produced is low presumably due to the presence of a strong RNA structure in the region of the translation start site (van den Born et al., 2005). After post-translational processing of the replicase proteins by virus-encoded proteases (Ziebuhr, Snijder, and Gorbalenya, 2000), the replication-transcription complex components are assembled on perinuclear membranes, and synthesis of (-) strand RNAs is initiated at the 3' end of the genomic RNA (Sawicki, Sawicki, and Siddell, 2007). The resulting genome-length (-) strand RNA (also known as the anti-genome) subsequently serves as a template for genome replication (Snijder, 2007).

The generation of the 5'-3' co-terminal subgenomic mRNAs is hypothesized to occur via a complex mechanism commonly referred to as discontinuous transcription (Pasternak, Spaan, and Snijder, 2006; Sawicki, Sawicki, and Siddell, 2007; Snijder and Meulenberg, 1998). Although the precise mechanism for discontinuous transcription is still subject to some debate, there is consensus about the fact that arterivirus sg mRNAs have co-terminal 5' and 3' ends, which are derived from the respective ends of the viral genome (Gorbalenya et al., 2006; Sawicki, Sawicki, and Siddell, 2007). Early studies with coronaviruses showed that the 5' common 'leader' sequence of sg (+) strands was identical to the 5' end of the genome (Lai, Patton, and Stohlman, 1982a). Subsequently, short, conserved AU-rich motifs were detected both at the 3' end of the common 'leader' sequence, and at the 5' ends of each transcription unit encoding a sg mRNA 'body', and were named 'leader' and 'body' transcription-regulating sequence (leader TRS and body...
TRS, respectively) (Baric, Stohlman, and Lai, 1983; Pasternak et al., 2001). Reverse
genetics studies with EAV showed that the leader TRS and body TRS base pair
(Pasternak et al., 2001). Base pairing between the (+) strand leader TRS and the (-)
strand body TRS facilitates the addition of the complement of the leader sequence to the
nascent (-) strand sg mRNA (Pasternak, Spaan, and Snijder, 2006). Recent RNA
structure probing analyses of the EAV 5' NCR indicated that the leader TRS is located in
the loop of a stable hairpin structure, which was named the leader TRS hairpin (LTH)
(Van Den Born, Gultyaev, and Snijder, 2004; van Marle et al., 1999). The LTH seems to
be a key structural feature for discontinuous sg RNA synthesis and is likely to be a
critical factor for leader TRS function (Van Den Born, Gultyaev, and Snijder, 2004; van
den Born et al., 2005). In addition, the location of the ORF1a start codon on the 3' side of
the LTH suggests that the structure of the LTH could decrease the efficiency of scanning
ribosomes reaching the translation start site (van den Born et al., 2005).

Two opposing models have been proposed for the generation of nidovirus sg
mRNAs. Although both models assume that the fusion of the 5' leader and 3' body
occurs co-transcriptionally, they are in disagreement about whether the discontinuous
step takes place during (+) or (-) strand synthesis. The initial model proposed that
discontinuous transcription takes place during (+) strand synthesis, and that the TRSs in
the anti-genome serve as promoters for continuation of transcription (Baric, Stohlman,
and Lai, 1983; Lai et al., 1984; Spaan et al., 1983). In this model, the initiation of
transcription from the 3' end of the anti-genome leads to the production of a leader primer
(leader TRS), which can subsequently base pair with one of several anti-body TRSs in
the anti-genome. Following base pairing, the transcript is extended to produce a complete sg mRNA (Pasternak, Spaan, and Snijder, 2006). This model was called the "leader-primed transcription" model, and was supported by the inability to detect (-) strand RNAs of sg length in coronavirus-infected cells (Fig. 2.2A) (Lai, Patton, and Stohlman, 1982b).

The "discontinuous extension of (-) strand RNA" model (Fig. 2.2B) was developed after subsequent detection of sg (-) strand RNA molecules in cells infected with TGEV, a coronavirus that infects pigs (Sawicki and Sawicki, 1995; Sethna, Hung, and Brian, 1989). According to this particular model, the discontinuous step occurs during (-) strand synthesis, and the resulting sg-length (-) strand RNA serves as a template for transcription. This model has subsequently been supported by data from studies of both corona- and arteriviruses (Pasternak et al., 2001; Zuniga et al., 2004).

**Proteases in Arteriviruses**

All arterivirus nonstructural polyproteins contain a cysteine protease (CP), a serine protease (SP) and papain-like cysteine proteases (PCP) in their 1a and 1ab polyproteins (van Dinten et al., 1996). Each genome has only one copy each of the CP and SP proteases, but the number of PCPs varies in the different arterivirus genomes (Ziebuhr, Snijder, and Gorbalenya, 2000).

The serine protease, also known as the 3C-like protease, is located in nsp4, and is considered the main protease because it processes the majority of the mature proteins, including those with the two most conserved replicase domains, the viral RdRp and RNA
Figure 2.2. Models for the generation of sg mRNAs in Nidoviruses. (A) Discontinuous (+) strand RNA synthesis according to the "leader-primed transcription model". (B) Discontinuous (-) strand RNA synthesis according to the "discontinuous extension of (-) strand RNA model". For the purposes of this illustration, the term "replication" refers to genome amplification from full length (-) strand RNA. The term "transcription" refers to the synthesis of sg RNAs. Figure modified from Pasternak, Spaan, and Snijder (2006).
helicase domains (Figure 2.3) (Snijder et al., 1996; Ziebuhr, Snijder, and Gorbale

The name "3C-like" was derived from the 3C proteases of the members of the virus family *Picornaviridae*, and indicates the position of this enzyme between two other domains in the 3B and 3D polyproteins of the genomes of these viruses (Rueckert and Wimmer, 1984). The tertiary fold of this protease resembles the bilobal β-barrel motif of the trypsin family of serine proteases, of which chymotrypsin is the prototype (Bazan and Fletterick, 1988). One feature that distinguishes the arteriviral main protease from that of the picornaviruses, is the replacement of the catalytic nucleophilic cysteine by serine in the catalytic triad His-Asp-Cys (Ziebuhr, Snijder, and Gorbale, 2000).

The cysteine protease resides in nsp2 and has similarities to both serine proteases and papain-like cysteine proteases. Its cleavage site is about 500 residues downstream of its catalytic domain and resembles that of serine proteases, while its Cys-His catalytic dyad resembles that of papain-like cysteine proteases (Ziebuhr, Snijder, and Gorbale, 2000). Studies with the EAV cysteine protease revealed that this protease cleaves at the nsp2/nsp3 junction (Snijder et al., 1995), but can also act as a serine protease cofactor for processing the nsp4/nsp5 junction (Wassenaar et al., 1997).

**Papain-like proteases**

One characteristic relating PCPs to papain is their catalytic Cys-His-Asn triad. The majority of the PCPs described in the literature, including those of coronaviruses, exhibit this triad (Guarne et al., 2000; Ratia et al., 2006; Sulea et al., 2006). Although
Figure 2.3. Overview of the proteolytic processing cascade of the EAV, LDV, and PRRSV replicase polyproteins. The boxes within the replicase polyproteins indicate the PCP, CP and 3C-like serine proteases, as well as the hydrophobic domains (HD), the viral RdRp, zinc finger (Z), and RNA helicase domain, and conserved domain of unknown function (C). A black line following the third HD in EAV indicates the position of the ribosomal frameshift site that separates ORF1a and ORF1b. The arrows and arrowheads indicate the cleavage sites of the respective proteases, with amino acid residues flanking the cleavage sites indicated below. The numbers above the individual domains indicate the nonstructural proteins produced (nsps). The inactive PCPα of EAV is marked with an asterisk. Figure modified from Ziebuhr, Snijder, and Gorbalenya (2000).
arteriviral PCPs only have a catalytic Cys-His dyad, they have been shown to function efficiently, which suggests that the third residue may be a supplemental residue that provides additional stability (Cygler and Mort, 1997; Herold, Siddell, and Gorbalenya, 1999; Ziebuhr, Snijder, and Gorbalenya, 2000).

A second characteristic relating PCPs to papain is their ability to fold into a variation of an $\alpha + \beta$ structure, which is conserved among this class of proteases (Ziebuhr, Snijder, and Gorbalenya, 2000). PCPs recognize and cleave one or two sites in the N-terminal half of coronavirus and arterivirus replicase polyproteins. Coronavirus PCPs are larger than arterivirus PCPs; the distance between the coronavirus PCP catalytic cysteine and histidine residues is twice as long as that in arterivirus PCPs. Arterivirus PCPs are the smallest PCPs so far identified. Coronavirus PCP motifs are located about 1300 residues downstream of the ORF1a N-terminus, and cleave upstream of their catalytic domain (Bonilla et al., 1995). In contrast, arterivirus PCP motifs are located 100-300 residues downstream of the ORF1a N-terminus and have been reported to cleave downstream of their catalytic domain (Snijder and Meulenberg, 1998; Ziebuhr, Snijder, and Gorbalenya, 2000).

The crystal structure of the SARS-CoV PCP has provided significant insight into its function (Figure 2.4) (Ratia et al., 2006). The monomeric form of this PCP forms a right-hand structure, consisting of separate palm, thumb and finger domains. The C-terminal region of the protein forms the finger domain, and is composed of a $\beta$-sheet that has four strands which are twisted and in an antiparallel configuration. The palm domain is composed of a six-stranded $\beta$-sheet, while the thumb domain is formed by four $\alpha$-
Figure 2.4. The structure of the SARS-CoV PCP. The right hand structure consists of separate palm, thumb and finger domains. The active site of the protease is situated in a cleft between the palm and thumb domains, and consists of a catalytic Cys-His-Asn triad in papain and coronavirus PCPs. Figure from Ratia et al. (2006).
helices (α4-7). The N-terminal region of the protein extends outwards from the thumb domain and adopts a fold that is similar to ubiquitin, a highly conserved cellular regulatory protein, which targets proteins for degradation. The active site of the protease is situated in a cleft between the palm and thumb domains, and consists of the catalytic Cys-His-Asn triad in papain and coronavirus PCPs, or the Cys-His dyad in arterivirus PCPs (Guarne et al., 1998; Ratia et al., 2006). The existence of a zinc binding domain which connects the left and right sides of this structure further likens these proteases to ubiquitin, as well as to herpesvirus-associated ubiquitin-specific protease (HAUSP), a cellular deubiquitinating enzyme (Ratia et al., 2006; Sulea et al., 2006). Moreover, recent studies have shown that the SARS-CoV PCP exhibits deubiquitinating activity (Barretto et al., 2005; Barretto et al., 2006).

ORF1a of EAV encodes an N-terminal PCP (PCPα) that is inactive due to a mutation in the cysteine residue of the catalytic site. An active PCP (PCPβ) is located in the region C-terminal to PCPα, and cleaves a Gly-Gly dipeptide, which is 31 aa downstream of the PCPβ active site His residue (Fig. 2.3) (Snijder, Wassenaar, and Spaan, 1992; Ziebuhr, Snijder, and Gorbalenya, 2000). There is no putative cleavage site for the inactive PCP, and a fusion protein of PCPα and β is produced. The possibility that PCPβ could back cleave at this site was not investigated. The ORF1a regions in the LDV and PRRSV genomes encode two active PCP domains, α and β, that cleave a Gln-Gln, and a Tyr-Gly dipeptide, respectively (Figure 2.3) (den Boon et al., 1995; Ziebuhr, Snijder, and Gorbalenya, 2000). SHFV is unique among the Nidoviruses in having three PCP motifs within the N-terminal region of ORF1a.
(Ziebuhr, Snijder, and Gorbalenya, 2000); designated $\alpha$, $\beta$, and $\gamma$. Alignments of the SHFV PCP motifs indicate that PCP$\beta$ and PCP$\gamma$ are similar in size, but that PCP$\alpha$ contains a large deletion (Figure 2.5).

Previous *in vitro* analyses suggested that cleavage of the SHFV PCP$\alpha$ cleavage site occurred efficiently (Maines and Brinton, unpublished data). However, the large deletion between the two catalytic sites of PCP$\alpha$ suggests that this PCP may not be functional (den Boon et al., 1995; Snijder and Meulenberg, 1998) (Maines and Brinton, unpublished data). The observation that cleavage occurs at one or both of the sites downstream of the SHFV PCP$\alpha$ suggests that back cleavage by either PCP$\beta$ or PCP$\gamma$ can occur. Since arterivirus PCPs have previously only been shown to cleave downstream of their catalytic domains, this upstream processing would represent an arterivirus protease activity that has not been previously reported.
Figure 2.5. **Alignment of the three SHFV PCPs.** Bold letters indicate the catalytic Cys and His residues. Shaded letters indicate the flanking amino acid residues of predicted cleavage sites. Dashed lines indicate missing residues and clearly show a large deletion between the catalytic residues of PCPα.
**Goals**

**Aim 1: Analysis of the activity of SHFV PCPα.** Wild type constructs consisting of the SHFV 5' NCR, the first 1722 nt of ORF1a, and the 3' NCR, and constructs with mutations in selected PCP catalytic cysteines will be used as templates for coupled *in vitro* transcription/translation. The cleavage patterns of polyproteins with mutations in the catalytic sites of individual or combinations of PCPs will be compared to investigate the activity of PCPα.

**Aim 2: Determination of which of the predicted cleavage sites are used.** Mutations will be introduced into each of the predicted cleavage sites one at a time. Mutant constructs will be transcribed and translated *in vitro* using a rabbit reticulocyte lysate, and cleavage patterns will be compared to determine whether the polyprotein is cleaved at those particular sites.

**Aim 3: Determination of which PCP cleaves which cleavage sites.** Combinations of individual or multiple cleavage sites and/or catalytic residues will be mutated to determine which sites are cleaved by which PCPs.
MATERIALS AND METHODS

Mutagenesis of an SHFV cDNA clone

An existing clone consisting of the first 1722 nt of the SHFV ORF1a region, and flanked by the SHFV 5' and 3' NCR was used as a starting template for mutagenesis. QuikChange Site-directed mutagenesis (Stratagene) was performed according to the manufacturer's instructions using primers which encoded mutations in the catalytic cysteines (Cys to Ala) or the predicted cleavage sites (Phe, Gly, Cys, Tyr, Thr, or Arg to Val).

Preparation of cDNA clones

To generate construct pFSm, the first 1731 nt of the SHFV ORF1a were amplified by PCR using a forward primer with a HindIII overhang and a reverse primer with an XbaI overhang. The resulting PCR product was purified using the QIA quick Gel Extraction Kit (Qiagen) and cloned into Topo-XL (Invitrogen). The plasmid DNA was then digested using HindIII and XbaI restriction enzymes, and the resulting SHFV fragment was cloned into expression vector pFLAG/myc (Sigma). To generate construct pTNT/FSm, the FSm fragment was amplified by PCR using a forward primer with a MluI overhang and a reverse primer with a SalI overhang. The resulting PCR product was purified using the QIA quick Gel Extraction Kit (Qiagen) and cloned into Topo-XL (Invitrogen). The plasmid DNA was then digested using MluI and SalI restriction enzymes, and the resulting fragment was cloned into expression vector pTNT (Promega). The 5' and 3' NCRs were inserted 5' and 3' of the FLAG/SHFVORF1a/myc sequence by
double PCR using appropriate primers. All correct clones were confirmed by sequencing.

**In vitro transcription**

pFSm plasmid DNA was amplified by PCR using a T7 tailed forward primer that annealed to the 5' end of the 5' NCR, and a reverse primer that annealed to the 3' end of the 3' NCR. pTNT/FSm plasmid DNA was amplified by PCR using a forward primer that annealed to the 5' end of the T7 promoter and a reverse primer that annealed to the 3' end of the T7 terminator sequence. Alternatively, pTNT/FSm plasmid DNA was also linearized at 37°C for 2 hr with BgIII enzyme (Roche) prior to use as a PCR template. PCR products or restriction digests were purified using the QIA quick Gel Extraction Kit (Qiagen) and the DNA was eluted in 10 µl RNase-free water. Eluted DNA was *in vitro* transcribed using T7 or SP6 polymerase (Ambion or Epicentre) for 3 hr at 37°C. The residual DNA was removed with DNase at 37°C for 15 min, and the RNA was purified using TRI Reagent (Molecular Research Center, Inc.). The purified RNA was stored at -80°C or used immediately.

**In vitro translation**

Purified mRNA was translated in wheat germ lysates at 25°C for 2 hr (Promega).

**In vitro transcription/translation**

Wild type and mutant DNA templates were amplified by PCR using a 5' primer which encoded a T7 promoter. Purified PCR products (Qiagen) served as templates for *in vitro* coupled transcription/translation reactions done using a rabbit reticulocyte, or wheat germ lysate (Promega). Some reactions were supplemented with microsomes from
canine pancreas (Roche). [\(^{35}\)S] cysteine was used to label resulting protein products, and reactions were incubated at 30°C for 90 minutes. Protein products were purified using CentriSpin columns (Princeton Separations) and separated by 12% SDS-PAGE. Gels were dried and visualized by autoradiography.

**Western blotting**

To visualize smaller, FLAG or myc-tagged processed protein products, proteins were separated by 15% SDS-PAGE, electrophoretically transferred to a nitrocellulose membrane, and blocked in milk for 1 hr. For detection of the FLAG epitope, the membrane was incubated with a 1:500 dilution of mouse or rabbit anti-FLAG primary antibody (Chemicon, Cell Signaling, Sigma, or Santa Cruz Biotechnology, Inc.) overnight at 4°C in the presence of blocking buffer. For detection of the myc epitope, the membrane was incubated in a 1:500 dilution of mouse anti-myc antibody (Santa Cruz Biotechnology, Inc.) overnight at 4°C in the presence of blocking buffer. The next day, the membrane was washed in TBS/Tween-20, incubated with a 1:2000 dilution of goat anti-mouse (Upstate) or goat anti-rabbit secondary antibody (Cell Signaling) for 1 hr, washed in TBS/Tween-20, and processed for enhanced chemiluminescence using a Super-Signal West Pico detection kit (Pierce, Rockford, IL).

**Estimation of protein molecular mass by relative electrophoretic mobility**

The relative electrophoretic mobilities (Petersen, Marfin, and Gubler) of standard proteins were determined and used to construct a calibration curve. The molecular masses of SHFV cleavage products were estimated using the calibration curve.
RESULTS

An SHFV construct consisting of the viral 5' NCR, the first 1722 nt of ORF1a, and the 3' NCR cloned into pCR 2.1 (Invitrogen), was generated by a previous student. The region of ORF1a in this construct contains the three viral PCP domains (α, β, and γ) and their predicted cleavage sites plus a short region 3' of the PCPγ cleavage site (Fig. 2.6). This construct was used to mutate the catalytic cysteines of each of the three SHFV PCPs. The catalytic cysteine residues of PCPα, β and γ (Cys115, Cys246 and Cys378, respectively) were first changed to alanine, yielding mα, mβ, mγ, and mβγ. To enable study of the activity of a single PCP, different combinations of two PCPs, or all three PCPs were next inactivated by mutations that changed the catalytic cysteine to an alanine, producing mαβ, mαγ, and mαβγ.

Prediction of PCP cleavage sites in the N-terminal region of SHFV ORF1a

The predicted cleavage patterns resulting from PCP cleavage of ORF1a of SHFV are illustrated in Figure 2.6. The unprocessed polyprotein is predicted to yield a band of ~63.1 kDa. Cleavage at one of the first two predicted cleavage sites is expected to yield peptides of 16.4/17.9 kDa and 46.7/45.2 kDa, depending on which of the two sites is used. Cleavage of the third and fourth predicted sites is predicted to yield peptides of 29.6 and 33.5 kDa. Cleavage at the third or fourth site, which are located between the catalytic residues of PCPβ, may inactivate PCPβ. Cleavage at the fifth predicted cleavage site is expected to yield peptides of 38.5 and 24.6 kDa, while cleavage at the
Figure 2.6. Locations of predicted PCP cleavage sites and predicted cleavage products resulting from SHFV PCP cleavage. The open box represents the N-terminal region of the SHFV ORF1a, and the black areas indicate the locations of the three $\alpha$, $\beta$, $\gamma$ protease motifs. The approximate locations of putative cleavage sites are indicated by dashed lines inside the box and numbers below the box. The numbers of cysteines and methionines for the various regions of the ORF1a fragment are indicated above the box. Solid lines below the box indicate predicted protein cleavage products. The estimated molecular masses of these products are shown above the lines, and the numbers of cysteines for each fragment are indicated below each line.
sixth predicted site is expected to yield a small peptide of 9.9 kDa and a large peptide of 53.1 kDa.

**Analysis of the activity of PCPα**

A large deletion between the two catalytic sites of PCPα suggests that this PCP may not be functional. To determine whether PCPα is an active PCP, each of the mutant cDNAs was *in vitro* transcribed and translated using rabbit reticulocyte lysates. The protein products produced were labeled by incorporation of $[^{35}S]$ cysteine, separated by 12% SDS-PAGE, and the resulting processed proteins were visualized by autoradiography. Reactions containing the wild type sequence (WT) and reactions containing no template (NT) served as positive and negative controls, respectively. In initial experiments, protein products were labeled by incorporation of $[^{35}S]$ methionine. This resulted in the inability to detect some of the cleavage products which did not contain methionines, as well as in the presence of background material that co-migrated in the region of 10-22 kDa, and made it impossible to visualize protein bands in this region. $[^{35}S]$ cysteine labeling allowed visualization of most cleavage products, since all predicted cleavage products contain at least one cysteine residue, whereas only two of them contain methionine residues. Finally, in an effort to further reduce background material, the transcription/translation reactions were purified using protein purification columns before separation on gels (Princeton Separations). The sizes of the cleavage products were estimated by relative migration distance compared to the migration of protein standards.
Mutations were introduced into the predicted catalytic cysteine of each PCP. These mutations consisted of changing the catalytic cysteine to an alanine. Mutants with the PCP\(\alpha\), \(\beta\), or \(\gamma\) cysteine substituted were designated \(m\alpha\), \(m\beta\), and \(m\gamma\), respectively. Mutants with two or three PCP cysteines substituted were designated \(m\alpha\beta\), \(m\alpha\gamma\), \(m\beta\gamma\), or \(m\alpha\beta\gamma\). The cleavage pattern observed with the \(m\alpha\) polyprotein was identical to that seen with the wild type polyprotein. Strong bands of \(~38.5\) kDa and \(~24.6\) kDa suggested that cleavage occurred at predicted site 5 (Fig. 2.7, lanes 2 and 3). The cleavage pattern observed with the \(m\beta\) polyprotein produced strong \(~53.2\), \(~38.5\) and \(~24.6\) kDa bands, which suggested that cleavage occurred at both sites 5 and 6 (Fig. 2.7, lane 4). The molecular weight of the top band was determined to be \(~53.2\) kDa, by calculating its relative electrophoretic mobility value (see Materials and Methods).

Cleavage of the \(m\gamma\) polyprotein yielded bands of \(~38.5\), \(~36.8\), and \(~24.6\) kDa, suggesting that cleavage occurred at predicted sites 1/2, 5 and 6 (Fig. 2.7, lane 5). Only the full length polyprotein fragment was detected with proteins that had mutations in the catalytic sites of both PCP\(\beta\) and PCP\(\gamma\) (\(m\beta\gamma\)) or all three PCPs (\(m\alpha\beta\gamma\)). These results confirmed that there were no active proteases present in either of these mutant polyproteins (Fig. 2.7, lanes 6 and 9). The cleavage pattern observed with a protein with the catalytic sites of both PCP\(\alpha\) and PCP\(\beta\) mutated (\(m\alpha\beta\)) was identical to that of the \(m\beta\) polyprotein (Fig. 2.7, lanes 4 and 7). Even though PCP\(\alpha\) was not mutated in the \(m\beta\) polyprotein, no protease activity attributable to this protease was observed. Also, proteins with mutations in both PCP\(\alpha\) and PCP\(\gamma\) (\(m\alpha\gamma\)) yielded a cleavage pattern identical to that observed with \(m\gamma\), (Fig. 2.7, lanes 5 and 8). Again, no protease activity
Figure 2.7. Cleavage products produced by polyproteins with mutations in the catalytic cysteines of individual or combinations of PCPs. Mutant cDNAs were \textit{in vitro} transcribed and translated by rabbit reticulocyte lysates in a coupled TNT system (Promega). Protein products were labeled by incorporation of \([^{35}\text{S}]\) cysteine and separated by 12\% SDS-PAGE. Positions of prestained molecular weight markers (Bio-Rad) are shown on the left. Arrows indicate the positions and the estimated sizes of bands observed. NT- no template; WT – wild type.
attributable to PCPα was observed. The faint bands detected in many of the lanes between the 36.8 and 24.6 bands may be due to premature translation termination or to inefficient cleavage at noncanonical cleavage sites. Collectively, these results indicate that PCPα is not an active protease.

Previous studies of coronaviral and arteriviral PCPs have reported that the autocatalytic function of these PCPs was enhanced in reticulocyte lysates that were supplemented with microsomal membranes (den Boon et al., 1995; Tibbles et al., 1996). Therefore, in an attempt to improve the efficiency of protein processing by the SHFV PCPs, microsomes from canine pancreas were added to the translation reactions. A comparison of the cleavage products generated in reactions done in the absence (Fig. 2.8, panel A), or presence (Fig. 2.8, panel B) of microsomes showed that they did not enhance the function of the SHFV PCPs. Consequently, all subsequent reactions were done in the absence of microsomes.

Identification of the cleavage sites used by individual SHFV PCPs

In order to determine which of the six predicted cleavage sites in the N-terminal region of SHFV ORF1a are cleaved, mutations were introduced into the individual predicted cleavage sites, consensus G-G/Y/V. The sequences of the five predicted PCP cleavage sites in SHFV ORF1a are YGG, TGG, FGY, CGV, FGG and RYG residues (Fig. 2.5), with bold letters indicating the two amino acids (+1 and +2) located immediately upstream of the predicted cleavage site. Previous studies showed that mutation of the +1 and +2 amino acids of PCP cleavage sites prevented their being
Figure 2.8. Comparison of PCP activity in the (A) absence, or (B) presence of microsomal membranes. Mutant cDNAs were *in vitro* transcribed and translated using rabbit reticulocyte lysates in a coupled system. Microsomes from canine pancreas were added to one set of reactions. Protein products were labeled by incorporation of $[^{35}\text{S}]$ cysteine and separated by 12% SDS-PAGE. Positions of prestained molecular weight markers (Bio-Rad) are shown on the left. Arrows indicate the approximate sizes of bands observed. NT – no template, WT – wild type.
cleaved (Bonilla, Hughes, and Weiss, 1997; van Dinten et al., 1999). Nucleotide substitutions that changed these two amino acids to valine were introduced. Initial attempts to change these amino acids to alanine failed. This was most likely due to the high GC content in the alanine codon as well as in the sequences flanking the cleavage sites, that made design of appropriate primers difficult. The mutant polyproteins were expressed in an *in vitro* coupled transcription/translation, and the cleavage patterns produced were analyzed after separation on gels. Polyproteins with either cleavage sites 1 or 2 mutated produced ~38.5 and ~24.6 kDa bands (Fig. 2.9, lanes 3 and 4), suggesting efficient cleavage at predicted site 5. A polyprotein with predicted cleavage site 3 mutated also yielded bands of ~38.5 and ~24.6 kDa (Fig. 2.9, lane 5), suggesting cleavage at site 5. A polyprotein with a mutation in predicted cleavage site 4 yielded bands of ~53.2, ~38.5 and ~24.6 kDa (Fig. 2.9, lane 6), suggesting that cleavage occurred at both sites 5 and 6. The observation that a mutation in site 4 yielded a different pattern from that in site 3 suggested the possibility that mutation of site 4 inactivated PCPβ. These two cleavage sites are only 6 residues apart, and are located in the middle of PCPβ. A polyprotein with cleavage site 5 mutated produced a band of ~53.2 kDa, suggesting cleavage at predicted site 6 (Fig. 2.9, lane 7). A polyprotein with predicted cleavage site 6 mutated produced bands of ~38.5 and ~24.6 kDa (Fig. 2.9, lane 8), suggesting cleavage at predicted site 5. A polyprotein with mutations in each of the 6 cleavage sites produced a full length ~63.1 kDa band (Fig. 2.9, lane 9). The efficiency of recovery of processed peptides from the protein purification columns, especially of the smallest peptides, as
Figure 2.9. Cleavage products produced by polyproteins with mutations in the predicted cleavage sites. Mutant cDNAs were *in vitro* transcribed and translated using rabbit reticulocyte lysates in a coupled system. Protein products were labeled by incorporation of $[^{35}\text{S}]$ cysteine and separated by 12% SDS-PAGE. Positions of prestained molecular weight markers (Bio-Rad) are shown on the left. Arrows indicate the positions and the approximate sizes of bands observed. NT- no template; WT – wild type.
well as the amount of total protein produced, varied between experiments. The results shown are representative of three or more experiments.

To further analyze the usage of the various cleavage sites, additional mutant polyproteins were generated that contained mutations in five out of the six predicted cleavage sites. Wild type and mutant constructs were transcribed and translated in vitro as previously described, and the resulting cleavage patterns were analyzed. This analysis revealed that when cleavage sites 1, 2, 3, or 4 were the only sites available for cleavage, only the unprocessed ~63.1 kDa polyprotein was detected (Fig. 2.10, lanes 3-6). These results suggest that initial cleavage can not occur at these sites in the polyprotein. When only site 5 was available about 50% of the full length polyprotein was cleaved at this site, yielding bands of ~38.5 and ~24.6 kDa (Fig. 2.10 lane 7). When only site 6 was available for cleavage, the polyprotein was efficiently cleaved at this site as indicated by the detection of a strong ~53.2 kDa band (Fig. 2.10, lane 8).

**Determination of which PCP cleaves which cleavage sites**

The data obtained show that PCPα is an inactive protease, but that the two cleavage sites (sites 1 and 2) immediately downstream of this protease are still cleaved. One of the other two proteases, PCPβ, and/or PCPγ must therefore be able to cleave at sites 1 and/or 2. To determine which of the two proteases performs this cleavage, polyproteins containing two different combinations of catalytic site and cleavage site mutations were generated. One mutant polyprotein contained mutations in the catalytic residues of PCPα and PCPβ as well as in the predicted cleavage sites 3 through 6 (αβ3-6),
Figure 2.10. Cleavage products generated from polyproteins with mutations in five out of six predicted cleavage sites. Mutant cDNAs were in vitro transcribed and translated using rabbit reticulocyte lysates. Protein products were labeled by incorporation of $[^{35}\text{S}]$ cysteine and separated by 12% SDS-PAGE. Positions of prestained molecular weight markers (Bio-Rad) are shown on the left. Arrows indicate the positions and the approximate sizes of bands observed. NT- no template; WT – wild type.
while the second mutant polyprotein contained mutations in the catalytic residues of PCPα and PCPγ as well as in the predicted cleavage sites 3 through 6 (αγ3-6) (Fig. 2.11A, upper and lower panels, respectively). Wild type and mutant constructs were transcribed and translated in vitro as previously described, and the resulting cleavage patterns were analyzed. Detection of a strong ~63.1 kDa band indicated that the majority of the mutant αβ3-6 polyprotein was not cleaved. However, the detection of a weak ~46.7 kDa band indicated that some cleavage did occur at cleavage sites 1 or 2 (Fig. 2.11B, lane 3). The majority of the mutant αγ3-6 polyprotein also was not cleaved and the ~46.7 band detected was fainter than that observed with the αβ3-6 polyprotein (Fig. 2.11B, lane 4). These results suggest that both the β and γ PCPs are able to cleave at the 1/2 site, but "back" cleavage by both of these PCPs is inefficient in the context of the intact polyprotein. Interestingly, cleavage of this site by the more distant PCPγ is somewhat more efficient than cleavage by the adjacent PCPβ.

Detection of terminal cleavage products by Western blotting of epitope-tagged polyproteins

Some of the small bands produced by the polyproteins were difficult to detect, due to background material that co-migrated in that region. Although the protein purification columns eliminated most of the background, they also often removed peptides that were smaller than 20 kDa. As an alternative method for detecting these smaller bands, polyproteins with N-terminal FLAG and C-terminal myc tags were generated. The pFSm construct was generated by inserting the 1731 nt of the SHFV 5'
Figure 2.11. Effect of combination catalytic and cleavage site mutations. (A) Schematic of mutant constructs indicating mutations in the catalytic residues of PCPα, PCPβ, as well as predicted cleavage sites 3-6 (upper panel), and in the catalytic residues of PCPα, PCPγ as well as the predicted cleavage sites 3-6 (lower panel). (B) Cleavage products generated from polyproteins αβ3-6 and αγ3-6. Mutant cDNAs were in vitro transcribed and translated using rabbit reticulocyte lysates in a coupled system. Protein products were labeled by incorporation of [35S] cysteine and separated by 12% SDS-PAGE. Positions of prestained molecular weight markers (Bio-Rad) are shown on the left. Arrows indicate the positions and the approximate sizes of bands observed. NT- no template; WT – wild type.
ORF1a into the multiple cloning site of pFLAG/myc (Sigma), using HindIII and XbaI restriction sites. This multiple cloning site was flanked by a FLAG sequence on the 5' side and by a myc sequence on the 3' side (Fig. 2.12A). The SHFV 5' and 3' NCRs were inserted 5' of the FLAG sequence and 3' of the myc sequence by double PCR using appropriate primers. The FLAG/SHFV/myc sequence was amplified using a T7-tailed primer that annealed to the 5' end of the 5' NCR, and a primer that annealed to the 3' end of the 3' NCR (see horizontal arrows in Fig. 2.12A). This PCR product served as a template in in vitro coupled transcription/translation reactions done in a wheat germ lysate. For detection of FLAG-tagged peptides, rabbit reticulocyte lysates could no longer be used, because antibodies to FLAG were made in rabbits. The resulting proteins were separated by 15% SDS-PAGE, and transferred to a nitrocellulose membrane. The membrane was blocked in milk, incubated with anti-FLAG and anti-myc primary antibodies, followed by incubation with secondary antibodies. The membrane was then processed for enhanced chemiluminescence using a Super-Signal West Pico detection kit (Pierce, Rockford, IL). No protein products were detected with either antibody using this method. In order to make sure that this construct was efficiently transcribed and translated, $[^{35}\text{S}]$ cysteine was added to the reaction and the products were visualized by autoradiography. Bands of ~38.5 and ~24.6 kDa characteristic of the SHFV ORF1a N-terminus wild type cleavage pattern were observed (Fig. 2.13). This suggested that the template FSm DNA was efficiently transcribed and translated.

In a further effort to detect the FLAG and myc tagged protein bands by Western blotting, the FSm construct was cloned into pTNT using restriction sites MluI and SalI, to
A.

Figure 2.12. Schematic of expression vectors pFLAG/myc and pTNT. (A) The expression vector pFLAG/myc was used to generate construct pFSm. The first 1731 nt of SHFV ORF1a were inserted into pFLAG/myc between a 5' FLAG and a 3' myc coding sequence, using restriction sites HindIII and XbaI. The SHFV 5' and 3' NCRs were inserted 5' of the FLAG sequence and 3' of the myc sequence by double PCR. Horizontal arrows indicate a forward T7 tailed primer that annealed to the 5' end of the 5' NCR, and a reverse primer that annealed to the 3' end of the 3' NCR. The PCR products generated using these primers served as DNA templates in in vitro coupled transcription/translation reactions done using a wheat germ lysate. (B) The expression vector pTNT was used to generate construct pTNT/FSm. The pFSm sequence was inserted into a multiple cloning site using restriction sites MluI and SalI. The inserted sequence was flanked by a T7 promoter, an SP6 promoter, and a 5' β-globin leader sequence on the 5' side, and a poly(A)$_{30}$ tail, and a T7 terminator sequence on the 3' side. Horizontal arrows indicate 5' and 3' primers that anneal to the 5' end of the T7 sequence and the 3' end of the T7 terminator sequence, respectively. These primers were used for generating PCR products that served as templates in in vitro transcription reactions. Alternatively, a BglII site located 5' of the T7 promoter could be used for linearizing the plasmid DNA prior to in vitro transcription.
Figure 2.13. Comparison of translation efficiency of pFSm with that of the wild type polyprotein. Wild type and FLAG/myc tagged (pFSm) SHFV cDNAs were in vitro transcribed and translated using rabbit reticulocyte lysates. Protein products were labeled by incorporation of $[^{35}S]$ cysteine and separated by 12% SDS-PAGE. Positions of prestained molecular weight markers (Bio-Rad) are shown on the left. Arrows indicate the positions and the approximate sizes of bands observed. NT- no template; WT – wild type.
produce pTNT/FSm. pTNT is an expression vector that contains a 5' β-globin leader sequence and a synthetic poly(A)$_{30}$ tail (Promega) (Fig. 2.12B). These elements have both been reported to enhance the expression of certain genes (Falcone and Andrews, 1991; Wakiyama, Futami, and Miura, 1997). Also, to maximize expression of the polyprotein, DNA templates were in vitro transcribed and translated in separate reactions. pTNT/FSm plasmid DNA could be used directly as a template for in vitro transcription with T7 polymerase, because pTNT encodes a T7 terminator sequence at the 3' end of the insert, or it could be linearized using the BglII restriction site (see Fig. 2.12B). Alternatively, plasmid DNA could also be amplified by PCR using a forward primer that annealed to the 5' end of the T7 promoter and a reverse primer that annealed to the 3' end of the T7 terminator sequence prior to in vitro transcription with either T7 or SP6 polymerase (indicated by horizontal arrows in Fig. 2.12B). DNA templates were purified using a PCR cleanup kit (Qiagen), in vitro transcribed (Ambion or Epicentre), and the resulting RNA was purified using TRI Reagent (Molecular Research Center, Inc.). The purified RNA was then used as a template in in vitro translation assays done in wheat germ lysates (Promega). Protein products were separated by 15% SDS-PAGE and electrophoretically transferred to a nitrocellulose membrane. The membrane was blocked in milk, incubated with anti-FLAG and anti-myc primary antibodies, followed by incubation with secondary antibodies. The membrane was then processed for enhanced chemiluminescence using a Super-Signal West Pico detection kit (Pierce, Rockford, IL). Most of the protein products detected were nonspecific, with no differences seen between lanes containing no template, and lanes containing the cDNA pFSm products (Fig
2.14A). As an alternative protocol, 4 µg of pTNT/FSm plasmid DNA was linearized at 37°C for 2 hr with BglII enzyme (Roche) in three separate reactions. The digested plasmid DNA in each of the reactions was purified using the QIA quick Gel Extraction Kit (Qiagen) and the DNA was eluted in 10 µl RNase-free water. DNA from each of the three eluates was *in vitro* transcribed separately using T7 polymerase (Ambion) for 3 hr at 37°C. The residual DNA was removed with DNase at 37°C for 15 min, and the RNA was purified using TRI Reagent (Molecular Research Center, Inc.). The purified RNA from the three reactions was pooled into one *in vitro* translation reaction done in a wheat germ lysate at 25°C for 2 hr (Promega). Proteins were separated by 15% SDS-PAGE, electrophoretically transferred to a nitrocellulose membrane, and blocked in milk for 1 hr. The membrane was incubated with a 1:500 dilution of mouse anti-FLAG primary antibody (Chemicon) overnight at 4°C in the presence of blocking buffer. The following day, the membrane was washed in TBS/Tween-20, incubated with a 1:2000 dilution of goat anti-mouse secondary antibody (Upstate) for 1 hr, washed in TBS/Tween-20, and processed for enhanced chemiluminescence using a Super-Signal West Pico detection kit (Pierce, Rockford, IL). The bands detected (Fig 2.14B) were the expected sizes for the FLAG-tagged N-terminal products generated by cleavage of a wild type polyprotein at sites 1/2 (~16 kDa, lower band) or 5 (~38.5 kDa, upper band). Two specific FLAG-tagged bands of ~38.5 and ~16 kDa were detected (Fig. 2.14B). However, no bands were detected with anti-myc antibodies.
Figure 2.14. Analysis of pTNT/FSm N-terminal cleavage products by Western blotting. (A) Representative blot of pTNT/FSm using anti-FLAG antibodies. No template (NT) and pFSm/FSm lanes exhibit the same pattern of nonspecific bands. (B) Optimized Western blot of pTNT/FSm using anti-FLAG antibodies. Arrows indicate the N-terminal FLAG-tagged ~38.5 and ~16.4 kDa cleavage products produced from a wild type polyprotein.
DISCUSSION

Cleavage of the wild type polyprotein produced strong bands of ~38.5 and ~24.6 kDa. According to the predicted processing scheme illustrated in Figure 2.6, these bands are characteristic of cleavage at site 5. The large deletion between the catalytic residues of SHFV PCPα suggested that this protease would be inactive. A polyprotein with a mutation in the catalytic cysteine of this PCP produced a cleavage pattern that was identical to the one observed with the wild type polyprotein. Furthermore, polyproteins with mutations in the catalytic cysteines of both PCPβ and PCPγ, or in the catalytic cysteines of all three PCPs yielded a single band characteristic of the full length 63.1 kDa polyprotein. Polyproteins with mutations in the catalytic cysteine of PCPβ, or in the catalytic cysteines of both PCPα and PCPβ produced identical cleavage patterns, with bands of ~53.2, ~38.5, and ~24.6 kDa. These data suggested that both sites 5 and 6 were cleaved, and that neither of these cleavages were attributable to PCPα. The fact that cleavage occurred at site 5, which is located downstream of an inactive PCPβ, shows that PCPγ can back cleave at this site. Polyproteins with mutations in the catalytic cysteine of PCPγ, or the catalytic cysteines of both PCPα and PCPγ also produced identical cleavage patterns, with bands of ~38.5, ~36.8, and ~24.6 kDa. These results suggested that cleavage occurred at sites 1/2, 5 and 6, and that none of these cleavages were attributable to PCPα. These data also suggested that all three of these cleavages were executed by PCPβ, and indicated that this PCP was capable of back cleavage at site 1/2 as well as forward cleavage at both its own site and that of PCPγ. Collectively, the cleavage
patterns observed in Figure 2.7 indicate that PCPα is not an active protease, and that both PCPβ, and PCPγ are capable of back cleaving at upstream sites.

Polyproteins with mutations in predicted cleavage sites 1 or 2 produced a wild type cleavage pattern, with bands of ~38.5 and ~24.6 kDa. This is not surprising, since site 5 is the most efficiently cleaved site, and since this site was still available for cleavage in these mutant polyproteins. The same ~38.5 and ~24.6 bands were observed when cleavage site 3 was mutated. However, a different cleavage pattern was observed when site 4 was mutated. A polyprotein with a mutation in this site produced bands of ~53.2, ~38.5, and ~24.6 kDa, suggesting that cleavage occurred at sites 5 and 6. The fact that polyproteins with mutations in site 3 or 4 produced different cleavage patterns was surprising, considering the fact that these two sites are only 6 residues apart and that they are located in the middle of an active protease. One explanation for this observation could be that mutation of site 4 inactivates PCPβ, but that mutation of site 3 does not. If PCPβ is inactive, then cleavage at both sites 5 and 6 would be done by PCPγ. Back cleavage at site 5 by PCPγ was observed when PCPβ was mutated (see Fig. 2.7, lanes 4 and 7). A polyprotein with a mutation in site 5 produced only one band of ~53.2 kDa, suggesting cleavage at site 6. A polyprotein with a mutation in site 6 produced the wild type cleavage pattern, suggesting cleavage occurred at site 5.

Polyproteins that could only be cleaved at sites 1, 2, 3, or 4 were not cleaved at all. This could indicate that PCPβ and PCPγ can only cleave at upstream sites once they have cleaved at downstream sites (sites 5 and 6). A polyprotein that could only be cleaved at site 5, was cleaved at this site, although with reduced efficiency. That
cleavage was observed at this site is not surprising, considering the fact that this site is cleaved efficiently in the wild type sequence. The reduced efficiency of cleavage observed at this site in a polyprotein that could only be cleaved at site 5 may be due to the reduced catalytic activity of PCPβ resulting from the mutation in site 4 (see Fig. 2.9, lane 6). A polyprotein that could only be cleaved at site 6 was cleaved very efficiently, probably by PCPγ.

When PCPβ was the only protease available for cleaving sites 1/2, it cleaved very inefficiently, if at all (see Fig. 2.11, lane 4). In contrast, when PCPγ was the only protease available to cleave at these two sites, it cleaved more efficiently (see Fig. 2.11, lane 3). One explanation for this result could be that mutating site 4 inactivated the catalytic activity of PCPβ. A second possibility could be that the numerous mutations present in polyproteins αβ3-6 and αγ3-6 caused the proteins to fold in such a way that facilitated better cleavage by PCPγ of the 1/2 sites.

A processing scheme is proposed based on the results obtained (Fig. 2.15). The data shown in Figure 2.7 suggest that site 5 is cleaved most efficiently and appears to be the first site cleaved in the PCP polyprotein, yielding peptides of 38.5 and 24.6 kDa (lanes 2 and 3). Cleavage at site 6 in the polyprotein is less efficient, since the 38.5 and 24.6 kDa peptides (site 5 cleavage) as well as a 53.2 kDa peptide (site 6 cleavage) were observed (lanes 4 and 7). Cleavage at site(s) 1/2 was also less efficient and appeared to follow cleavage at site 6 since a 36.8 kDa as well as a 38.5 and 24.6 kDa band rather than a 46.7 kDa band was observed (lanes 5 and 8). The possible presence of a 22.1 kDa band
Figure 2.15. Putative SHFV PCP processing cascade. The cleavage products are shown in the order of the efficiency of their cleavage based on the data obtained in this study. The open box represents the N-terminal region of the SHFV ORF1a, and the black areas indicate the locations of the three α, β, γ protease motifs. The approximate locations of putative cleavage sites are indicated by dashed lines inside the box and numbers below the box. The numbers of cysteines and methionines for the various regions of the ORF1a fragment are indicated above the box. Solid lines below the box indicate predicted protein cleavage products. The estimated molecular masses of these products are shown above the lines, and the numbers of cysteines for each fragment are indicated below each line.
could not be ruled out because it might co-migrate with the 24.6 kDa band. However, this would be generated by cleavage at sites 5 and 1/2. Finally, exclusive but inefficient cleavage at site(s) 1/2 occurs when those two sites are the only ones available for cleavage, resulting in a band of 46.7 kDa (Fig. 2.11).

The SHFV nsp1α, β, and γ proteins resulting from cleavage at sites 1/2, 5, and 6 are composed of 163, 187, and 128 amino acids, respectively. An analysis of the amino acid sequences of these three proteins by alignment using the ClustalW program and using the PredictProtein program to predict structures and domains indicated that the nsp1β and nsp1γ protein were more similar to each other than to the nsp1α protein. In addition to differences observed in sequence and length, nsp1α was observed to contain a transmembrane domain in the region containing the predicted catalytic Cys residue. Neither nsp1β or γ were predicted to have a transmembrane domain. The detection of a transmembrane domain in nsp1α suggests that PCPα may be membrane associated and might not localize to the nucleus as has been reported for the EAV nsp1 (Tijms, van der Meer, and Snijder, 2002). Although immunofluorescence analyses localized the nsp1 of EAV to the nucleus early in infection, no canonical nuclear localization sequence (NLS) was predicted in the EAV nsp1 sequence. However, the EAV nsp1 was shown to interact with the cell protein p100 and its nuclear localization was suggested to be facilitated by its interaction with p100, which efficiently shuttles to and from the nucleus (Tijms and Snijder, 2003; Tijms, van der Meer, and Snijder, 2002).

The N-terminal region of the EAV nsp1 protein has been predicted to form a C3H zinc finger motif. Mutational analyses in an EAV infectious clone showed that the
presence of this zinc finger domain was essential for sg mRNA synthesis, but not for PCP function (Tijms et al., 2001). The SHFV nsp1α domain harbors four cysteine residues in its N-terminal region that are predicted to form a C₄ α + β zinc finger motif. A C₆ zinc finger motif (Schjerling and Holmberg, 1996) is predicted to form in this region in the context of the uncleaved PCP polyprotein. No zinc finger motif was predicted in nsp1β or γ.

All three protease regions harbor one or more predicted sites for phosphorylation by protein kinase C and casein kinase II. In addition to sites for these two kinases, PCPβ also has a cAMP-dependent phosphorylation site, while PCPγ has a tyrosine phosphorylation site. Once phosphorylated, these proteins may function by binding to components of a cell signaling pathway. Both PCPα and PCPγ also have putative N-linked glycosylation sites. One site is located at the C-terminus of PCPα and sites are located at both the N-terminus and C-terminus of PCPγ. Both phosphorylation and glycosylation can change the conformation of a protein which could facilitate interactions with different cellular binding partners. Finally, each of the three nsp1 proteins harbors three or more myristylation sites. Myristylated proteins are localized to cellular membranes and are often targeted to signal transduction pathways.

In the context of the uncleaved polyprotein, multiple Cys residues in the region surrounding the catalytic domain of PCPβ are predicted to form disulfide bonds with very high probability (9 out of a possible score of 10). These disulfide bonds are not predicted to form in an nsp1α-nsp1β polyprotein that would result from cleavage at site 5. The observation that site 5 is cleaved very efficiently by PCPβ in the polyprotein suggests that
these disulfide bonds may assist in forming a functional protease fold. The loss of these predicted disulfide bonds in PCPβ once site 5 is cleaved, suggests that PCPβ may be destabilized in nsp1αβ and therefore inactive. If this is the case, subsequent cleavage at site(s) 1/2 within nsp1α/β would be effected \textit{in trans} by PCPγ. A delay in cleavage of the 1/2 site suggests the possibility that the nsp1αβ precursor might have different functions than the individual nsp1α and nsp1β proteins.

The inability to consistently recover the small protein fragments (<20 kDa) from the protein purification columns limited some of the interpretations that could be drawn from the data. Epitope tagging of the N- and C-termini of a polyprotein was attempted to facilitate the detection of the N-terminal ~16 and C-terminal 9.9 kDa protein bands (see Fig. 2.6). When sufficient protein levels were generated, FLAG-tagged protein bands of ~16 and ~38.5 kDa were detected. The presence of the ~16 kDa band confirms that cleavage can occur at site 1 or 2 in a wild type polyprotein.

The myc tag was never detected by Western blotting. One explanation for this could be that this tag was folded into the protein preventing antibody binding. Another explanation could be that these reactions produced insufficient amounts of protein for detection of myc by Western blotting due to lower antibody affinity or antigen antigenicity. Nevertheless, the data obtained with a number of mutant polyproteins was sufficient to confirm the inactivity of PCPα and to demonstrate that one or both of the sites located downstream of PCPα could be cleaved by either PCPβ or PCPγ.
REFERENCES


