STRUCTURAL AND FUNCTIONAL ANALYSIS OF CORONAVIRUS CYSTEINE PROTEASE NSP5

By

Christopher Colin Stobart

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Microbiology and Immunology

December 2013

Nashville, Tennessee

Approved:

H. Earl Ruley, Ph.D.

Christopher R. Aiken, Ph.D.

Andrew J. Link, Ph.D.

Mark R. Denison, M.D.

To my family, friends and wife.

Thank you.

ACKNOWLEDGEMENTS

Inspiration is perhaps one of the greatest motivators of success. These last five years have been some of the most fulfilling, turbulent, stressful and rewarding experiences that I have ever endured. It is for these reasons that I must acknowledge those whom have been there when I needed assistance and encouragement and offered inspiration and guidance when I was lost:

The Denison Lab

These past five years have afforded many ups and downs. However, the assistance, support and encouragement that I have received from my lab family have made this time both tolerable and enjoyable. All of you know what impact you have made on my life and I wish you all the best in the your future pursuits. I have many wonderful memories, which I will likely carry with me for the rest of my life.

- Mark, my graduate mentor*:* Thank you for giving me the opportunity to work with you these past few years and the support and training necessary to make me the scientist that I am today. You have been there to offer advice, support, and encouragement through the highs and lows and I have benefited immensely because of it. I am a much better writer, thinker, presenter, teacher and scientist because of the feedback and direction that I have received from you. Thank you.
- Michelle: I personally want to thank you, not just for ordering the countless thousands of dollars worth of reagents that were ultimately lost to ineptitude and

scientific failure or making sure I was being a good lab member, but also for being there to listen when I needed advice or simply had to vent.

- Xiaotao: Quite possibly, no single person in the lab has had more of a direct influence on the success of graduate students in our lab than you and I am no exception. Your kindness, scientific knowledge and outstanding abilities at the bench have offered me both direction and support when my experiments weren't working or I was unable to recover an unwilling virus. There is a reason that you are on so many of our papers and I wish the best for you and your family in the future.
- Lance: Although you departed the lab mid-way through my training, I want to thank you for offering your assistance and advice on both experiments and approaches as well as being a good friend in those late afternoons when everyone else left and made us feel like we lacked a social life.
- Mark Gadlage and Sunny (Alice): When I first joined the lab, the two of you were both welcoming and supportive as I was learning the various tools and techniques that I would need. Thank you both for being who you are and helping me get adjusted to being a graduate student in the lab.
- Dia: It was hard for me to decide where I should acknowledge you. You have been a close friend, lab partner and roommate. Thank you for helping me bridge the gap between my lab life and my social life. You have been a great friend to my wife and I and we wish you the best in your future pursuits.

- Lindsay: I truly appreciate your candor and enthusiasm having taken the reins of the nsp5 project. You are a good friend and I wish you continued success in your future endeavors.
- Clint, Brett, Megan, Nicole, Wayne and Chris Peek: All of you over these last few years have offered advice, encouragement, laughs, and joy in what at times seemed to be a neverending monotonous story. Thank you.

My Graduate Dissertation Committee

We have had a fun time and I have learned a lot from you all. You have made me think objectively about my work and not allow me to BS my way through graduate school. Thank you for your candidness, advice and encouragement. I really appreciate it.

- Earl Ruley, my chair: There are three distinct things that I can personally thank you for: 1. How to think critically about science; 2. Realizing that synthesizing the right questions is the most important step of the scientific method; 3. Knowing how to speak to Aunt Mable or any other non-scientific muggle on the planet. Thank you for the hours of advice and encouragement. You are a great teacher and mentor and are at the heart of our graduate training program. I wish you the best.
- Chris Aiken: Thank you for challenging me to aim high and think outside of the box. The direction of my graduate project was largely influenced by this mentality and in hindsight; it has made me a better scientist for it.
- Ellen Fanning: As the sole member of my committee from an outside department, I want to thank you for being willing and for contributing your time and

knowledge to my training. I know that I am a far way from completely understanding biochemistry but I think your direction and assistance has made me come a far way as well. You will be greatly missed.

- Andy Link: For as much as Chris and Earl liked to challenge my direction and focus, you were there to offer encouragement and support. Thank you for all of the laughs, your guidance and willingness to support me over these last few years.

My Research Collaborators

My research focus has shifted dramatically since the day I started as a graduate student. The feedback and support that I have gotten from Purdue and UNC have offered me the flexibility and resources to keep moving forward.

- Andy Mesecar: Thank you for welcoming me to your lab and offering a biochemical and biophysical view during our collaboration.
- Aimee Eggler: When I visited West Lafayette, you and your family welcomed me into your house. You offered advice, encouragement and a large amount of your time to help me with my project. Thank you for everything and I wish you the best of luck at Villanova.
- Sakshi Tomar and Katie Molland: In my limited interactions with the two of you, you have helped keep my project moving and offered a unique perspective. I really appreciate it.
- Eric Donaldson: Although we never formally had collaboration, you helped me a lot early on with learning structural modeling, bioinformatics and assisting me in any way you could. Thank you and good luck at the NIH.

Vanderbilt Division of Pediatric Infectious Diseases

Thank you for being welcoming and helpful over these past few years. I have great friendships with several of you and I hope the division maintains its strength and camaraderie for years to come.

Vanderbilt Graduate Program in Microbiology and Immunology

The M & I program has transformed considerably since I first joined it as a department. However, I feel that I have been well trained and prepared to think critically about science and offered many tools to help me succeed as I leave here. Thank you all.

My Faculty and Mentors at Xavier University

The advice and support that I received from my academic and research faculty and mentors at my undergraduate institution, Xavier, are the reason that I ultimately opted for graduate school (and a life in academia) rather than medical school (and a life as a physician). You are all amazing people and have given me countless hours of your time and overwhelming support for my life choices and directions. Thank you for being there for me and helping me see the light when I needed it most.

The Nashville Office of The Princeton Review

Thank you Stephanie, Lauren, Dallas, Stacy and the other members of the office staff and the teachers whom I have had the privilege and opportunity to work with since I moved

to Nashville. The opportunity to teach and inspire students is a large part of my career ambitions and TPR Nashville has significantly contributed to it.

My Friends

All of my friends know whom you are and have been there to make me feel human when I have been pounding away the hours in the lab. Thank you for your support, encouragement, and most especially, the laughs.

My Family

My family has been supportive of my vision longer than anyone else on this list. I want to thank them all for being understanding, listening to me when I needed them, and offering a great reprieve when I needed to get away from the lab and Nashville. I love my family very much and really appreciate all that they have done and continue to do for me. Thank you all.

My Wife

Michelle, you are the best thing to have happened to me during my time as a graduate student. You have endured long nights keeping me company while babysitting viruses, early mornings while listening to me study or work on papers and the ups and downs of graduate school life. Thank you for always being there and supporting me in my career goals and ambitions. You have made my life as a graduate student much more enjoyable and I love you very much.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

BACKGROUND AND LITERATURE REVIEW

Introduction

The recent emergence of positive-strand RNA viruses including West Nile Virus (WNV), Chikungunya virus, and SARS-CoV demonstrates a need for understanding positive-strand RNA virus biology, evolution, and pathogenesis. Coronaviruses (CoVs) are positive-strand RNA viruses whose genomes range from 27 to 32 kb in size and infect a wide range of hosts. Currently, six human coronaviruses have been identified including the emergence of four viruses in the last ten years. During the 2002 – 2003 outbreak of Severe Acute Respiratory Syndrome coronavirus (SARS-CoV), over 8000 individuals contracted the disease, which was associated with a 10% mortality rate (Drosten et al., 2003). Recently, Middle East Respiratory Syndrome coronavirus (MERS-CoV) has been identified in patients in the Middle East and has been associated with nearly a 50% mortality rate (Zaki et al., 2012). These coronavirus outbreaks highlight the importance of understanding mechanisms of viral emergence to facilitate vaccine and antiviral development (Drosten et al., 2003; Zaki et al., 2012).

A key hallmark of positive-strand RNA virus infection is the early translation of the viral genome in the host cytoplasm and subsequent maturation processing steps mediated by viral and host proteases. During infection, the coronavirus genome is translated into two large polyproteins that must be proteolytically processed by virus-

encoded proteases (Brierley et al., 1989; Lee et al., 1991; Perlman and Netland, 2009). The coronavirus protease nonstructural protein 5 (nsp5 or 3CLpro) is responsible for catalyzing the maturation processing at 11 cleavage sties and is required for replication (Lee et al., 1991; Perlman and Netland, 2009). Despite a wealth of biochemical and biophysical data at the initiation of this research, very little was known about the mechanism of nsp5 protease action in the context of virus infection in cells. Furthermore, studies to develop nsp5-specific inhibitors have been largely focused on substrate peptide analogs and have not yet resulted in effective antivirals (Jacobs et al., 2013).

This project focused on identifying the key structural and functional determinants of coronavirus nsp5 protease that govern activity and regulate specificity, studies in Chapter II describe the recovery and characterization of MHV nsp5 temperature-sensitive and second-site suppressor mutants and the biochemical characterization of the mutant proteases *in vitro*. Studies in Chapter III evaluate the extent of nsp5 functional conservation through substitution of other coronavirus nsp5 proteases into the background of murine hepatitis virus (MHV). The implications for this work are discussed in Chapter IV and future directions are examined in Chapter V.

Coronavirus classification

Coronaviruses are members of the order *Nidovirales*, family *Coronaviridae*, subfamily *Coronavirinae,* and genus *Coronavirus*. They were originally named due to their virion structure resembling that of the corona of the sun (**Figure 1.1)**. The first coronaviruses were identified in 1932 in cloaca samples of chicken (Hudson and

Beaudette, 1932). These viruses were later identified as infectious bronchitis virus (IBV). Initial classification of coronaviruses was performed by serotype analyses. However, with the advent of facile sequencing and *in silico* analyses, coronaviruses classification has transformed with the continual discovery of new coronavirus species. Recently, coronaviruses have been reorganized into four distinct phylogenetic genera based upon genomic organization and phylogenetic analyses: alphacoronaviruses, betacoronaviruses, gammacoronaviruses, and deltacoronaviruses (**Figure 1.1**) (Woo et al., 2012).

Known alphacoronaviruses include human coronaviruses 229E and NL63, canine and feline coronaviruses, porcine viruses epidemic diarrhea virus (PEDV) and transmissible gastroenteritis (TGEV), as well as several bat coronaviruses (BtCoVs) including M*iniopterus* BtCoV1 and BtHKU8, *Rhinolophus* BtHKU2, and S*cotophilus* BtCoV 512. Among the known betacoronaviruses are four human coronaviruses (HKU1, OC43, SARS-CoV, and MERS-CoV), murine hepatitis virus (MHV), and three distinct BtCoV species (*Rousettus* BtHKU9, *Pipistrellus* BtHKU5, and *Tylonycteris* BtHKU4). To date, all six known human coronaviruses have been either alpha- or betacoronaviruses and only recently has the first non-avian coronavirus been identified outside of these phylogenetic groups (Mihindukulasuriya et al., 2008). Gammacoronaviruses and Deltacoronaviruses largely consist of avian coronaviruses including infectious bronchitis virus (IBV), bulbul HKU11, thrush HKU12, munia HKU13 as well as a beluga whale coronavirus SW1 (Mihindukulasuriya et al., 2008; Woo et al., 2012).

The majority of all coronaviruses that have been identified to date have been found circulating asymptomatically in bats. Studies from our lab and others have shown that SARS-CoV is most likely derived from bats (Becker et al., 2008; Lau et al., 2005). In

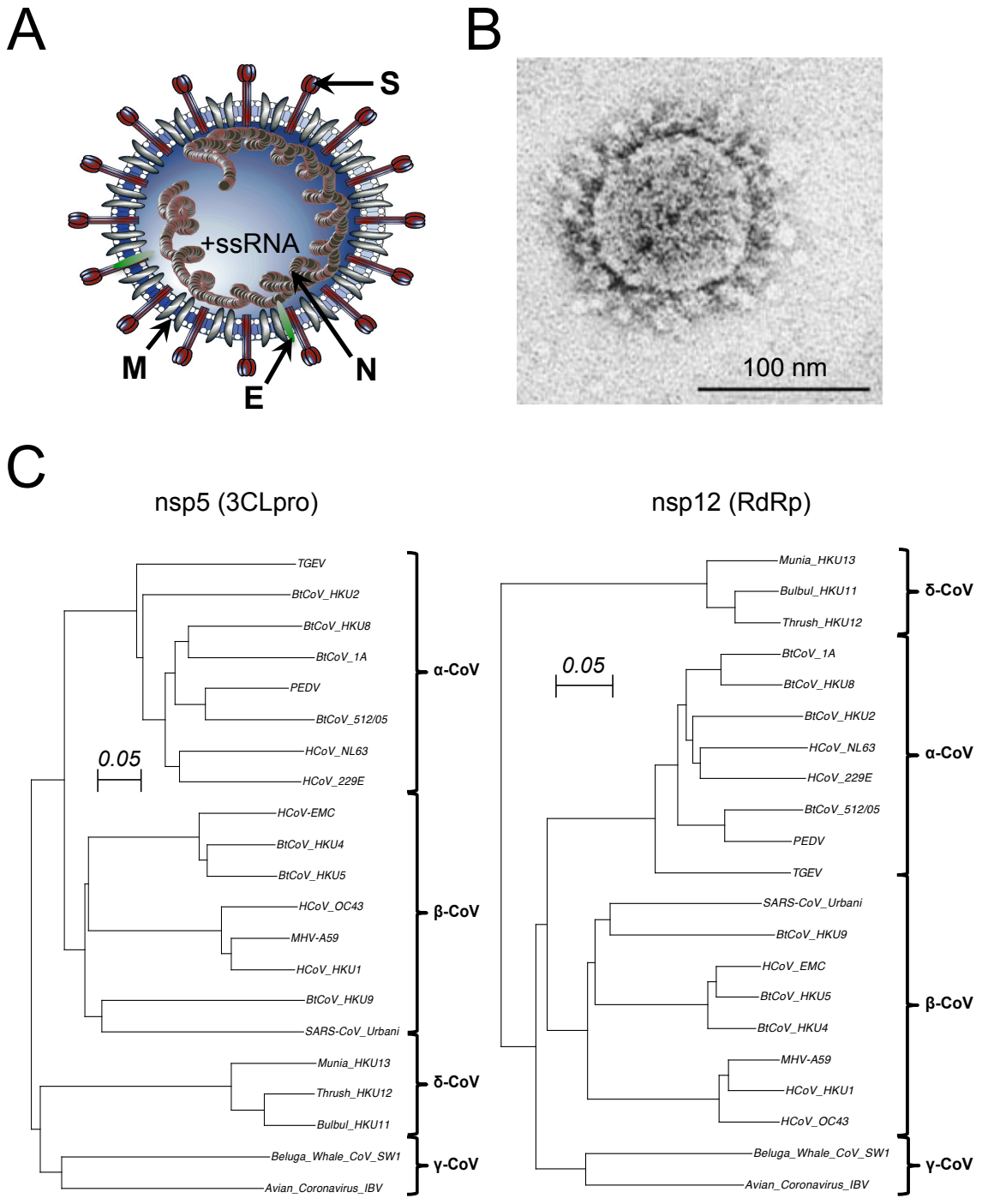


Fig. 1.1. Coronavirus virion structure and phylogeny. A) Coronavirus virion schematic exhibiting the structural proteins spike (S), membrane (M), envelope (E) and nucleocapsid (N), which are responsible for forming and maintaining virion structure and genomic (+ssRNA) stability. Image modified from Jen Sparks. **B)** A negative-stain electron micrograph of SARS-CoV, the horizontal bar represents the image scale. Image adapted from Ksaizak et al., 2003. **C)** Phylogenetic trees based on the primary peptide sequences of nonstructural proteins 5 (nsp5 or 3CLpro) and 12 (nsp12 or RdRp). The trees were generated using a bootstrap alignment in ClustalX. Phylogenetic groups are denoted by brackets.

addition, phylogenetic analysis of the recently emerged MERS-CoV coronavirus suggests that the virus may also be derived from circulating myotic species (Anthony et al., 2013; Kindler et al., 2013). These data stress the importance for ongoing surveillance and understanding the mechanisms and limitations for coronavirus emergence and pathogenesis.

Coronavirus emergence and human disease

Coronaviruses infect a wide range of animal hosts and cause illnesses ranging from bronchiolitis to gastroenteritis. In November 2002, an outbreak of a respiratory pathogen was identified in the Guangdong province of China (Drosten et al., 2003). By the end of July 2003, the virus had spread to 37 countries infecting over 8000 individuals with a mortality rate of 9.6%. At the time of the identification of the causative agent as severe acute respiratory syndrome (SARS-CoV) coronavirus, only two other human coronaviruses were known (**Table 1.1**). Coronaviruses 229E and OC43 were first identified in the 1960's and were associated with common colds and pneumonia (Cavallaro and Monto, 1970; McIntosh et al., 1967a; McIntosh et al., 1967b). The emergence of a new virulent human coronavirus led to global efforts to identify the zoonotic source of the virus and to understand the mechanisms of emergence, virulence and replication. During the SARS-CoV epidemic, the virus was identified in a number of animal hosts including raccoon dogs (*Nyctereuteus* sp.), ferret badgers (*Melogale sp.*), and palm civets (*Paguna* sp.) (Becker et al., 2008; Lau et al., 2005; Normile, 2005; Vijayanand et al., 2004). However, later studies showed that the most likely origin of

zoonotic spread were Chinese horseshoe bats (*Rhinolophus* sp.) (Becker et al., 2008; Lau et al., 2005).

Table 1.1. Human coronavirus identification, classification and disease

Human Coronaviruses			
Virus	***Group***	***Year Identified***	***Clinical Disease***
HCoV-229E	α-CoV	1967	colds, pneumonia
HCoV-OC43	β-CoV	1969	colds, pneumonia
SARS-CoV	β-CoV	2002	severe acute respiratory syndrome
HCoV-NL63	α-CoV	2004	bronchiolitis, colds, pneumonia
HCoV-HKU1	β-CoV	2005	colds, pneumonia
MERS-CoV	β-CoV	2012	pneumonia, kidney failure

Increased surveillance during the SARS-CoV epidemic resulted in identification of HCoV-NL63 from a 7-month old child presenting with bronchiolitis and conjunctivitis in the Netherlands. The virus was similar to HCoV-229E and had spread worldwide by the end of 2004 (van der Hoek et al., 2004). A year later, a fifth coronavirus was discovered in Hong Kong in two pneumonia patients and was subsequently named HCoV-HKU1 (Woo et al., 2005). Surveillance studies worldwide have shown persistent circulation of human coronaviruses 229E, OC43, NL63, and HKU1 associated with upper and lower respiratory illnesses of varying severity (Dijkman et al., 2012; Mackay et al., 2012; Prill et al., 2012). In September of 2012, a sixth human coronavirus was identified in Qatar and Saudi Arabia, and has since been named Middle East Respiratory Syndrome coronavirus (MERS-CoV) (Zaki et al., 2012). Efforts at identifying the source of the virus have suggested that MERS-CoV, like SARS-CoV, has a bat origin (Anthony et al., 2013; Kindler et al., 2013). Phylogenetic analysis of MERS-CoV and known bat

coronaviruses have shown that MERS-CoV shares 97% identity to the RNA-dependent RNA polymerase of a bat coronavirus (Anthony et al., 2013). Ongoing surveillance of coronaviruses in nature has led to a continual expansion of the breadth of coronaviruses and potential for zoonotic spread and virus emergence.

Coronavirus replication

Coronaviruses are enveloped and initiate infection through engagement of viral class I fusion spike (S) attachment proteins to specific host cell surface receptors (**Figure 1.2**). Receptor usage varies greatly among coronaviruses and includes carcinoembyronic antigen adhesion molecule 1, CEACAM1a (MHV), aminopeptidase N (229E), angiotensin-converting enzyme 2, ACE2 (NL63 and SARS-CoV), 9-*O*-acetylated sialic acid (OC43) and DPP4 (MERS-CoV) (Perlman and Netland, 2009; Raj et al., 2013; Williams et al., 1991). Following entry by either direct membrane-envelope fusion or receptor-mediated endocytosis, the 20 – 22 kb replicase gene (ORF1) is translated in the host cell cytoplasm into two large polyproteins (pp1a and pp1ab) by host ribosomes (**Figure 1.3**) (Kooi et al., 1991; Lee et al., 1991). Polyprotein 1a includes nonstructural proteins 1 – 11 (nsps 1 – 11) and pp1ab includes nsps 1 – 16 via a -1 ribosomal frameshift between nsp10 and nsp12 (Brierley et al., 1989; Lee et al., 1991; Perlman and Netland, 2009). The replicase polyproteins are comprise the replication machinery of the virus including a single-strand RNA binding protein (nsp7), a primase (nsp8), the RNA-dependent RNA polymerase (nsp12), helicase (nsp13), an exonuclease with proofreading activity (nsp14), a uridylate-specific endoribonuclease (nsp15), and a ribose-2'-*O*-

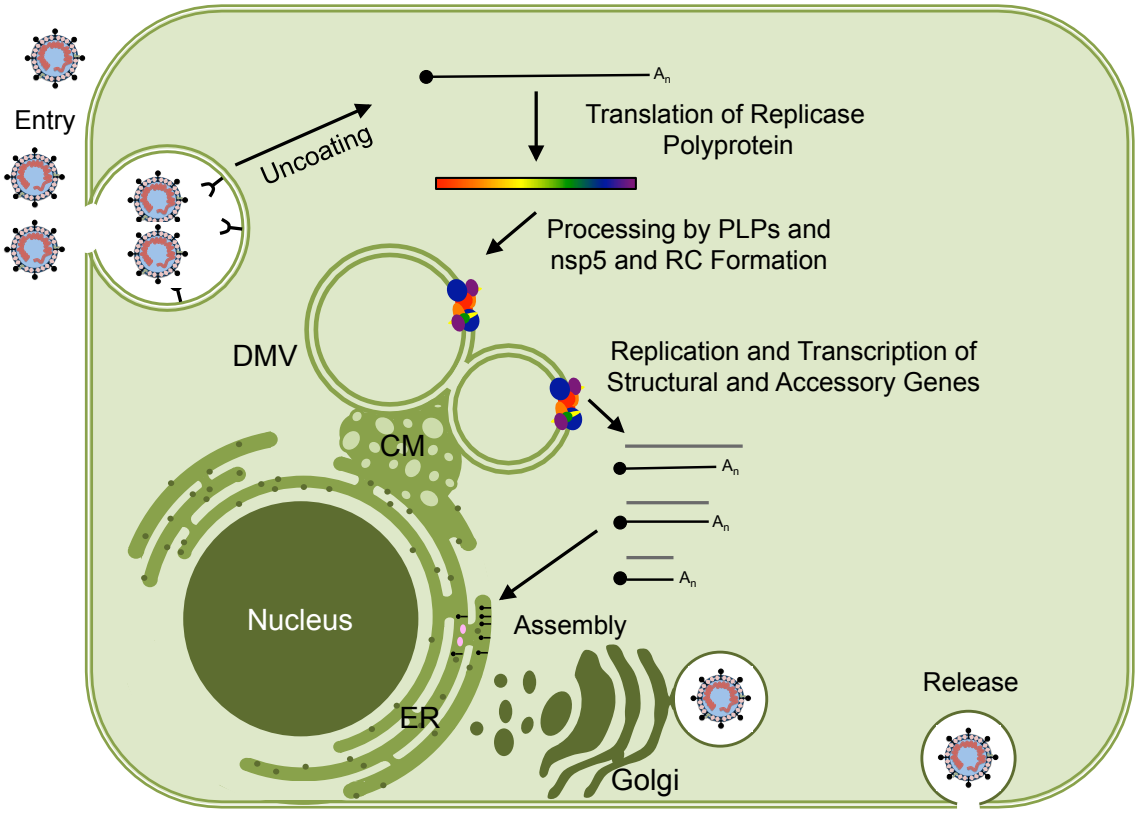


Fig. 1.2. Coronavirus replication cycle. Coronaviruses initiate infection and entry through receptor-mediated endocytosis or direct membrane fusion. Upon entry into the host cytoplasm, the replicase gene encoding viral replication machinery is translated by host ribosomes to yield two polyproteins (pp1a and pp1ab). The replicase polyproteins are proteolytically processed by virus-encoded papain-like proteases and nonstructural protein 5 (nsp5) to yield up to 16 nonstructural proteins (nsps) that assemble to form replication complexes (RCs) on host membrane-derived viral structures including double membrane vesicles (DMVs) and convoluted membranes (CMs). The +ssRNA genome is then used as a template to create genomic and subgenomic RNA to be used for generation of structural and accessory gene transcripts. The viral genome and structural and accessory proteins are then trafficked to the ER-Golgi Intermediate Compartment (ERGIC) where they assemble into virions. Virions are released through a non-lytic vesicle via the secretory pathway.

methyltransferase (nsp16) (Perlman and Netland, 2009). Other nonstructural proteins have been associated with host-membrane alterations (nsps 3, 4, and 6), host mRNA degradation and immune inactivation (nsp1), deubiquitylating activity (nsp3), and proteolytic processing of the polyproteins (nsp3 and nsp5) (Perlman and Netland, 2009; Ziebuhr et al., 2000). During and after translation, two or three virus-encoded proteases process polyproteins 1a and 1ab to yield mature nsps 1 – 16 (Ziebuhr et al., 2000). One or two papain-like proteases in nsp3 are responsible for processing nsps 1 – 3, and a cysteine protease, nsp5, the focus of my research, is responsible for processing nsps 4 – 16 (Ziebuhr et al., 2000).

During the early stages of proteolytic processing and replication complex formation, nsps 3, 4, and 6 are proposed to modify host ER membranes, resulting in intracellular virus-induced structures referred to collectively as a reticulovesicular networks and are composed of double membrane vesicle (DMV) and convoluted membrane (CM) structures (Knoops et al., 2008; van Hemert et al., 2008). Genomic replication is proposed to occur in association with these structures and involves two distinct stages: full genome replication and transcription of subgenomic mRNAs. Genome replication consists of generation of a full-length negative-sense genome intermediate, which is used as a template for generating positive-strand genome copies. Translation of all structural and accessory proteins including hemagglutinin-esterase (HE), spike (S) attachment, envelope (E), membrane (M), nucleocapsid (N) and additional accessory proteins occurs from subgenomic mRNAs (Pasternak et al., 2006; Sawicki and Sawicki, 1998).

After genomic replication and transcription have occurred, virions assemble in the endoplasmic reticulum-Golgi intermediate compartment (ERGIC) and Golgi. Nucleocapsid (N) associates with the genomic RNA and is transported to the sites of virion assembly by an unknown mechanism. The virion structure is assembled by interactions between M and E proteins and the ERGIC membrane. HE and S are integrated into the virion membrane and the virions are released via the secretory pathway.

Proteolytic processing of coronavirus replicase polyproteins

A common feature of all positive-strand RNA virus replication is the translation of one or more polyproteins that must be proteolytically processed by virus- or host-derived proteases to yield mature viral proteins. The genomes of the order *Nidovirales* (including genera *Coronaviridae*, *Arteriviridae, Mesoniviridae, and Roniviridae)* differ considerably in size ranging from MHV at 31.5 kb to equine arteritis virus (EAV) of a little less than 13 kb. However, all known nidovirus genomes include two or more virus-encoded proteases that appear indispensible for viral replication (Ziebuhr et al., 2000). The first two-thirds of the nidovirus genome consists of two large overlapping replicase reading frames (ORF1a and ORF1ab), which are translated to yield two large polyprotein precursors, pp1a and pp1ab (**Figure 1.3**), ranging in size from 3100 to 7200 amino acids in length. Similar to many other positive-strand RNA viruses, these polyproteins are cleaved by proteinases, which resemble structurally the cellular proteases of papain and chymotrypsin (Ziebuhr et al., 2000). In MHV, two separate papain-like protease domains

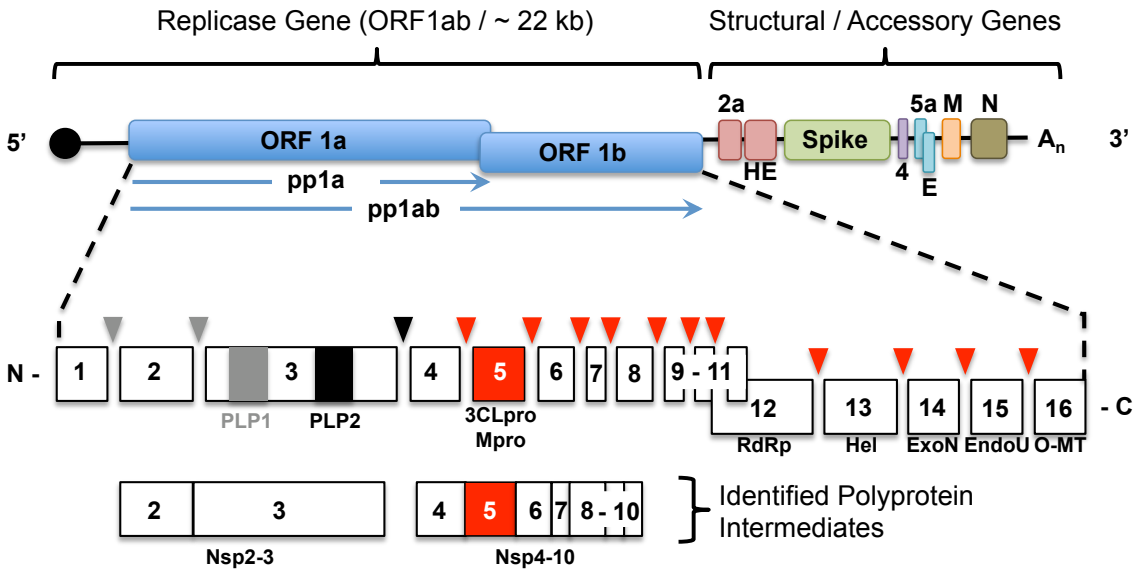


Fig. 1.3. Murine hepatitis virus (MHV) translation and polyprotein processing. The MHV genome is 32 kb in length and consists of 7 genes. The replicase gene (blue) consists of two reading frames (ORF1a and ORF1ab), which overlap at a ribosomal frameshift and are translated to yield polyproteins, pp1a and pp1ab, encoding nonstructural proteins (nsps 1 – 16). Maturation cleavage events (arrows) are mediated by three viral protease domains, papain-like proteases PLP1 (light gray) and PLP2 (black) and nsp5 (red). Known polyprotein intermediates are shown. RdRp, RNA-dependent RNA polymerase; Hel, Helicase; ExoN, exoribonuclease; EndoU, endoribonuclease; O-MT, *O*-methyltransferase; HE, hemagglutinin-esterase; E, envelope, M, membrane, and N, nucleocapsid.

(PLP1 and PLP2) are responsible for processing nsps 1 – 3. Whereas, the chymotrypsin-like cysteine protease, nsp5, is responsible for catalyzing 11 separate maturation cleavages between the C-terminus of nsp4 and nsp16.

The papain-like proteases of coronaviruses are responsible for recognizing and processing at two or three cleavage sites in replicase polyproteins. All known papain-like proteases use a Cys-His catalytic dyad and an α+β-like fold similar to the cellular papain enzyme (Ziebuhr et al., 2000). Some betacoronaviruses (Group 2b, such as SARS-CoV) and all known gamma- and deltacoronaviruses encode a single papain-like protease (PL_{pro}), while all known alphacoronaviruses and remaining betacoronaviruses (Group 2a, such as MHV) encode two separate PLP domains (PLP1 and PLP2). Inactivation of PLP1, although tolerated, causes impairment of viral RNA replication (Graham and Denison, 2006). Similarly, deletion and substitution of the PLP1 and PLP2 cleavage sites in MHV is tolerated, but results in considerable replication defects (Gadlage and Denison, 2010; Gadlage et al., 2010; Graham and Denison, 2006).

Coronavirus nsp5 protease, also referred to as 3CLpro or Mpro, is indispensible for virus replication and catalyzes proteolytic processing at 11 different cleavage sites, including its own autoproteolytic liberation. Processed intermediates, such as an nsp4-10 and nsp7-10 precursor, are detected during coronavirus infection, indicating a specific order and hierarchy in nsp5 proteolytic events (Sparks et al., 2008; Stobart et al., 2012). Additionally, several cleavage events have been shown to be dispensable for replication including processing between the nsp9/nsp10 and nsp10/nsp12 (Deming et al., 2007).

Studies by Grum-Tokars et al., have suggested cleavage at the nsp4-nsp5 cleavage site is most efficient and alteration of either terminus of nsp5 reduces protease activity

(Grum-Tokars et al., 2008). Consequently, efficient autoproteolysis and early nsp5 activity likely requires a hydrophobic milieu for optimum processing (Okamoto et al., 2010; Pinon et al., 1997). Nsp5 is located between two clear hydrophobic membrane-spanning regions of nsp4 and nsp6. Adjacent to the N-terminus of nsp5 are the C-terminal hydrophobic loops of nsp3 and the hydrophobic membrane-associated nsp4 and adjacent to the C-terminus of nsp5 is an additional membrane spanning and associated protein, nsp6. The localization of nsp5 to membranes and its association with the replication complex is likely dependent upon its interaction with neighboring proteins nsp4 and nsp6, however much remains unknown of this interaction.

Coronavirus protease nsp5 structure

Coronavirus nsp5 is a 28 kDa protein ranging in size from 300 to 306 amino acids. Nsp5 (3CLpro) has been crystallized for at least seven different coronavirus species from three different genera including SARS-CoV, NL63, 229E, Bat HKU4, HKU1, IBV and recently MERS-CoV (Anand et al., 2002; Anand et al., 2003; Bacha et al., 2008; Xue et al., 2007; Xue et al., 2008; Zhao et al., 2008). In addition, structures of SARS-CoV nsp5 have been resolved for both dimeric and monomeric forms, and complexed with a variety of peptidomimetic inhibitors. All known crystal structures demonstrate tertiary and quaternary structure conservation despite considerable variation in primary sequences between coronaviruses (**Figure 1.4**).

Nsp5 protease consists of three distinct domains (**Figure 1.5)**. Domains 1 (MHV residues 1 – 100) and 2 (MHV residues 101 – 199) constitute a chymotrypsin-like fold.

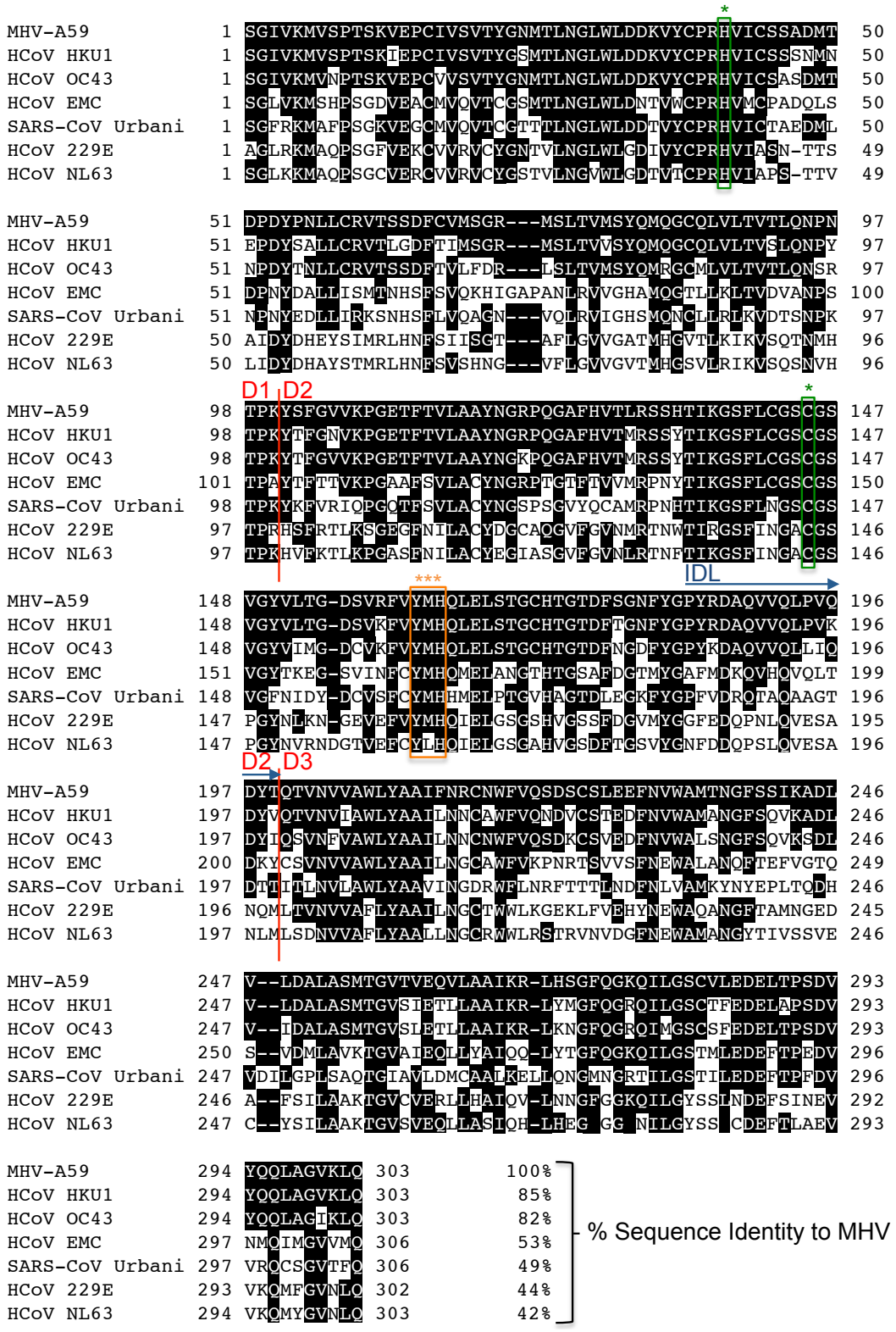


Fig. 1.4. Primary sequence alignment of MHV and human coronavirus nsp5 sequences. Residues that identical to MHV are shaded. Domain breaks (red), the interdomain loop (IDL; blue), catalytic dyad residues (green) and substrate binding pocket residues (orange) are denoted. Percent total sequence identity to MHV is shown at the end of the sequence reads.

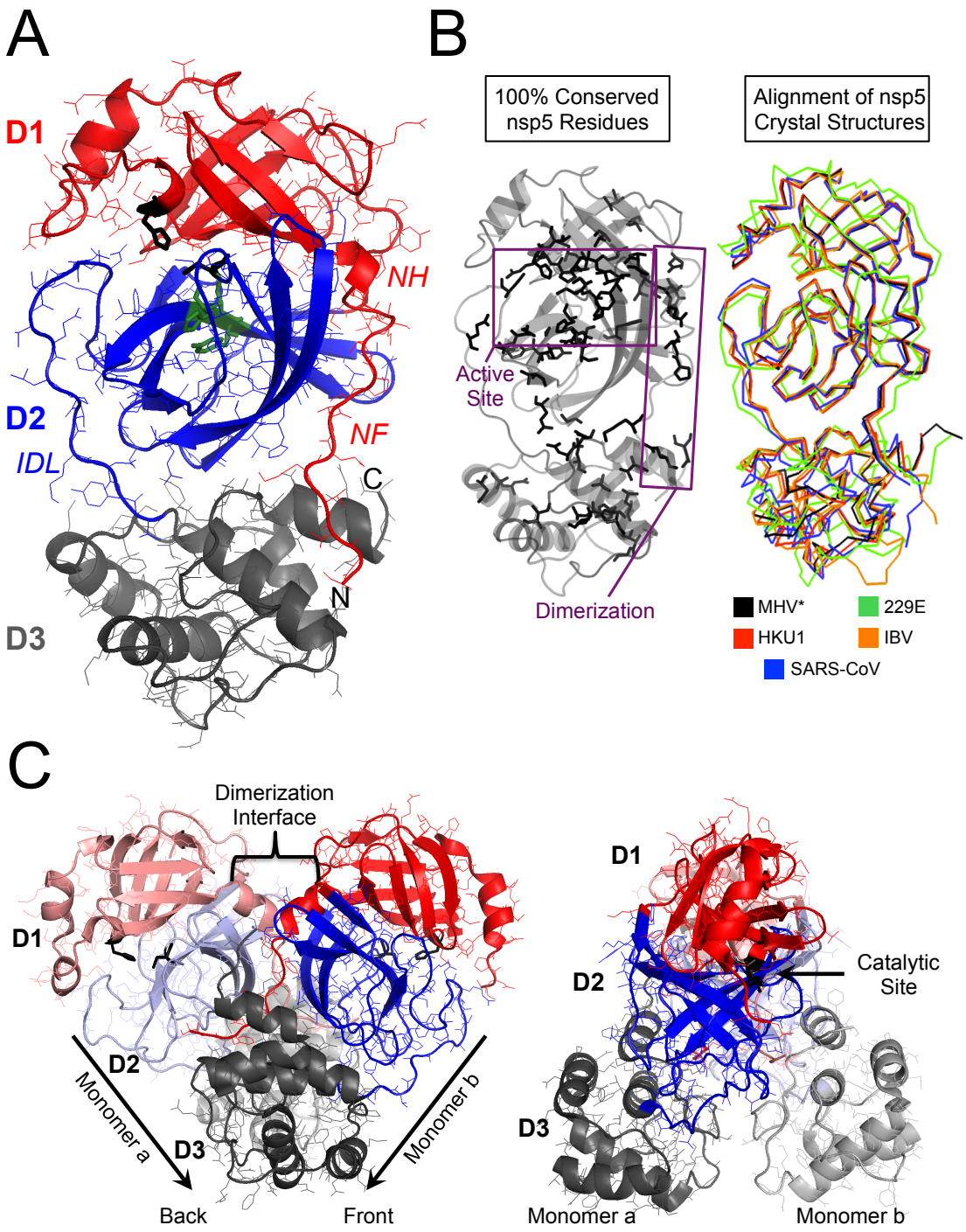


Fig. 1.5. Structural analysis of coronavirus protease nsp5. A) Monomeric crystal structure of HKU1 nsp5 (PDB code 3d23; Zhao, 2008). Domains 1 (D1; red), 2 (D2; blue), and 3 (D3; gray) are labeled and color-coded, the catalytic dyad (black) are shown in bold, and the N-terminal finger (NF), N-terminal helix (NH) and interdomain loop (IDL) are labeled in italics. **B)** Residues that 100% conserved across coronavirus nsp5 protease sequences (left) are shown including outlines of the active site and dimerization regions (purple) and an overlay of known nsp5 crystal structures of HKU1, 229E (1P9T; Xue, 2006), IBV (2Q6D; Xue, 2007) and SARS-CoV (2H2Z; Anand, 2002) with a structural model generated in Modeller from HKU1 are shown. **C)** Dimeric structure of SARS-CoV nsp5 (2GT7; Lee, 2007) with two different views.

Mutagenesis in MHV, IBV, 229E and feline infectious peritonitis virus (FIPV) demonstrated that catalysis is mediated by a His/Cys dyad spanning the interface between domains 1 and 2 (Lu et al., 1996; Lu and Denison, 1997; Ziebuhr et al., 2000). Substrate binding is coordinated through interaction with a conserved Tyr-X-His motif and several conserved residue contacts surrounding the interface between the domains. A long conserved interdomain loop (MHV residues 184 to 199) connects the domains 1 and 2 to domain 3. Domain 3 (MHV residues 200 to 303) consists of five predominantly anti-parallel α-helices and represents a unique structural element not observed in any other viral or cellular chymotrypsin-like proteases.

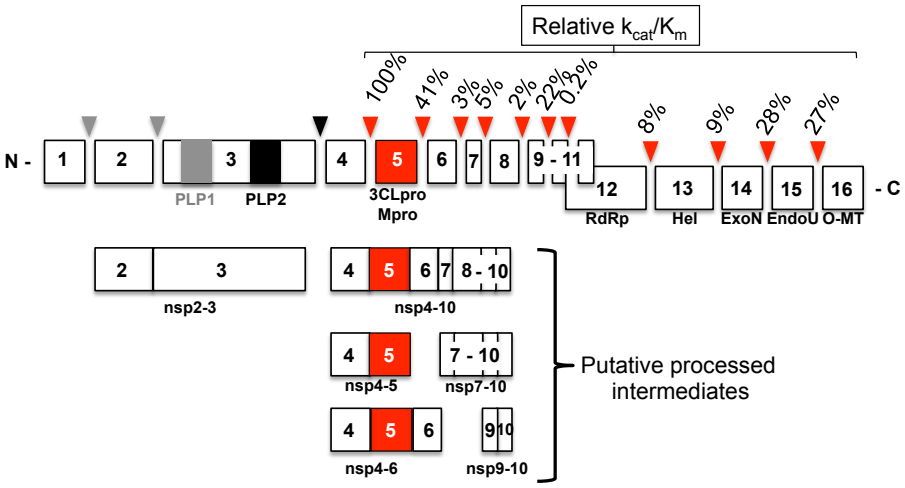
Analyses of expressed and purified 3CLpro have demonstrated that a mixture of monomeric (~35%) and dimeric (~65%) states exists with continual monomer-dimer exchange (Chen et al., 2006; Shi and Song, 2006). The dimer structure is assembled with two nsp5 monomers oriented perpendicular to one another and interfacing along a N-terminal extension or finger of domain 1 and part of domain 3. The active sites of the two monomers are oriented away from the dimerization interface and some *in vitro* studies have suggested that only one monomer may be functionally active in the dimer (Anand et al., 2002; Anand et al., 2003; Chen et al., 2006; Zhao et al., 2008). Dimerization of nsp5 appears to be required for full proteolytic activity and specificity *in vitro*. Experiments have shown that disrupting the dimerization interface *in vitro* result in partial or complete protease inactivation. Deletion of N-terminal finger residues 1 -3 in SARS-CoV results in a 50% reduction in catalytic activity despite retaining dimerization (Hsu et al., 2005). However, deletion of residues 1 – 4 in SARS-CoV resulted in a 100-fold reduction in catalytic activity and a subsequent loss of dimerization. A recent finding has suggested

that dimerization may not be required for residual catalytic activity, as the N-terminal autoproteolytic cleavage appears to occur in a monomeric form *in vitro* (Chen et al., 2010; Muramatsu et al., 2013).

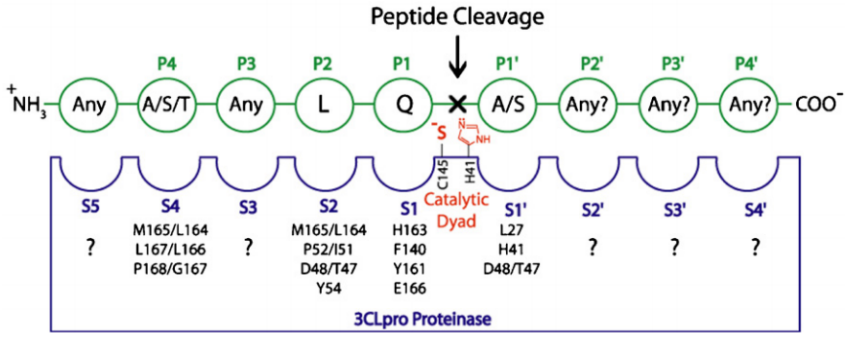
Coronavirus nsp5 substrate specificity

Coronavirus nsp5 (3CLpro) is responsible for recognizing and processing at 11 distinct cleavage sites (nsps 4 – 16 in MHV). However, the mechanism and order of nsp5-mediated processing during infection is unknown. Protease cleavage sites are designated at P5 – P5' sites where cleavage occurs between the P1 and P1' amino acid residues. Nsp5 almost exclusively cleaves following a P1-Gln, with the sole exception being a P1-His in the nsp13-nsp14 cleavage site of HCoV-HKU1 (**Figure 1.6**) (Chuck et al., 2011; Ziebuhr et al., 2000). There is a preference for small non-charged residues at the P1' position, but Cys may also be tolerated. In addition, there is a strong preference for Leu at P2, but Phe, Met and Ile may also be utilized. Less conservation is observed at the P3, P4, P2' and P3' sites. Overall, the consensus nsp5 cleavage site among known coronaviruses is P3-LQ†(S/G/A/V)-P2' (Grum-Tokars et al., 2008; Ziebuhr et al., 2000). Catalytic activity has been shown *in vitro* to be most active towards the autoproteolytic N-terminal cleavage site (nsp4-nsp5) (Grum-Tokars et al., 2008). Several studies have assessed which cleavage events are essential for virus replication. Deming *et al*, demonstrated that elimination of the nsp7-nsp8 and nsp8-nsp9 cleavage sites resulted in loss of virus recovery; elimination of the nsp9-nsp10 site was tolerated with a reduction in virus replication kinetics (Deming et al., 2007). Other unpublished studies from our lab

A



B



C

	Coronavirus nsp5 Cleavage Sites			
Junction	**MHV**	**HKU1**	**OC43**	**SARS-CoV**
nsp4/5	TSFLQ SG	TSFLQ SG	TSFLQ SG	SAVLQ SG
nsp5/6	GVKLQ SK	GVKLQ SK	GIKLQ SK	GVTFQ SK
nsp6/7	VSQIQ SR	VSQIQ SK	VSQFQ SK	VATVQ SK
nsp7/8	NTVLQ AL	STVLQ AL	NTVLQ AL	RATLQ AI
nsp8/9	TVVLQ NN	NAVMQ NN	ATVLQ NN	AVKLQ NN
nsp9/10	TVRLQ AG	TIRLQ AG	TVRLQ AG	TVRLQ AG
nsp10/12	GSQFQ SK	SVAVQ SK	DTTVQ SK	EPLMQ SA
nsp12/13	SAVLQ SV	SAVMQ SV	SAVMQ SV	HTVLQ AV
nsp13/14	NPRLQ CT	LPRLH CT	ETKVQ CS	VATLQ AE
nsp14/15	FTRLQ SL	FTTLQ SL	FTKLQ SL	FTRLQ SL
nsp15/16	YPRLQ AA	YPKMQ AT	YPRLQ AA	YPKLQ AS

Fig. 1.6. Nsp5 cleavage site specificity and processing. A) MHV polyprotein 1ab (pp1ab) consists of nonstructural proteins 1 – 16, which are processed by viral proteases PLP1 (red), PLP2 (black) and nsp5 (red). The relative activity as measured by the k_{cat}/K_m towards each cleavage site is shown (Grum-Tokars et al. 2008 and Fan et al. 2005). Predicted putative processed intermediates are shown below. **B)** A canonical nsp5 cleavage site substrate is shown with interacting residues for SARS nsp5 (3CLpro) in the active site (adopted from Grum-Tokars et al. 2008). **C)** Table displaying the coronavirus nsp5 recognition sites for MHV, HKU1, OC43, and SARS-CoV. Differences from the analogous MHV recognition site are highlighted in red.

have shown that the nsp13-nsp14 and nsp14-nsp15 sites were also dispensable for virus replication (Eckerle, unpublished results).

Intermolecular interactions of nsp5

Several studies have demonstrated that intermolecular associations of nsp5 with other replicase polyproteins are critical for nsp5 activity. Early coronavirus studies sought to determine complementation groups to identify and comprehend the different phenotypic regions of the coronavirus genome (Baric et al., 1990; Sawicki et al., 2005). These studies suggested that nsp4-10 constituted a unique complementation group or cistron (I) (Sawicki et al., 2005). Subsequent studies have shown that an intermediate nsp4-10 precursor may have one or more unique functions prior to nsp5-mediated processing (**Figure 1.7**). Stokes et al, demonstrated that a temperature-sensitive (*ts)* mutant (MHV-Br*ts*31) containing a *ts* allele in nsp3 had reduced processing of the nsp4-10 precursor (p150) (Stokes et al., 2010). Another study showed that a *ts* mutation in nsp10 resulted in decreased nsp5-mediated polyprotein processing (Donaldson et al., 2007). Lastly, the direct mutagenesis of the IBV nsp15-nsp16 cleavage site, resulted in compensatory mutations (P166S or P166L) in the nsp5 substrate-binding pocket (Fang et al., 2010). Collectively, these studies suggest that long-distance communication between nsp5 and other elements of the replicase gene are critical for nsp5 specificity and function. In Chapter III, I will describe data that begins to evaluate the conservation of nsp5 activity between coronaviruses and the suitability of different nsp5 proteases to function in a common isogenic coronavirus background.

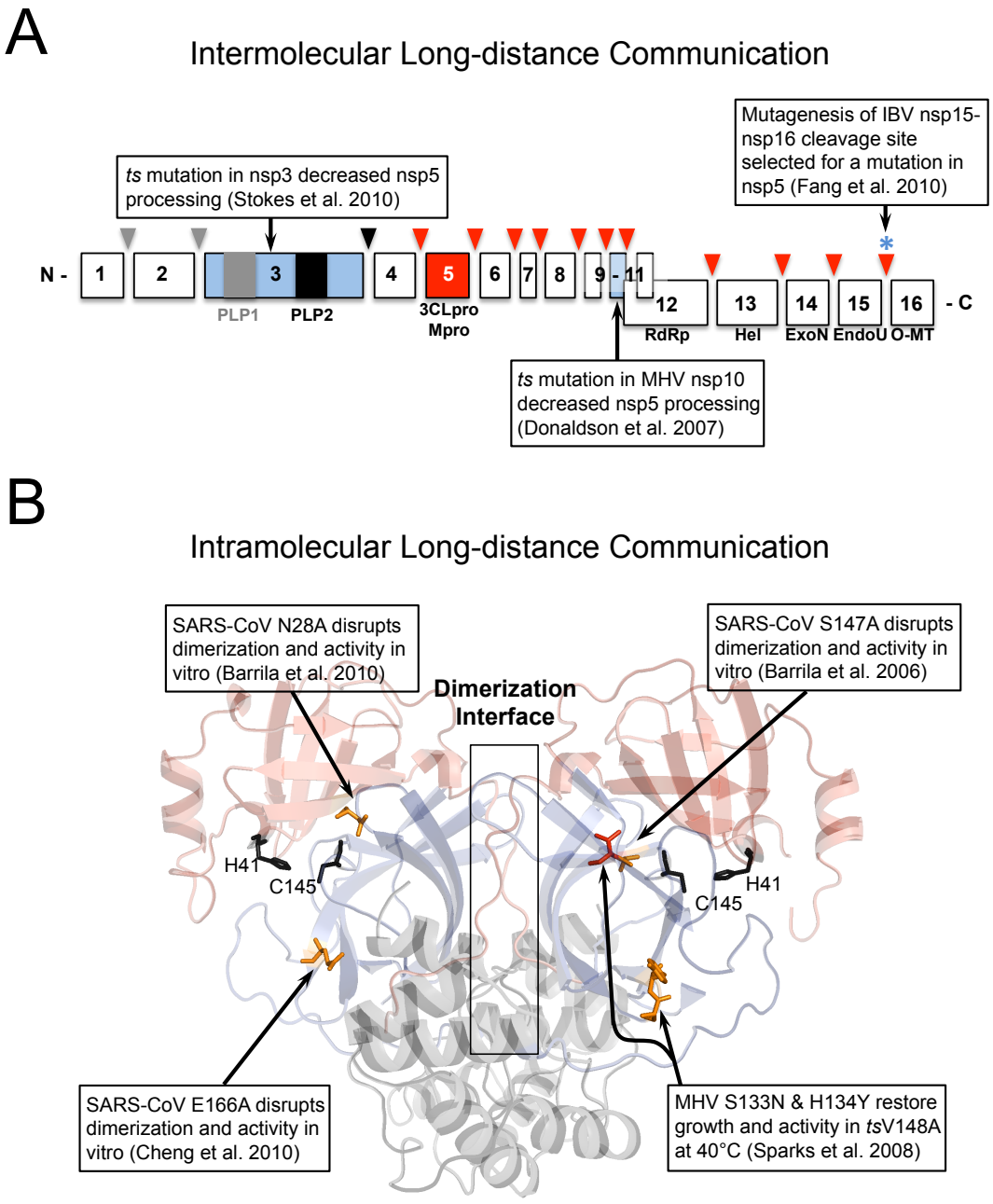


Fig. 1.7. Inter- and intramolecular long-distance communication and nsp5 activity. A) Published studies demonstrating intermolecular long-distance communication between nsp5 and other elements (light blue) of the replicase polyprotein. **B)** Published studies demonstrating intermolecular long-distance communication between residues (gold and red [*ts*]) of coronavirus nsp5 proteases.

Intramolecular long-distance communication and nsp5 function

Studies from our lab and others have demonstrated that nsp5 amino acid residues distinct from active site, substrate binding pockets, or predicted dimerization determinants may be important for nsp5 proteolytic activity. Substitution of alanine at SARS nsp5 residue S147 (S147A) resulted in a 150-fold decrease in protease activity. Although the residue S147 is 9Å from the dimer interface, the mutation S147A disrupts dimerization of nsp5 *in vitro*, consistent with intramolecular long-distance communication, or interaction between non-adjacent residue determinants (Barrila et al., 2006). Two follow-up studies showed that positions N28 and E166, when mutated to alanine, were also associated with disruption of dimerization and loss of protease activity (Barrila et al., 2010; Cheng et al., 2010). In 2008, Sparks et al reported the identification of a temperature-sensitive mutation, V148A, in domain 2 of MHV nsp5 which was associated with decreased viral replication and altered polyprotein processing at non-permissive temperature of 40°C (Sparks et al., 2008). The MHV *ts*V148A phenotype was suppressed by second-site mutations S133N or H134Y, which were greater than >16Å from the original *ts* allele. Introduction of the suppressor alleles in modeled structures predicted adjacent side-chain modifications, but there were no predicted changes that propagated to V148A. The mechanism of long-distance communication and its role in nsp5 stability, specificity and function remains largely unknown. In Chapters II and III will describe data, which provides strong evidence that coronavirus nsp5 activity is governed by complex intramolecular long-distance communication and that these residue interactions have likely diverged between coronaviruses.

Development of nsp5-specific coronavirus inhibitors

To date, there are no commercially available coronavirus vaccines or antivirals despite the clear demonstrated potential for emergence of a pandemic SARS-like coronavirus. Coronavirus nsp5 is indispensible for virus replication and has been a primary target for coronavirus inhibitor design (Lu et al., 1996; Lu and Denison, 1997). The coronavirus nsp5 protease active site consists of a cysteine residue (Cys145 in MHV) which acts as the nucleophile group and a histidine (His41 in MHV) that as an acid-base donor (Anand et al., 2002; Anand et al., 2003; Ziebuhr et al., 2000). Numerous approaches have been used to develop a pan-coronavirus nsp5 inhibitor. Peptidomimetic inhibitors resembling the consensus P5 – P3' cleavage site residues have been developed and have shown sub-μM inhibitory concentrations (Jacobs et al., 2013). Hexapeptidyl chloromethyl ketone inhibitors have recently been a focus of inhibitor design. Several inhibitors of this class have been co-crystallized with SARS-CoV nsp5 and are currently being optimized for broad neutralization. These compounds consist of a peptide chain bearing a reactive warhead group that interacts covalently with the nucleophilic Cys. Warhead groups that have been used include aldehydes, halo-methyl ketones, epoxy-ketones, trifluoromethyl ketones, and several different Michael acceptors (Jacobs et al., 2013). Other non-peptidic inhibitors have been developed including cinanserin, which demonstrated a reported IC_{50} of 5 μM. Despite encouraging results in early screens of compounds, many of these compounds will not be practical due to toxicity and off-target effects. An eventual goal of my research will be to identify key structural and functional determinants of nsp5, which may be targeted for inhibitor design.

Coronavirus reverse genetics and recombination

Coronaviruses employ the largest and most complex RNA genomes. However, numerous groups have developed reverse genetic approaches for manipulation of the coronavirus genome. The first coronavirus reverse genetics methods employed targeted RNA recombination and selection for a virus with increased fitness at elevated temperatures (Koetzner et al. 1992; Masters et al. 1994). Subsequent systems have been developed using bacterial artificial chromosomes (BAC) or vaccinia virus-driven expression systems. The initial development of a robust cDNA cassette-based system in TGEV has led to similar reverse genetic infectious clone systems being developed for MHV, SARS-CoV, IBV, MERS-CoV and several recently identified Bat coronaviruses. For MHV, the RNA genome is reverse-transcribed and subcloned as 7 cDNA fragments in cloning vectors (**Figure 1.8**). Once manipulated, the 7 cDNA fragments named A – G are restriction digested and ligated together to yield a full-length cDNA copy of the viral genome. The cDNA copy is then transcribed *in vitro* to yield infectious genomic RNA, which is subsequently electroporated into permissive cells for virus recovery and further analysis.

Coronaviruses are known to exhibit some of the highest known rates of recombination, which have been attributed to polymerase pausing and RNA template switching (Makino et al. 1986; Sawicki et al. 2007). Early studies evaluating recombination through mapping of complementation groups using temperature-sensitive lesions throughout the genome of MHV estimated a frequency of approximately 1% per 1.3 kb or nearly 25% of the entire genome (Baric et al. 1990; Fu et al. 1992). Sawicki *et*

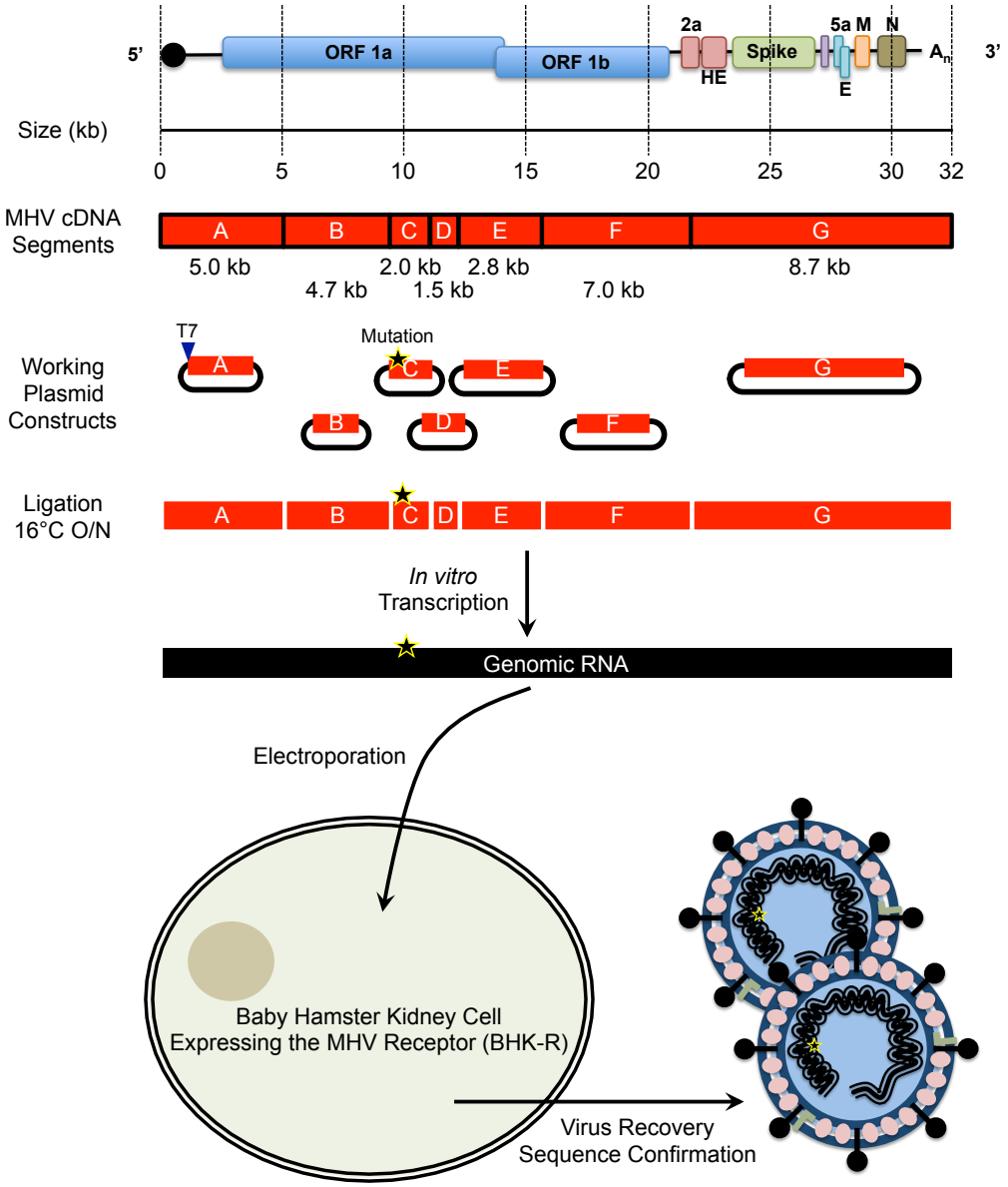


Fig. 1.8. MHV Infectious Clone, Reverse Genetics and Virus Recovery. The MHV genome (31.6 kb) has been reverse transcribed and subcloned into separated into 7 cDNA fragments in plasmid constructs. Alterations and mutations (star) to the genome are made in the plasmid constructs and sequence confirmed before the fragments are digested and ligated *in vitro* at 16°C overnight (O/N). After ligation, the cDNA is washed, precipitated and re-suspended for *in vitro* transcription. The resulting MHV infectious clone RNA genome is electroporated into baby hamster kidney cells expressing the MHV receptor (BHK-R) and these cells are combined with permissive murine delayed brain tumor (DBT-9) cells to recover virus. Recovered virus is sequenced to confirm the presence of the alterations and for any other alterations to the genome.

al (2005) and Stokes *et al* (2010) later demonstrated using complementation analysis that the coronavirus genome consists of at least five different complementation groups or cistrons (named cistrons 0, I, II, IV and VI) that contribute to unique viral phenotypes. Cistron I comprised of nsps 4 – 10 has been identified in several studies as a processed intermediate and has one or more related functions prior to maturation processing by nsp5 (Sawicki et al. 2005, Donaldson et al. 2007, Stokes et al. 2010, Xiao et al. 2012).

These data suggest that the defined regions of the coronavirus genome, which share a common function or phenotype, may have increased capacity for recombination. However, no studies have directly tested the potential or limitations for coronavirus recombination within or across genera. More specifically, it is unknown if the extensive interactions of nsp5 with 11 cleavage sites as well as with alleles in nsp3 and nsp10 constitute a barrier to genetic exchange between coronaviruses. If so, this would constitute limitations in over half of the coronavirus genome.

Summary

Coronavirus nsp5 protease is indispensible for virus replication, is a primary target of inhibitor design, and is responsible for catalyzing the maturation processing of nonstructural proteins 4 – 16. Upon beginning this project, very little was known about the regulatory mechanisms underlying nsp5 stability, specificity or function. Numerous crystal structures of the protease had been resolved and considerable focus had been placed on the generation of competitive inhibitors. However, there remains a dearth of

knowledge in regards to the connection between nsp5 structure and function in the context of a replicating virus.

I hypothesize that nsp5 activity is regulated by complex networks of interacting residues within nsp5 and between nsp5 and other elements of the replication complex. Furthermore, I hypothesize that these networks are highly adapted to the individual frameworks of the individual viral backgrounds due to the critical role of nsp5 in coronavirus replication, and have diverged in function between even closely-related coronaviruses. In the succeeding chapters, I will describe significant contributions I have made to understanding how nsp5 functions during virus replication, key structural and functional similarities and differences between the nsp5 protease of different coronaviruses, and how nsp5 may affect viral evolution and zoonotic emergence. Chapter II describes two temperature-sensitive (*ts*) alleles and four second-site suppressor alleles that are independent of known catalytic determinants and collectively regulate nsp5-mediated processing and impact viral replication. Biochemical data generated in collaboration with the Mesecar Lab demonstrates that one of the *ts* substitutions, V148A, destabilizes the protease structure in vitro and that the common second-site suppressor allele H134Y suppresses the *ts* phenotype. Chapter III explores the intra- and intermolecular regulation of coronavirus nsp5 through the design of chimeric nsp5 substitution mutants and introduction of MHV *ts* alleles into the backgrounds of closely-related nsp5 proteases. Chapters IV and V discuss the significance of these findings and the future directions of this work.

CHAPTER II

IDENTIFICATION OF TEMPERATURE-SENSITIVE AND SUPPRESSOR ALLELLES CRITICAL FOR LONG-DISTANCE COMMUNICATION AND CORONAVIRUS NSP5 PROTEASE ACTIVITY

Introduction

Positive-strand RNA viruses are responsible for a wide range of diseases in a broad range of vertebrate hosts including new and emerging viruses such as SARS-CoV, West Nile Virus, and Chikungunya virus. Rapid evolution, zoonotic movement, and potential lethality of positive-strand RNA virus infections demonstrate the necessity for developing novel strategies to prevent and treat current and new diseases. Emerging coronaviruses such as SARS-CoV and MERS-CoV pose a continual threat due to the lack of coronavirus vaccines, the lack of commercially available antiviral therapies and a dearth of information related to the mechanisms and limitations of viral emergence. An early step of Positive-strand RNA virus replication is the essential processing of translated polyproteins by host or virus-derived proteases. RNA virus proteases consequently have been primary targets for development of antiviral inhibitors, with most protease inhibitors targeted to active sites or substrate binding sites (Grum-Tokars et al., 2008; Lu et al., 2011; Nguyen et al., 2011; Poordad and Khungar, 2011). Yet, due to the potential for viral escape, it is critical to identify novel non-catalytic, non-substrate binding determinants of protease activity as putative targets for inhibition that are evade viral resistance.

Coronaviruses encode the largest known positive-strand RNA genomes ranging in size from 26 to 32 kb in length. Murine hepatitis virus strain A59 (MHV) is an

established model for study of coronavirus replication and pathogenesis. The 32 kb genome of MHV consists of seven genes, of which the replicase gene (22 kb) encodes 16 nonstructural proteins (nsp1-16) (Gorbalenya et al., 1989; Lee et al., 1991). Translation of the replicase genes yields two separate polyprotein products: polyprotein 1a (pp1a, nsps 1 – 11) or, via a ribosomal frameshift, pp1ab (nsps 1 – 16) (Brierley et al., 1989; Lee et al., 1991; Perlman and Netland, 2009). Maturation processing of pp1a and pp1ab is mediated by three virus-encoded proteases: two papain-like protease (PLP1 and PLP2) subunits of nsp4 that are responsible for cleavages of nsps 1 – 3, and an nsp5 protease, also known as 3CLpro or Mpro, that mediates maturation cleavages of nsps 4 to 16 and is required for virus replication (Perlman and Netland, 2009; Ziebuhr et al., 2000).

Coronavirus nsp5 is a cysteine protease that has been identified in all known coronaviruses, and resembles the nsp4 protease of distantly related arteriviruses (Barrette-Ng et al., 2002; Perlman and Netland, 2009; Zhao et al., 2008). The crystal structure of nsp5 has been solved for divergent coronaviruses from alpha- (α-CoV), beta- (β-CoV), and gammacoronaviruses (γ-CoV) including NL63, 229E, SARS-CoV, human HCoV-HKU1 and infectious bronchitis virus (IBV). Nsp5 protease crystal structures demonstrate conservation of tertiary structure despite numerous differences in primary sequences (Anand et al., 2002; Anand et al., 2003; Bacha et al., 2008; Xue et al., 2007; Xue et al., 2008; Zhao et al., 2008). The X-ray crystal structure of MHV nsp5 has yet to be determined; however the structure of the closely related HCoV-HKU1 nsp5 (84% sequence identity) has been resolved to 2.5 Å (Zhao et al., 2008). All determined nsp5 protease crystal structures exhibit a three-domain structure. Domains 1 and 2 (D1 and D2, respectively) form a chymotrypsin-like fold with the His41-Cys145 catalytic dyad and

substrate binding sites spanning the interface (Anand et al., 2002; Anand et al., 2003; Bacha et al., 2008; Lu and Denison, 1997; Xue et al., 2007; Xue et al., 2008). In contrast, domain 3 (D3) is structurally distinct among chymotrypsin-like enzymes, and also displays more divergence in both sequence and structural organization between coronaviruses. *In vitro* analysis indicates that domain 3 is potentially critical for stabilization of the chymotrypsin-like fold and may also be essential in mediating dimerization between nsp5 monomers (Anand et al., 2002; Lu and Denison, 1997; Shi et al., 2008; Shi and Song, 2006). Recent structural and biochemical studies have demonstrated that nsp5 dimerization is required for proteolytic activity *in vitro* (Chen et al., 2006; Chen et al., 2008; Shi et al., 2008).

Early efforts to design coronavirus inhibitors targeting nsp5 have focused on competitive substrate or peptidomimetic analogs. Numerous putative inhibitor substrates have been co-crystallized with nsp5 protease and studies evaluating substrate specificity have assisted with optimizing inhibitor constructs. However, to date, the best candidates for nsp5 inhibition have exhibited IC_{50} concentrations in the low μM range (Jacobs et al., 2013). The inability to develop inhibitors with cross-specificity and low inhibitory concentrations highlights the need to evaluate other non-active site, non-peptidomimetic substrates for coronavirus inhibition.

Over the last ten years, several amino acid residues have been identified that may regulate nsp5 activity but which are distinct from active site cavity, substrate-binding pocket, or dimerization interface determinants. Alanine substitution of SARS-CoV nsp5 Ser147, a conserved serine residue in MHV and HKU1, disrupts dimerization and impairs nsp5 proteolytic activity, despite no apparent direct connectivity to the dimerization

interface (Barrila et al., 2006). Other studies have shown that nsp5 activity may be altered by changes in replicase nonstructural proteins nsp3 and nsp10 (Donaldson et al., 2007; Stokes et al., 2010). Our laboratory experimentally confirmed the first temperature-sensitive mutation (*ts*V148A) in MHV nsp5 (Sparks et al., 2008). This mutation results in impairment of viral growth and nsp5 activity at the non-permissive temperature of 40ºC (Sparks et al., 2008). Growth of the *ts*V148A mutant at 37ºC led to the emergence of two independent second-site suppressor mutations (S133N and H134Y), which suppressed the *ts*V148A phenotype. While the V148 residue is adjacent to the catalytic C145 residue, neither S133N nor H134Y have predicted direct interactions with catalytic or substrate residues or identified pathways for propagation of structural changes. In 2006, Sawicki *et al* identified another putative *ts* allele in nsp5 (F219L) at the base of unique domain 3, and distant from the dimerization interface, active site cavity, and substrate binding regions (Sawicki et al., 2005). Collectively, these findings support the hypothesis that residues independent from catalytic and substrate-binding sites are important for regulating nsp5 protease activity. However, long-distance communication mechanisms between these determinants and their role in regulation of protease activity remain unclear.

In this chapter, I test the hypothesis that MHV nsp5 is governed by long-distance communication that spans the protease structure across all three domains and is not limited to determinants within the active site or dimerization interface. I describe the role of MHV nsp5 domain 2 residues S133 and H134 and putative *ts* allele F219 in domain are described. A previously unknown MHV nsp5 *ts* mutation in domain 2 (*ts*S133A) is identified and the *ts* phenotype of F219L is experimentally confirmed. Both of these

mutations lead to profound replication and protein-processing defects at non-permissive temperatures. During passage at non-permissive temperatures, a series of second-site suppressor mutations emerged that were able to suppress the *ts* phenotypes of *ts*S133A and *ts*F219L. These alleles were largely distant from the cognate *ts* alleles, the nsp5 active site cavity, and the dimerization interface. A single non-synonymous mutation resulting in a H134Y substitution suppressed the *ts* phenotype of all three independent *ts* alleles (S133A, V148A, and F219L) in domain 2 and 3. The *ts* mutations reduced but did not abolish nsp5 protease activity during virus infection following a shift to non-permissive temperature, while individual and combined suppressor mutations restored nsp5 activity to an extent that directly correlated with increased replication.

In collaboration with the Mesecar lab at Purdue University, MHV nsp5 proteases containing the *ts* mutation V148A with and without the second-site suppressor mutation H134Y were purified and evaluated for thermal stability and secondary structure. These studies suggest that the temperature-sensitive lesions likely destabilize the protease structure at elevated temperatures and that the H134Y mutation was selected as it hyperstabilizes domain 2. These data collectively demonstrate the presence of multiple interconnected long-distance communication nodes in nsp5 and suggest novel mechanisms of regulation of nsp5 activity during CoV replication. These findings establish the experimental basis further characterization in other coronavirus backgrounds described in Chapter III.

S133A is a novel temperature-sensitive mutant of MHV nsp5

Initial studies by Sparks et al described two second-site suppressor mutations that suppressed the temperature-sensitive phenotype of MHV *ts*V148A; H134Y and S133N (Sparks et al., 2008). The residues H134 and S133 are greater than 20Å from V148, from the active site cavity, and from the S1 substrate-binding site in the crystal structure of HKU1 nsp5. The mechanism by which these residues suppress the cognate *ts*V148A virus remains unclear. Furthermore, substitution of the H134Y or S133N alleles alone in the isogenic wild-type (WT) MHV background had little to no effect on virus replication at 37°C. Therefore, I sought to determine the roles of these residues in nsp5 activity. Alanine substitutions were engineered at the S133 and H134 residues in MHV nsp5. S133A and H134A viral mutants were recovered at 30°C and initial sequencing from the initial passage (P1) 30°C virus stocks confirmed the presence of mutations S133A or H134Y, with no other sequence changes in nsp5. These viruses were compared to recombinant WT MHV and *ts*V148A for virus titers at 30°C and 40°C, and the efficiency of plating (EOP) was calculated by plaque assay (titer 40°C / titer 30°C) (**Table 2.1**). WT MHV exhibited an EOP of 3.3, a 3.3-fold increase in visible plaques at 40°C compared to 30°C. The previously described *ts*V148A virus exhibited an EOP of 3×10^{-5}, confirming the *ts* phenotype reported by Sawicki et al (Sparks et al., 2008). The engineered H134A mutant had an EOP of 10^{-1}, which is similar to previously reported engineered H134Y mutation in the WT MHV background (Sparks et al., 2008). However, the S133A mutant had an EOP of 2.7×10^{-5}, consistent with an independent *ts* phenotype. Surprisingly, these data demonstrate that reversion of *ts*V148A had resulted in identification of an

allele at S133 that with an Asn substitution could suppress the V148A *ts* phenotype (S133N), yet with a conservative Ala substitution resulted in new *ts* phenotype (S133A). The H134A virus was not further examined and *ts*S133A was studied in subsequent experiments.

Identification of two different suppressor mutants of *ts*S133A

To test for revertants or suppressors of *ts*S133A, DBT-9 cells were initially infected with *ts*S133A at 37°C, followed by a subsequent shift to 40°C. Virus recovered at the non-permissive temperature was used for a plaque assay at 40°C. Ten virus plaques were isolated and expanded at 40°C in T25 flasks. Reverse transcription (RT) PCR sequencing of the nsp5 coding region from all ten clones confirmed the retention of the original S133A ($_{10605}$AGC to GCC) engineered mutation, as well as two separate patterns of second-site non-synonymous mutations: H134Y ($_{10608}$CAT to $_{10608}$TAT), and T129M / H134Y ($_{10593}$ACG to $_{10593}$ATG; $_{10608}$CAT to $_{10608}$TAT). The S133A had not reverted in all sequenced clones and no other mutations were detected in the nsp5 coding sequence. The S133A / H134Y plaque isolate had an EOP of 0.5 and the S133A / T129M / H134Y plaque isolate had an EOP of 2, suggesting almost complete suppression of the S133A *ts* phenotype (**Table 2.1**).

To evaluate the contribution of the second-site mutations to suppression of the *ts* phenotype, we engineered the different combinations of identified mutations into the isogenic MHV background. Since both suppressor mutants contained H134Y and there were subtle differences in the EOP values, we tested the independent contribution of the

T129M substitution by introducing the S133A and T129M mutations in the absence of the H134Y mutation. All of the engineered recombinant viruses were readily recovered at 30°C, and sequencing confirmed that the engineered changes were present and no other mutations had arisen in nsp5. The EOP of the recombinant S133A/H134Y and S133A/T129M/H134Y mutants were nearly identical to the cognate biologically recovered mutant, demonstrating that the identified changes in nsp5 were necessary and sufficient for the phenotypic reversion (**Table 2.1)**. However, the engineered recombinant S133A/T129M mutant virus exhibited an EOP of 1×10^{-3}, approximately 33-fold greater than *ts*S133A but still significantly *ts* compared to either the S133A/H134Y or S133A/T129M/H134Y viruses. These data demonstrated that H134Y was sufficient for suppression of the *ts*S133A, T129M was unable to suppress the *ts* phenotype by itself, and the combination of T129M and H134Y was additive or synergistic, suggesting two distinct mutations that could have arisen sequentially or concurrently in the same or different genomes.

Table 2.1. Virus titers and EOP for WT, *ts*S133A and suppressor viruses

Wild-type (WT), *ts*S133A and Suppressor Viruses			
Virus	***Virus Titer (PFU/mL) at:***		***EOP***
	30°C	***40°C***	
WT MHV	9×10^{7}	3×10^{8}	3×10^{0}
*ts*S133A	6×10^{6}	2×10^{2}	3×10^{-5}
S133A/H134Y	2×10^{7}	1×10^{7}	5×10^{-1}
S133A/T129M/H134Y	3×10^{7}	4×10^{7}	2×10^{0}
S133A/T129M	4×10^{7}	5×10^{4}	1×10^{-3}
H134A	5×10^{8}	4×10^{8}	7×10^{-1}

Replication kinetics of *ts*S133A and Suppressor Mutants

To evaluate the kinetics of viral infection and predict overall growth fitness, I compared the replication of WT and mutant viruses at permissive (30°C) and non-permissive temperatures (40°C) during single cycle infections. Based on earlier single-cycle growth studies of MHV replication, replicate plates of DBT-9 cells were infected with WT and recombinant mutant viruses at 30°C and a multiplicity of infection (MOI) of 1 PFU/cell (Gadlage and Denison, 2010; Gadlage et al., 2010). At 6 h p.i. one replicate for each virus was transferred to the non-permissive temperature of 40ºC and one remained at 30ºC, with supernatant samples obtained from 0 to 30 h p.i. during regular intervals for determination of viral titers. WT virus demonstrated onset of exponential growth at 30°C between 10 and 12 h p.i. and achieved peak virus titers of ~10^8 PFU/cell by 30 h p.i. (**Figure 2.1**), consistent with previous studies (Sparks et al., 2008). In contrast, *ts*S133A and all three recombinant suppressor mutant viruses demonstrated identical growth curves, with an approximate 10-fold reduction in titers compared to WT MHV from 12-24 h p.i., but achieving titers identical to WT by 30 h p.i.

When replicating with a shift from 30°C to 40°C at 6 h p.i., WT exhibited onset of exponential growth within 2 h after shift and reached peak titers between 12 and 16 h p.i., albeit at lower titers than at 30°C due to rapid destruction of the monolayer. *ts*S133A virus demonstrated profoundly impaired growth for 10 h after temperature shift. Both recombinant S133A/H134Y and S133A/T129M/H134Y mutant viruses exhibited onset of exponential growth and peak titers similar to WT following the shift to the non-

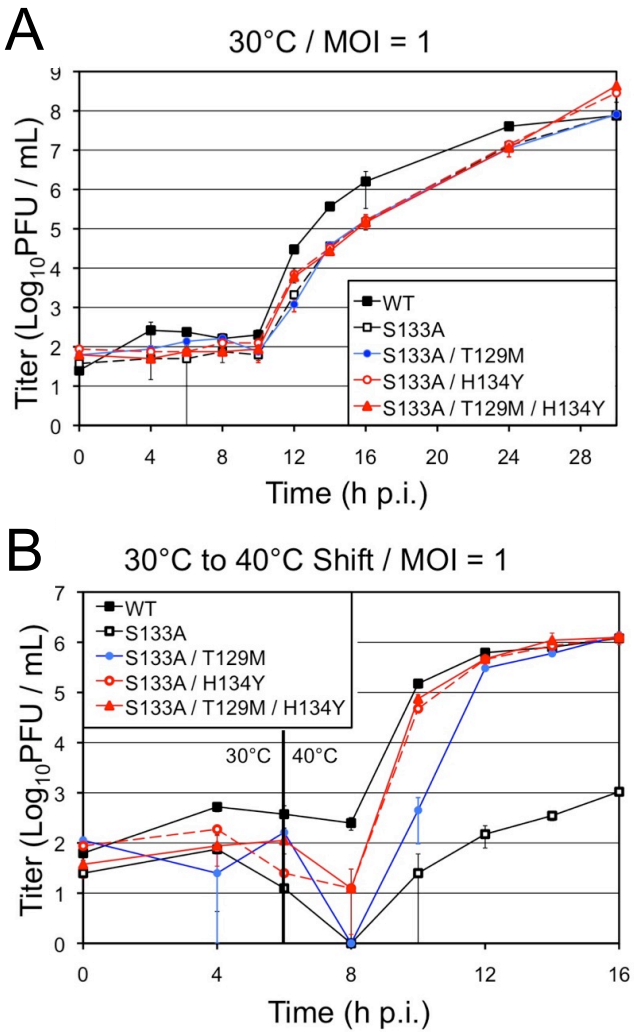


Fig. 2.1. Analysis of replication of tsS133A and suppressor mutants. A and B) Growth analysis of the WT MHV, recombinant tsS133A, suppressor mutants (S133A/H134Y and (S133A/T129M/H134Y), and engineered mutant (S133A/T129M) grown at 30°C (A) or grown at 30°C and then shifted to 40°C at 6 h p.i. on DBT-9 cells (B). The virus titers were determined by plaque assay on DBT-9 cells at 30°C.

permissive temperature, consistent with the EOP data. However, the S133A/T129M virus showed a significant delay in exponential growth before achieving titers similar to WT by 6 h after temperature shift. Collectively, these results indicate that the S133A mutation confers a subtle replication defect at 30°C that is not further impaired nor complemented by the suppressor mutations, and that H134Y alone or in combination with T129M is sufficient for suppression of the *ts*S133A growth phenotype. The S133A/T129M virus, while demonstrating improved growth over *ts*S133A, remained impaired compared to mutants containing H134Y, a result also consistent with the EOP data, and supporting either that T129M arose first to be superseded by emergence of H134Y or arose second by conferring a subtle growth advantage of the combination over H134Y alone.

Recovery of recombinant *ts*F219L

Experimentation with *ts*S133A and previously described *ts*V148A demonstrated that the H134Y substitution was able to suppress two distinct and independently derived nsp5 *ts* alleles. These data indicate an important role for intra- or intermolecular communication involved in regulation of nsp5 activity. Yet, both of these *ts* alleles and their suppressor alleles are located in domain 2, and thus provide no insight into potential long-distance communication between domains. I next sought to determine if the H134Y mutation would emerge as a suppressor for another putative *ts* allele in domain 3. The $_{10864}$UUU to CUU mutation resulting in the F219L substitution was predicted as a *ts* allele by Sawicki *et al*. using partial genome sequencing and reversion analysis of the *ts* mutant

Alb *ts*16 (Sawicki et al., 2005). However, this was not confirmed as the sole mutation by complete genome sequencing or by reverse genetic analysis. In addition, primary reversion rapidly occurs in biological mutants with a single nucleotide polymorphism, therefore possible second-site suppressors would not be identified. We introduced the F219L codon change as a two-nucleotide change ($_{10864}$UUU to CUA) in the isogenic MHV clone, which would require a two-nucleotide alteration for reversion to Phe219 (UUU or UUC). The engineered recombinant F219L mutant was obtained at 30°C, and complete genome sequencing confirmed the $_{10864}$UUU to CUA mutations as the only changes in the genome. The recombinant F219L mutant had an EOP of 3×10^{-5}, confirming that the F219L substitution alone was sufficient to confer a *ts* phenotype (**Table 2.2**). The *ts*F219L mutant exhibited titers and plaque morphology indistinguishable from WT MHV at 30°C (data not shown).

Identification of second-site suppressor mutations of tsF219L

To test for *ts*F219L phenotypic revertants or suppressor mutations, DBT-9 cells were infected with recombinant *ts*F219L mutant virus at 40°C. Unfortunately, no CPE or productive infection occurred at 40°C despite multiple attempts and prolonged incubation. Accordingly, we initiated infection at 30°C for 6 h followed by shift to 37°C for 24 h. This stock was then passaged at 37°C with shift to 40°C, followed by passage and selection of 10 plaques at 40°C. Sequencing of the nsp5 coding region of ten plaque clones confirmed retention of the engineered $_{10864}$CUA (F219L) codon. However all 10 isolated plaque cloned viruses demonstrated one of two distinct patterns of second-site

mutations in nsp5 in the presence of *ts*F219L: H134Y ($_{10608}$CAT to $_{10608}$TAT)/H270 duplication (H270HH) ($_{11016}$CAT to $_{11016}$CATCAT) with an EOP of 1.2; or H134Y (CAT to TAT)/E285V ($_{11061}$GAA to $_{11061}$GTT) with an EOP of 2.9.

To evaluate the contribution of the H134Y, E285V and H270HH changes to suppression of *ts*F219L, the identified mutations were introduced with and without F219L, alone or in the combination seen in the recovered viruses. Furthermore, although no F219L/E285V/H270HH mutant was identified among the sequenced plaques, we also engineered this combination to test for the capacity of this combination to suppress the *ts* phenotype**.** In total, nine different virus genomes were engineered: F219L/E285V; F219L/H134Y; F219L/H270HH; F219L/H134Y/E285V; F219L/H134Y/H270HH; F219L/H270HH/E285V; H134Y/E285V; H134Y/H270HH; and H270HH/E285V. All viruses were recovered at 30°C with the engineered mutations detected and confirmed by sequencing across nsp5. The EOP was determined for each of the viruses by plaque assay on DBT cells at 30°C and 40°C (**Table 2.2**). The second-site substitution combinations identified by reversion analysis were sufficient to suppress the F219L *ts* phenotype: H134Y/E285V, EOP = 2.9; and H134Y/H270HH; EOP = 1.2. When the second-site suppressor alleles were tested alone with F219L, the results showed that the individual substitutions either minimally or partially suppressed *ts*F219L: H134Y, EOP = 2×10^{-5}; E285V, EOP = 1×10^{-5}; and H270HH, EOP = 2×10^{-3}. When second-site substitutions were introduced in the WT background, either alone or in combination, there was no effect on EOP, suggesting that the changes were not responsible for any replication defects in the presence of F219L. The non-biologically derived combination of F219L/E285V/H270HH also completely suppressed the *ts*F219L phenotype (F219L /

E285V / H270HH, EOP = 1.0). These EOP results demonstrated that the biologically identified second-site substitution combinations were both necessary and sufficient to suppress the *ts*F219L phenotype. Further, the results showed that H134Y did emerge as a suppressor allele for domain 3 *ts*F219L, but that in contrast to *ts*V148A and *ts*S133A, suppression of *ts*F219L required at least one other substitution in domain 3 in combination with H134Y.

Table 2.2. Virus titers and EOP for WT, *ts*F219L and suppressor viruses

Wild-type (WT), *ts*F219L and Suppressor Viruses			
Virus	***Virus Titer (PFU/mL) at:***		***EOP***
	30°C	***40°C***	
WT MHV	9×10^7	3×10^8	3×10^0
*ts*F219L	1×10^8	3×10^3	3×10^{-5}
F219L/H134Y/E285V	1×10^7	4×10^7	3×10^0
F219L/H134Y/H270HH	2×10^7	2×10^4	1×10^0
F219L/H270HH/E285V	3×10^7	3×10^7	1×10^0
F219L/E285V	1×10^7	2×10^2	2×10^{-5}
F219L/H134Y	1×10^8	1×10^3	1×10^{-5}
F219L/H270HH	8×10^6	1×10^4	2×10^{-3}

Replication kinetics of *ts*F219L and suppressor mutants

The *ts*F219L and recombinant suppressor viruses were grown in DBT cells at 30°C, or with a shift from 30°C to 40°C at 6 h p.i. (**Figure 2.2**). At 30°C, all of the recombinant mutant viruses showed replication kinetics and virus yield indistinguishable from WT MHV. Following shift to 40°C at 6 h p.i. however, *ts*F219L showed no further replication for 10 h. In contrast, the mutant viruses containing any two of the H134Y,

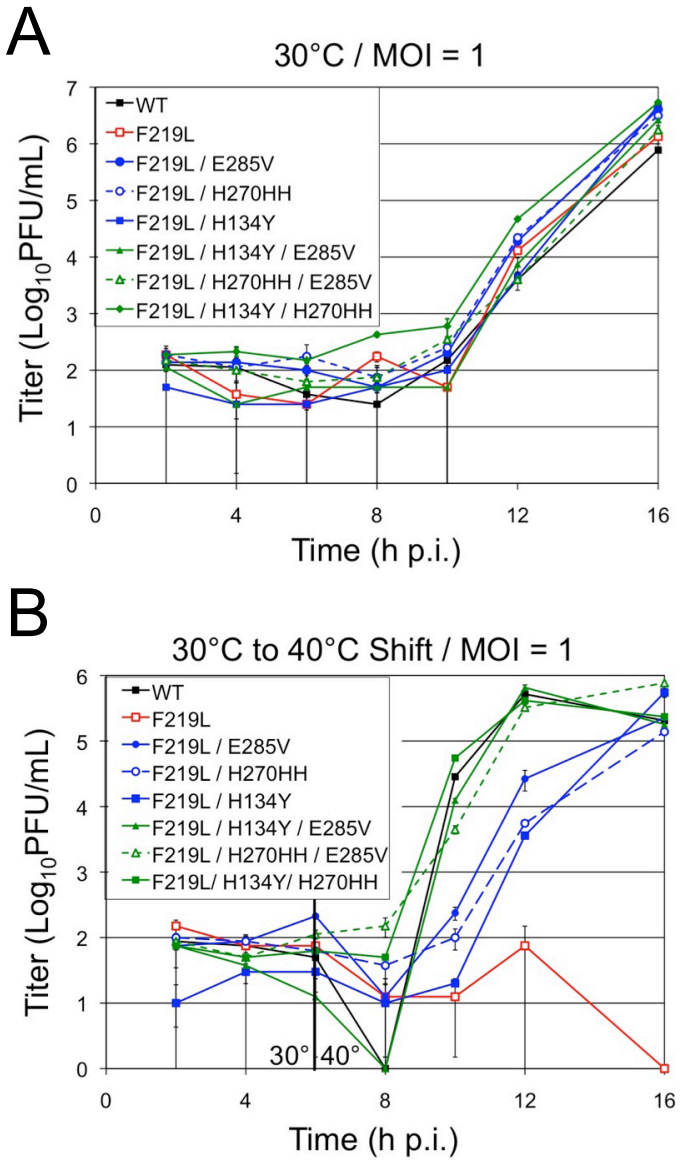


Fig. 2.2. Analysis of *ts*F219L and suppressor mutant EOP and replication. A and B) Growth analysis of WT MHV, recombinant tsF219L, suppressor mutants (F219L/H134Y/E285V and F219L/H134Y/H270HH), and engineered mutants (F219L/H270HH/E285V, F219L/E285V, F219L/H270HH, and F219L/H134Y) grown at 30°C (A) or grown at 30°C and then shifted to 40°C at 6 h p.i. (B) on DBT-9 cells. The virus titers were determined by plaque assay on DBT-9 cells at 30°C.

E285V and H270HH substitutions demonstrated growth after shift to 40°C similar to WT. In contrast, all single suppressor residues expressed with F219L demonstrated a 4 h lag before exponential growth compared to the double mutants, but ultimately achieved similar peak titers. The titers at 10 h p.i. were consistent with the EOP data (10^{-3} to 10^{-5} compared to WT) and overall demonstrated that the individual mutations were capable of improved viral replication compared to the *ts*F219L virus yet still were *ts* compared to WT. Thus, in contrast to *ts*V148A and *ts*S133A, suppression of *ts*F219L appears to require a combination of at least two second-site mutations. These data may justify the tight *ts*F219L phenotype at 40°C, as well as for the necessity of sequential passage of the *ts*F219 mutant virus at 30°, 37° and 40°C to recover phenotypic revertants. In addition, these results identify H134Y as a second-site suppressor for a third *ts* allele in nsp5, this one in a domain 3. Overall, the results show that cooperation of H134, E285, and H270 in nsp5 is necessary for efficient virus replication, as well confirming communication between nsp5 domains 2 and 3.

*ts*S133A and *ts*F219L exhibit temperature-sensitive impaired activity

To directly compare the nsp5 protease activity of *ts*V148A, *ts*S133A, *ts*F219L and second-site suppressor mutants, DBT cells were infected at an MOI of 5 PFU/cell and incubated at 30°C. At 6 h p.i., replicate monolayers were maintained at 30°C or transferred to 40°C and infected cells were radiolabeled with [^{35}S]Met-Cys. Lysates from infected, radiolabeled cells were immunoprecipitated with antisera specific for nsp2, nsp5, and nsp8, to test for processing of nsp2 by PLP1 and of nsp5 and nsp8 by nsp5.

Immunoprecipitation with antibodies specific for nsp2 of all lysates from WT and mutant-infected cells with labeling at 30°C and 40°C resulted in detection of mature processed nsp2, demonstrating that at both permissive and non-permissive temperatures there was similar translation of pp1a (nsp1 – 11) and normal PLP1 activity. Recently, Stokes et al. reported that a *ts* mutation in nsp3 resulted in a significant decrease in nsp5-mediated processing (Stokes et al., 2010). In our study, I detected the nsp4-10 precursor polyprotein (150 kDa), which demonstrates that PLP2 is functional and is not inhibited by the nsp5 mutations (**Figure 2.3**). The presence of mature nsp2 and the p150 bands at non-permissive temperature indicates that both PLP1 and PLP2 domains of nsp3 are active and are not affected by the nsp5 mutations. Although I did not test the processing of nsp3 directly, detection of both p150 and mature nsp2 is consistent with normal processing of N- and C-termini of nsp3. Immunoprecipitation of cells infected with WT MHV by nsp5-specific antibodies at both 30°C and 40°C resulted in detection of mature processed nsp5, as well as co-immunoprecipitation of nsp8, and two distinct small protein bands at 10 and 12 kDa, consistent with the predicted migration of nsp7 and nsp9. Immunoprecipitation with nsp8-specific antibodies detected nsp8 as well as probable co-precipitation of nsp5 and the 10 and 12 kDa proteins. These results show that expression and processing of pp1a proteins nsp5, nsp7, nsp8 and nsp9 are accelerated at 40°C in cells infected with WT MHV.

The temperature-sensitive viruses *ts*S133A and *ts*F219L, as well as the previously described *ts*V148A, exhibited profoundly impaired processing of nsp5 and nsp8 at 40°C when compared to 30°C, indicating a specific defect in processing by the nsp5 protease. This was consistent with decreased detection of the reciprocal co-immunoprecipitating

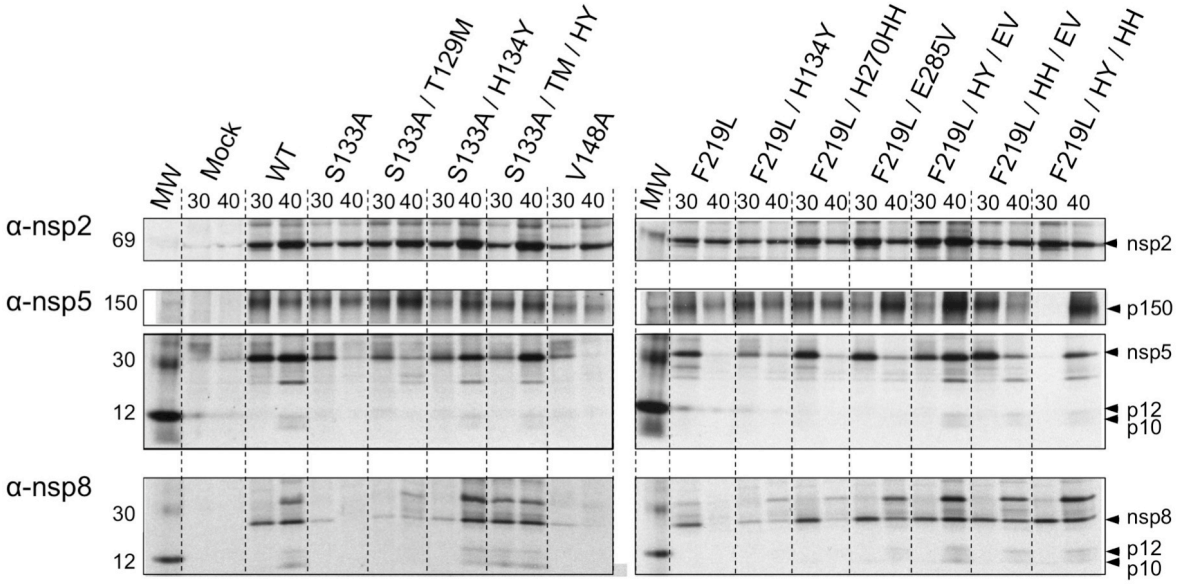


Fig. 2.3. Proteolytic processing of WT, tsS133A, tsF219L, and suppressor viruses. DBT-9 cells were 35S radiolabeled during viral infection or mock infection. Cellular lysates were harvested from cells infected withWTMHV,tsS133A mutants, tsF219L mutants, and previously described tsV148A and from mock-infected cells. Labeled proteins were immunoprecipitated using antiserum specific for nsp2, nsp5, and nsp8. The temperature during virus infection is indicated above the lanes (30°C [30] or a temperature shift from 30° to 40°C at 6 h p.i. [40]). One hundred microliters of lysate was used for all immunoprecipitations. The positions of putative viral proteins are shown to the right of the gels based upon the predicted size, and identified bands are labeled. The positions of molecular weight standards (MW) are shown to the left of the gels, and sizes are shown in kilodaltons. Protein expression profiles were resolved by SDS-PAGE and visualized by autoradiography. α-nsp2, anti-nsp2 antibody.

protein (nsp8 or nsp5) as well as decreased detection of the 10 and 12 kDa proteins. Although the detection of nsp5-processed proteins was profoundly decreased, I was unable from multiple replicate experiments (>5) to demonstrate complete loss of nsp5 activity. The results suggest that the S133A, F219L and V148A mutations do not directly affect the catalytic or substrate binding functions of nsp5, but rather modify protease activity in other ways. Alternatively, it is possible that the residual processing might be the result of nsp5 expressed and folded into active forms or complexes prior to the temperature shift, and thus still retain residual activity.

Viruses carrying biological and engineered suppressor mutations demonstrated restoration of processing by nsp5 that directly correlated with the degree of recovery of EOP and virus growth. Single second-site mutants S133A/T129M, F219L/H134Y, F219L/H270HH, and F219L/E285V showed an increase in detectable processed nsp5 and nsp8 only, while the double second-site suppressors restored WT-like patterns of processed proteins. Collectively, the results show a direct correlation of detection of proteins processed by nsp5 (nsp5 and nsp8) and the extent of restored growth fitness in culture and EOP analysis in plaque assay. Further, these findings indicate that the impairment in growth at non-permissive temperatures is not due to complete inactivation of nsp5 protease activity.

Analysis of *ts* and suppressor alleles in coronavirus nsp5 structures

To evaluate the structural connection between the *ts* and second-site suppressor mutations, the distances between the combinations of *ts* and suppressor residues were

determined by analyzing the crystal structure of the nsp5 protease of a closely related human coronavirus, HKU1 (**Figure 2.4**). The structure of MHV nsp5 has not been determined despite many efforts. However, MHV and HKU1 nsp5 proteases exhibit 84% sequence identity, and share all of the same amino acids at the residue positions reported in this study with the exceptions of H134 and H270 (Y134 and Y270 in HKU1). All of the second-site suppressor residue positions in HKU1 nsp5 were greater than 10 Å distance from the *ts* residues in the monomeric structure with the sole exception that the H134 residue is 3.8 Å from the juxtaposed S133. Measurement of the distance between residues in different monomers of the dimeric structure of both SARS-CoV and HKU1 nsp5 demonstrated that no two residues from this study were closer than 15.0 Å. Measurement of the distance between the *ts* and second-site suppressor mutants and the catalytic dyad residues, H41 and C145, showed that only V148A was within 10 Å of either residue. Collectively, these data demonstrate that the relationship between the *ts* and suppressor mutations cannot be explained by direct interactions between residues, and that nsp5 dimerization does not provide direct inter-monomer associations between the residues identified in this study.

Modeling of the S133A and F219L mutations on the structure of HKU1 nsp5 failed to predict any clear pathways of side-chain remodeling or perturbation between the *ts* residues and the protease active site (data not shown). In contrast, analysis of residue conservation using an alignment of 130 non-redundant coronavirus nsp5 amino acid sequences identified a series of 100% identical residues that span the regions of nsp5 between each of the *ts* residues (S133, V148, and F219) and the common second-site suppressor residue (H134) (**Figure 2.5**). These findings indicate that the structural and

Distances in Monomer

Distance across Dimer

	H41	C145	T129	S133	H134	V148	F219	H270	E285
H41	55.3	9.5	24.2	21.6	19.4	15.8	46.5	40.9	34.6
C145	46.2	37.2	18.4	19.6	17.1	8.6	42.5	36.9	28.1
T129	38.3	28.9	22.4	10.8	9.9	13.8	19.4	17.5	7.9
S133	49.3	40.0	30.7	40.0	3.8	20.1	25.2	15.0	14.3
H134	47.6	38.2	30.5	39.2	39.5	17.1	28.8	21.6	20.9
V148	40.1	30.9	25.2	36.1	34.6	25.4	34.8	34.1	24.2
F219	46.3	38.9	24.2	27.2	31.6	27.8	31.1	13.6	12.2
H270	40.9	32.8	30.6	35.0	38.1	40.8	26.9	24.3	14.8
E285	34.6	28.4	16.2	22.1	24.8	26.2	21.7	22.2	15.0

Fig. 2.4. Distances between MHV temperature-sensitive and suppressor alleles. Distances between the *ts*, suppressor mutant, and catalytic dyad residues (H41 and C145) are shown in angstroms (Å) within the monomeric structure (top right half of table) and across the dimeric structure (bottom left half of the table). Calculations were performed by measurement between alpha carbons of each amino acid residue position in the structure of HKU1 nsp5 (Rao et al. 2008).

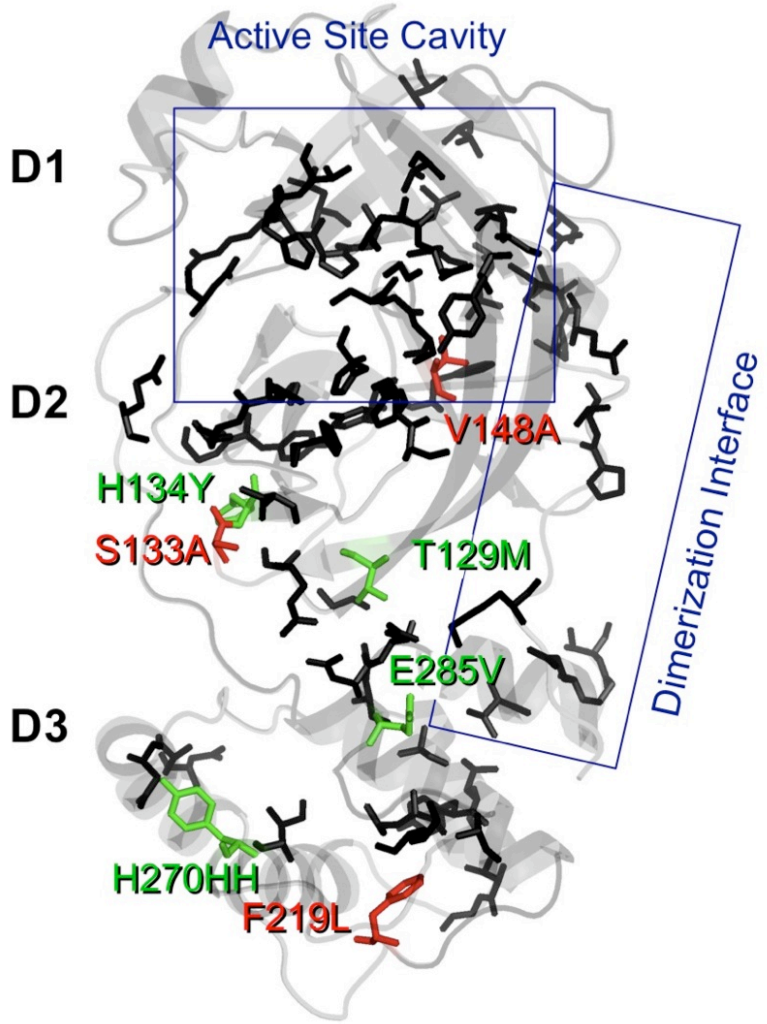


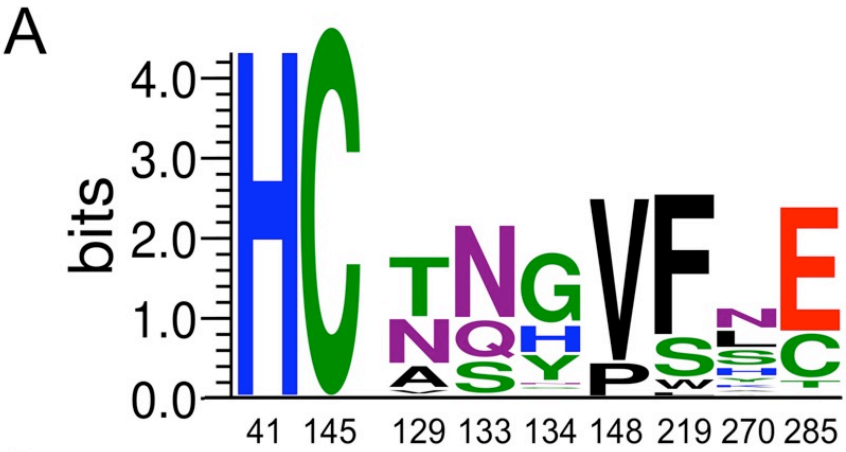
Fig. 2.5. Structural relationship of *ts* and suppressor alleles to conserved MHV alleles. A conservation map of 100% identical nsp5 residues (black) across 130 non-redundant CoV nsp5 sequences is shown on the HKU1 nsp5 protease monomer structure (Rao et al 2008). The location of identified *ts* alleles (red) and suppressor alleles (green) are indicated.

functional perturbations on nsp5 protease of the *ts* mutations may span long distances across the protease structure through yet to be identified cooperative interactions.

Residue conservation of *ts* and suppressor alleles is clade-specific

To evaluate the conservation of the *ts* and suppressor alleles, a sequence logo (Crooks et al., 2004) was generated using a non-redundant alignment of 130 coronavirus nsp5 amino acid sequences, and the *ts* and suppressor alleles for the 17 coronavirus species recognized by the International Committee of Taxonomy of Viruses (ICTV) were analyzed (**Figure 2.6**). From two to eight different residues occupy each position across available CoV sequences. Surprisingly, the H134Y common second-site suppressor mutation selected for a tyrosine that is already present in several betacoronaviruses. Further, the S133N second-site suppressor first reported by Sparks *et al.* is common as an Asn in many coronaviruses (Sparks et al., 2008). Conservation of distinct *ts* and suppressor alleles within the three genera suggest that there may be select combinations of alleles that are necessary for nsp5 activity. The role of these determinants remains unknown, however these data suggest that non-conserved alleles located in nsp5 domains 2 and 3, as well as potentially unidentified determinants in domain 1 are critical for regulating nsp5 structure, specificity, stability or function. The presence of similar alleles in closely-related human coronaviruses HKU1 and OC43 may indicate similar functional regulation and likely a strong evolutionary link between the three betacoronaviruses. These findings further suggest that alterations in structure attributed to these residues could have analogous combinations in other CoVs.

A



B

MHV Residue Position:	129	133	134	148	219	270	285
Genus: Alphacoronavirus							
α-CoV1 (TGEV)	N	Q	G	V	F	N	C
HCoV 229E	N	N	W	P	W	N	N
HCoV NL63	N	N	F	P	W	H	C
PEDV	N	N	Y	P	W	H	T
BtCoV HKU2	N	T	H	P	W	A	T
BtCoV 512/05	N	N	Y	P	W	H	C
BtCoV 1A	I	N	K	P	F	A	T
BtCoV HKU8	T	N	Y	P	F	A	T
Genus: Betacoronavirus							
MHV-A59	T	S	H	V	F	H	E
β-CoV1 (HCoV OC43)	T	S	Y	V	F	K	E
HCoV HKU1	T	S	Y	V	F	Y	E
BtCoV HKU9	T	N	G	P	F	L	E
BtCoV HKU4	N	N	S	V	F	H	E
BtCoV HKU5	N	Q	H	V	F	H	E
SARS-CoV Urbani	A	N	H	V	F	L	E
Genus: Gammacoronavirus							
Avian Coronavirus (IBV)	T	N	G	V	S	S	E
Beluga Whale CoV SW1	T	N	G	V	F	A	E

Fig. 2.6. Coronavirus protein sequence conservation of *ts* and suppressor alleles. A) A sequence logo of conservation of *ts* and suppressor mutations across an alignment of 130 non-redundant CoV nsp5 sequences was generated using WebLogo (13). The height of each letter corresponds to the relative conservation of that amino acid at the position and the height of the column corresponds to the sequence conservation at the position. The residue numbers are relative to the MHV amino acid positions. **B)** The *ts* and suppressor residues of coronavirus species by CoV genus are shown for each MHV residue position.

Expression and purification of codon-optimized WT MHV nsp5

In order to identify the functional role of the MHV *ts* and second-site mutations, MHV nsp5 was expressed and purified for biochemical analysis. Wild-type MHV nsp5 sequence was codon-optimized (3CLpro) for bacterial expression in *E. coli* and cloned into a pET-15b expression plasmid construct (**Figure 2.7**). The N-terminal end of the nsp5 sequence is flanked by a tobacco etch virus (TEV) cleavage site and a His tag for efficient purification and subsequent cleavage to yield mature native nsp5. Initial attempts at expression and purification of MHV nsp5 demonstrated that the protease retained proteolytic activity and autoprocessed the TEV cleavage site preventing Ni-affinity chromatography for purification. The TEV cleavage site (NLYFQ†G/S) resembles the nsp5 consensus site (SATLQ†SG) at the P2 – P1' and may explain the inadvertent auto-processing that occurred. Efforts were then directed to optimize a native nsp5 purification approach based upon studies described by Grum-Tokars et al describing substantial differences in nsp5 protease activity when tagged either amino- or carboxy-terminally.

WT MHV nsp5 was overexpressed in *E. coli* BL21 DE3 cells following induction with IPTG at 37°C. The final purification procedure devised consists of a two-step process. First, the bacterial lysate is flowed through a hydrophobic interaction chromatography column (phenyl sepharose column) followed by elution from a DEAE anion exchange column. Upon purification, the two clear distinct bands are visualized in a Coomassie gel of approximately the same sizes of the monomeric (~30 kDa) and dimeric (~60 kDa) nsp5 proteases. The purity of the protease is approximately 95% after

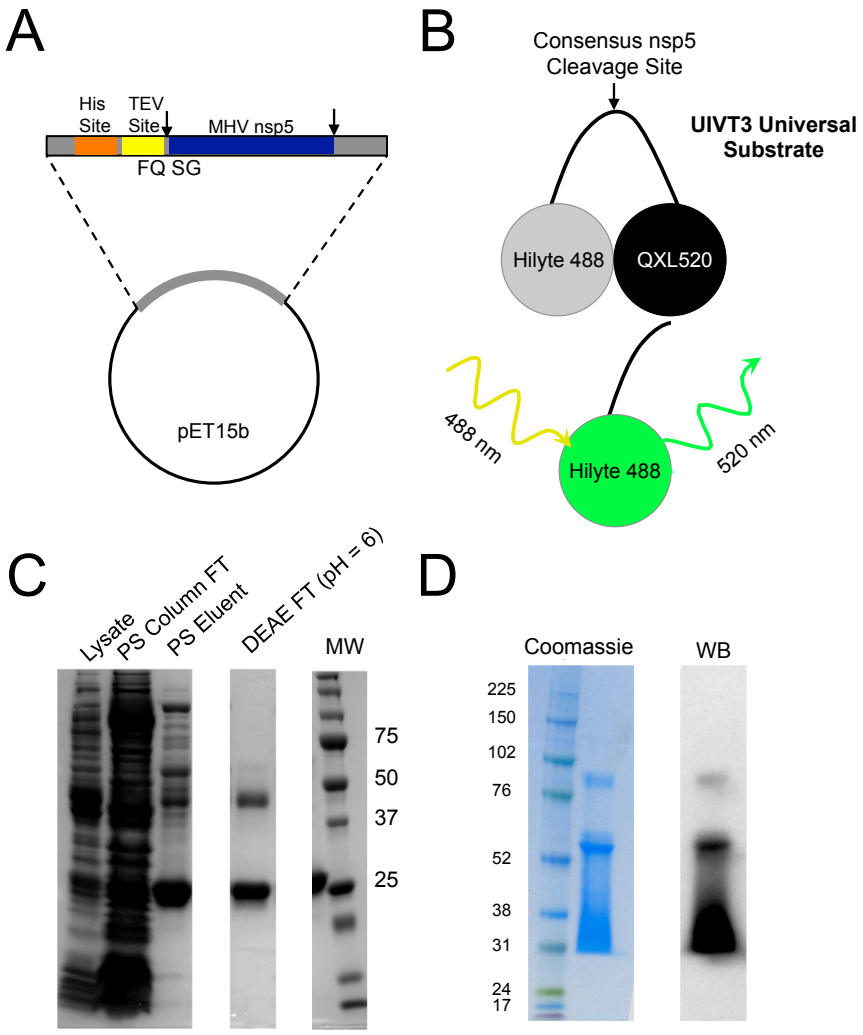


Fig. 2.7. Expression, purification and activity of wild-type MHV nsp5. A) Expression construct for codon-optimized expression of MHV nsp5 in *E. coli* BL21 DE3 cells. The amino-terminus of nsp5 is flanked by a TEV site and a His tag. **B)** FRET-based substrate (UIVT3) containing a consensus nsp5 cleavage site connecting a donor-quencher pair. **C)** Two-step purification of WT MHV on phenyl sepharose (PS) and DEAE anion exchange columns. Protein products are displayed on an SDS-PAGE gel Coomassie stained. MW, molecular weight marker in kilodaltons. **D)** Coomassie and western blots of purified wild-type MHV nsp5 protease. The western blot was performed using nsp5-specific antisera. Data acquired in collaboration with Sakshi Tomar and Aimee Eggler (Purdue University)

the two-step protocol. Protease activity was assayed using a FRET-based peptide substrate (UIVT3) containing a universal nsp5 consensus site flanked by a donor-quencher pair (HilyteFluor488-ESATLQ†SGLRKAK-QXL520-NH_2) which measures activity by emission if cleaved at 520 nm (following excitation at 488 nm). The fractions isolated following purification showed strong proteolytic activity against UIVT3 indicative of active and mature nsp5 protease.

Introduction of MHV *ts* and suppressor alleles for purification

To evaluate the role of the *ts* and second site suppressor alleles identified in MHV nsp5, I introduced the following combinations of mutations into the codon-optimized pET-15b expression construct: H134Y, V148A, V148A/H134Y, S133A, S133A/H134Y, F219L, F219L/H134Y, and F219L/H134Y/E285V (**Table 2.3**). The S133A, V148A and F219L mutations were introduced into the expression construct as they all conferred *ts* phenotypes during MHV infection. The H134Y was introduced as it had been selected as a second-site suppressor independently for all three viruses. However, during reversion analysis of the *ts*F219L virus, the H134Y mutation was insufficient to complete recover viral replication and processing at the non-permissive temperature of 40°C. Consequently, an additional combination containing both the H134Y and E285V was also introduced as this combination yielded a wild-type-like efficiency of plating (EOP) indicative of no apparent temperature-sensitivity, similar replication kinetics, and restored polyprotein processing.

The same approach for expression and purification of WT was employed for the following constructs: F219L, F219L/H134Y, F219L/H134Y/E285V, H134Y, V148A, S133A/H134Y. However, only H134Y and S133A/H134Y showed good expression after induction using IPTG at 37°C (**Figure 2.8**). A band migrating at the predicted size of monomeric nsp5 was detected in the lysed bacterial pellet for all three F219L mutant constructs and V148A indicating that a considerable fraction of the nsp5 expressed was likely mobilized to inclusion bodies. It was then decided to select one *ts* construct and second-site combination for optimization and purification.

Table 2.3. Oligonucleotide primers for codon-optimized MHV nsp5 mutagenesis

Oligonucleotides for generation of codon-optimized MHV nsp5 constructs			
Mutation	***Sequences***	***Type***	***Constructs***
S133A	5’ CTGCGTAGC**GCA**CACACCATTAAAG 3’ 5’ CTTTAATGGTGTG**TGC**GCTACGCAG 3’	*ts* Allele	S133A S133A/H134Y
H134Y	5’ CTGCGTAGCAGC**TAT**ACCATTAAAG 3’ 5’ CTTTAATGGT**ATA**GCTGCTACGCAG 3’	Suppressor Allele	H134Y V148A/H134Y S133A/H134Y F219L/H134Y F219L/H134Y/E285V
V148A	5’ TCC**GCC**GGCTATGTTCTGACTGGT 3’ 5’ ACCAGTCAGAACATAGCC**GGC**GGA 3’	*ts* Allele	V148A V148A/H134Y
F219L	5’ ACTGG**CTG**GTCCAGAGCGATTCTT 3’ 5’ AAGAATCGCTCTGGAC**CAG**CCAGT 3’	*ts* Allele	F219L F219L/H134Y F219L/H134Y/E285V
E285V	5’ GTTCTG**GTT**GATGAGCTGACC 3’ 5’ GGTCAGCTCATC**AAC**CAGAAC 3’	Suppressor Allele	F219L/H134Y/E285V

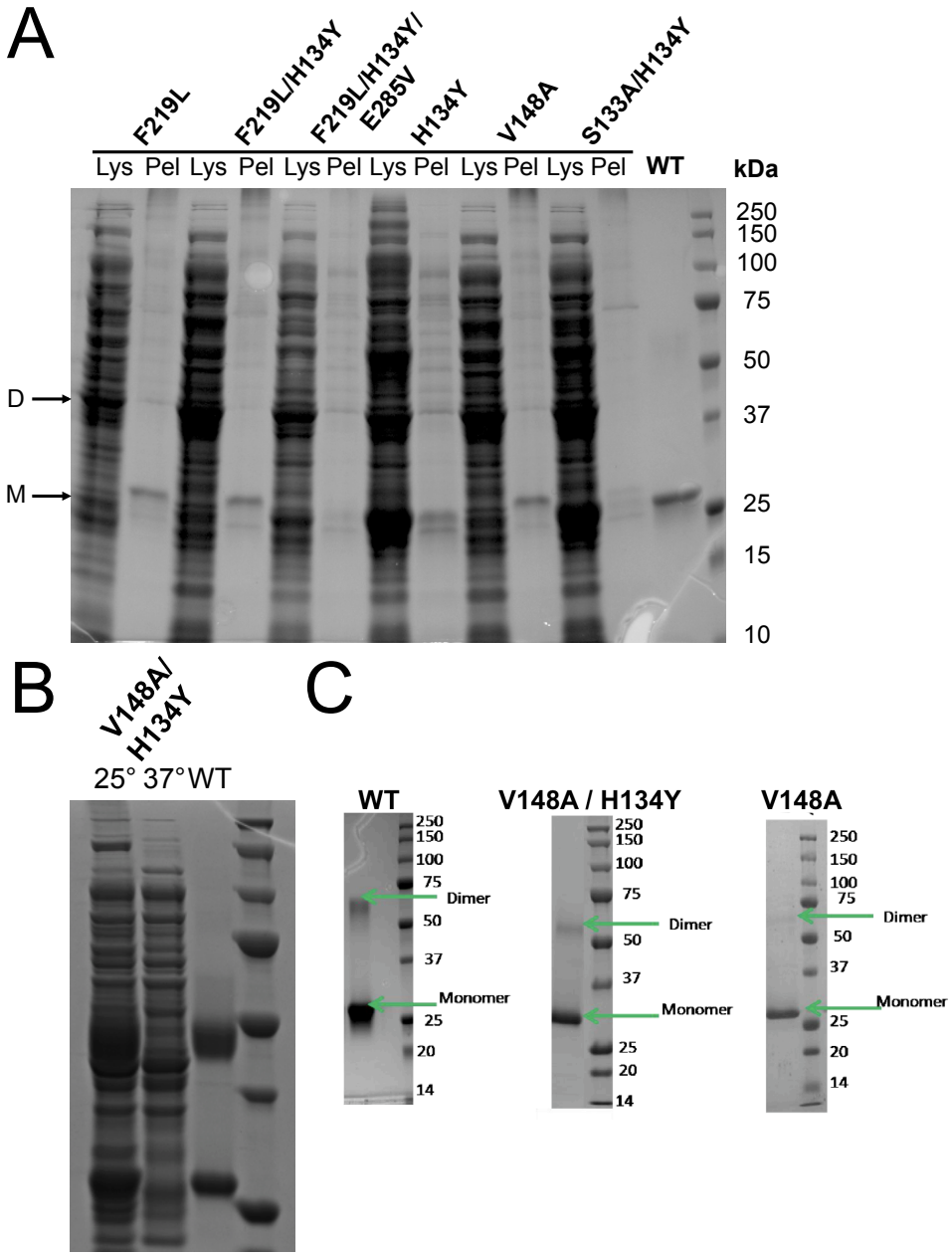


Fig. 2.8. Expression and purification of MHV nsp5 containing *ts* and suppressor mutations. A) Coomassie stain of bacterial lysates (Lys) and pellet fractions (Pel) of nsp5 containing the mutations F219L, F219L/H134Y, F219L/H134Y/E285V, H134, V148A, S133A/H134Y, and wild-type (WT) MHV. Arrows denoting the putative monomer (M) and dimer (D) bands are shown at left. **B)** Bacterial lysates of MHV nsp5 containing the V148A/H134Y mutations expressed under 0.15 mM IPTG induction at either 25° or 37°C. A wild-type (WT) nsp5 control is also provided for comparison. **C)** Purified WT, V148A/H134Y and V148A on Coomassie gel. The monomer and dimers are marked by arrows and molecular weights are in kilodaltons. Data acquired by Sakshi Tomar (Purdue).

Purification of V148A, H134Y, and V148A/H134Y nsp5 proteases

The temperature-sensitivity and suppressor phenotypes are reproducible and profound in the context of a replicating virus, however it remains unknown whether these effects are isolated to the protease or are associated with alterations to critical nsp5 allostery with other components of the viral genome. To evaluate whether these effects are internal to nsp5 or are associated with allosteric interactions with other viral elements, the *ts* lesion V148A and second-site suppressor mutation H134Y were individually and in combination (V148A/H134Y) expressed for purification. Sparks et al demonstrated that the *ts*V148A virus exhibited reduced replication kinetics and nsp5-mediated processing at the non-permissive temperature of 40°C. Initially, the same purification protocol used for WT MHV was employed for expression and purification of the mutant constructs. However, very little yield of the V148A and V148A/H134Y nsp5 products was recovered after IPTG induction at 37°C (**Figure 2.8**). Since the cognate viruses were temperature-sensitive, the temperature for expansion was lowered to 25°C to evaluate whether the temperature was affecting nsp5 bacterial mobilization. Upon shifting to 25°C, appreciably more expression was detected in the lysates. Further expression and purification of other mutated MHV nsp5 constructs will utilize these expression and bacterial growth conditions.

V148A reduces nsp5 protease activity at elevated temperatures

To evaluate the role of temperature on the MHV nsp5 containing V148A and V148A and H134Y in combination, a thermal inactivation assay was developed (**Figure 2.9**). During the assay, 10 μM of purified MHV nsp5 proteases containing V148A and V148A/H134Y as well as a wild-type control were incubated at temperatures ranging from 30°C to 50°C over 2 h. At individual time points, aliquots of the enzymes were mixed with the UIVT3 peptide substrate to evaluate the percentage of activity remaining relative to before incubation.

At 30°C, the wild-type MHV virus exhibited an approximately 1 log greater titer during virus replication over either the V148A or V148A/H134Y. However, the replication kinetics of all three viruses were nearly identical over an infection at an MOI of 0.01 PFU/mL. Incubation of the corresponding proteases at 30°C resulted in no appreciable loss of protease activity over the entire time frame for WT or the V148A-modified protease (**Figure 2.9**). Interestingly, the V148A/H134Y protease displayed a subtle drop in residual activity to approximately 50% of the original amount by the end of the time course before reaching a constant level of activity (**Table 2.4**). At 35°C, both WT and the suppressor combination of V148A/H134Y exhibited sustained activity with no apparent loss over the 2 h incubation period. Whereas, the V148A-modified protease displayed a small, but sustained decrease in activity over the entire survey associated with a half-life of approximately 130 min. At the non-permissive temperature for virus of 40°C, WT and the suppressor combination, V148A/H134Y, exhibit a slight but subtle decline in activity over time with half-lives exceeding 2 h. In contrast, V148A nsp5

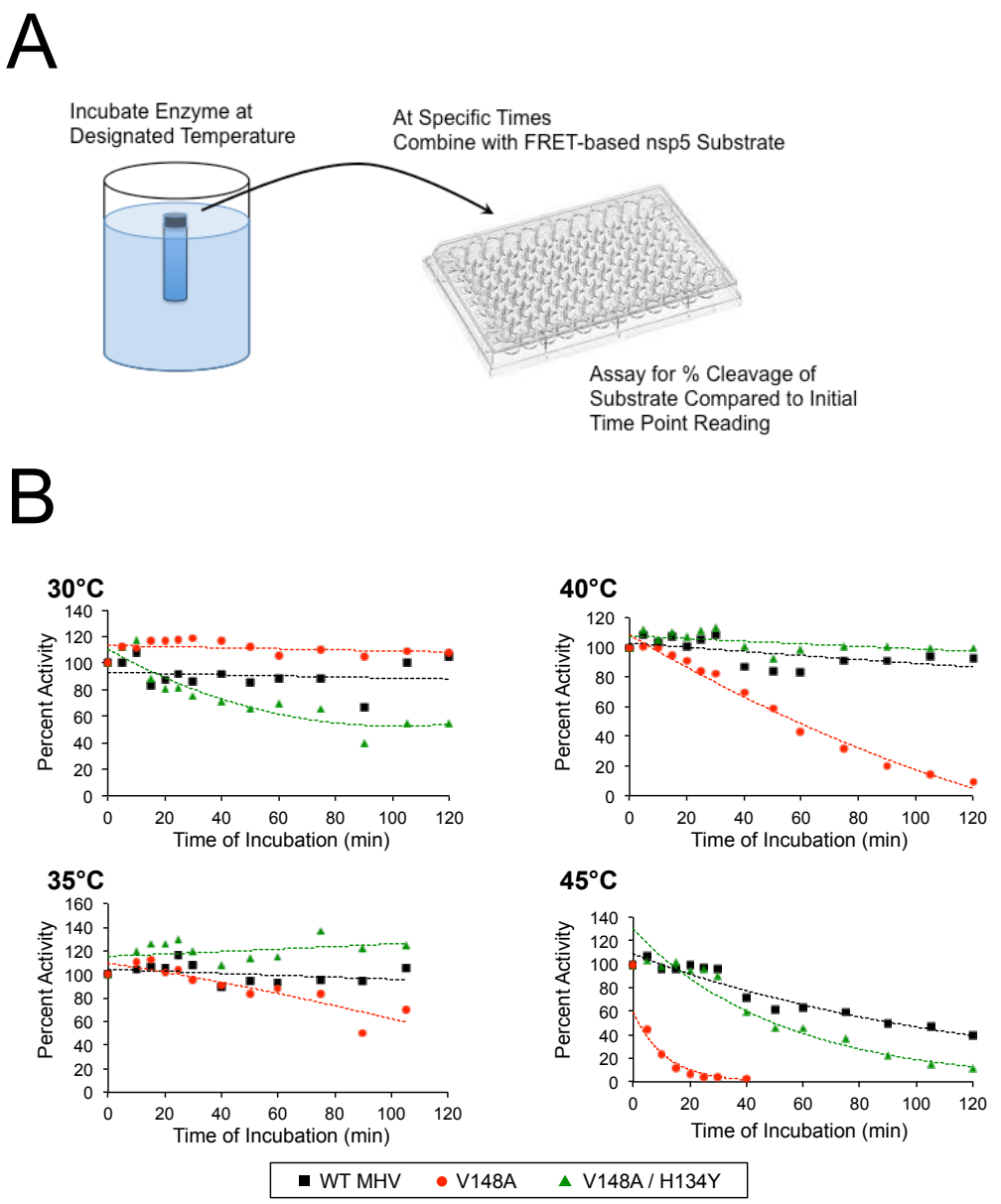


Fig. 2.9. Thermal inactivation of WT, V148A and V148A/H134Y nsp5 proteases. A) Thermal inactivation assay experimental design. **B)** Evaluation of wild-type (WT; black), V148A (red) and V148A/H134Y (green) MHV nsp5 protease activities during a 2 h incubation in temperatures ranging from 30° to 45°C. Percent activity is measured activity by cleavage of the UIVT3 universal substrate compared to activity at 0 min post-incubation. Data acquired by Sakshi Tomar (Purdue).

exhibits a steady linear decline in activity associated with a half-life of 55 min. When incubated at 45°C above biological conditions, WT and V148A/H134Y both exhibit a subtle loss of activity over 2 h time with an approximate half-lives of 90 and 75 min respectively. However, V148A nsp5 demonstrates a profound sensitivity to the temperature with an abrupt decline in activity and a half-life of approximately 2 - 5 min at the temperature. Collectively, these data correlate directly with the viral EOP data and indicate that the mutation V148A results in a decline in nsp5 protease activity at elevated temperatures, whereas the second-site suppressor mutation H134Y partially restores protease activity at these temperatures. Further, these data suggest that the inherent defect associated with these mutations in the context of a virus is not likely due to disruption of critical allosteric interactions with other viral elements.

Table 2.4. Approximate half-life of MHV nsp5 proteases *in vitro*

Half-life (in min) of MHV nsp5 proteases			
	WT MHV	***V148A***	***V148A/H134Y***
Virus EOP*:	3.3	2×10^{-5}	1.1
30°C	> 120 min	> 120 min	> 120 min
35°C	> 120 min	130 min	> 120 min
40°C	> 120 min	55 min	> 120 min
45°C	90 min	3 min	45 min
50°C	7 min	<3 min	< 3min

*Sparks et al. 2008 *J Virol*

V148A destabilizes and H134Y restabilizes MHV nsp5 secondary structure

To test whether the decrease in half-life is associated with structural instability, the proteases were subjected to circular dichroism (CD) spectroscopy. CD spectroscopy

is used to evaluate the secondary structure of proteins. Gross chains in secondary structure will result in modifications in the ellipticity or signal at defined wavelengths. WT, V148A, V148A/H134Y and H134Y MHV nsp5 proteases were evaluated by CD spectroscopy for differences in protease secondary structure. All four proteases exhibited a similar spectral curve at room temperature (25°C), which is comparable to previously reported nsp5 CD spectra and is reminiscent of largely β-sheet structure (**Figure 2.10**). These data indicate that the secondary structure of the proteases is indistinguishable at room temperature.

To assess the affect of temperature of protease secondary structure stability, the CD signal at a wavelength of 222 nm (selected due to its robust signal in the CD spectra) was evaluated with increasing temperature. Before 40°C, there were no apparent loss of signal among any of the proteases. However, at approximately 40°C, an inflection occurs in the spectral signal of V148A and it reaches half of its signal at approximately 47.5°C. The V148A/H134Y protease exhibits an approximate T_m of 49.5°C indicative of a subtle increase in structural stability over protease exhibiting the original *ts* allele. WT MHV exhibits a T_m of approximately 52.0°C. These data correlate strongly with the viral EOP data as well as the thermal inactivation assay data and suggest that phenotype associate with the *ts* lesion V148A is due to structural instability. Surprisingly, the H134Y demonstrated a higher T_m (54.0°C) than WT indicative of a hyperstable protease. Further study will need to evaluate the thermal stability of this protease *in vitro*.

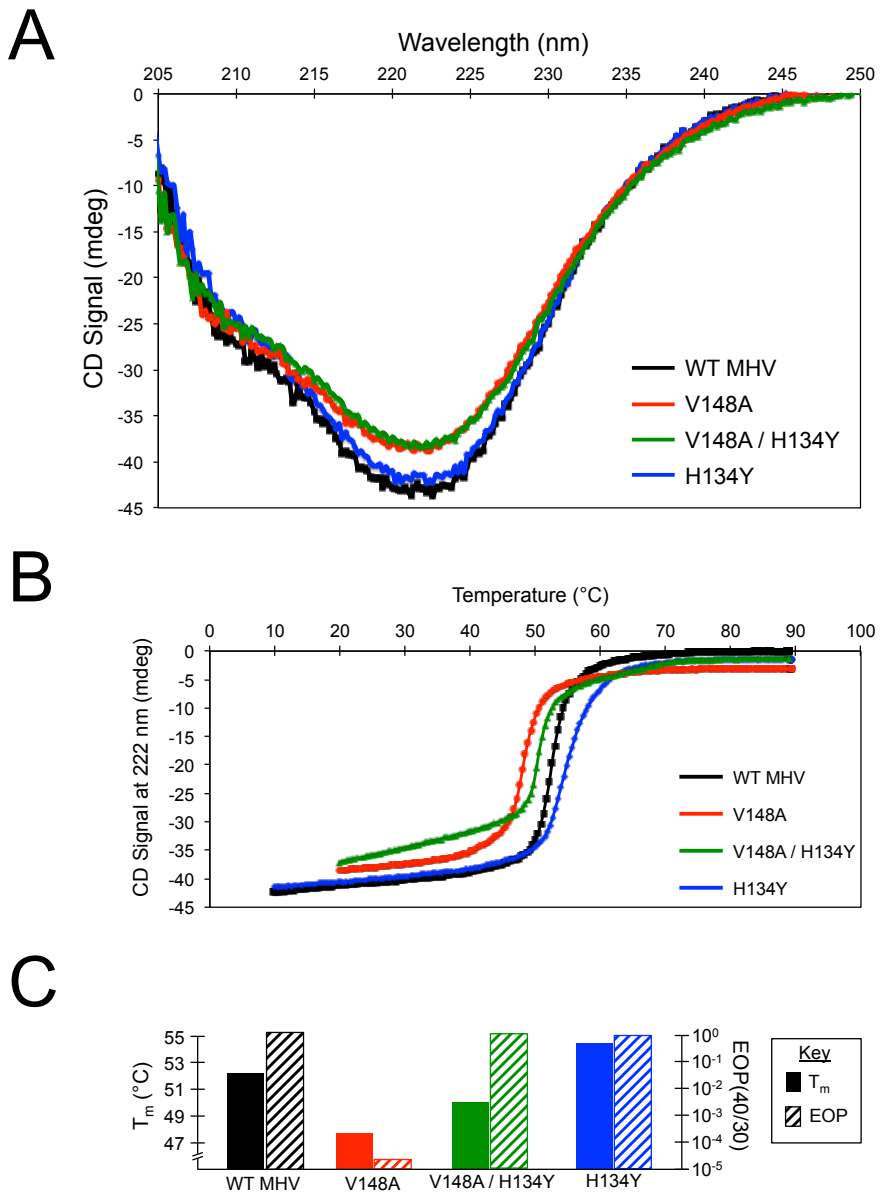


Fig. 2.10. Circular dichroism (CD) spectroscopy and thermal stability of MHV nsp5 proteases. A) CD spectra of WT (black), V148A (red), V148A/H134Y (green) and H134Y (blue) MHV nsp5 proteases. **B)** Evaluation of secondary-structure stability as measured by ellipticity at 222 nm in millidegrees (mdeg) over a temperature range of 0° to 100°C. **C)** Melting temperatures as measured by unfolding at a wavelength of 222 nm. The viral EOP (ratio of titers at 40° over 30°C) is provided for comparison. Data acquired by Sakshi Tomar (Purdue).

Summary

This study identified a novel temperature-sensitive lesion (S133A) in domain 2 of MHV and confirmed an additional temperature-sensitive lesion (F219L) in domain 3. These *ts* alleles were suppressed by a series of largely non-adjacent alleles spanning domains 2 and 3, which recovered viral replication kinetics and nsp5-mediated polyprotein processing at the non-permissive temperature of 40°C. Analysis of nsp5 structures and sequences demonstrated that communication between these alleles cannot be predicted by structural modeling and that none of the alleles described are conserved across known coronavirus sequences, but appear to be conserved within phylogenetic groups or clades. Codon-optimized MHV nsp5 proteases were expressed, purified and evaluated for thermal stability and sustained activity over time. These studies demonstrated that the *ts* allele V148A which was originally described by Sparks et al as having reduced viral replication and altered polyprotein processing results in a reduction in viral protease half-life and a decrease in thermal stability. These findings correlate directly with the viral replication kinetics and indicate that the defect associated with the *ts* mutations is intrinsic to nsp5 rather than due to unknown allosteric interactions with other viral elements. These studies provide a potential biochemical mechanism of action. These findings provide the basis for studies described in Chapter III. The implications of these findings and application of these data to other areas of coronavirus nsp5 protease regulation and activity are discussed in detail in Chapter IV.

CHAPTER III

CHIMERIC EXCHANGE OF CORONAVIRUS NSP5 PROTEASES (3CLPRO) IDENTIFIES COMMON AND DIVERGENT REGULATORY DETERMINANTS OF PROTEASE ACTIVITY

Introduction

Human coronaviruses (CoVs) are enveloped, positive-strand RNA viruses responsible for causing upper and lower respiratory illnesses of varying severities ranging from the common cold to severe acute respiratory syndrome (SARS). In 2002 – 2003, SARS-CoV spread to at least 32 countries and resulted in the deaths of nearly 10% of the approximately 8000 individuals infected. Recently, a new coronavirus strain (HCoV-EMC) identified in the Middle East and Europe has been associated with up to 65% mortality in infected individuals (Zaki et al., 2012). Phylogenetic and molecular analysis of known coronaviruses supports a zoonotic origin of SARS-CoV and EMC from bats (Anthony et al., 2013; Perlman and Zhao, 2013). However, it remains unclear what elements of coronavirus replication are conserved and more important the mechanisms and limitations of emergence of new coronavirus zoonoses. Study of SARS-CoV and HCoV-EMC has been aided by the ability of the viruses to replicate in culture and by identification of virus receptors. However, for many human and zoonotic coronaviruses, analysis is limited by the inability to cultivate the viruses from primary specimens. Thus, it is important to develop approaches to study coronavirus inhibition in the background of replicating viruses, as well as to understand the conservation of viral proteins as targets.

Coronaviruses are members of the order *Nidovirales*, family *Coronaviridae* and subfamily *Coronavirinae*. Among the viruses in *Coronavirinae*, four main genera have

recently been designated (Woo et al., 2012): (i) Alphacoronaviruses (α-CoVs) which contain the human coronaviruses 229E and NL63, (ii) Betacoronaviruses (β-CoVs) containing human coronaviruses SARS-CoV, HKU1, EMC and OC43, and (iii) Gammacoronaviruses (γ-CoV) and (iv) Deltacoronaviruses (δ-CoV), from which no current human coronaviruses have been identified. Coronaviruses genomes range in size from 27 to 32 kb, with genome organization similar among all four genera (Woo et al., 2012). However, co-evolution of proteins within and between diverse phylogenetic groups is unknown (Woo et al., 2009). Murine hepatitis virus (MHV), a β-CoV, is a well-established model system for the study of coronavirus replication and pathogenesis. The genome of MHV is approximately 32 kb in size and encodes seven genes including a replicase gene and six structural and accessory genes (Gorbalenya et al., 1989; Lee et al., 1991). The 22 kb viral replicase gene is translated into two large polyproteins (pp1a or pp1ab) containing 16 non-structural protein domains (nsps1-16) including two papain-like protease activities (PLP1 and PLP2) in nsp3, and a chymotrypsin-like cysteine protease (nsp5 / 3CLpro / Mpro) that mediate all proteolytic processing of the CoV nsps (Brierley et al., 1989; Lee et al., 1991; Perlman and Netland, 2009; Ziebuhr et al., 2000).

Coronavirus nsp5 proteolytic processing occurs at 11 cleavage sites between nsp4-16 and is required for virus replication. Nsp5 protease is conserved in all known coronaviruses and is comprised of 3 structural domains (D1-D3), of which D1-D2 form a chymotrypsin-like fold containing the catalytic Cys-His catalytic dyad and substrate binding pocket (Anand et al., 2002; Anand et al., 2003; Bacha et al., 2008; Xue et al., 2008; Zhao et al., 2008). Nsp5 cleaves after a P1-glutamine in all studied a, b and g coronaviruses (with the exception of a single site in HCoV-HKU1). Numerous studies

have demonstrated significant intra- and intermolecular associations within and between the coronavirus nsp5 protease and the rest of the replicase gene. We reported three different temperature-sensitive intramolecular alleles in domains 2 and 3 of MHV nsp5 as critical determinants of nsp5 protease activity and viral replication (Sparks et al. 2008; Stobart et al. 2012). Studies by others have also demonstrated that nsp5 activity was altered or aborted by mutations in nsp3 and nsp10 (Stokes et al. 2010; Donaldson EF et al. 2007). These data collectively suggest a close and crucial co-evolution between coronavirus nsp5 protease and the rest of the replicase gene.

Coronaviruses are known to exhibit some of the highest known rates of recombination, which have been attributed to polymerase pausing and RNA template switching (Makino et al. 1986; Sawicki et al. 2007). The first coronavirus reverse genetics methods employed targeted RNA recombination and selection for a virus with increased fitness at elevated temperatures (Koetzner et al. 1992; Masters et al. 1994). Early studies evaluating the rates of coronavirus recombination through mapping of complementation groups using temperature-sensitive lesions throughout the genome of MHV estimated a frequency of approximately 1% per 1.3 kb or nearly 25% of the entire genome (Baric et al. 1990; Fu et al. 1992). Sawicki *et al* (2005) and Stokes *et al* (2010) later demonstrated using complementation analysis that the coronavirus genome consists of at least five different complementation groups or cistrons (named cistrons 0, I, II, IV and VI) that contribute to unique viral phenotypes. Cistron I comprised of nsps 4 – 10 has been identified in several studies as a processed intermediate and has one or more related functions prior to maturation processing by nsp5 (Sawicki et al. 2005, Donaldson et al. 2007, Stokes et al. 2010, Xiao et al. 2012). These data suggest that the defined regions of

the coronavirus genome, which share a common function or phenotype, may have increased capacity for recombination. However, no studies have directly tested the potential or limitations for coronavirus recombination within or across genera. More specifically, it is unknown if the extensive interactions of nsp5 with 11 cleavage sites as well as with alleles in nsp3 and nsp10 constitute a barrier to genetic exchange between coronaviruses. If so, this would constitute limitations in over half of the coronavirus genome.

In this study, I address two specific questions: 1) Does the nsp5 protease establish a barrier to genetic exchange between closely or distantly related coronaviruses? 2) How conserved is nsp5 protease activity and specificity across coronavirus genera and species? To answer these questions, I engineered artificial recombinant viruses of the β-CoV MHV by exchanging complete nsp5 coding sequences from closely-related human b-CoVs HKU1 and OC43 and distantly-related α- and β-CoVs SARS-CoV, Bat-HKU4, NL63 and 229E. In this study, I demonstrate that artificial recombination of nsp5 from closely related HKU1 and OC43 is readily tolerated in MHV for efficient replication, but confers a profound loss in competitive fitness. In contrast, more distantly related coronavirus nsp5 proteases could not be recovered, even with exchange of the species-specific cleavage sites. These results suggest that co-evolution of nsp5 with cleavage sites and other proteins is extensive within subgroups of genera, and may represent a profound barrier to recombination in replicase gene across genera and even between more divergent subgroups (such as 2a and 2b) Further nsp5 may constitute a profound but not absolute barrier to emergence of recombinants between closely related species due to loss of fitness. On the other hand, the studies also demonstrate that it is possible within

subgroups to establish replicating platform coronaviruses such as MHV for study of nsp5 function and inhibition from difficult or non-cultivatable coronaviruses such as HCoV-HKU1. This approach provides a rapid new system to test protease inhibitors against coronaviruses of bats such as HKU4 or other animals before they emerge as zoonotic infections.

Generation of synthetic recombinant MHV nsp5 coronaviruses

To test the differences between nsp5 proteases across coronavirus phylogenetic groups in a replicating virus, I generated artificial recombinant genomes of MHV-A59 with chimeric replicase genes by exchanging the complete nsp5 coding sequence of α-coronaviruses (α-CoVs) 229E and NL63, and β-coronaviruses (β-CoVs) HKU1, OC43, SARS-CoV and bat coronavirus HKU4. Previous studies have noted that the tertiary structure of nsp5 and cleavage site recognition sequences are largely conserved across all known coronavirus strains (Anand et al., 2002; Anand et al., 2003; Bacha et al., 2008; Woo et al., 2009; Xue et al., 2008; Zhao et al., 2008). However, primary peptide sequence alignments across coronaviruses show considerably less conservation between coronavirus groups. Peptide alignments of closely-related β-CoVs MHV, HKU1 and OC43 nsp5 exhibit 80 to 84% sequence identity to each other and occupy the same 2a subgroup. In contrast a-CoVs 229E, NL63, and more distant β-CoVs SARS-CoV as well as bat coronavirus HKU4, show no greater than 53% identity (**Figure 3.1)**. Bat CoV-

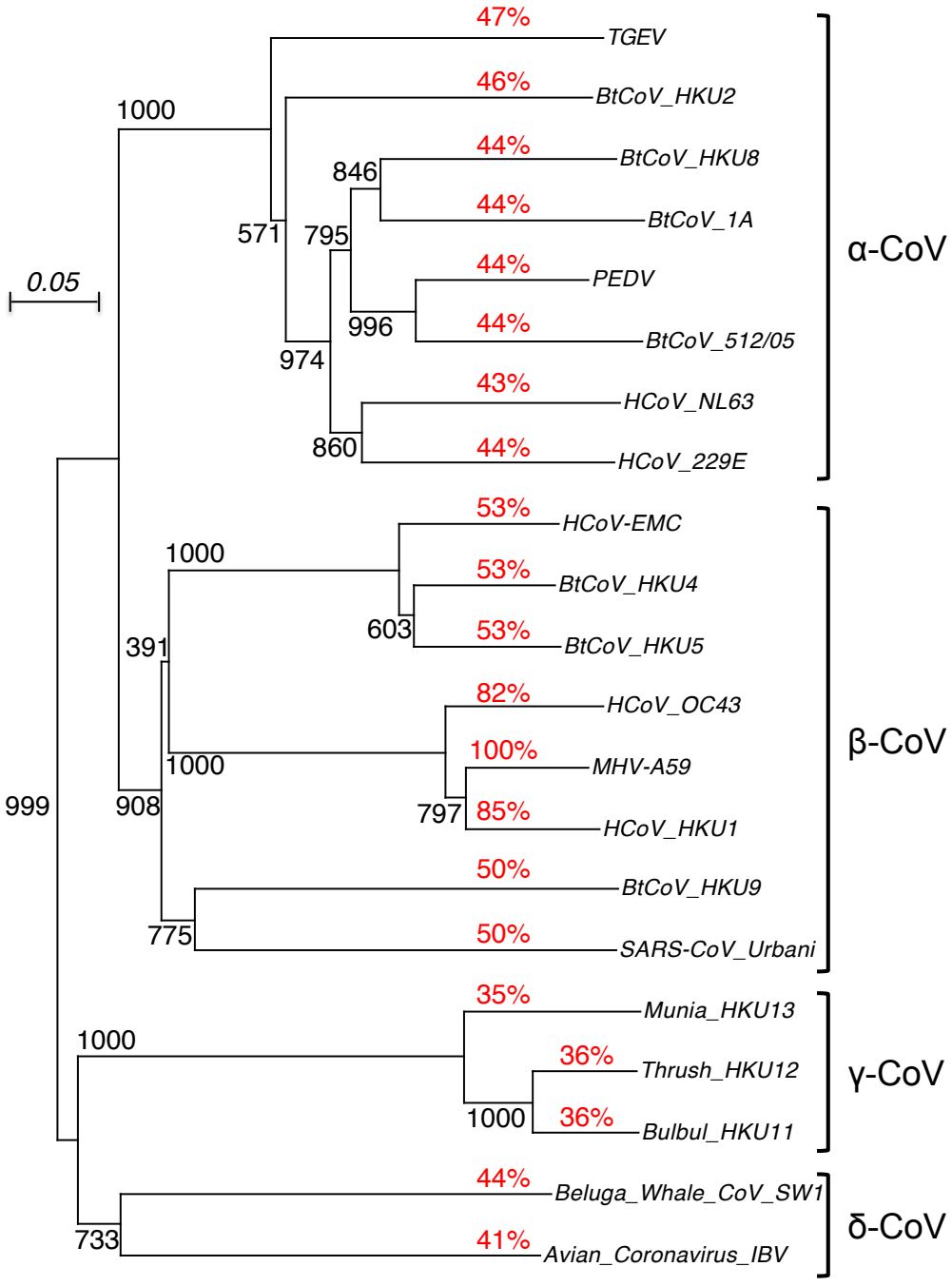


Fig. 3.1. Phylogeny of coronavirus nsp5 protease. Phylogenetic tree generated using a bootstrap alignment of coronavirus nsp5 sequences among species recognized by the International Committee on Taxonomy of Viruses (ICTV). Bootstrap values and phylogenetic genera are identified and the percent conservation among amino acid sequences compared to MHV is shown in red.

HKU4 nsp5 was selected for study since it exhibits the highest identity (53%) among coronaviruses outside of β-CoV group 1 (2a) (containing MHV, HKU1, and OC43), as well as showing high sequence homology (81% identity) to the recently identified human coronavirus EMC (Zaki et al., 2012).

Chimeric MHV-A59 viruses encoding nsp5 from HKU1 (H5-MHV) and OC43 (O5-MHV) were recovered at 37°C and exhibited cytopathic effects and recovery kinetics similar to wild-type MHV (**Figure 3.2**). The complete genomes of H5- and O5-MHV were sequenced with the chimeric nsp5 intact and no additional mutations identified. In contrast, MHV with nsp5 from SARS-CoV, NL63, 229E, or Bat-HKU4 could not be recovered following 3 or more independent attempts. Further, testing of total cell RNA by PCR did not detect transcripts of M, N or E subgenomic transcripts. These results demonstrated that only nsp5 from closely related of human coronaviruses HKU1 and OC43 were compatible within the background of MHV, whereas more distant chimeric substitutions were not tolerated. I therefore used these viruses to test the functional similarity in replication, fitness, cleavage site usage and response to known MHV nsp5 temperature sensitive mutations.

Chimeric H5-MHV and O5-MHV exhibit a subtle defect in replication kinetics

To compare the replication of the chimeric H5- and O5-MHV viruses to wild-type MHV, replicate plates of DBT cells were infected at low multiplicity of infection (MOI 0.01 PFU/cell) or high MOI (1 PFU/cell) (**Figure 3.3**). In low MOI infection, WT-MHV

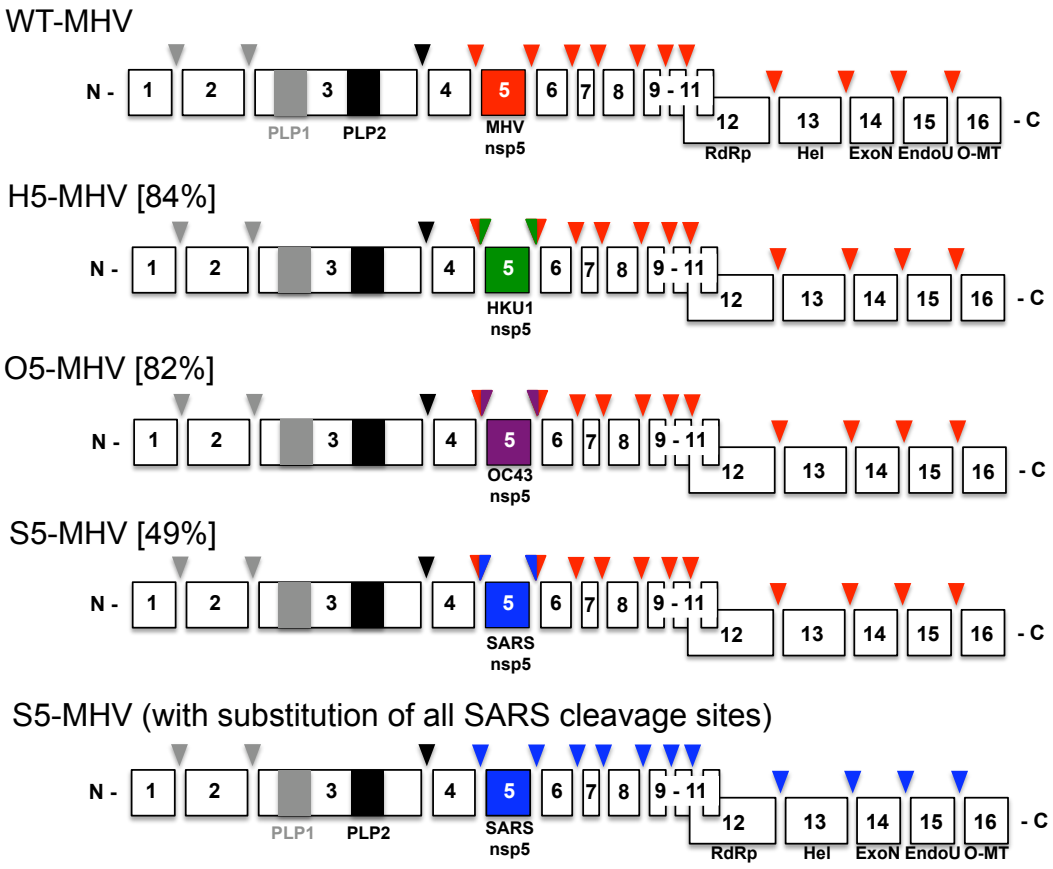


Fig. 3.2. Generation of chimeric synthetic recombinant MHV nsp5 viruses. HKU1 (H5-MHV; green), OC43 (O5-MHV; purple), SARS-CoV (S5-MHV; blue), NL63 (not shown), 229E (not shown) and Bat HKU4 (not shown) nsp5 coding regions were substituted into the background of wild-type (WT) MHV. An additional attempt at recovery of the S5-MHV was made by substitution of all 11 cleavage sites (P5 – P2') into the MHV background.

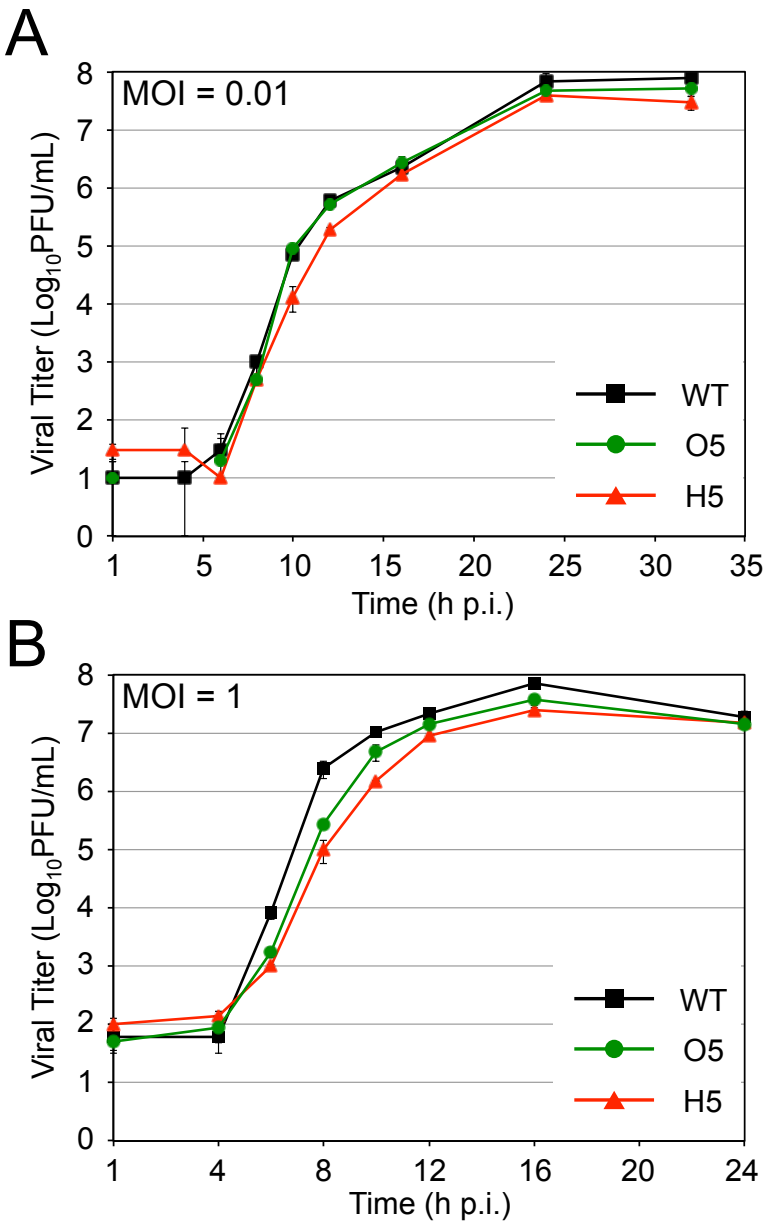


Fig. 3.3. Replication kinetics of H5-MHV and O5-MHV at high and low MOIs. Virus infections using wild-type (WT) MHV, H5-MHV, or O5-MHV were carried out in DBT-9 cells at 37°C and multiplicities of infection (MOI) of either 0.01 PFU/cell (**A**) or 1 PFU/cell (**B**). Samples were acquired in triplicate and titers determined by plaque assay in duplicate per sample.

and the O5-MHV chimeric virus exhibited indistinguishable replication, while H5-MHV displayed a delay in replication with an approximate 10-fold reduction in virus titers 10 h p.i. compared to WT MHV. During single-cycle infection (MOI 1 PFU/cell), both H5- and O5-MHV displayed approximately 10-fold and 30-fold reduced titers for O5-MHV and H5-MHV, respectively, compared to WT-MHV at 8 h p.i. These results suggested a subtle mismatch between the protease and virus that might be amplified in competitive environment.

H5- and O5-MHV are profoundly impaired in competitive fitness

To evaluate fitness cost of chimeric nsp5 in direct competition WT and chimeric viruses, or both chimeric viruses were combined at varying ratios and used at a total MOI of 0.01 PFU/cell to infect DBT cell monolayers. At approximately 30% monolayer involvement in CPE, total cell RNA was used to generate cDNA amplicons containing the nsp5-coding region. The amplicons were then treated using restriction enzymes that recognize unique sites in the nsp5 sequence of HKU1 or OC43 and quantitated following electrophoresis and densitometry (Graham et al., 2012). (**Figure 3.4**).

Following co-infection with MHV and H5-MHV at an initial ratio of 1:1, WT-MHV represented >75% of supernatant infectious virus after passage 1 (P1), and little or no chimeric virus was detected by P3 (**Figure 3.5**). Even when H5-MHV or O5-MHV were given a 10:1 advantage over WT-MHV in the infecting inoculum, WT equaled or dominated the chimeric mutants by P2 and was dominant in both cases by P3. I next directly compared H5-MHV and O5-MHV to determine if there were differences in

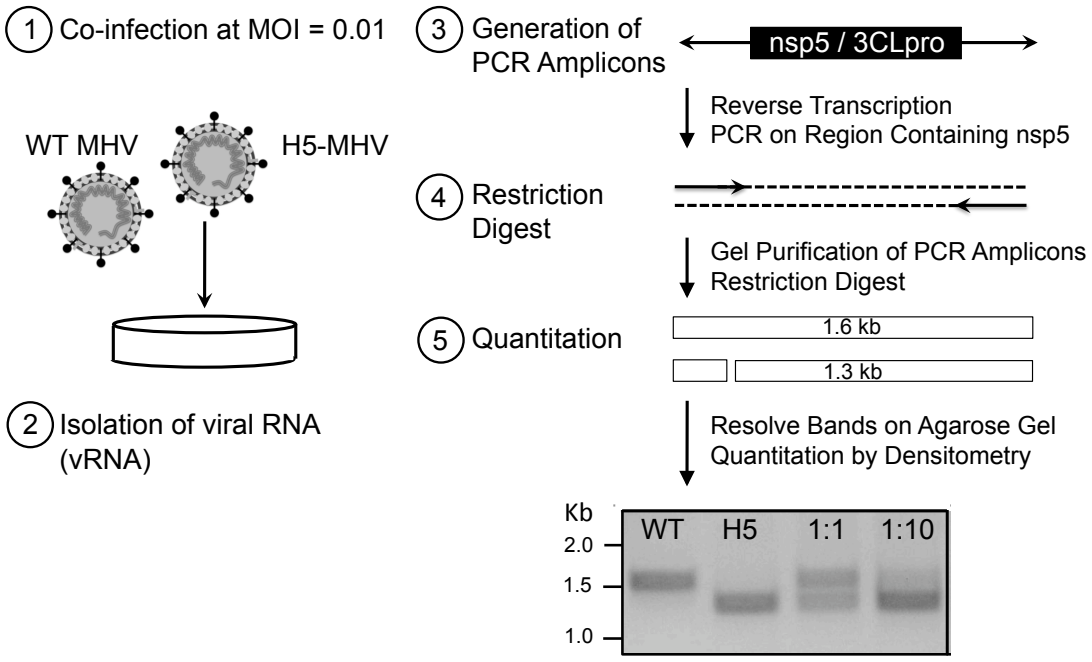


Fig. 3.4. Competitive fitness assay. Confluent monolayers of DBT-9 cells were infected with WT- or H5-MHV and viral RNA isolated. cDNA amplicons containing the nsp5 coding region of pure WT-MHV, H5-MHV or mixtures of either 1:1 or 1:10 of WT:H5 were generated and digested with an HKU1 nsp5-specific restriction enzyme and resolved and quantified by gel electrophoresis.

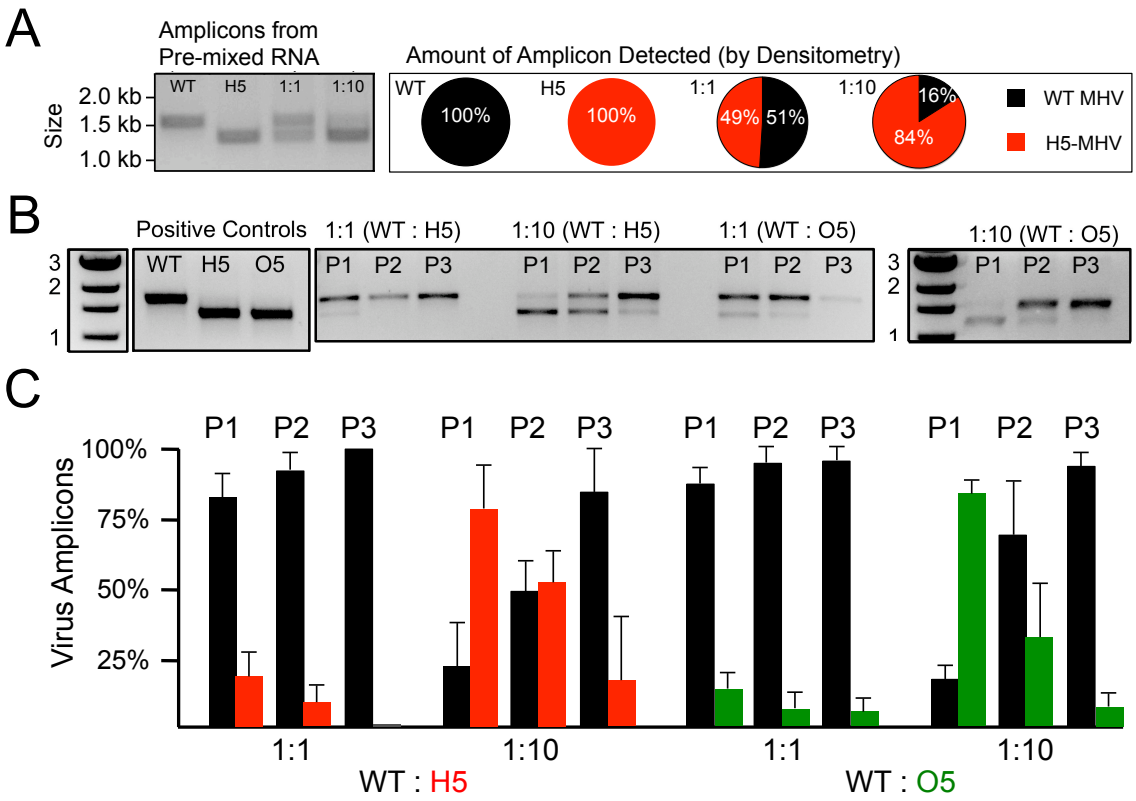


Fig. 3.5. Fitness analysis of WT MHV compared to H5- and O5-MHV. (A) Confluent monolayers of DBT-9 cells were infected with WT- or H5-MHV and viral RNA isolated. cDNA amplicons containing the nsp5 coding region of pure WT-MHV, H5-MHV or mixtures of either 1:1 or 1:10 of WT:H5 were generated and digested with an HKU1 nsp5-specific restriction enzyme and resolved by gel electrophoresis (left) and quantified by gel electrophoresis (right). (**B**) Co-infections of either 1:1 or 1:10 of either WT- and H5-MHV or WT- and O5-MHV were carried out at 37°C and passaged three times (P1 – P3) and the generated cDNA amplicons subjected to either HKU1 nsp5-specific or OC43 nsp5-specific restriction digests and resolved by gel electrophoresis. (**C**) Quantification of the ratio of WT to H5-MHV or O5-MHV as determined by the average of three replicates at each designated passage.

competitive fitness in the chimeric viruses (Fig. 4). At a 1:1 ratio for co-infection, the relative amounts of H5:O5 detected were approximately 1:1 (53%:47%) while at P2 and P3, O5-MHV appeared to outcompete H5-MHV (**Figure 3.6**). To directly test this, O5-MHV was given a 10:1 advantage to determine whether it would drive H5-MHV to extinction. However, by passage 3, a small amount of H5-MHV was still detected indicating no absolute fitness advantage for O5-MHV over H5-MHV. These results demonstrated that while the replication defects of chimeric H5-MHV and O5-MHV were mild in single infection, introduction of even a closely related nsp5 resulted in profound loss of competitive fitness compared to WT. In contrast, there was a detectable but not absolute fitness advantage of one of the chimeric nsp5 proteins over the other. Thus nsp5 of very closely related viruses can mediate all required activities for replication, but in artificial recombinants cannot recapitulate the precise roles and interactions of the parent nsp5.

In vitro nsp5 processing of polyprotein cleavage sites

I next determined whether there are differences in the activity of nsp5 at replicase polyprotein cleavage sites that might explain differences in replication and fitness of chimeric viruses and the inability to recover more distantly related nsp5 chimeric MHV. Coronavirus nsp5 protease cleavage sites exhibit some conserved elements. The P1 position is a glutamine (Q) in all known cleavage sites (except for the nsp13-nsp14 cleavage site in HKU1) and the P2 and P1' positions exhibit small variation in amino acid usage (**Table 3.1**). The native nsp5 proteases MHV, HKU1, OC43, and SARS-CoV (SARS) were expressed and purified as previously described and were mixed with FRET-

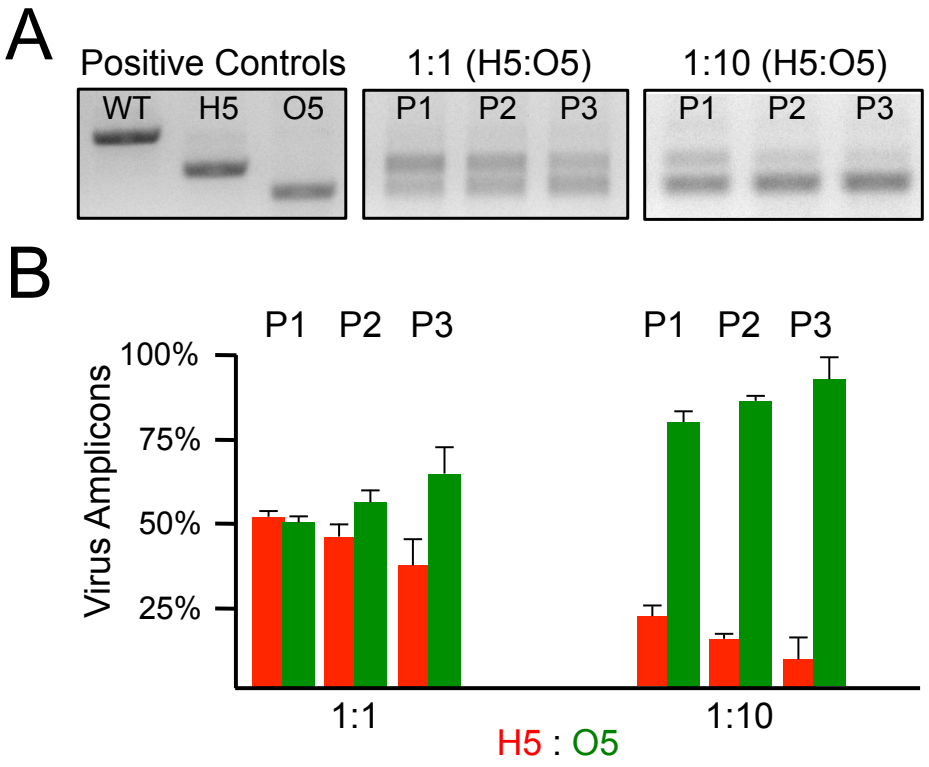


Fig. 3.6. Determination of the Relative Fitness of H5- Compared to O5-MHV. (A) Co-infections of 1:1 of H5- and O5-MHV were carried out at 37°C and passaged three times (P1 – P3) and the generated cDNA amplicons subjected to an OC43 nsp5-specific restriction digests and resolved by gel electrophoresis. (**B**) Quantification of the ratio of H5-MHV to O5-MHV as determined by the average of three replicates at each designated passage.

based substrates to assess the relative rates of enzymatic cleavage *in vitro* (Jacobs et al., 2013). A universal substrate (UIVT3) containing a coronavirus nsp4/nsp5 consensus cleavage site was used to verify that all proteases exhibit proteolytic activity (**Figure 3.7**) (Grum-Tokars et al., 2008). MHV and OC43 nsp5 exhibited the greatest relative rates of cleavage, followed by HKU1, and with SARS-CoV nsp5 the least active. The result directly correlated with the ability to recover chimeric virus and the fitness rank of the recovered chimeric viruses.

Table 3.1. Coronavirus nsp5 cleavage sites

	Coronavirus nsp5 Cleavage Sites				
Junction	*MHV*	*HKU1*	*OC43*	*SARS-CoV*	*Relative* k_{cat} / K_m*
nsp4/5	TSFLQ SG	TSFLQ SG	TSFLQ SG	SAVLQ SG	100%
nsp5/6	GVKLQ SK	GVKLQ SK	GIKLQ SK	GVTFQ SK	41%
nsp6/7	VSQIQ SR	VSQIQ SK	VSQFQ SK	VATVQ SK	3%
nsp7/8	NTVLQ AL	STVLQ AL	NTVLQ AL	RATLQ AI	5%
nsp8/9	TVVLQ NN	NAVMQ NN	ATVLQ NN	AVKLQ NN	2%
nsp9/10	TVRLQ AG	TIRLQ AG	TVRLQ AG	TVRLQ AG	22%
nsp10/12	GSQFQ SK	SVAVQ SK	DTTVQ SK	EPLMQ SA	0.2%
nsp12/13	SAVLQ SV	SAVMQ SV	SAVMQ SV	HTVLQ AV	8%
nsp13/14	NPRLQ CT	LPRLH CT	ETKVQ CS	VATLQ AE	9%
nsp14/15	FTRLQ SL	FTTLQ SL	FTKLQ SL	FTRLQ SL	28%
nsp15/16	YPRLQ AA	YPKMQ AT	YPRLQ AA	YPKLQ AS	27%

*Modified from Grum-Tokars, 2008.

To determine activity towards specific coronavirus nsp5 cleavage sites, FRET-based substrates were generated containing P8 – P8' peptide sequences for the nsp4/nsp5 and nsp7/nsp8 cleavage sites of MHV and SARS (Felber et al., 2004). The nsp4/nsp5 cleavage site represents a very early step in the autoproteolytic processing of nsp5, is cleaved most efficiently *in vitro*, and is required for virus replication (Chen et al., 2010; Grum-Tokars et al., 2008; Lu et al., 1996; Lu and Denison, 1997). Cleavage between

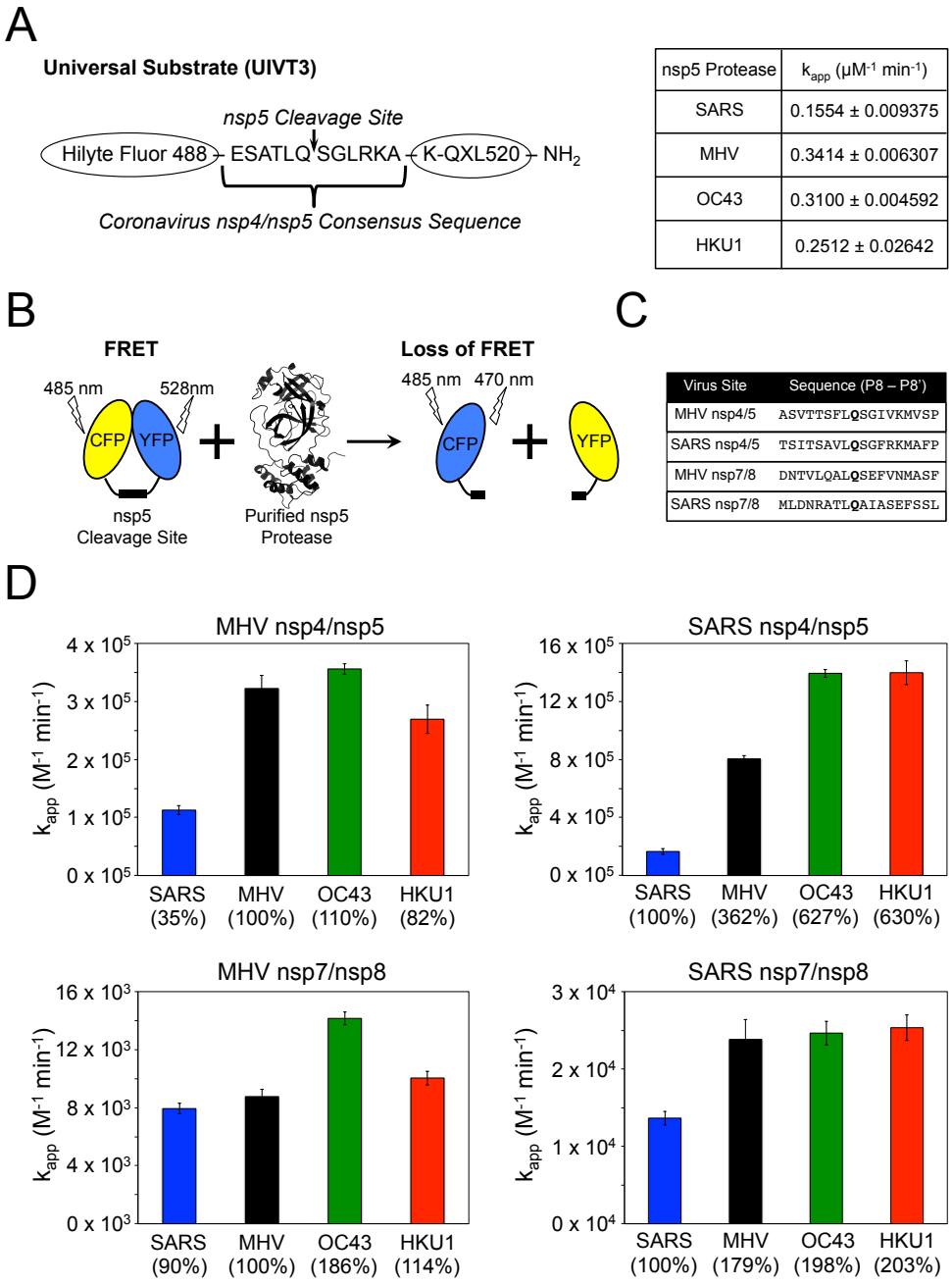


Fig. 3.7. Relative Rates of Coronavirus nsp5 processing *in vitro*. (A) FRET based substrate containing the coronavirus nsp4/nsp5 consensus sequence and relative rates of cleavage by coronavirus nsp5 proteases. **(B)** The nsp5 cleavage site peptide sequences are engineered between a CFP-YFP FRET pair. Upon cleavage, increase in donor emission at 485 nm is quantified over time. (**C**) P8 – P8' peptide sequences of MHV and SARS nsp4/5 and nsp7/8 cleavage sites. (**D**) Relative rates of processing by purified coronavirus nsp5 proteases MHV (black), OC43 (red), HKU1 (green) and SARS-CoV (blue). Percentage of activity relative to the cognate proteases is shown. Data obtained by Katie Molland and Sakshi Tomar (Purdue University).

nsp7/8 also is required for virus viability (Deming et al., 2007). In SARS-CoV the relative *in vitro* activity of nsp5 at the nsp7/nsp8 site is approximately 5% of the activity at the nsp4/5 cleavage site (Grum-Tokars et al., 2008). When the purified nsp5 proteins were tested against these constructs, the rates of enzymatic cleavage towards the nsp7/8 cleavage sites by all four proteases were considerably less than that towards the nsp4/5 cleavage sites, consistent with published studies (Grum-Tokars et al., 2008). SARS-CoV nsp5 had the lowest cleavage rates of all cleavage site constructs. The relative rates of enzymatic cleavage of the SARS nsp4/5 peptide construct by MHV, HKU1, and OC43 nsp5 were much greater than that of SARS-CoV nsp5. Neither the HKU1 nor OC43 protease showed an appreciable defect in processing the MHV nsp4/5 or 7/8 cleavage sites. The relative rates of enzymatic cleavage of the MHV nsp4/5 peptide construct by MHV and OC43 nsp5 were similar and were greater than that of the proteases of HKU1 and SARS-CoV. The OC43 nsp5 showed the fastest rate of cleavage of the MHV nsp7/8 peptide construct. In summary, the results of specific cleavage events by the nsp5 proteins also are consistent with the observed ability and order of recovery and fitness. However, only two of 11 cleavage sites were tested, and may not completely represent the complexity required nsp5 functions during replication. To test whether cleavage sites alone are the only determinant of successful nsp5 genetic exchange, we attempted to recovery SARS nsp5 / MHV chimera that also contained all 11 SARS-CoV cleavage sites (from P5 to P2'). However, even complete cleavage site substitution did not allow recovery of the SARS-nsp5 chimeric MHV. Overall our results show that the ability to efficiently process at cleavage sites from another CoV is necessary, but is not sufficient for viability and fitness of genetically exchanged nsp5.

MHV *ts* mutations differ in their phenotype in chimeric H5- and O5-MHV

To evaluate whether functional similarities and differences within a phylogenetic group are conserved, a known MHV functional pathway was selected for introduction into the chimeric backgrounds of H5- and O5-MHV. The identification of three different temperature-sensitive mutations (S133A, V148A, and F219L) located in MHV domains 2 and 3 was reported to cause greatly reduced viral replication kinetics and diminished polyprotein processing at a non-permissive temperature of 40°C (Sparks et al., 2008; Stobart et al., 2012). HKU1 and OC43 nsp5 share the same amino acids and codons at the three temperature-sensitive alleles, but contain a Tyr rather than a His at the 134 and 270 positions that were selected second-site suppressor residues in MHV (Sparks et al., 2008; Stobart et al., 2012). The direct mechanism by which these mutations impact the protease remains unknown. To further evaluate the functional differences between closely- and more distantly related coronaviruses, the temperature-sensitive mutations S133A, V148A, and F219L were introduced into the background of the chimeric H5- and O5-MHV viruses. Identical two or three nucleotide mutations were engineered with S133A ($_{10605}$AGU to GCC), V148A ($_{10650}$GUU to GCC) and F219L ($_{10864}$UUU to CUG) introduced into the HKU1 and OC43 nsp5 coding sequences of H5- and O5-MHV. Viruses were assembled as previously described, however recovery and characterization of the viruses was done at a temperature of 32°C. As previously shown, the temperature-sensitive mutations did not appear to adversely affect the viruses at 30°C (Sparks et al., 2008; Stobart et al., 2012). Since the permissive DBT-9 cell line is more tolerant of 32°C, this temperature was selected as a permissive temperature instead. Three different

chimeric viruses containing MHV temperature-sensitive alleles were successfully recovered at 32°C: H5-V148A (H5-MHV containing the V148A mutation), O5-V148A (O5-MHV containing the V148A mutation), and O5-S133A (O5-MHV containing the S133A mutation). Neither chimeric virus containing the F219L mutation was successfully recovered and introduction of the S133A mutation into the H5-MHV background did not permit recovery. Sequencing of the nsp5 coding regions of HKU1 and OC43 confirmed the sole presence of the introduced S133A and V148A mutations.

To assess temperature-sensitivity of the recovered viruses, the titers were determined at 32°C, 37°C, and 40°C for wild-type MHV, H5-MHV, O5-MHV, and each of the recovered mutants and the efficiency of plating (a ratio of titers determined by plaque assay) was determined for 40°C compared to 32°C and 37°C (**Table 3.2**). Similar to previous studies, WT-MHV exhibited a slight increase in detected titers at 40°C relative to either lower temperature of 32°C (EOP = 1.4) or 37°C (EOP = 0.9) (Sparks et al., 2008; Stobart et al., 2012). Both H5- and O5-MHV viruses exhibited similar subtle advantages in detected titers at the elevated temperature consistent with not being temperature-sensitive (*ts*). Analysis of the EOP values for the recovered mutants containing the MHV *ts* alleles yielded varying results. Introduction of the V148A mutation did not confer an EOP indicative of temperature-sensitivity in either the HKU1 or OC43 chimeric virus backgrounds (EOP > 1 for either ratio of 40°C/32°C or 40°C/37°C). Surprisingly, introduction of the S133A mutation into the OC43 background (O5-S133A) resulted in an EOP of 6×10^{-3} when comparing detected titers at 40°C relative to 32°C and an EOP of 1×10^{-2} for the ratio of titers at 40°C to 37°C consistent with a temperature-sensitive phenotype.

Based on previous studies, the sensitivity of viruses to changes in temperature has been shown to be independent of reduced or delayed virus replication kinetics even at permissive temperatures (Stobart et al., 2012). To evaluate the replication kinetics of the recovered MHV *ts* alleles in the backgrounds of the H5- and O5-MHV chimeras, infections were carried out in confluent flasks of DBT-9 cells at a MOI of 1 PFU/cell at 32°C and with a temperature shift at 6 h p.i. to 40°C (**Figure 3.8**). Congruent with earlier studies, most of the viruses surveyed exhibited similar replication kinetics to WT MHV at 32°C (Sparks et al., 2008; Stobart et al., 2012). However, the O5-S133A viruses exhibited a clear delay in logarithmic growth of approximately 2 hours similar to H5-MHV suggesting a subtle replication defect at the permissive temperature. After shifting to the non-permissive temperature of 40°C at 6 h p.i., the H5- and O5-V148A mutants displayed similar replicative kinetics to the H5- and O5-MHV backgrounds consistent with a non-*ts* phenotype and supporting the EOP data. Consistent with the MHV *ts* mutant viruses, the O5-S133A showed a clear delay and reduction in virus replication kinetics after shifting to the non-permissive temperature (Stobart et al., 2012). These data correlate directly with the calculated EOP (6×10^{-3}) and are consistent with a *ts* phenotype at 40°C.

Table 3.2. Temperature-sensitivity of H5- and O5-MHV containing MHV *ts* alleles

***EOP and Titers of H5- and O5-MHV viruses containing MHV ts* alleles**					
Virus	*Virus Titers (PFU/mL)**			*EOP (Ratio of Titers)*	
	32°C	37°C	40°C	40°/32°	40°/37°
WT-MHV	5.1×10^7	8.3×10^7	7.2×10^7	1.4	0.9
H5-MHV	2.0×10^7	1.2×10^8	1.1×10^8	5.7	1.0
O5-MHV	1.1×10^7	2.0×10^7	2.0×10^7	1.7	1.0
H5-V148A	3.1×10^6	2.0×10^7	8.7×10^6	2.8	0.4
O5-V148A	1.2×10^7	3.7×10^7	2.1×10^7	1.7	0.6
O5-S133A	1.5×10^7	7.5×10^6	8.8×10^4	6×10^{-3}	1×10^{-2}

*Average titer, N > 4

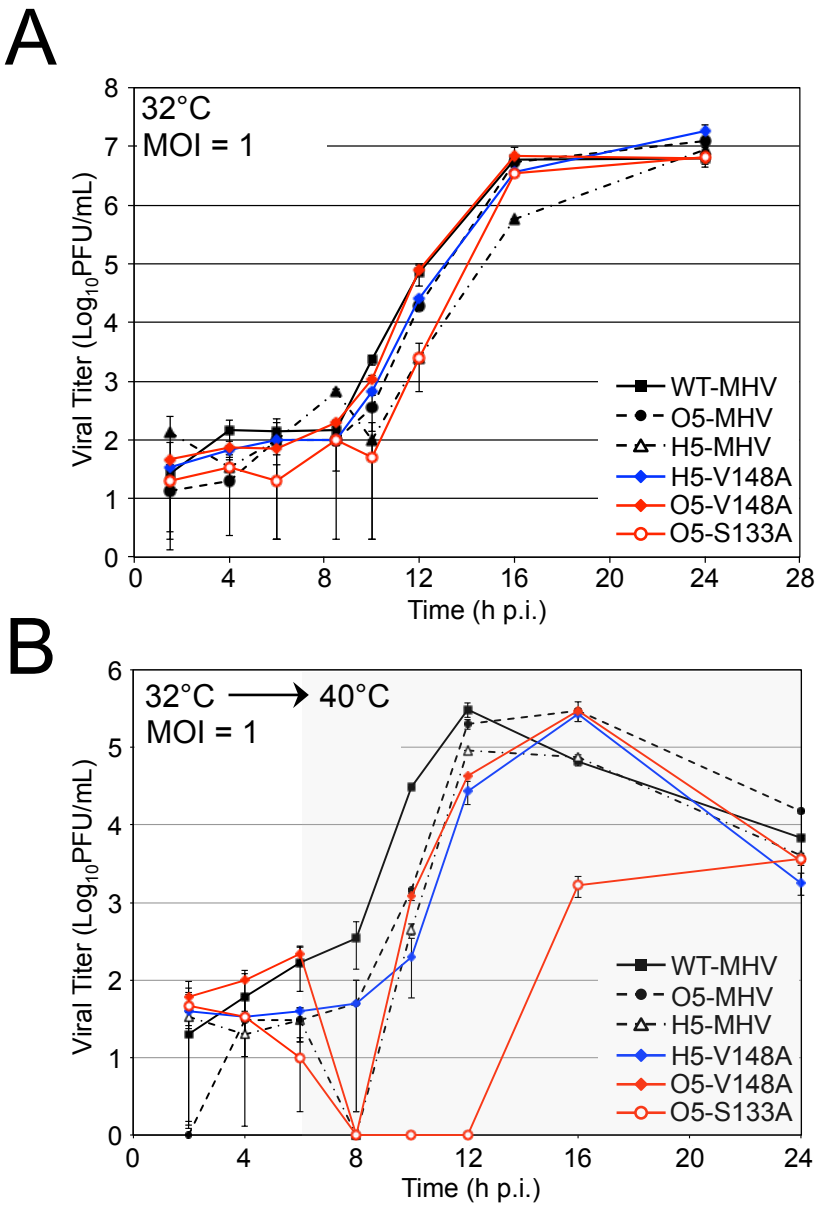


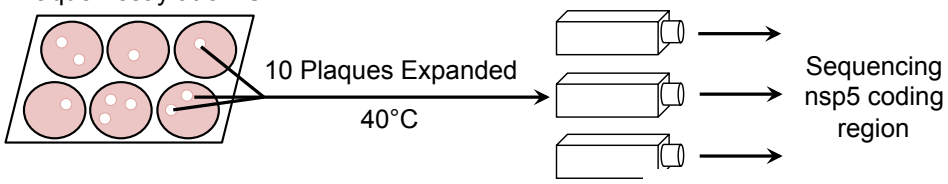
Fig. 3.8. Replication kinetics of H5- and O5-MHV viruses at 32°C and 40°C. A & B) DBT-9 cells were infected at an MOI of 1 PFU/cell. Samples were acquired from triplicate infections at various time points and the temperature either remained at 32°C (**A**) or was shifted to 40°C (**B**) at 6 hours post infection (h p.i.). Titers determined by plaque assay in duplicate per sample.

Expansion of the MHV temperature-sensitive viral mutants at 40°C resulted in the emergence of five separate second-site suppressor mutations in tandem that spanned domains 2 and 3 of MHV. To determine whether OC43 shares similar pathways of *ts* suppression, a plaque assay was performed at 32°C and 10 plaques were expanded directly at 40°C (**Figure 3.9**). Once signs of cytopathic effects (CPE) were visualized and the infection reached approximately 30 to 50% syncytial involvement, the RNA was isolated and the viral nsp5 coding regions were sequenced. Of the 10 plaques, one of the viruses failed to expand at the non-permissive temperature, 8 of the viruses contained the original *ts* S133A mutation in addition to either an A116V (8/9 clones) or N8Y (1/9 clones) and one of the viruses exhibited a primary reversion (A133S). Neither the A116 nor the N8 residues had previously been selected for by any of the MHV *ts* viruses. These data suggest that similar to MHV, OC43 may also utilize pathways of suppression throughout Domain 2, but that these pathways may not be identical to MHV.

Nsp5 allele Y134 provides resistance to the *ts* phenotype in OC43

The selection of an Y134 allele in MHV conferred partial or complete resistance to the *ts* phenotype associated with mutations V148A, S133A, or F219L. To test the hypothesis that the Y134 allele may play a similar role in the wild-type backgrounds of HKU1 and OC43, the Y134H mutation was engineered into the H5- and O5-MHV backgrounds in the presence of the S133A and V148A mutations. Only one virus was successfully recovered: O5-V148A/Y134H (O5-MHV containing V148A and Y134H mutations). Despite multiple attempts at recovery, the H5-V148A/Y134H, H5-

A



B

Codon Changes	Mutation	Plaques	Notes
AGU to GCC	S133A	8 / 9	Original *ts* Allele
GCU to GUU	A116V	7 / 9	Located in D2
AAU to UAU	N8Y	1 / 9	Located in D1 / Dimerization Interface
GCC to UCC	A133S	1 / 9	Primary Reversion

C

Virus	Virus Titers (PFU / mL)		EOP
	32°C	40°C	40°/ 32°
WT MHV	9.0×10^7	1.5×10^7	6.0
MHV *ts* V148A	3.0×10^6	2.9×10^2	9.7×10^{-5}
O5-V148A / Y134H	2.0×10^5	5.0×10^1	2.5×10^{-4}

D

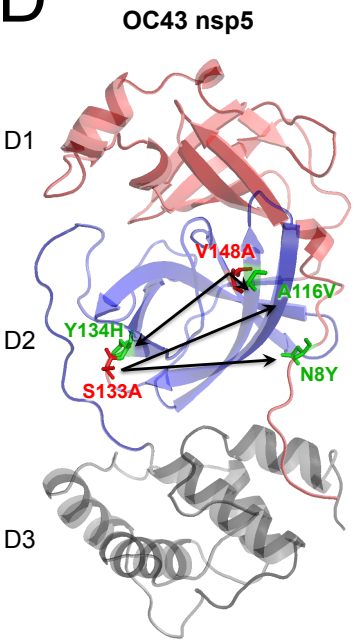


Fig. 3.9. Reversion analysis and temperature-sensitivity of O5-MHV mutants. (**A**) 10 plaques were selected from a plaque assay of O5-S133A at 32°C and were expanded in 25 cm^2 flasks at 40°C. At ~30 – 50% monolayer involvement in syncytia, the viral RNA was isolated and the nsp5 coding region sequenced. (**B**) Of the 9 clones sequenced, 8 contained the original S133A mutation and a second-site mutation at either A116V (7/9 clones) or N8Y (1/9) and 1 clone exhibited a primary reversion. (**C**) The location of the original S133A *ts* allele (red) and the identified second-site suppressor mutations (green) on a model of OC43 nsp5 generated using the crystal structure of HKU1 nsp5 (PDB code 3D23) and Modeller. (**D**) Virus titers for WT-MHV, MHV *ts*V148A, and O5-MHV virus containing V148A and Y134H mutations were determined by plaque assay on DBT-9 cells at either 32°C or 40°C. The efficiency of plating (EOP) was determined by the ratio of the average titers ($N \geq 2$) for either 40°C over 32°C.

S133A/Y134H, and O5-S133A/Y134H viruses could not be recovered. Sequencing of the nsp5 coding region of the O5-V133A/Y134H confirmed the presence of both the V148A ($_{10650}$GUU to G<u>CC</u>) and Y134H ($_{10608}$CAU to <u>U</u>AU) mutations and no other changes. To evaluate the ability of the viruses to infect and expand at the non-permissive temperature, a plaque assay was conducted at 32°C and 40°C using the MHV *ts* V148A virus and the O5-V148A/Y134H to determine the efficiency of plating for the viruses. The MHV V148A virus exhibited an EOP of 1.1×10^{-3}, consistent with previously published findings of temperature-sensitivity at the 40°C. Surprisingly, introduction of the Y134H mutation into the O5-V148A background resulted in an EOP similar to that of V148A (**Figure 3.9)**. These data suggest that the Y134H mutation confers sensitivity to elevated temperatures in the presence of the MHV *ts* alleles in the background of the O5-MHV virus and plays an important role in both MHV and OC43 proteases.

Summary

In this chapter, alpha- and betacoronavirus nsp5 proteases were substituted into the wild-type background of MHV. Only closely-related HKU1 and OC43 substitutions were tolerated and were associated with reduced viral replication kinetics and substantial fitness costs. Analysis of nsp5-mediated processing *in vitro* by MHV, HKU1, OC43, and SARS-CoV purified nsp5 proteases indicated that critical differences exist in protease substrate recognition and activity. These data indicate that the nsp5 protease and the remainder of the replicase gene may have tightly coevolved and suggests that genetic

exchange of nsp5 or regions containing nsp5 cleavage sites may be limited to closely-related viral species. Introduction of the MHV *ts* alleles S133A and V148A into the backgrounds of H5- and O5-MHV demonstrated that the function of these alleles may not be completely conserved indicating that even closely-related coronavirus nsp5 proteases may be structurally and functionally different. These studies shed light on the genetic relationship between coronavirus nsp5 proteases, highlight the putative limits of genetic exchange and identify key alleles that are functionally conserved between closely related betacoronaviruses MHV and OC43. The implications of these findings are discussed in more detail in Chapter IV and expand on themes and data discussed in Chapters II.

CHAPTER IV

DISCUSSION

Temperature-sensitivity and regulation of MHV nsp5

Previous studies with other viruses as diverse as HIV, Sindbis virus, poliovirus and vaccinia virus have reported *ts* mutations in virus proteases that affect protein processing, RNA synthesis, and virus capsid assembly (Chen et al., 2010; Graham et al., 2012; Jung et al., 2012; Vijgen et al., 2006). Typically these studies have tested abrogation and reversion of targeted functions to define broad strategies by which viruses may complement or suppress the *ts* phenotype. Most of the *ts* alleles of viral proteases have been proposed to occur at conserved residues, and in some cases of mutagenic scanning, to occur in pairs of conserved and structurally adjacent residues. In contrast, the data presented in Chapter II demonstrates that multiple non-conserved, structurally distant residues in the CoV nsp5 protease participate in communication within and between domains, function cooperatively to suppress ts phenotypes, and are important for nsp5 activity during virus polyprotein expression and processing. This conclusion is based on several lines of evidence: 1) The initially reported *ts*V148A mutant virus selected revertants at S133N and H134Y, both greater than 15 Å from V148A, the catalytic dyad, or from the other S133A or H134Y in the probable dimeric form (Sparks et al., 2008). 2) S133A constitutes a distinct *ts* mutant, and selected for partial suppression at T129M, 11 Å distant, along with directly juxtaposed H134Y. 3) The independent domain 3 *ts*F219L is greater than 40 Å from the active site cavity but had the same *ts* phenotype as S133A and V148A. 4) *ts*F219L also selected for suppression at

H134Y in domain 2 (28.8 Å distant), but of the second-site revertants identified, always in combination with a substitution in D3. 5) Engineering the combination of F219L with H270HH and E285V changes resulted in equal or greater suppression *ts*F219L than the combinations identified in the biological second-site mutant viruses. 6) The *ts* and suppressor mutations and combinations in nsp5 engineered recombinant genomes are necessary and sufficient to account for the observed replication and protein processing phenotypes. 7) The suppressor mutations alone in absence of the *ts* mutation have little or no effect on virus replication. 8) Finally, both the *ts* residues and second-site revertant residues are variable among nsp5 of over 226 unique nsp5 sequences. Thus the results support the hypothesis that the residues identified by our iterative mutagenesis and recovery approach are interconnected, communicating, and cooperative during nsp5 function in replication. These results have significant implications for understanding how nsp5 regulates the complex cascade of proteolytic events required for successful replication.

Since the SARS epidemic ten years ago, numerous studies have evaluated coronavirus nsp5 protease as a prime target for coronavirus inhibition due to its complex and essential role in the ordered processing and subsequent formation of the viral replication complex. Despite significant effort, however, the goal of developing a cross-reactive coronavirus nsp5 inhibitor effective at sub-uM concentrations has not been met, highlighting the need to understand the key differences between coronavirus nsp5 molecules and their relationship with the rest of the viral replication machinery. With the goal of better understanding the cross-reactivity of coronavirus proteases and their functional differences in the context of a replicating virus, we set out to substitute the

closely related coronavirus nsp5 proteases of HKU1 and OC43 into the MHV background. We have shown that permissive substitution of HKU1 and OC43 was associated with a clear fitness cost and that substitution of more distant coronaviruses was not tolerated. The recovery and characterization of chimeric nsp5 proteases provides a new tractable and safer system for the targeted study of nsp5 inhibitors directed against human coronavirus proteases in the context of a replicating virus. These findings demonstrate that the nsp5 protease and the 11 cleavage sites spanning the length of the replicase gene (nsps 4 – 16) of coronaviruses are structurally and functionally distinct and supports a tight co-evolution that limits viral emergence by recombination.

Long-distance communication in nsp5

The experiments in Chapter II were based on the observation that three independent and physically distant *ts* alleles in domains 2 and 3 resulted in emergence of the identical, structurally distant, revertant H134Y. The emerging data continued to identify structurally separate suppressor alleles and combinations, all of which would require additional pathways for propagation of communication. This cannot be accounted for by interactions across the dimeric structure, since the known orientation of monomers in solved dimer structures of SARS-CoV, IBV and HKU1 nsp5 allowed for calculation of distances between any pairs of residues (same or different) are greater in the dimer than the monomer. Thus other explanations for long-distance communication must be considered. Compared to a cellular protease with perhaps a single substrate or choice of substrates depending on environment, the CoV nsp5 must orchestrate the ordered

processing of 11 cleavage sites. In addition to having to choose between closely related and competing cleavage sites, nsp5 confronts dynamically changing substrate and cleavage site concentrations and structures, since cleavage at one site likely leads to altered folding of the remaining intermediate precursor and presentation of new cleavage sites. Any explanation for nsp5 intra- or inter-molecular communication must account for these constantly evolving variables. A straightforward explanation would be one in which nsp5 interacts with substrate cleavage sites and the intermediate replicase polyproteins, and those interactions result in allosteric communication through residues such as those identified in this study, with structural changes propagated throughout nsp5. This could occur by two or more possible mechanisms. The first would be a classical "induced fit allostery" in which interaction with the substrate results in modifications that increase the affinity for that substrate (Whitley and Lee, 2009). This would amplify any selection imparted by direct enzyme substrate specificity and favor a particular hierarchy. It would also suggest a highly adaptable enzyme that could fit itself to a variety of alternative cleavage sites on changing structures. A second model would be one variously referred to as "ensemble" or "equilibrium allostery", where the enzyme exists as a non-negligible equilibrium of variable states, which in the case of nsp5, might represent conformations favoring different cleavage sites or intermediates (Whitley and Lee, 2009). In this model, interactions with substrates or other proteins would signal through allosteric interactions resulting in a shift toward a conformation more favorable for the immediate environment. This would allow for a rapidly shifting and potentially mixed environment of substrate and protease. Differentiating between these models in a biochemical *in vitro* system may

be difficult, since only in the infected cell can the dynamic of all polyprotein forms and local concentrations be reproduced.

However, neither of these models identifies the mechanism by which local changes are propagated between communication nodes in protein structure. Propagation of changes could occur and be amplified by a direct mechanical process, or by altering protein flexibility. Recently, approaches have been developed to predict the most energetically favorable and direct pathways for connecting starting and ending nodes in allosteric communication, using combinations of structural data and multiple sequence alignments of all residues and probable direct interactions (Atilgan et al., 2004; Pyrc et al., 2010; Tang et al., 2007). This has been applied to known allosteric interactions, specifically with myosin (Tang et al., 2007). Application of such an approach with nsp5 might be informative in predicting possible propagation from the identified *ts* and second-site alleles to each other, the active site cavity, substrate binding residues and dimerization interfaces. This in turn might identify additional residues in a communication network. Ranganathan and coworkers have proposed a model by which proteins may consist of a series of functional sectors that are connected in tertiary structure and exhibit unique roles in the protein (Pyrc et al., 2010). In both S1A serine proteases and HIV protease, regions distant from known catalytic and functional determinants have been shown to be critical for protease activity (Perryman et al., 2004; Pyrc et al., 2010). Nearly all of the biological suppressor mutations identified in this study contain combinations of residues rather than a single suppressor mutation, suggesting that the capacity of a single residue to complement or mitigate nsp5 cleavage defects of the *ts* mutation is greatly limited. These data suggest that active long-distance

communication between suppressor residues and functionally active residues is critical to restoration of nsp5 function. Finally, alteration in protein flexibility could provide an explanation for how nsp5 functions within the polyprotein and in rapidly changing partially processed intermediates. If the residues we identified in this study function as hinges, pivots, or axes in a flexible structure, then changes at those residues would be immediately manifested as changes in protein flexibility, possibly altering the range or type of substrate of catalysis possible. A most likely scenario is some combination of signal propagation, amplification and protein flexibility modification.

Phenotype and mechanism of MHV nsp5 *ts* perturbations

The phenotype of the nsp5 *ts* mutant viruses may be due to incomplete folding, thermal instability or flexibility, inability to dimerize, failure to recognize and cleave specific cleavage sites, or disruption of allostery in the replication complex. Analyses of the *ts* mutant viruses indicated a defect in nsp5-mediated processing and virus replication at the non-permissive temperature. Yet, mature nsp5 cleavage products were still detected for all three *ts* mutant viruses at non-permissive temperature, albeit at profoundly reduced levels, suggesting that defects in viral growth at the non-permissive temperature were not the result of complete loss of nsp5 activity, but may be the result of a shift in enzyme activity after temperature change. These findings indicate that the source of the *ts* phenotype is not due to reduced catalysis, but on choice of substrate. We previously reported that nsp5 activity was not present in *ts*V148A at non-permissive temperature. However, we did not directly investigate nsp5 cleavage in that study, and thus in this

study, we determined that *ts*V148A functions like *ts*S133A and *ts*F219L in exhibiting more impaired processing of nsp8 than nsp5 at non-permissive temperature, but not complete ablation of nsp5 activity (Sparks et al., 2008). This is similar to the effect of mutations introduced in the interdomain loop (IDL) between domains 2 and 3 of the arterivirus nsp4, a smaller but structural orthologue of CoV nsp5 (van Aken et al., 2006). Several of the recovered IDL mutant viruses of EAV did not abolish activity, but rather altered the substrate specificity. It is possible that the *ts* mutants of nsp5 are acting in a similar manner.

Non-conserved residues in nsp5 function.

Analysis of a CoV nsp5 MSA indicates that all three currently identified residues that resulted in *ts* mutations (S133, V148, and F219) were non-conserved residues with at least two different amino acids occupying each position. Similarly, at each of the residues that were involved in a second-site reversion, at least three different amino acids occupy those positions in other CoVs. Both the H134 and S133 second site revertants had changes (Y134, N133) that were already present in numerous other CoV nsp5 sequences. An MSA of 226 unique CoV nsp5 sequences demonstrates that there is variability at every residue represented in our study and much less variability within defined CoV groups. Specifically, the betacoronavirus phylogenetic group 1 (Group 2a) CoVs, including MHV and HCoVs HKU1 and OC43, show variability only at H134 (Y,V respectively) and H270 (Y, K respectively), while the phylogenetic group 2 (Group 2b) CoVs show variability only at H134 (Y). In contrast, alphacoronavirus phylogenetic

group 2 (Group 1b) CoVs show variability at all residues except P148. While the results have to be interpreted in light of the overall nsp5 sequence variability or conservation, it is possible to speculate that these residues may represent an evolutionarily adaptive system co-evolving with other interacting proteins or cleavage sites. Recently, it was reported that a mutation at the P1 position in the cleavage site between nsp15-16 of IBV resulted in a debilitated virus whose phenotype was compensated by a mutation in nsp5 (Vijgen et al., 2005). This supports the concept of co-evolution of nsp5 with its cleavage sites. A system of mutationally variable residues could account for the rapid emergence of second-site revertants of *ts* viruses with combinations of more than one second-site mutation that also function in combinations not derived during virus reversion analysis.

Models for testing nsp5 long-distance communication.

The findings of this study highlight a common obstacle in confirming and understanding the role of long-distance communication in protein function. In the absence of a specific mechanism for how long-distance communication alters protein activity, several approaches may be proposed to address the implications of these proposed nsp5 communication pathways. First, future studies may express, purify, and characterize MHV nsp5 *in vitro* to identify the impact of each individual *ts* and second-site suppressor mutation or combinations of mutations to understand the role of each mutation in the function of the protease. Study of the structure or function of the *ts* and second-site mutants, ideally at different temperatures, may provide new insight into possible structural perturbations attributed to the mutations at the non-permissive

temperature. However, the optimum temperature and conditions for crystallization may prevent study of the protease structure in its non-functional, *ts* state. Second, comparative modeling of available nsp5 structures and use of a computational bioinformatics approaches may provide new insight into other potential long-distance communication nodes or pathways spanning nsp5 structure and its 11 respective cleavage sites. Third, iterative mutagenesis of the second-site residue alleles might identify additional interacting nodes that expand the interactions detailed in this study. Lastly, introduction of these mutations into other CoVs would provide evidence whether these pathways of perturbation are a conserved feature of CoV protease structure. Collectively, these findings strongly suggest that small local changes at a single residue of CoV nsp5 may have broad and major implications on both protease activity and viral replication.

Coronavirus nsp5 proteases differ in activity and have tightly co-evolved

Earlier studies have emphasized the close nsp5 recognition site homology between coronaviruses and the high density of conserved residues in the protease substrate-binding pocket (Fan et al., 2005; Goetz et al., 2007; Hegyi and Ziebuhr, 2002). However, the inability to recover the SARS-CoV nsp5 substitution chimera suggests that structural conservation of nsp5 may be limited and that the subtle differences in the substrate-binding pocket and other determinants of coronavirus proteases result in major functional differences. Analysis of substrate specificity by Chuck et al., indicated that clear preferences for specific amino acids in the nsp5-recognition sequence exist between even closely-related coronaviruses (Chuck et al., 2011). In this study, the relative rates of

processing of MHV and SARS nsp4/5 and 7/8 cleavage sites by all four proteases varied among all four sites assayed *in vitro*. This finding is consistent with other published cleavage site assays suggesting that the tertiary structure conservation across coronaviruses of different groups and clades does not directly equate to functional conservation (Chuck et al., 2011; Fan et al., 2005; Goetz et al., 2007; Hegyi and Ziebuhr, 2002). These major differences in the relative rates of processing of the nsp5 cleavage sites may represent differences in the timing and ordered hierarchy of nsp5 cleavage events between coronaviruses.

Surprisingly, for all four cleavage site constructs surveyed, the purified SARS protease was consistently the slowest enzyme with a greater than 3-fold reduction in rate relative to MHV and six-fold relative to HKU1 and OC43 at the SARS nsp4/5 site. A similar study reported that NL63 exhibited a 2.5-fold increased rate towards the SARS nsp4/5 site relative to the SARS-CoV protease (Chuck et al., 2011). The relative rates of processing by OC43 were consistently among the highest surveyed towards each cleavage site construct which supports earlier findings that have suggested that the OC43 nsp5 protease may be the most versatile in accepting changes in the recognition peptide sequence (Chuck et al., 2011). Similarly, the higher relative rate of processing by OC43 compared to HKU1 may account for the higher fitness of O5-MHV in direct competition to H5-MHV. These findings, however, do not account for the known role of the intracellular milieu and differences in cell environment affecting protease activity during virus infection (Okamoto et al., 2010). Further studies will be needed to evaluate other cleavage sites and the role of the chemical environment in the activity of the proteases *in vitro*. These differences in protease activity and the inability to recover a SARS-CoV

nsp5 substitution chimera with substitution of all 11 SARS-CoV cleavage sites strongly supports the hypothesis that the processing of the replicase gene and the formation of the viral replication complex is dependent on the compatibility between the protease and replicase polyproteins.

Conserved pathways of nsp5 protease function are clade-specific

Recently, our lab reported the identification of three temperature-sensitive alleles (S133A, V148A, and F219L) in MHV that resulted in reduced viral titers and impaired nsp5-mediated processing at non-permissive temperatures (Sparks et al., 2008; Stobart et al., 2012). Growth of these viruses at the non-permissive temperature resulted in the emergence of five second-site suppressor alleles (T129M, S133N, H134Y, H270HH [duplication], and E285V), which arose largely in combinations and compensated for the *ts* phenotype. Surprisingly, the Y134 allele was independently identified as a suppressor allele for all three *ts* mutations. These studies were limited, however, to the scope of MHV. In this study, we show that the *ts* alleles S133A and V148A also result in a temperature-sensitive phenotype in the nsp5 backgrounds of OC43, although V148A necessitated the addition of the Y134H mutation. The inability to recover the H5-S133A mutation and the temperature-sensitivity of the O5-S133A virus highlights the importance of this region and specific residue in the stability and function of betacoronavirus-1 nsp5 proteases. The selection of the Y134 allele in the wild-type strains of HKU1 and OC43 and the inability to tolerate an Y134H mutation in the presence of either S133A or V148A mutations in HKU1 support the hypothesis that this

allele represents a critical structural adaptor residue that is conserved among betacoronaviruses but may function in subtly different ways.

The inability to recover HKU1 and OC43 mutants containing the domain 3 F219L mutation supports the findings in MHV that this residue and region of the protease play an important role (albeit not yet entirely understood) in the coronavirus nsp5 structure. This is the first study to identify temperature-sensitive mutations in the nsp5 from a human coronavirus. These mutations indicate that there may be conserved functional pathways between the betacoronavirus-1 strains of MHV, HKU1, and OC43. Combined, these data provide new insight into conserved betacoronavirus functional pathways and reveal new putative targets for coronavirus inhibitor design.

Nsp5 represents a genetic restriction to coronavirus recombination

Numerous studies have postulated that recombination readily occurs in coronaviruses based upon bioinformatic analyses and the presence of regions of particularly high homology throughout the genome of closely related coronaviruses (Lau et al., 2011; Vijgen et al., 2006; Woo et al., 2009; Woo et al., 2012; Woo et al., 2006). The emergence of HCoV-EMC in Saudi Arabia provides a recent example of the potential detriment to human health created by coronavirus evolution and emergence. It is known that HCoV-EMC and SARS-CoV virus likely evolved from bat coronaviruses based upon close homology with known bat coronaviruses sequences (Becker et al., 2008; Lu and Liu, 2012; Perlman and Netland, 2009; Perlman and Zhao, 2013). SARS-CoV exhibits sequence homology to several identified bat coronavirus genomes including

Bat SARS Rp3, Bat Coronavirus 273.2005 and Bat SARS HKU3, and recent findings from our group and others have now shown that SARS-CoV appears to have emerged directly from bats (Becker et al., 2008; Perlman and Netland, 2009). HCoV-EMC also exhibits sequence homology to known bat coronaviruses HKU4 and HKU5 (Zaki et al., 2012), suggesting that emergence of an HCoV directly from an animal reservoir is not an event isolated to SARS-CoV. Bioinformatic studies evaluating the time of emergence of BCoV, OC43, and PHEV suggest a recent common ancestor approximately 100 to 200 years ago (Vijgen et al., 2006). Additional studies evaluating the genetic diversity of human coronavirus HKU1 and OC43 have suggested that the three HKU1 and four OC43 genotypes likely emerged due to recombination events within the replicase gene (Lau et al., 2011; Woo et al., 2010; Woo et al., 2009; Woo et al., 2006). Additionally, other regions of the virus genomes also show a high degree of structural and functional homology (Dijkman et al., 2012).

Upon cursory examination, our studies indicate that the nsp5 protease of closely-related coronaviruses and the 11 cleavage sites of the replicase polyprotein may not represent a block to recombination and could support the feasibility of recombination described in these studies. However, no known human coronaviruses displaying recombination between HKU1 and OC43 have been identified to date despite several clinical surveys showing individuals co-infected with both viruses (Cui et al., 2011; Mackay et al., 2012; Prill et al., 2012),

The recovery of the H5- and O5-MHV chimeric viruses in our investigations, although permissive to direct nsp5 substitution, resulted in a direct drop in virus fitness compared to the wild-type MHV strain indicating that there is a strong selective pressure

against nsp5 recombination even within these closely related coronavirus backgrounds. Furthermore, the inability to recover chimeric MHV viruses containing the more-distant human coronavirus nsp5 proteases of SARS-CoV, 229E, NL63 or with the bat coronavirus HKU4 (which is now known to exhibit 82% sequence identity to human coronavirus EMC) suggests that clear differences in the structure and function of the nsp5 proteases and potentially cleavage sites of these coronaviruses limit viable recombination events within the replicase gene. More specifically, we hypothesize that the limitation may due to one or more of the following reasons: i) misfolding of the protease or polyprotein in the chimeric background, ii) disruption of dimerization due to inaccessibility of the dimerization interface in the background of the chimeric MHV polyprotein, iii) inability to recognize or process one or more critical nsp5 cleavage sites, or iv) disruption of essential interaction with other members of the replication complex. More direct study of coronavirus recombination is necessary to identify other key coronavirus features that permit or restrict viable recombination events. Collectively, these data demonstrate that although human coronaviruses HKU1 and OC43 share a common ancestor and many similar structural and functional features, nsp5 protease and possibly its cleavage sites restrict the potential for coronavirus emergence by recombination events even between closely related coronaviruses.

CHAPTER V

SUMMARY AND FUTURE DIRECTIONS

Introduction

Before the work described in Chapters II – IV, considerable progress had been made in evaluating coronavirus nsp5 protease structure, activity and specificity *in vitro*. Early virus studies demonstrated that nsp5 was required for virus replication and that it could be targeted with available inhibitors *in vitro* (Lu et al., 1996; Lu and Denison, 1997; Ziebuhr et al., 2000). From the limited studies of nsp5 in the context of a replicating virus, many key questions remain regarding the critical steps of nsp5 activity during viral infection (**Figure 5.1**). Upon translation, nsp5 must first properly fold (1) and stabilize itself (2). Once folded and stable, it must dimerize (3) to attain full function before cleaving itself completely out and processing the remaining cleavage sites (4). Lastly, it must coordinate with the rest of the replication complex (5). However the mechanism of all of these crucial steps remains unclear. Several key questions remained unanswered: 1) How does nsp5 cleave itself from the replicase polyproteins pp1a and pp1ab? 2) Which regions or motifs are essential for nsp5 folding and stability? 3) How does nsp5 orchestrate the ordered processing of up to 11 distinct cleavage sites? 4) In the context of virus inhibitor design, how conserved is nsp5 function and can a pan-coronavirus nsp5 inhibitor be developed? 5) Which determinants of nsp5 structure are critical for nsp5 specificity? 6) How does nsp5 associate with other members of the coronavirus replication complex? 7) Does nsp5 play a role in down-

regulation of host cell responses or acquisition of host resources? My contributions to answering some of these critical questions are described in Chapters II – IV. In this chapter, I will discuss the implications of my work and future directions in the field to better understand the role of nsp5 protease during viral infection. To begin to address these critical remaining questions, my work began by building on the *ts*V148A mutant described by Jen Sparks.

Coronavirus nsp5 *ts* viruses

In Chapter II, a series of two temperature-sensitive alleles (*ts*) and four second-site suppressor alleles, all distant from known catalytic regions, are identified and described in work stemming from the initial characterization of the *ts*V148A virus. To understand the second-site suppressor alleles of MHV V148A, alanine mutagenesis was applied to generate S133A and H134A viruses. Quite surprisingly, it was discovered that the same residue capable of suppression of a *ts* phenotype (in the case of V148A) happened to also be a *ts* allele when mutated to an alanine. Furthermore, reversion analysis of the *ts*S133A resulted in the emergence of two different patterns of suppression: S133A/H134Y and S133A/T129M/H134Y. In a follow-up to a paper published by Sawicki et al., a putative domain 3 *ts* allele (F219L) was introduced into MHV and shown to also demonstrate temperature-sensitivity. Reversion analysis of *ts*F219L resulted in the identification of an additional two patterns of reversion: F219L/H134Y/E285V and F219L/H134Y/H270HH.

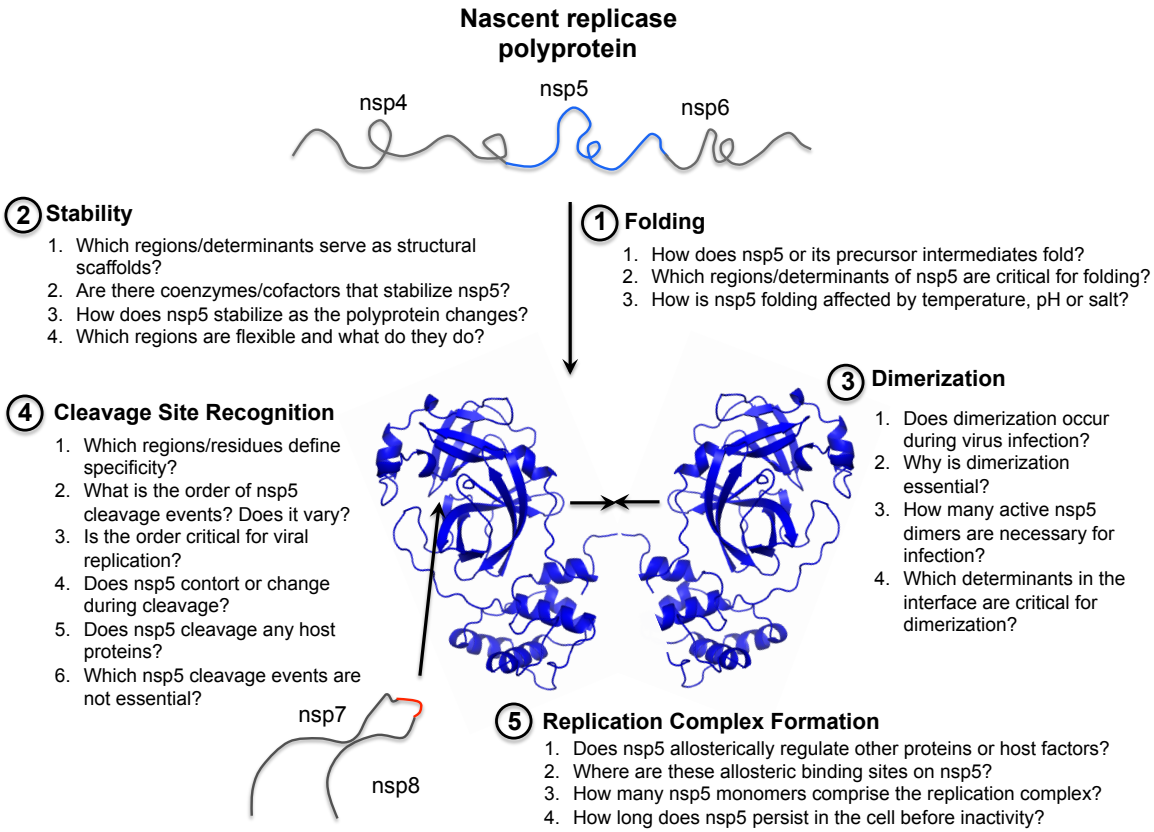


Fig. 6.1. Model of coronavirus nsp5 activity and key questions. The diagram demonstrates what currently is known regarding nsp5 activity during viral infection. Key questions that this work and others are actively working on are shown in the diagram.

All three *ts* mutations (S133A, V148A, F219L) resulted in reduced viral replication kinetics at the non-permissive temperature of 40°C, altered nsp5-mediated polyprotein processing, and were capable of partial or full suppression by a common suppressor allele H134Y. These findings identified new critical determinants and regions, previously unknown, which are important regulators of nsp5 activity and of structural stability. Additionally, these residues, which are non-conserved, highlight crucial long-distance communication nodes spanning domains 2 and 3 in the structure of nsp5. The significance and discussion of these results are described in Chapter IV. Future directions will focus on defining the role of these mutations and the structural basis for their interactions.

Biochemical analysis of MHV nsp5 *ts* lesions

In work described in Chapter II in collaboration with the Mesecar lab at Purdue, the MHV *ts* mutation V148A and the second-site suppressor mutation H134Y were introduced into an expression vector and purified for *in vitro* characterization to understand the mechanism by which the V148A lesion disrupts nsp5 function and how the H134Y suppresses the V148A-associated defect at non-permissive temperatures. Initial recovery of the V148A and V148A/H134Y proteases required lower bacterial incubation temperatures. Although viewed initially as a technical obstacle only, this early finding indicated that there were subtle differences in the protease prior to purification. To assess the temperature-sensitivity of the proteases, a thermal inactivation assay was established whereby the enzymes would be incubated at varied temperatures and

analyzed for residual proteolytic activity over time. These data demonstrated that at temperatures of at least 40°C, the V148A protease demonstrated substantially reduced residual activity over time relative to either WT MHV nsp5 or the suppressed V148A/H134Y. Subsequent analysis using circular dichroism suggested that the secondary structure of the V148A nsp5 mutant was considerably less stable than WT-MHV indicated that the V148A mutation was disrupting nsp5 structural stability. Furthermore, the H134Y exhibited increased stability relative to WT and may highlight a critical interacting node for nsp5 folding or stabilization. These findings (which are discussed in more detail in Chapter IV) highlight key structural motifs and regions, which are important for maintaining nsp5 structure and stability, and may show critical regions involved in nsp5 folding. All of these studies were carried out in the context of MHV. An important future direction is to define their relevance to the nsp5 of other coronaviruses including distantly related alpha- and gammacoronaviruses.

Coevolution of intra- and intermolecular networks in coronavirus nsp5

In Chapter III, the conservation of nsp5 structure and function was examined by substitution of the nsp5 coding region from alpha- and betacoronaviruses NL63, 229E, SARS-CoV, Bat HKU4, HKU1 and OC43 into the background of MHV. Despite appreciable tertiary structure conservation and similar nsp5 cleavage sites, only closely-related betacoronavirus HKU1 (H5-MHV) and OC43 (O5-MHV) nsp5 proteases permitted virus recovery. Interestingly, these viruses exhibited similar replication kinetics (albeit slightly lower than WT MHV) and no adaptive mutations. However, in

competitive fitness assays, neither virus demonstrated comparable fitness to WT indicating a clear fitness cost with genetic exchange of the nsp5 protease even between closely-related species. Analysis of the MHV, HKU1, OC43 and SARS-CoV nsp5 proteases in vitro demonstrated that clear differences in activity and specificity exist among betacoronavirus nsp5 proteases. Introduction of the MHV *ts* alleles into the chimeric H5- and O5-MHV backgrounds showed that subtle structural differences exist and that the function of these alleles although similar is not necessarily conserved. Collectively, these data demonstrated that coronavirus nsp5 proteases are highly coevolved and that genetic exchange, deemed common between coronaviruses, may be limited in substitution of nsp5 or its associated cleavage sites to closely-related species.

Generation of chimeric MHV viruses with more distant coronavirus nsp5 proteases

In Chapter III, I described my efforts to introduce the nsp5 proteases of alpha- and betacoronaviruses into the background of MHV. Only closely-related betacoronaviruses HKU1 and OC43 permitted direct substitution into the background of MHV. Substitutions of SARS-CoV, HCoV-229E, HCoV-NL63, and Bat Coronavirus HKU4 were not tolerated despite at least 3 attempts with each of the designed MHV constructs. In addition, several approaches were used to try to recover chimeric MHV with SARS nsp5. In summary, these approaches include the introduction of SARS nsp5 cleavage sites in various mixtures, using MHV F219L *in trans* to aid initial replication, expression of MHV WT nsp5 *in trans,* and introduction of the ExoN mutations (knockout of nsp14-associated proofreading activity). These approaches are summarized in this section in the

hopes that future generations of graduate students will know what has been tried in recovery of this virus and potential other directions to pursue.

Initial efforts at recovery of the SARS virus consisted of trying to swap only the SARS nsp5 protease into the background of MHV. Xiaotao initially carried out these studies. Despite multiple attempts, no signs of infection were observed. As I described in Chapter III, PCR using oligonucleotides specific for subgenomic RNA of N, M, E and S failed to amplify detect any signs of late stages of infection of this SARS chimera. Once direct recovery was deemed unlikely, we decided to substitute all 11 SARS-CoV cleavage sites (P5 – P2') into the MHV infectious clone fragment sequences. The nsp5 cleavage sites are distributed throughout the MHV infectious clone fragments as shown in Table 5.1. Several attempts using all of these sites combined with the SARS-CoV nsp5 protease (with and without its flanking nsp4 and nsp6 cleavage sites) were unsuccessful. In addition, I have also tried to the following fragment combinations: C alone, D alone, E alone, F alone, C+D+F, and C+D+E alone. In each of these cases, there was no detectable signs of replication.

Table 5.1. Locations of nsp5 cleavage sites in MHV infectious clone fragments

MHV Infectious Clone Fragment	***Nsp5 Cleavage Sites***
C	nsp4/5 and nsp5/6 (2)
D	nsp6/7, nsp7/8 and nsp8/9 (3)
E	nsp9/10 and nsp10/12 (2)
F	nsp12/13, nsp13/14, nsp14/15 and nsp15/16 (4)

In addition to cleavage site substitutions, we have also tried virus recovery of the SARS-CoV chimera with MHV nsp5 protease transfected and stably expressed in DBT-9 cells. Xiaotao observed initial replication of the SARS-CoV virus when nsp5 was

transfected in DBT-9 cells, however the virus was not able to be passaged and the replication remained at a low level. This observation was not able to be repeated. The DBT-9 cell line stably expressing nsp5 protease is available and stored for potential use. In an additional experiment, the chimera recovery was attempted in cells that had been previously infected with MHV F219L at 30°C. Upon addition of the cells, the flask was shifted to 40°C to inactivate the MHV F219L virus. No changes in CPE were observed following the temperature-shift.

We have begun experiments to try to express both SARS-CoV nsp5 and MHV nsp5 in a common virus background, swap domains of SARS-CoV nsp5 into MHV, and swap individual cleavage sites for virus recovery. All of these represent future potential directions. However, it will be critical to identify which sites are fundamentally processed different between the proteases and identify critical allostery or intermolecular regulation within the background of MHV that may have been altered or disrupted upon chimeric substitution.

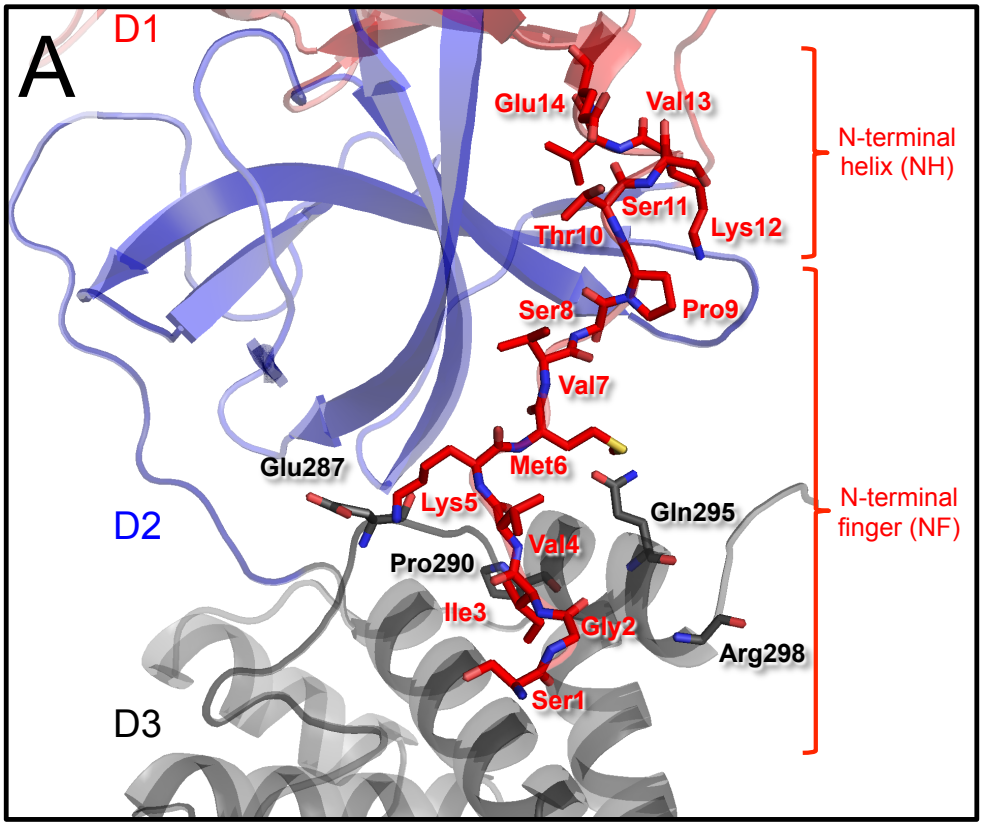
Future studies: Dimerization and nsp5 activity

Coronavirus nsp5 protease is responsible for recognizing and processing 11 different cleavage events including its own autoproteolytic maturation cleavage. *In vitro* studies have shown that expressed and purified coronavirus nsp5 protease exists as a mixture of monomeric and dimeric forms. Functional assays suggest that the dimer is the predominant form comprising approximately 65% of the detectable protease and is the primary functional active form. Early studies have suggested that dimerization was

critical for nsp5 protease function. However, a recent publication has shown that the monomeric form may be capable of the amino-terminal autocleavage event prior to dimerization. No studies have tested this hypothesis in the context of a replicating virus. X-ray crystallographic studies and biochemical analyses have demonstrated that the dimeric form is structurally stable and have identified several important interacting motifs and determinants.

The amino terminus of coronavirus nsp5 is comprised two distinct functional motifs, an amino terminal finger (NF) and amino terminal helix (NH) (**Figure 5.2**). Studies using purified SARS-CoV nsp5 have demonstrated that the first ten amino acids (Ser1 – Thr10) comprising the amino-terminal finger are critical for dimerization and nsp5 activity. Several inter- and intramonomer interactions are predicted based on structural models including Arg4-Glu290 and Met6-Arg298. Deletion of residues 1 – 3 in SARS-CoV nsp5 resulted in a 50% reduction of catalytic activity while retaining dimerization. In contrast, deletion of the first four residues resulted in a loss of dimerization and 2 log reduction in enzymatic activity *in vitro*. Residues Thr10 – Glu14 comprise a small α-helical region (termed the amino terminal helix, NH) and alanine substitution of Gly11 resulted in a complete loss of dimerization and enzymatic activity. Many of the residues in the NF are predicted to interact with regions of the alpha-helical carboxy terminus of domain 3.

The coronavirus nsp5 amino-terminal motifs and domain 3 have been shown to play an important role in biochemical and *in vitro* studies. However the interactions and predicted functions of these regions have yet to be tested in the context of a replicating



B

	N-terminal Finger (NF)			N-terminal Helix (NH)
	1	4	6	11
MHV	S	I	VKMVSP	TSKVE
SARS-CoV	S	F	RKMAFP	SGKVE
229E	A	L	RKMAQP	SGFVE
IBV	S	F	KKLVSP	SSAVE

Fig. 5.2. Amino-terminus of nsp5 Protease. A) Structure of MHV nsp5 protease (modeled on HKU1 structure [PDB code 3D23] showing the amino-terminal residues. Putative dimerization determinants in domain 3 (D3) are also shown. The regions for the amino-terminal helix (NH) and amino-terminal finger (NF) are also designated. **B)** Alignment of NF and NH residues from MHV, SARS-CoV, 229E and IBV. Select residue numbers are shown.

virus. Consequently, several key questions exist: 1) Does active mature nsp5 dimers exist during virus infection and how are they formed? 2) Does nsp5 dimerize before processing its amino-terminus or is autoproteolysis required for dimerization? 3) Which residues and predicted interactions are critical for nsp5 dimerization and enzymatic activity? 4) How structurally and functionally conserved are the dimerization interfaces of coronavirus nsp5 proteases? 5) Does domain 3 mediate nsp5 dimerization?

To address these questions, I propose the following approaches. First, mutations introduced into the dimerization interface, which are known to abrogate dimerization in vitro, may be introduced for virus recovery in the MHV, H5-MHV and O5-MHV viral backgrounds. Second, nsp5 may be immunoprecipitated and run on a native gel in the presence and absence of heat to evaluate whether active dimers can be detected by size and gel mobility. Third, based on studies evaluating cleavage site specificity in vitro, cleavage sites may be swapped at the nsp4/5 and nsp5/6 sites with less efficiently processed sites to identify whether: (a) reordering of sites prevents viral recovery; (b) alteration of polyprotein intermediates may be detected among recovered viruses; (c) unique pathways of reversion exist that are intrinsically tied to specific sites or processing events. Collectively, these approaches may shed new light on the role of the amino-terminal region of nsp5 and the potential for dimerization during viral infection.

Future studies: Role of the domain 2-domain 3 interdomain loop (IDL)

Nsp5 is structurally conserved among coronaviridae and is similar to the nsp4 protease of distantly-related arteriviruses. Domains 2 and 3 of coronavirus nsp5 and

arterivirus nsp4 are connected by a conserved interdomain loop (IDL) structure. Equine arteritis virus (EAV) nsp4 is 204 amino acids in length and uses a catalytic triad (H39-D65-S120) for replicase polyprotein cleavage. Mutagenesis of the seven residue EAV nsp4 IDL resulted in reduced viral fitness or lethality and alteration in proteolytic processing. The MHV IDL consists of approximately 15 residues (Pro184-Thr199) and remains structurally conserved among all known nsp5 crystal structures (**Figure 5.3**). Two residues in the MHV IDL, D187 and Q192, are 100% conserved in all known nsp5 sequences. Predictive modeling of these residues suggests that they play an important role in stabilizing the horseshoe structure at the amino terminus of the loop. However, no viral, structural, or biochemical studies have evaluated the functional role neither of the IDL nor of these residues. It has been proposed that the IDL may function as a flexible linker with domain 3 or a hinge to permit active protease function. These hypothesis have yet to be tested. Consequently, many key questions remain: 1) What is the role of the domain 2-domain 3 IDL? 2) How does alteration of the length or flexibility of the IDL impact nsp5 function *in vitro* and during virus infection? 3) Which residues or determinants of the nsp5 IDL are indispensible for protease activity and/or virus viability?

During study of the MHV *ts* and suppressor alleles described in Chapter II, I initiated mutagenesis studies of the residues of the MHV IDL in order to evaluate its functional role in nsp5 activity and virus replication. In all, nine different mutations were attempted in addition to mutagenesis for several others (**Table 5.1**). The following virus recoveries attempted were: P184A, D187A, D187Q, Q192A, Q192N, P194A, Q196A, Y198A and T199A. Four viruses were successfully recovered and sequence confirmed

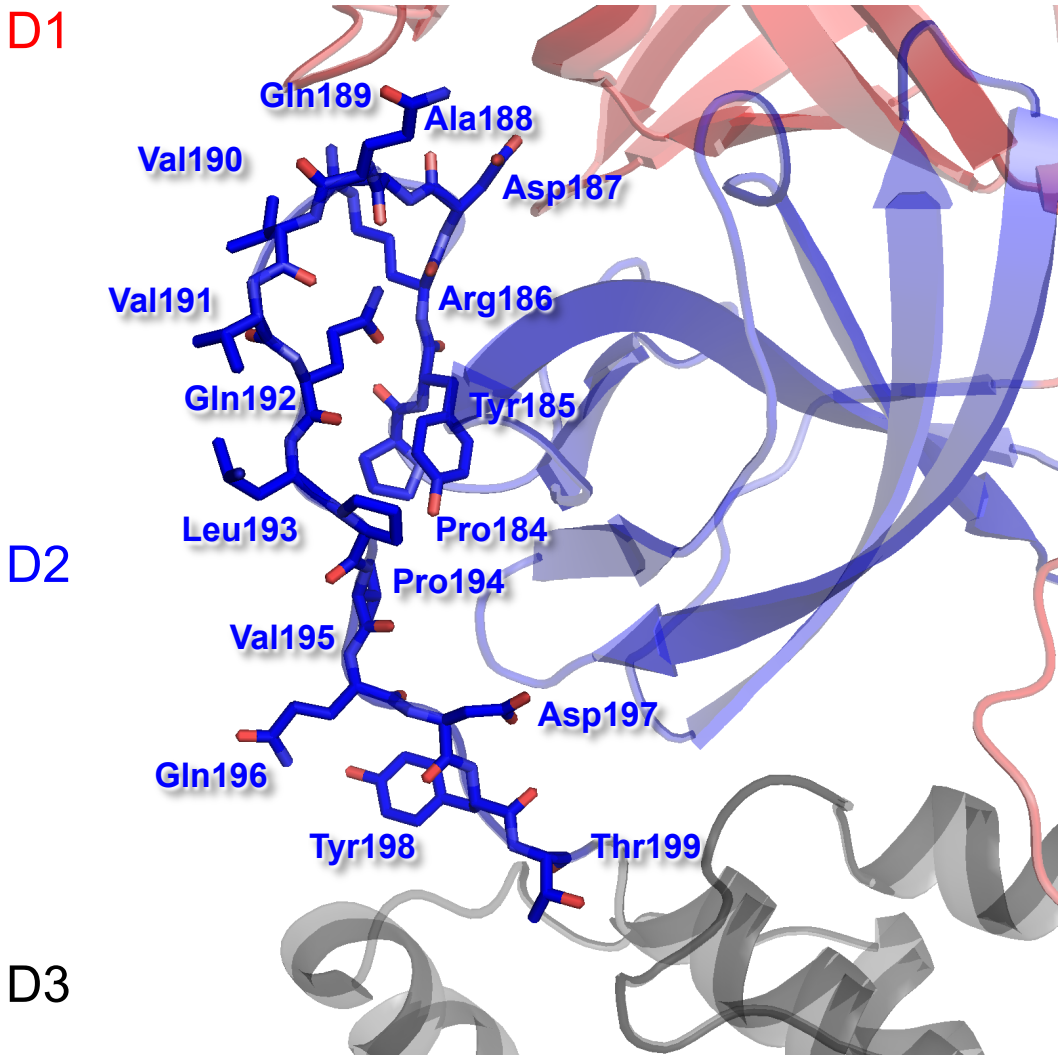


Fig. 5.3. Coronavirus nsp5 domain2-domain3 interdomain loop. The interdomain loop residues (184 – 199) are shown and labeled on a modeled structure of MHV nsp5. Domains 1 -3 are color coded and labeled.

confirmed: P184A, P194A, Q196A and T199A. However, no further analysis of these viruses was attempted. The inability to recover mutants of residues D187 and Q192 demonstrate an important functional role of this loop structure. Ongoing and future work may evaluate the replication of these viruses and continue to mutagenize and alter the length of the nsp5 IDL. Efforts at identifying critical determinants and region may want to focus exclusively at the following areas: (a) 100% conserved residues D187 and Q192, (b) the horseshoe loop which is structurally conserved in all known nsp5 crystal structures, (c) the linker regions connecting the loop to domains 2 and 3, (d) the role of length and flexibility in nsp5 function. In addition to mutagenesis, other potential approaches to studying the loop include biochemical characterization (thermal stability, structural flexibility by Rosetta predictions or cryo-NMR characterization, and catalysis upon mutagenesis) and chimeric exchange to evaluate conservation of the structure and its function.

Table 5.2 Site-directed mutagenesis of the MHV IDL

Progress towards mutating the MHV D2-D3 IDL		
Mutation	*Cloning Status*	*Virus Recovered?*
P184A	Cloning complete	Yes
Y185A	Mutation inserted	
D187A	Cloning complete	No
D187E	Cloning complete	Not Attempted
D187Q	Cloning complete	
Q189A	Needs to be reattempted	
V190I	Needs to be reattempted	
Q192A	Cloning complete	No
Q192N	Mutation inserted	
P194A	Cloning complete	Yes
Q196A	Cloning complete	Yes
Y198A	Cloning complete	No
T199A	Cloning complete	Yes

Future studies: Role of nsp5 domain 3

MHV domain 3 is comprised of five alpha helices and has no known homolog to any viral or cellular proteases. Domain 3 is indispensible for nsp5 activity and several residues in the carboxy terminus are essential for dimerization and subsequent proteolytic activity. Domain 3 is considerably more structurally divergent and has been proposed as a key determinant of nsp5 specificity. However, the functional role of domain 3 remains unknown. Consequently, several key questions exist: 1) What functional role does nsp5 domain 3 play in protease structure and activity? 2) Is the function of domain 3 conserved across all coronavirus nsp5 proteases? 3) Which helices are essential for nsp5 protease activity? 4) Are there key interactions between domains 2 and 3? If so, which residues mediate these interactions?

Future studies: nsp5 folding and flexibility

Coronavirus nsp5 protease is comprised of three distinct domains, each of which are approximately 100 amino acids in length. Domains 1 and 2 are β-barrels which constitute a chymotrypsin-like fold and domain 3 is a unique helical element of unclear function. Despite numerous crystal structures of nsp5, no studies to date have evaluated the mechanism of protease folding and flexibility. Upon translation in the context of the polyproteins pp1a and pp1ab, it has been postulated that nsp5 is associated with membranes due to its location between neighboring membrane-associated proteins nsps 3, 4, and 6. Several studies have demonstrated that nsp5 requires a hydrophobic milieu

for proper function, which support this hypothesis. Interestingly, nsp5 appears to process the replicase polyproteins in a distinct order (as evidenced by readily detectable intermediate polyprotein products) and biochemical analyses of nsp5-mediated processing in vitro indicate that nsp5 has appreciably higher specificities for some sites over others. These data indicate that nsp5 exhibits a distinct hierarchy.

Structural predictive modeling of nsp5 dynamics using a freeware platform El Nemo indicated that the junction between domains 2 and 3 linked by the IDL may be extremely dynamic. The discovery of three different temperature-sensitive mutations in these domains, which potentially destabilize the protease structure, may also highlight critical folding motifs or platforms that have been disrupted. Further biochemical analysis will be needed to answer this question. As previously mentioned, it remains unknown how nsp5 orchestrates its autoproteolytic cleavage and how it forms active dimers. Despite numerous crystal structures available and advances in structural modeling, very little continues to be known about nsp5 protease folding and flexibility raising several key questions: 1) Which intramolecular associations are essential for nsp5 folding and stability? 2) Does the nsp5 IDL regulate protease flexibility? If so, for what purpose? 3) How does nsp5 interact with neighboring transmembrane spanning regions in nsp4 and nsp6 and what role do these interactions have in nsp5 folding or flexibility? 4) Does the flexibility of the replicase polyproteins or nsp5 localized flexibility in the context of nsps 3, 4, and 6 modulate its cleavage site specificity and apparent processing hierarchy?

Future studies: Development and testing of coronavirus nsp5 inhibitors

Since the first studies using protease inhibitor E64d showed inhibition of coronavirus replication, considerable energy has been put into the development of broadly-neutralizing coronavirus nsp5 inhibitors. Surprisingly, nearly all nsp5 inhibitors developed to date have been peptidomimetic active site competitive inhibitors. The data presented in Chapter IV demonstrated however that clear functional differences exist between coronaviruses, which may make development of a single silver bullet inhibitor for coronaviruses unfeasible. Furthermore, practically all inhibitors that have shown low μM inhibitory concentrations have yet to be tested against a replicating virus. Consequently, several key questions exist: 1) Do any current top inhibitors demonstrate cellular toxicity or efficacy towards replicating virus? 2) Are there any non-active site regions that are structurally conserved and may be targeted for inhibitor design? 3) What mechanisms will representative viruses of different coronavirus genera use to evade nsp5 inhibitors? Will the pathways of virus escape be the same? Will they be genus specific? 4) Can a combination of inhibitors be used to provide synergistic inhibition of coronaviruses and limit viral escape?

Ongoing collaboration with the Mesecar lab has led to testing of nsp5-specific inhibitors using the H5- and O5-MHV chimeric viruses described in Chapter III. Particular aims of this research are to evaluate the efficacy of these inhibitors and identify potential mechanisms of viral escape. These studies will provide direct feedback for optimization and characterization of future inhibitors being developed.

Future studies: Role of nsp5 in host inactivation and immune evasion

The specificity of nsp5 is known to exhibit considerable plasticity based on *in vitro* biochemical analyses. During virus infection, nsp5 is readily detectable by immunoprecipitations. However, no studies to date have identified any host cell targets of coronavirus nsp5. Studies expressing nsp5 in cells have shown that it modulates the host cell immune response specifically decreasing NF-κB signaling and interferon expression. Recent data from Carolyne Coyne's lab using coxsackievirus B 3C protease (a positive-strand RNA virus related to polio; 3Cpro is protease similar to our nsp5 protease) has identified a number of interferon signaling pathway protein targets that are cleaved during infection including MAVS and TRIF (Mukherjee et al. 2011, PLoS Pathog). In 2004, a study published in Denmark described an online algorithm similar to the one developed for identifying 3C pro targets for coronavirus 3CLpro. The algorithm compares protein amino acid sequences to a consensus nsp5 cleavage site and assigns a score to all putative sites ranging from 0 to 1 based on the increasing likelihood of cleavage. Using this algorithm to predict putative nsp5 cleavage targets and a list provided by the Coyne lab, we identified several host protein targets that may have major implications on host detection and immune activation to virus infection including OAS, IRAK, Ubiquitin-binding protein 1 (UBP1), and ApolipoproteinB 100 (ApoB-100) (**Table 5.2**).

Intriguingly, ApoB-100 is a cholesterol carrier molecule found predominately in hepatic cells and has 9 putative nsp5 cleavage sites. A study of a patient with chronic Hepatitis C infection demonstrated that chronic viral infection elicited an acquired

apolipoprotein B deficiency (Gupte et al. 2006). Despite nearly 50 years of study, the pathogenesis of disease associated with MHV infection remains unknown. These data may in part explain the pathogenesis of the virus and progression of disease in the murine model. The use of host cell membranes by coronaviruses has been extensively described, which applies a motive for the virus. Yet, the mechanisms by which coronaviruses acquire or utilize these membranes are unknown. To date, no studies have reported any nsp5 host targets and identification of a host target may shed new light on the role of protease during virus infection. Therefore, many key questions remain: 1) Are there any nsp5 protease host cell targets during virus infection? If so, what are they and do they play a role in virus evasion, acquisition of host cell resources, or the viral replicative cycle? 2) Is there a connection between cleavage of ApoB-100 and the pathogenesis and disease associated with MHV infection? 3) Where is nsp5 protease localized during virus infection and does its localization impact its ability to cleave host targets?

Concluding remarks

In this dissertation, I have described the contributions I have made to understand the structure and function of coronavirus nsp5 protease. There are many unanswered questions remaining. The emergence of new coronaviruses and current dearth of information related to the mechanism of nsp5 action demand attention for the development of inhibitors and understanding the limitations of virus evolution.

Table 5.3 Putative host cell targets of coronavirus nsp5 protease

Progress towards mutating the MHV D2-D3 IDL (only hits above priority score of 0.7 shown)	
Protein Target	*Highest Priority Score (# of predicted sites)*
BATF2	0.932 (2)
UBP1	0.932 (2)
RIP1	0.927 (2)
NAP1	0.919 (3)
ATF6B	0.916 (1)
IFI6	0.890 (1)
INI2	0.890 (1)
CUT	0.889 (2)
OCT1	0.874 (1)
APLB100	0.866 (9)
VIM	0.866 (1)
IRAK1	0.859 (1)
CYP1B1	0.851 (1)
SCAD	0.828 (1)
EIF4G	0.822 (3)
LPL	0.819 (1)
SLC25A28	0.813 (1)
OAS1	0.793 (1)
p130CAS	0.790 (2)
17AH	0.782 (1)
P53BP3	0.782 (3)
IRF8	0.766 (1)
RPOLIII_C1	0.765 (3)
SIRPA	0.742 (2)
P53BP1	0.738 (2)
STAT1	0.736 (1)
IKKε	0.735 (3)
TFIID_1	0.730 (1)
TRIF	0.720 (4)

CHAPTER VI

MATERIALS AND METHODS

Viruses, cells, and antisera. Recombinant wild-type (WT) MHV strain A59 (GenBank no. AY910861) was used as a MHV wild-type control. Delayed brain tumor cells (DBT-9), which are naturally permissive for MHV infection, and baby hamster kidney 21 cells expressing the MHV receptor (BHK-MHVR) were used for all experiments (Yount et al., 2002). Dulbecco's modified Eagle medium (DMEM) (Gibco) was supplemented with 10% heat-inactivated fetal calf serum (FCS) for all experiments described. BHK-MHVR medium was supplemented with G418 (0.8 mg/mL; CellGro) to maintain selection for MHVR expression. All biochemical experiments were carried out using rabbit polyclonal antibodies previously described in the literature. The antisera used include nsp2-specific (VU154), nsp5-specific (VU6), and nsp8-specific (VU123) antibodies (Bost et al., 2000; Lu et al., 1996; Sims et al., 2000).

Mutagenesis of MHV cDNA C fragment. Assembly of the complete MHV genome is generated through the ligation of seven cDNA fragments (A-G) digested from individual plasmids as previously described by Yount *et al.* (Yount et al., 2002). All viruses were engineered by inserting the specified amino acid substitution into the MHV infectious clone (MHVic) C fragment containing the nsp5 sequence, which was constructed by PCR and cloned into the XL-pSMART vector (Yount et al., 2002). Sense and antisense primers were designed to be overlapping with nucleotide changes within the middle of the primer. The primers used for mutagenesis for experiments described in Chapters II,

III, and IV. All mutant plasmids were sequence confirmed prior to ligation and MHVic assembly.

Recovery of MHV mutant viruses. The MHV nsp5 mutant viruses were engineered through the infectious cDNA assembly strategy as described previously by Yount *et al*. (Yount et al., 2002). In brief, the seven cDNA fragments were digested, gel-purified, and ligated overnight at 16°C. Transcription of the extracted ligated DNA, as well as N cDNA which encodes the nucleocapsid protein, was performed by using the mMachine T7 transcription kit (Ambion) under conditions previously described in detail (Yount et al., 2002). The transcribed genome and N-gene were electroporated into BHK-MHVR cells and the electroporated cells were placed into a sub-confluent flask of DBT-9 cells and incubated at either 30°C for potentially *ts* viruses or 37°C.

Cloning and recovery of chimeric and mutant viruses. Viruses were assembled and recovered using the MHV infectious clone protocol previously described by Yount et al (Yount et al., 2002). The nsp5 nucleotide sequences for human coronaviruses HKU1 (GenBank accession number NC006577), OC43 (NC005147), and SARS-CoV Urbani (AY278741) were each substituted into the MHV cDNA genomic constructs (BioBasic) and sequence confirmed prior to attempted virus recovery (Rota et al., 2003; Vijgen et al., 2005; Woo et al., 2005). Using the assembly protocol previously described, the genomic cDNA fragments were ligated, transcribed, and electroporated into BHK-MHVR cells which were then added to a sub-confluent flask of DBT-9 cells at 37°C (Yount et al., 2002).

RNA extraction and sequencing. A confluent monolayer of DBT-9 cells in a T25 flask was infected with viral mutant stocks at a multiplicity of infection (MOI) of 10 plaque-forming units (PFU) per mL and grown to approximately 30 – 50% involvement in syncytia. Supernatant was removed from each T25 containing isolated mutant virus and stored at -20°C. The cells were harvested in TRIzol reagent (Invitrogen) for isolation of total RNA. Reverse transcriptase PCR (RT-PCR) was performed using SuperScript III RT (Invitrogen) and random hexamers (Applied Biosystems) at 55°C for 1 h and the resulting cDNA was PCR amplified using Easy-A high-fidelity PCR cloning enzyme (Stratagene) and MHV genome oligonucleotides covering the nsp5 region (nucleotides 10160 – 11799). Amplified regions were gel purified and analyzed by sequencing.

Isolation and expansion of suppressor mutants. In order to isolate potential suppressor mutant viruses, two different types of plaque assays were performed. Reversion analysis of the *ts* viruses described in Chapter II was performed by infecting DBT-9 cells in duplicate using 6-well plates with serial dilutions of potential revertant samples with a 1 h adsorption period. The overlay contained a one to one mixture of 2% agar and 2X Dulbecco's Modified Eagle Medium. The plates were incubated for 48 h so plaques were easily visible. Ten plaques were picked for each virus sample and re-suspended in gel saline. Isolated plaques for each virus were used to infect separate T25s for expansion at 40°C. Flasks were removed from non-permissive temperatures when 70-95% of cells were involved in syncytia, then RNA was isolated as described above. The suppressor mutants identified in Chapter III were acquired by plaque assay by infecting DBT-9 cells

in duplicate using 6-well plates with serial dilutions and a 1 h adsorption period at 32°C. The overlay contained a 1:1 mixture of 2% agar and 2X Dulbecco's Modified Eagle Medium. The plates were incubated for 48 h until plaques were easily visible. Ten plaques were picked for each virus sample and re-suspended in gel saline. Isolated plaques for each virus were used to infect separate T25s for expansion at 40°C. Flasks were removed from non-permissive temperatures when 70-95% of cells were involved in syncytia, then RNA was isolated as described above.

Analysis of viral replication kinetics and nsp5 processing. Confluent monolayers of DBT-9 cells in 60-mm dishes were infected at an MOI of 0.01 or 1 PFU/mL and 5 PFU/mL for growth analysis and immunoprecipitation analyses, respectively. In the temperature shift experiments, the cells were shifted from 30°C or 32°C to 40°C at 6 h p.i. During the growth analysis, samples of supernatant were acquired and pre-warmed media added back to maintain a fixed volume on the cells. Virus titers were determined by plaque assay in duplicate or triplicate. Immunoprecipitation experiments were carried out as previously described (Sparks et al., 2008). Eluted proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 4 to 12% polyacrylamide gradient Bis-Tris gels (Nu-PAGE/Invitrogen) and analyzed by autoradiography. A ^{14}C-labeled high-molecular-weight standard (New England Biolabs) and a full range rainbow marker (RPN 800E/ Invitrogen), were used as protein mass standards.

Sequence alignments and modeling of MHV nsp5 structures. A multiple sequence alignment of coronavirus nsp5 sequences and a phylogenetic tree were generated using ClustalX and a bootstrap alignment from 1000 trials. A model of 100% conserved residues was determined using Consurf (Eswar et al., 2008) and a multiple sequence alignment (MSA) of 226 available CoV nsp5 sequences. A sequence logo was generated using the WebLogo server and the MSA of 226 available CoV sequences (Crooks et al., 2004). The X-ray crystal structure of HCoV-HKU1 nsp5 (Protein Data Bank [PDB] code 3D23) was used as a structural model of comparison (Zhao et al., 2008). Structural models were generated using Modeller (Stokes et al., 2010) and MacPyMol (DeLano Scientific). Other CoV nsp5 X-ray structures used for alignment were SARS (PDB 2H2Z), HCoV-229E (PDB 1P9T), and IBV (PDB 2Q6D) (Anand et al., 2003; Xue et al., 2007; Xue et al., 2008). Distance calculations were determined by measurement of the alpha carbon to alpha carbon in the structure of HCoV-HKU1 nsp5.

Assay of virus fitness. Confluent monolayers of DBT-9 cells were co-infected in triplicate at ratios of 1:1 or 1:10 to a total MOI of 0.01 PFU/cell in 25 cm^2 flasks. Upon reaching 30 to 50% involvement in syncytia, the supernatant was stored at -80°C and the monolayer treated with TRIzol reagent for RNA acquisition. For subsequent passages, the virus supernatant was thawed at 4°C and 5 μL were added to each of three confluent 25 cm^2 flasks. RNA was isolated and reverse-transcribed as previously described (Stobart et al., 2012). Amplicons using oligonucleotides flanking the nsp5 coding region were generated and treated with HKU1- (HincII) and OC43-sequence specific (BsiWI)

restriction enzymes, which resulted in a single cut. The bands were resolved on a 0.8% agarose gel containing ethidium bromide and quantified by densitometry.

Creation of FRET-based substrates. The genes coding for cyan and yellow fluorescent proteins (CFP and YFP) joined by a multiple cloning site were synthesized (BioBasic) and inserted into a vector for expression in *E. coli*. Complimentary oligonucleotides representing P8 to P8' for each nonstructural protein cleavage site were synthesized (BioBasic) and inserted between the CFP and YFP via recombinant cloning following restriction enzyme digestion of the multiple cloning site (Oliner et al., 1993). Following confirmation of intended clones by DNA sequencing, fluorescent resonance energy transfer (FRET)-based substrates consisting of cyan and yellow fluorescent proteins linked by the nonstructural protein cleavage sites were expressed and purified as previously described (Felber et al., 2004).

Analysis of 3CL kinetic activity against FRET-based substrates. The nsp5 proteases of SARS-CoV, HKU1, OC43, and MHV were expressed and purified using an approach previously described (Jacobs et al., 2013). Assays were conducted in black 96-well plates (Corning) using a Synergy 5 plate reader (Biotek). Assays using the universal nsp5 cleavage site substrate were performed as described previously (Grum-Tokars et al., 2008). Emission of the FRET substrates at 485 nm was monitored over a time-course varying from 30 minutes to 6 hours depending on the substrate used. Based on the cleavage rates of each substrate, enzyme concentrations varied from 100 to 500 nM. Initial cleavage velocities were determined for each substrate at 1, 2, and 4 μM and

normalized using a fluorescent extinction coefficient empirically calculated for each substrate. For each substrate/enzyme combination, data were adjusted for enzyme concentration and reaction volume, then plotted versus substrate concentration to yield a line whose slope represented a unique k apparent (k_{app}) value.

APPENDIX A: LIST OF PUBLICATIONS

Stobart CC, Sexton NR, Munjal H, Lu X, Molland KL, Tomar S, Mesecar AD, and Denison MR. Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. (In Press; *Journal of Virology* 2013)

Stobart CC and Denison MR. 2012. "MS.570 | Coronavirus picornain-like cysteine proteinase" in Handbook of Proteolytic Enzymes (3rd Ed). ND Rawlings and G Salvesen, Eds. Elsevier Press. UK.

Stobart CC, Lee AS, Lu X, and Denison MR. 2012. Temperature-sensitive mutants and revertants in the coronavirus nsp5 protease (3CLpro) define residues involved in long-distance communication and regulation of protease activity. *J. Virol.* 86:4801-10.

Gadlage MJ, Beachboard DC, Cox RJ, Doyle JD, Sparks JS, **Stobart CC**, and Denison, MR. 2010. Murine Hepatitis Virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. *J. Virol.* 84:280-90.

APPENDIX B

TEMPERATURE-SENSITIVE MUTANTS AND REVERTANTS IN THE CORONAVIRUS NONSTRUCTURAL PROTEIN 5 PROTEASE (3CLpro) DEFINE RESIDUES INVOLVED IN LONG-DISTANCE COMMUNICATION AND REGULATION OF PROTEASE ACTIVITY

Christopher C. Stobart, Alice S. Lee, Xiaotao Lu and Mark R. Denison
Journal of Virology 2012, 86(9):4801-4810

JVI
Journals.ASM.org

Temperature-Sensitive Mutants and Revertants in the Coronavirus Nonstructural Protein 5 Protease (3CLpro) Define Residues Involved in Long-Distance Communication and Regulation of Protease Activity

Christopher C. Stobart,[b,c] Alice S. Lee,[a,c] Xiaotao Lu,[a,c] and Mark R. Denison[a,b,c]

Departments of Pediatrics[a] and Pathology, Microbiology and Immunology,[b] and The Elizabeth B. Lamb Center for Pediatric Research,[c] Vanderbilt University Medical Center, Nashville, Tennessee, USA

Positive-strand RNA virus genomes are translated into polyproteins that are processed by viral proteases to yield functional intermediate and mature proteins. Coronaviruses (CoVs) carry genes that encode an nsp5 protease (also known as 3CLpro or Mpro) responsible for 11 maturation cleavages. The nsp5 structure contains two chymotrypsin-like domains (D1 and D2) and a unique domain (D3), and forms functional dimers. However, little is known of interactions or communication across the structure of the protease during nsp5 activity. Using reverse genetic mutagenesis of the CoV murine hepatitis virus (MHV) nsp5, we identified a new temperature-sensitive (*ts*) mutation in D2 of nsp5 (Ser133Ala) and confirmed a *ts* residue in D3 (Phe219Leu). Both D2-*ts*S133A and D3-*ts*F219L were impaired for viral replication and nsp5-mediated polyprotein processing at the nonpermissive temperature. Passage of *ts*S133A and *ts*F219L at the nonpermissive temperature resulted in emergence of multiple second-site suppressor mutations, singly and in combinations. Among the second-site mutations, a D2 His134Tyr change suppressed the *ts* phenotype of D2-*ts*S133A and D3-*ts*F219L, as well as the previously reported D2-*ts*V148A. Analysis of multiple CoV nsp5 structures, and alignment of nonredundant nsp5 primary sequences, demonstrated that *ts* and suppressor residues are not conserved across CoVs and are physically distant (>10 Å) from each other, from catalytic and substrate-binding residues, and from the nsp5 dimer interface. These findings demonstrate that long-distance communication pathways between multiple residues and domains of nsp5 play a significant role in nsp5 activity and viral replication, suggesting possible novel targets for non-active site inhibitors of nsp5.

Positive-strand RNA viruses are responsible for prevalent and epidemic diseases in a wide range of vertebrate hosts, as well as new and emerging viruses, such as severe acute respiratory syndrome coronavirus (SARS-CoV), West Nile virus, and Chikungunya virus. The rapid evolution, host species movement, and diseases of positive-strand RNA viruses demonstrate the need to develop novel strategies to prevent and treat present and new diseases caused by these viruses. A key determinant of positive-strand RNA viruses is the requirement for processing of translated polyproteins by virus gene-encoded proteases. RNA virus proteases therefore have been high-profile targets for development of antiviral agents, with most protease inhibitors targeted to active sites or substrate-binding sites (21, 26, 32, 35). However, due to the potential for viral escape mutants, it is critical to identify additional noncatalytic, non-substrate-binding determinants of protease activity as potential targets for inhibition that are less prone to development of resistance.

To date, five CoVs have been shown to be associated with human respiratory diseases of different degrees of severity: human coronavirus HCoV-229E, HCoV-OC43, HCoV-NL63, HCoV-HKU1, and SARS-CoV (10, 15, 30, 31, 44, 45). CoVs contain the largest known positive-strand RNA genomes, ranging from 26 to 32 kb in length. Murine hepatitis virus (MHV) strain A59 is an established model for study of CoV replication and pathogenesis. The 32-kb genome of MHV contains seven genes, with the replicase gene (22 kb) encoding 16 nonstructural proteins (nsp1 to nsp16) (Fig. 1A) (20, 25). The replicase gene is translated into polyprotein 1a (pp1a; nsp1 to nsp11) or, via a ribosomal frameshift, pp1ab (nsp1 to nsp16) (9, 25, 33). MHV encodes two papain-like proteases (PLP1 and PLP2) responsible for cleavages of nsp1 to nsp3, and an nsp5 protease, also known as 3CLpro or Mpro, that mediates maturation cleavages of nsp4 to nsp16 and is required for virus replication (33, 50).

The CoV nsp5 is a cysteine protease present in all known CoVs and is structurally similar to the nsp4 protease of distantly related arteriviruses (6, 33, 49). The crystal structure of nsp5 has been solved for divergent CoVs from every genus, including SARS-CoV, infectious bronchitis virus (IBV), human HCoV-HKU1, and human HCoV-229E. Comparison of solved nsp5 structures demonstrates conservation of tertiary structure despite numerous differences in primary sequences (1, 2, 5, 46, 47, 49). The X-ray crystal structure of MHV nsp5 has yet to be determined; however, the structure of the closely related HCoV-HKU1 nsp5 (84% sequence identity) has been resolved to 2.5 Å (Fig. 1B and C) (49). The nsp5 proteases of all CoVs exhibit a three-domain structure, with domains 1 and 2 forming a chymotrypsin-like fold containing the His41-Cys145 catalytic dyad and substrate-binding sites (Fig. 1B) (1, 2, 5, 28, 46, 47). In contrast, domain 3 is unique to the CoV nsp5 protease among chymotrypsin-like enzymes and also shows more divergence in both sequence and structural organization between CoVs. *In vitro* studies indicate that domain 3 is important for stabilization of the chymotrypsin-like fold and may

Received 7 November 2011 **Accepted** 9 February 2012
Published ahead of print 15 February 2012
Address correspondence to Mark R. Denison, mark.denison@vanderbilt.edu.

doi:10.1128/JVI.06754-11

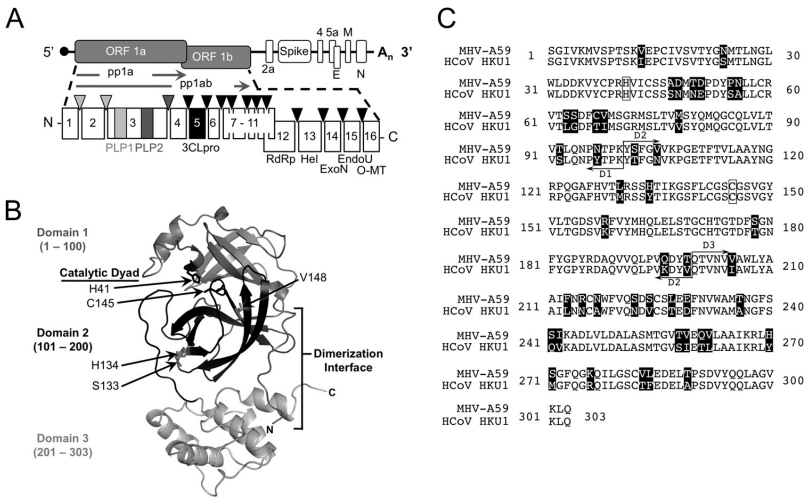


FIG 1 MHV protease nsp5 structure and sequence alignment. (A) Genome and polyprotein processing of MHV. The replicase gene (gray) consists of two open reading frames (ORF1a and ORF1ab) overlapping at a ribosomal frameshift and is translated to yield two polyprotein products, pp1a and pp1ab, encoding nonstructural proteins (nsps) nsp1 to nsp16. Maturation cleavages (arrows) are mediated by three viral proteases, papain-like proteases PLP1 (light gray) and PLP2 (dark gray) and nsp5 (black). M, matrix; E, envelope; N, nucleocapsid; RdRp, RNA-dependent RNA polymerase; Hel, helicase; ExoN, exoribonuclease; EndoU, endoribonuclease; O-MT, *O*-methyltransferase. (B) Modeled structure of a MHV nsp5 monomer based on the crystal structure of HKU1 (49) shows domain 1 (D1), domain 2 (D2), and domain 3 (D3). *ts* and second-site suppressor residues described by Sparks et al. are also identified (40). N and C denote the amino and carboxy termini. The dimerization interface is labeled, and functional residues discussed are shown in black and are labeled. (C) Sequence alignment of MHV-A59 and HCoV HKU1 with nonconserved residues highlighted, catalytic residues boxed, and domain separations identified.

also be important in mediating dimerization between nsp5 monomers (1, 28, 37, 38). Structural and biochemical studies demonstrate that nsp5 dimerization is required for proteolytic activity *in vitro* (11, 12, 37).

Recently, amino acid residues that may regulate nsp5 activity but that are distinct from the active site cavity, substrate-binding pocket, or dimerization interface have been identified. Alanine substitution at SARS-CoV nsp5 Ser147, a conserved serine residue in MHV and HKU1, disrupts dimerization and impairs nsp5 proteolytic activity, despite being greater than 9 Å from the dimerization interface (7). Our laboratory identified a temperature-sensitive (*ts*) mutation (*ts*V148A) in MHV nsp5 that impairs virus growth and nsp5 activity at 40°C (Fig. 1B) (40). Growth of the *ts*V148A mutant at 40°C resulted in emergence of second-site mutations (S133N and H134Y), which suppressed the *ts*V148A phenotype. While V148 is adjacent to the catalytic C145, neither S133N nor H134Y have predicted direct interactions with catalytic or substrate-binding residues or identified pathways for propagation of structural changes. Sawicki et al. identified a putative *ts* allele in nsp5 (F219L) at the base of unique domain 3, and distant from the dimerization interface, active site cavity, and substrate-binding regions (36). Finally, other studies have shown that nsp5 activity may be altered by changes in replicase nonstructural proteins nsp3 and nsp10 (14, 41). Together, these findings support the hypothesis that residues distant from catalytic and substrate-binding sites are important for regulating nsp5 protease activity. However, mechanisms of communication between residues and regulation of protease activity are unknown.

In this study, we tested the roles of MHV nsp5 domain 2 residues S133 and H134, and domain 3 F219, on nsp5 activity. The experiments identified a previously unknown MHV nsp5 *ts* mutation in domain 2 (*ts*S133A) and confirmed the *ts* phenotype of F219L, both of which result in profound growth and protein-processing defects at nonpermissive temperatures. Under passage at nonpermissive temperatures, a series of second-site mutations emerged that were able to suppress the *ts* phenotypes of *ts*S133A and *ts*F219L while being physically distant from the *ts* allele, the nsp5 active site cavity, and the dimerization interface. A single nonsynonymous mutation resulting in a H134Y substitution suppressed the *ts* phenotype of all three independent *ts* alleles in domains 2 and 3. The *ts* mutations reduced but did not abolish nsp5 protease activity during virus infection following a shift to the nonpermissive temperature, while individual and combined suppressor mutations restored nsp5 activity to an extent that directly correlated with increased replication. These results demonstrate the presence of multiple interconnected long-distance communication nodes in nsp5 and suggest novel mechanisms of regulation of nsp5 activity during CoV replication.

MATERIALS AND METHODS

Viruses, cells, and antisera. Recombinant wild-type (WT) MHV strain A59 (GenBank accession no. AY910861) was used as a MHV wild-type control. Delayed brain tumor cells (DBT-9), which are naturally permissive for MHV infection, and baby hamster kidney 21 cells expressing the MHV receptor (BHK-MHVR) were used for all experiments (48). Dulbecco's modified Eagle medium (DMEM) (Gibco) was supplemented with 10% heat-inactivated fetal calf serum (FCS) for all experiments described. BHK-MHVR medium was supplemented with G418 (0.8 mg/ml; CellGro) to maintain selection for MHVR expression. All biochemical experiments were carried out using rabbit polyclonal antibodies previously described in the literature. The antisera used include nsp2-specific

TABLE 1 Oligonucleotide primers used in the mutagenesis of MHV cDNA C fragment plasmids

Mutation(s) made	Primer sequence (5 → 3)[a]
S133A	ACG CTT CGT AGT GCC CAT ACC ATA AAG
S133A/H134Y	ACG CTT CGT AGT GCC TAT ACC ATA AAG
S133A/T129M	TTC CAT GTT ATG CTT ATG CTT CGT AGT GCC CAT ACC ATA
S133A/H134Y/T129M	TTC CAT GTT ATG CTT CGT AGT GCC TAT ACC ATA AAG
H134A	CTT CGT AGT AGC GCT ACC ATA AAG GGC
H134Y	ACG CTT CGT AGT AGC TAT ACC ATA AAG GGC TCC
F219L	AAC AGA TGC AAC TGG CTA GTG CAA AGT GAT AGT
H270HH	GCT ATT AAG AGG CTG CAT CAT TCT GGA TTC CAG
E285V	GGT AGT TGT GTG CTT GTT GAT GAG ACA CCA AGT

[a] The nucleotide changes are underlined.

(VU154), nsp5-specific (VU6), and nsp8-specific (VU123) antibodies (8, 27, 39).

Recovery of MHV mutant viruses. The MHV nsp5 mutant viruses were engineered through the infectious cDNA assembly strategy described previously by Yount et al. (48). In brief, the seven cDNA fragments were digested, gel purified, and ligated overnight at 16°C. Transcription of the extracted ligated DNA, as well as N cDNA which encodes the nucleocapsid protein, was performed by using the mMachine T7 transcription kit (Ambion) under conditions previously described in detail (48). The transcribed genome and N gene were electroporated into BHK-MHVR cells, and the electroporated cells were placed into a subconfluent flask of DBT-9 cells and incubated at either 30°C for potentially *ts* viruses or 37°C.

Mutagenesis of MHV cDNA C fragment. Assembly of the complete MHV genome is generated through the ligation of seven cDNA fragments (A to G) digested from individual plasmids as previously described by Yount et al. (48). All viruses were engineered by inserting the specified amino acid substitution into the MHV infectious clone (MHVic) C fragment containing the nsp5 sequence, which was constructed by PCR and cloned into the XL-pSMART vector (48). Sense and antisense primers were designed to be overlapping with nucleotide changes in the middle of the primer. The primers used for mutagenesis are listed in Table 1. The sequences of all mutant plasmids were confirmed prior to ligation and MHVic assembly.

RNA extraction and sequencing. A confluent monolayer of DBT-9 cells in a T25 flask was infected with viral mutant stocks at a multiplicity of infection (MOI) of 10 PFU per ml and grown until approximately 30 to 50% of the cells had formed syncytia. Supernatant was removed from each T25 flask containing isolated mutant virus and stored at −20°C. The cells were harvested in TRIzol reagent (Invitrogen) for isolation of total RNA. Reverse transcriptase PCR (RT-PCR) was performed using SuperScript III RT (Invitrogen) and random hexamers (Applied Biosystems) at 55°C for 1 h, and the resulting cDNA was PCR amplified using Easy-A high-fidelity PCR cloning enzyme (Stratagene) and MHV genome oligonucleotides covering the nsp5 region. Amplified regions were gel purified and analyzed by sequencing.

Isolation and expansion of suppressor mutants. Confluent monolayers of DBT-9 cells were initially infected with the temperature-sensitive viruses and incubated at either 40°C (S133A) or 30°C (F219L) with increases in temperature to 37°C and subsequently to 40°C. A plaque assay was then performed using 10-μl and 100-μl aliquots of viral stocks. DBT-9 cells were infected in duplicate using 6-well plates with serial dilutions of potential revertant samples with a 1-h adsorption period. The overlay contained a 1:1 mixture of 2% agar and 2× Dulbecco's modified Eagle medium. The plates were incubated for 48 h so the plaques were easily visible. Ten plaques were picked for each virus sample and resuspended in gel saline. Isolated plaques for each virus were used to infect separate T25 flasks for expansion at 40°C. The flasks were removed from nonpermissive temperatures when 70 to 95% of cells had formed syncytia, and then RNA was isolated as described above.

Analysis of viral growth and nsp5 processing. Confluent monolayers of DBT-9 cells in 60-mm dishes were infected at an MOI of 1 PFU/cell for growth analysis or 5 PFU/cell for immunoprecipitation analysis. In the temperature shift experiments, the cells were shifted from 30°C to 40°C at 6 h postinfection (p.i.). During the growth analysis, samples of supernatant were acquired and prewarmed medium was added back to maintain a fixed volume of medium for the cells. Virus titers were determined by plaque assay in duplicate. Immunoprecipitation experiments were carried out as previously described (40). Eluted proteins were resolved by sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) on 4 to 12% polyacrylamide gradient Bis-Tris gels (Nu-PAGE; Invitrogen) and analyzed by autoradiography. A ^{14}C-labeled high-molecular-weight standard (New England BioLabs) and a full range rainbow marker (RPN 800E; Invitrogen) were used as protein mass standards.

Sequence alignments and modeling of MHV nsp5 structures. The X-ray crystal structure of HCoV-HKU1 nsp5 (Protein Data Bank [PDB] code 3D23) was used as a structural model of comparison (49). Structural models were generated using Modeler (16) and MacPyMol (DeLano Scientific). Other CoV nsp5 X-ray structures used for alignment and comparison were SARS (PDB code 2H2Z), HCoV-229E (PDB code 1P9T), and IBV (PDB code 2Q6D) (2, 46, 47). A query of nsp5 sequences in GenBank resulted in the identification of 432 complete coronavirus nsp5 amino acid sequences. A perl script was used to generate a representative sequence from repetitive sequences and eliminate redundancies as a means of unbiasing the data set; a final alignment of 130 unique, nonredundant sequences was prepared. A model of 100% conserved residues was determined using Consurf (4) and a sequence logo was generated using the WebLogo server (13) with the prepared nonredundant CoV nsp5 sequence data set. Distance calculations were determined by measuring from alpha carbon to alpha carbon in the structure of HCoV-HKU1 nsp5.

RESULTS

Domain 2 S133A is a novel temperature-sensitive mutant of MHV nsp5. We previously identified two second-site mutations that suppressed the temperature-sensitive phenotype of MHV *ts*V148A: H134Y and S133N (40). Both H134 and S133 are greater than 20 Å from V148, from the active site cavity, and from the S1 substrate-binding site. The mechanism by which these residues complement or suppress *ts*V148A remains unclear. Further, substitution of the H134Y or S133N residue alone in the isogenic cloned wild-type (WT) MHV background had little or no effect on virus replication at 37°C. We therefore sought to determine whether these residues had critical roles in nsp5 activity by engineering alanine substitution mutations in the MHV genome. S133A and H134A mutant viruses were recovered at 30°C and sequencing from the initial passage (passage 1 [P1]) 30°C virus stocks confirmed the presence of mutation S133A or H134A, with no other sequence changes in nsp5. The viruses were compared with recombinant WT MHV and with *ts*V148A for virus titer at 30°C and 40°C, and the efficiency of plating (EOP) was calculated (titer at 40°C/titer at 30°C) (Fig. 2A). WT MHV had an EOP of 3.3, a 3.3-fold increase in visible plaques at 40°C compared to 30°C. The previously described *ts*V148A virus exhibited an EOP of 3×10^{-5}, confirming the *ts* phenotype reported by Sparks et al. (40). The engineered H134A mutant had an EOP of 10^{-1}, which is similar to that of the previously reported engineered H134Y mutation in the WT MHV background (40). In contrast, the S133A

A

Virus Strain	Titer at:		EOP (40/30)
	30°C	40°C	
WT MHV	9×10^7	3×10^8	3×10^0
S133A	6×10^6	2×10^2	3×10^{-5}
S133A / H134Y*	2×10^7	1×10^7	5×10^{-1}
S133A / T129M / H134Y*	3×10^7	4×10^7	2×10^0
S133A / T129M^	4×10^7	5×10^4	1×10^{-3}
H134A	5×10^8	4×10^8	7×10^{-1}

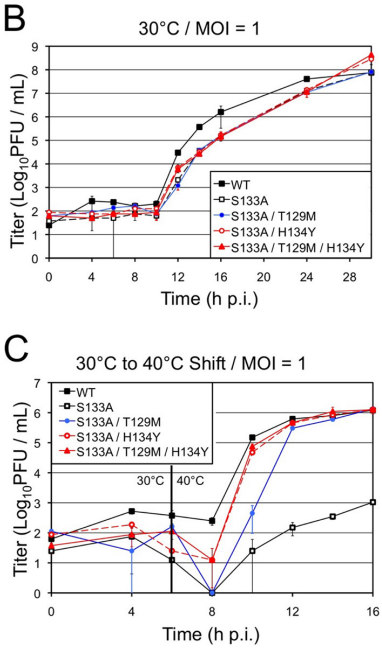


FIG 2 Analysis of replication of *ts*S133A and suppressor mutants. (A) Virus titers and EOP data for WT MHV and S133A mutant viruses determined by plaque assay at 30°C and 40°C. The virus titers were calculated in duplicate by infection of DBT-9 cells. Recombinant second-site suppressor mutants (*) and engineered recombinants whose mutations were artificially recombined (^) are indicated. (B and C) Growth analysis of the WT MHV, recombinant *ts*S133A, suppressor mutants (S133A/H134Y and S133A/T129M/H134Y), and engineered mutant (S133A/T129M) grown at 30°C (B) or grown at 30°C and then shifted to 40°C at 6 h p.i. on DBT-9 cells (C). The virus titers were determined by plaque assay on DBT-9 cells at 30°C.

mutant had an EOP of 2.7×10^{-5}, consistent with an independent *ts* phenotype (Fig. 2A). Interestingly, reversion of *ts*V148A had resulted in identification of an allele at S133 that with an Asn substitution could suppress the V148A *ts* phenotype (S133N), but with a conservative Ala substitution resulted in a new *ts* phenotype (S133A). H134A was not further examined, and *ts*S133A was studied in subsequent experiments.

Phenotypic reversion of *ts*S133A at the nonpermissive temperature reveals second-site suppressor mutations in domain 2. To test for revertants or suppressors of S133A, DBT-9 cells were infected with *ts*S133A at 37°C, followed by a shift to 40°C. Recovered supernatant virus was used to infect cells at 40°C, and 10 virus plaques were isolated and expanded at 40°C. RT-PCR sequencing of the nsp5 coding regions from all 10 clones confirmed the retention of the S133A ($_{10605}$AGC to GCC with the mutation shown underlined) engineered mutation, as well as two distinct patterns of second-site nonsynonymous mutations resulting in amino acid substitutions: H134Y ($_{10608}$CAT to $_{10608}$TAT) and T129M/H134Y ($_{10593}$ACG to $_{10593}$ATG; $_{10608}$CAT to $_{10608}$TAT). No primary reversion at S133A was identified, and no other mutations were detected in the nsp5 coding sequence. The S133A/H134Y plaque isolate had an EOP of 0.5, and the S133A/T129M/H134Y plaque isolate had an EOP of 2, suggesting almost complete suppression of the S133A *ts* phenotype (Fig. 2A). To test the contribution of the second-site mutations to suppression of the *ts* phenotype, we engineered different combinations of mutations into the isogenic cloned MHV background. Since both biologically derived suppressor mutants contained H134Y and there were subtle differences in the EOP values, we tested the independent contribution of the T129M substitution by introducing the S133A and T129M mutations in the absence of the H134Y mutation. All of the engineered recombinant mutants were readily recovered at 30°C, and sequencing confirmed that the engineered changes were present and no other mutations had arisen in nsp5. The EOPs of the recombinant S133A/H134Y and S133A/T129M/H134Y mutants were identical to the cognate biologically recovered mutant, demonstrating that the identified changes in nsp5 were necessary and sufficient for the phenotypic reversion (Fig. 2A). In contrast, the engineered recombinant S133A/T129M mutant showed an EOP of 10^{-3}, 2 log units greater than *ts*S133A but still significantly *ts* compared to either the S133A/H134Y or S133A/T129M/H134Y mutant. The results demonstrated that H134Y was sufficient for suppression of the *ts*S133A, T129M was unable to suppress the *ts* phenotype by itself, and the combination of T129M and H134Y was additive or synergistic, suggesting that they are two distinct mutations that could have arisen sequentially or concurrently in the same or different genomes.

Growth of *ts*S133A and second-site suppressor mutant viruses demonstrate differential effects of single and multiple suppressor mutations. The capacity to generate a visible plaque is one indicator of viral fitness but has limited ability to predict overall growth fitness. We therefore compared the capability of WT and mutant viruses to replicate at permissive (30°C) and nonpermissive temperatures (40°C) in single-cycle growth experiments. On the basis of earlier single-cycle growth studies of MHV replication, we infected replicate plates of DBT-9 cells with WT and recombinant mutant viruses at 30°C and a multiplicity of infection (MOI) of 1 PFU/cell (18, 19). At 6 h p.i., one replicate plate for each virus was transferred to the nonpermissive temperature of 40°C and one plate was maintained at 30°C, with supernatant samples obtained from 0 to 30 h p.i. at regular intervals for determination of viral titers. WT virus incubated at 30°C demonstrated onset of exponential growth between 10 and 12 h p.i. and achieved peak virus titers of $\sim 10^8$ PFU/cell by 30 h p.i. (Fig. 2B), consistent with previous studies (40). In contrast, *ts*S133A and all three recombinant suppressor mutant viruses demonstrated identical growth curves, with a 1-log-unit reduction in titer compared to the titer of WT MHV from 12 to 24 h p.i., but achieving titers identical to WT virus by 30 h p.i.

When the replicate virus-infected monolayers were shifted from 30°C to 40°C at 6 h p.i. (Fig. 2C), the WT virus-infected monolayers showed onset of exponential growth within 2 h after

shift and achieved peak titers between 12 and 16 h p.i., albeit at lower titers due to rapid destruction of the monolayer. The *ts*S133A mutant virus showed profoundly impaired growth for 10 h after the temperature shift. Both recombinant S133A/H134Y and S133A/T129M/H134Y mutant viruses demonstrated onset of exponential growth and peak titers similar to those of the WT virus following the shift, consistent with EOP analysis. However, the S133A/T129M virus showed a significant delay in exponential growth before achieving titers similar to those of the WT virus by 6 h after the temperature shift. These results collectively indicate that the S133A mutation confers a subtle replication defect at 30°C that is not further impaired nor complemented by the suppressor mutations and that H134Y alone or in combination with T129M is sufficient for suppression of the *ts*S133A growth phenotype. The S133A/T129M virus, while demonstrating improved growth over *ts*S133A, was still impaired compared to mutants containing H134Y, a result consistent with the EOP data, and supporting either conjecture that T129M arose first to be superseded by emergence of H134Y or arose second by conferring a subtle growth advantage of the combination over H134Y alone.

Recovery and reversion of recombinant nsp5 *ts*F219L. The experiments with *ts*S133A, in combination with our previous studies, demonstrated that the H134Y substitution was able to suppress two distinct and independently derived nsp5 *ts* alleles, *ts*V148A and *ts*S133A, suggesting an important role for intra- or intermolecular communication involved in regulation of nsp5 activity. However, both of these *ts* alleles are in domain 2 as were the suppressor mutations, and thus could not provide insight into potential long-distance communication between domains. We therefore next sought to determine whether H134 would emerge as a suppressor for another putative *ts* F219L allele in domain 3. The $_{10864}$UUU-to-CUU mutation resulting in the F219L substitution was predicted as a *ts* allele by Sawicki et al. using partial genome sequencing and reversion analysis of the *ts* mutant Alb *ts*16 (36). However, this was not confirmed in this study as the sole mutation by complete genome sequencing or by reverse genetic analysis. In addition, primary reversion occurs rapidly in biological mutants with a single-nucleotide polymorphism; therefore, possible second-site suppressors could not be identified. We engineered the F219L codon change as a two-nucleotide mutation ($_{10864}$UUU to CUA) in the isogenic MHV clone, which would require a two-nucleotide change for reversion to Phe219 (UUU or UUC). The engineered recombinant F219L mutant was recovered at 30°C, and complete genome sequencing confirmed the $_{10864}$UUU-to-CUA mutations as the only changes in the genome. The recombinant F219L mutant had an EOP of 3×10^{-5}, confirming that the F219L substitution alone was sufficient to confer a *ts* phenotype (Fig. 3A). The titer and plaque morphology of the *ts*F219L mutant were indistinguishable from those of WT MHV at 30°C (data not shown).

Identification of second-site suppressor mutations of recombinant *ts*F219L. To select for phenotypic revertants or suppressor mutations of *ts*F219L, DBT-9 cells were infected with recombinant *ts*F219L mutant virus at 40°C. However, no cytopathic effect (CPE) or productive infection occurred at 40°C, despite multiple attempts and prolonged incubation. Consequently, we initiated infection at 30°C for 6 h, followed by a shift to 37°C for 24 h. This stock was then passaged at 37°C with a shift to 40°C, followed by passage and selection of 10 plaques at 40°C. Sequencing of 10 plaque clones confirmed retention of the engineered $_{10864}$CUA (F219L) codon. However, all 10 isolated plaque cloned viruses demonstrated one of two patterns of second-site mutations in nsp5 in the presence of *ts*F219L: (i) H134Y ($_{10608}$CAT to $_{10608}$TAT)/H270 duplication (H270HH) ($_{11016}$CAT to $_{11016}$CATCAT) with an EOP of 1.2 or (ii) H134Y (CAT to TAT)/E285V ($_{11061}$GAA to $_{11061}$GTT) with an EOP of 2.9. To test the contribution of the H134Y, E285V, and H270HH changes to suppression of *ts*F219L, the identified mutations were engineered with and without F219L, alone or in the combination seen in the recovered viruses. In addition, although no F219L/

A

Virus Strain	Titer at: 30°C	Titer at: 40°C	EOP (40/30)
WT MHV	9×10^7	3×10^8	3×10^0
F219L	1×10^8	3×10^3	3×10^{-5}
F219L / H134Y / E285V*	1×10^7	4×10^7	3×10^0
F219L / H134Y / H270HH*	2×10^7	2×10^4	1×10^0
F219L / H270HH / E285V^	3×10^7	3×10^7	1×10^0
F219L / E285V^	1×10^7	2×10^2	2×10^{-5}
F219L / H134Y^	1×10^8	1×10^3	1×10^{-5}
F219L / H270HH^	8×10^6	1×10^4	2×10^{-3}

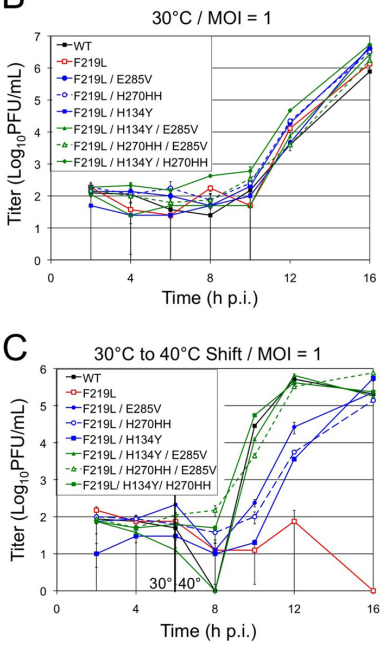


FIG 3 Analysis of replication of *ts*F219L and second-site suppressor mutants. (A) Virus titers and EOP data for WT MHV and F219L mutant viruses determined by plaque assay at 30°C and 40°C. Titers were calculated in duplicate by infection of DBT-9 cells. Recombinant second-site suppressor mutants (*) and engineered recombinants whose mutations were artificially recombined (^) are indicated. (B and C) Growth analysis of WT MHV, recombinant *ts*F219L, suppressor mutants (F219L/H134Y/E285V and F219L/H134Y/H270HH), and engineered mutants (F219L/H270HH/E285V, F219L/E285V, F219L/H270HH, and F219L/H134Y) grown at 30°C (B) or grown at 30°C and then shifted to 40°C at 6 h p.i. (C) on DBT-9 cells. The virus titers were determined by plaque assay on DBT-9 cells at 30°C.

E285V/H270HH mutant was identified among the sequenced plaques, we engineered this combination to test for the capacity of the combination to also suppress the *ts* phenotype. In total, nine genomes were engineered: F219L/E285V, F219L/H134Y, F219L/H270HH, F219L/H134Y/E285V, F219L/H134Y/H270HH, F219L/H270HH/E285V, H134Y/E285V, H134Y/H270HH, and H270HH/E285V. All viruses were recovered at 30°C with the engineered mutations detected and confirmed by sequencing across nsp5. The EOP was determined for each of the viruses by plaque assay on DBT-9 cells at 30°C and 40°C (Fig. 3A). The second-site substitution combinations identified by reversion analysis were sufficient to suppress the F219L *ts* phenotype: H134Y/E285V, EOP = 2.9; and H134Y/H270HH; EOP = 1.2. When the second-site suppressor alleles were tested alone with F219L, the results showed that the individual substitutions either minimally or partially suppressed *ts*F219L: H134Y, EOP = 2×10^{-5}; E285V, EOP = 1×10^{-5}; and H270HH, EOP = 2×10^{-3}. When the second-site substitutions were introduced in the WT background, either alone or in combination, there was no effect on EOP, suggesting that the changes were not responsible for any replication defects in the presence of F219L. The nonbiologically derived combination of F219L/E285V/H270HH also completely suppressed the *ts*F219L phenotype (F219L/E285V/H270HH, EOP = 1.0). The EOP results confirmed that the biologically identified second-site substitution combinations were both necessary and sufficient to suppress the F219L *ts* phenotype. Further, the results showed that H134Y did emerge as a suppressor allele for domain 3 *ts*F219L, but that in contrast to *ts*V148A and *ts*S133A, suppression of *ts*F219L required at least one other substitution in domain 3 in combination with H134Y.

Growth of *ts*F219L and revertants. The *ts*F219L and recombinant suppressor mutant viruses were grown in DBT-9 cells at 30°C or beginning at 30°C with a shift to 40°C at 6 h p.i. (Fig. 3B and C). At 30°C, the growth kinetics and virus yields for all mutant viruses were indistinguishable from those of WT MHV. Following the shift to 40°C at 6 h p.i., *ts*F219L showed no further replication for 10 h. In contrast, the mutant viruses containing any two of the H134Y, E285V, and H270HH substitutions showed growth after the shift to 40°C similar to WT. In contrast, all single suppressor residues expressed with F219L demonstrated a 4-h lag before exponential growth compared to the double mutants, but ultimately, they achieved peak titers similar to those of the double mutants. The titers at 10 h p.i. were consistent with the EOP data (10^{-3} to 10^{-5} compared to WT) and overall demonstrated that the individual mutations were capable of improved viral replication compared to the *ts*F219L virus yet still were *ts* compared to WT (Fig. 3A and C). Thus, in contrast to *ts*V148A and *ts*S133A, suppression of *ts*F219L appears to require a combination of at least two second-site mutations. This may be the explanation for the tight *ts*F219L phenotype at 40°C, as well as for the necessity of sequential passage of the *ts*F219 mutant virus at 30°C, 37°C, and 40°C to recover phenotypic revertants. In addition, these results identify H134Y as a second-site suppressor for a third *ts* allele in nsp5, this one in a domain 3. Overall, the results demonstrate cooperation of H134, E285, and H270 in nsp5 for efficient virus replication, as well confirming communication between nsp5 domains 2 and 3.

The *ts*S133A and *ts*F219L mutant viruses have temperature-sensitive impaired processing by nsp5. The nsp5 protease is present in all CoVs and is responsible for 11 maturation cleavage events in the replicase polyprotein (nsp4 through nsp16). To directly compare the nsp5 protease activity of *ts*V148A, *ts*S133A, *ts*F219L, and second-site suppressor mutants, DBT-9 cells were infected with WT and mutant viruses at an MOI of 5 PFU/cell and incubated at 30°C. At 6 h p.i., replicate monolayers were maintained at 30°C or transferred to 40°C, and infected cells were radiolabeled with [^{35}S]Met-Cys. Lysates from infected, radiolabeled cells were immunoprecipitated with antibodies specific for nsp2, nsp5, and nsp8 to test for processing of nsp2 by PLP1 and of nsp5 and nsp8 by nsp5.

Immunoprecipitation with antibodies specific for nsp2 of all lysates from WT and mutant virus-infected cells with labeling at 30°C and 40°C resulted in detection of mature processed nsp2, demonstrating that at both permissive and nonpermissive temperatures, there was similar translation of pp1a (nsp1 to nsp11) and normal PLP1 activity (Fig. 4). Recently, Stokes et al. reported that a *ts* mutation in nsp3 resulted in a significant decrease in nsp5-mediated processing (41). In our study, we detected the nsp4-to-nsp10 precursor polyprotein (150 kDa), which demonstrates that PLP2 is functional and is not inhibited by the nsp5 mutations (Fig. 4). The presence of mature nsp2 and the p150 bands at the nonpermissive temperature indicate that both PLP1 and PLP2 domains of nsp3 are active and are not affected by the nsp5 mutations. Although we did not test the processing of nsp3 directly, detection of both p150 and mature nsp2 is consistent with normal processing of N and C termini of nsp3. Immunoprecipitation of cells infected with WT MHV by nsp5-specific antibodies at both 30°C and 40°C resulted in detection of mature processed nsp5, as well as coimmunoprecipitation of nsp8, and two distinct small protein bands at 10 and 12 kDa, which is consistent with the predicted migration of nsp7 and nsp9. Immunoprecipitation with nsp8-specific antibodies detected nsp8 as well as probable coprecipitation of nsp5 and the 10- and 12-kDa proteins. These results show that expression and processing of pp1a proteins nsp5, nsp7, nsp8, and nsp9 are accelerated at 40°C in cells infected with WT MHV.

The temperature-sensitive viruses *ts*S133A and *ts*F219L, as well as the previously described *ts*V148A, exhibited profoundly impaired processing of nsp5 and nsp8 at 40°C compared to 30°C, indicating a specific defect in processing by the nsp5 protease. This was consistent with decreased detection of the reciprocal coimmunoprecipitating protein (nsp8 or nsp5) as well as decreased detection of the 10- and 12-kDa proteins. Although the detection of nsp5-processed proteins was profoundly decreased, we were unable from multiple replicate experiments (more than 5 experiments) to demonstrate complete loss of nsp5 activity. The results suggest that the S133A, F219L, and V148A mutations do not directly affect the catalytic or substrate-binding functions of nsp5 but rather modify protease activity in other ways. Alternatively, it is possible that the residual processing might be the result of nsp5 expressed and folded into active forms or complexes prior to the temperature shift, and thus, the protein still retains residual activity.

Viruses carrying biological and engineered suppressor mutations demonstrated restoration of processing by nsp5 that directly correlated with the degree of recovery of EOP and virus growth. Single second-site mutants S133A/T129M, F219L/H134Y, F219L/H270HH, and F219L/E285V showed an increase in detectable processed nsp5 and nsp8 only, while the double second-site suppressors restored WT-like patterns of processed proteins. Collectively, the results show a direct correlation of detection of proteins

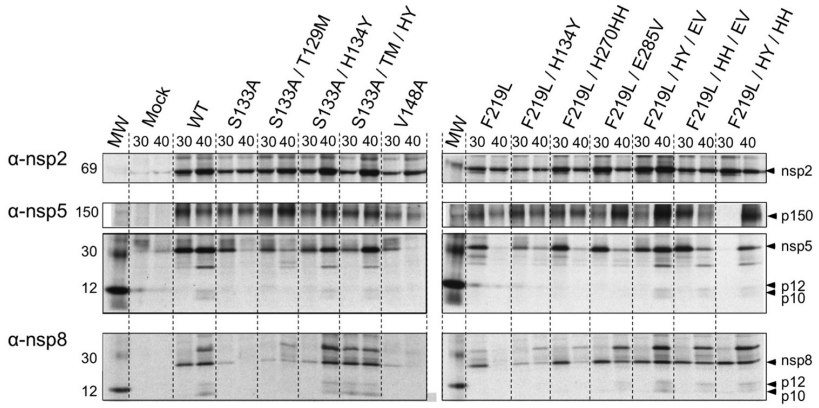


FIG 4 Proteolytic processing of WT, *ts*S133A, *ts*F219L, and suppressor viruses. DBT-9 cells were ^{35}S radiolabeled during viral infection or mock infection. Cellular lysates were harvested from cells infected with WT MHV, *ts*S133A mutants, *ts*F219L mutants, and previously described *ts*V148A and from mock-infected cells (40). Labeled proteins were immunoprecipitated using antiserum specific for nsp2, nsp5, and nsp8. The temperature during virus infection is indicated above the lanes (30°C [30] or a temperature shift from 30° to 40°C at 6 h p.i. [40]). One hundred microliters of lysate was used for all immunoprecipitations. The positions of putative viral proteins are shown to the right of the gels based upon the predicted size, and identified bands are labeled. The positions of molecular weight standards (MW) are shown to the left of the gels, and sizes are shown in kilodaltons. Protein expression profiles were resolved by SDS-PAGE and visualized by autoradiography. α-nsp2, anti-nsp2 antibody.

processed by nsp5 (nsp5 and nsp8) and the extent of restored growth fitness in culture and EOP analysis in plaque assay. Further, these findings indicate that the impairment in growth at nonpermissive temperatures is not due to complete inactivation of nsp5 protease activity.

Analysis of *ts* and suppressor mutations in nsp5 structures. To evaluate the structural relationships between the *ts* and second-site suppressor mutations, the distances between the combinations of *ts* and suppressor residues were determined by analyzing the crystal structure of the nsp5 protease of a closely related human CoV, HKU1. The structure of MHV nsp5 has not been determined; however, MHV and HKU1 nsp5 proteases exhibit 84% sequence identity (Fig. 1C) and share all of the same amino acids at the residue positions reported in this study with the exceptions of H134 and H270 (Y134 and Y270 in HKU1). All of the second-site suppressor residue positions in HKU1 nsp5 were greater than 10 Å distance from the *ts* residues in the monomeric structure with the sole exception that the H134 residue is 3.8 Å from the juxtaposed S133 (data not shown). Measurement of the distance between residues in different monomers of the dimeric structure of both SARS-CoV and HKU1 nsp5 demonstrated that no two residues from this study were closer than 15.0 Å. Measurement of the distance between the *ts* and second-site suppressor mutants and the catalytic dyad residues, H41 and C145, showed that only V148A was within 10 Å of either residue. Collectively, these data demonstrate that the relationship between the *ts* and suppressor mutations cannot be explained by direct interactions between residues and that nsp5 dimerization does not provide direct intermonomer associations between the residues identified in this study. Modeling of the S133A and F219L mutations on the structure of HKU1 nsp5 failed to predict any clear pathways of side chain remodeling or perturbation between the *ts* residues and the protease active site (data not shown). In contrast, analysis of residue conservation using an alignment of 130 nonredundant CoV nsp5 amino acid sequences identified a series of 100% identical residues that span the regions of nsp5 between each of the *ts* residues (S133, V148, and F219) and the common second-site suppressor residue (H134) (Fig. 5). These findings indicate that the structural and functional perturbations on nsp5 protease of the *ts* mutations may span long distances across the protease structure through yet to be identified cooperative interactions.

Residue conservation of *ts* and suppressor alleles is group specific. To evaluate the variability at the *ts* and suppressor alleles,

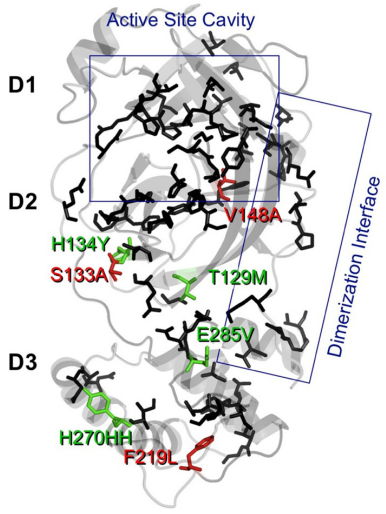


FIG 5 Conserved, *ts*, and suppressor alleles in a solved CoV nsp5 structure. A conservation map of 100% identical nsp5 residues (black) across 130 nonredundant CoV nsp5 sequences is shown on the HKU1 nsp5 protease monomer structure. The locations of identified *ts* alleles (red) and suppressor alleles (green) are indicated.

B

MHV Residue Position:	129	133	134	148	219	270	285
Genus: Alphacoronavirus							
α-CoV1 (TGEV)	N	Q	G	V	F	N	C
HCoV 229E	N	N	W	P	W	N	N
HCoV NL63	N	N	F	P	W	H	C
PEDV	N	N	Y	P	W	H	T
BtCoV HKU2	N	T	H	P	W	A	T
BtCoV 512/05	N	N	Y	P	W	H	C
BtCoV 1A	I	N	K	P	F	A	T
BtCoV HKU8	T	N	Y	P	F	A	T
Genus: Betacoronavirus							
MHV-A59	T	S	H	V	F	H	E
β-CoV1 (HCoV OC43)	T	S	Y	V	F	K	E
HCoV HKU1	T	S	Y	V	F	Y	E
BtCoV HKU9	T	N	G	P	F	L	E
BtCoV HKU4	N	N	S	V	F	H	E
BtCoV HKU5	N	Q	H	V	F	H	E
SARS-CoV Urbani	A	N	H	V	F	Q	E
Genus: Gammacoronavirus							
Avian Coronavirus (IBV)	T	N	G	V	S	S	E
Beluga Whale CoV SW1	T	N	G	V	F	A	E

FIG 6 Coronavirus protein sequence conservation of *ts* and suppressor alleles. (A) A sequence logo of conservation of *ts* and suppressor mutations across an alignment of 130 nonredundant CoV nsp5 sequences was generated using WebLogo (13). The height of each letter corresponds to the relative conservation of that amino acid at the position, and the height of the column corresponds to the sequence conservation at the position. The residue numbers are relative to the MHV amino acid positions. (B) The *ts* and suppressor residues of coronavirus species by CoV genus are shown for each MHV residue position. α-CoV1, alphacoronavirus 1; TGEV, transmissible gastroenteritis virus; PEDV, porcine epidemic diarrhea virus; BtCoV, bat coronavirus.

a sequence logo (13) was generated using the nonredundant alignment of 130 CoV nsp5 amino acid sequences, and the *ts* and suppressor alleles for 17 coronavirus species were analyzed (Fig. 6A and B). From two to eight different residues occupy each position across available CoV sequences (Fig. 6A). Surprisingly, the H134Y common second-site suppressor mutation selected for a tyrosine that is already present in several betacoronaviruses. Further, the S133N second-site suppressor first reported by Sparks et al. is common as an Asn in many coronaviruses (40). Conservation of distinct *ts* and suppressor alleles within the three genera suggest that there may be select combinations of alleles that are necessary for nsp5 activity (Fig. 6B). These findings further suggest that alterations in structure attributed to these residues could have analogous combinations in other CoVs.

DISCUSSION

Previous studies with other viruses as diverse as HIV, Sindbis virus, poliovirus, and vaccinia virus have reported *ts* mutations in virus proteases that affect protein processing, RNA synthesis, and virus capsid assembly (3, 23, 24, 29). Most of the *ts* alleles of described viral proteases have occurred at individual conserved residues or in pairs of conserved and structurally adjacent residues. For the HIV protease, residues distant from known catalytic and functional determinants have been shown to be critical for protease activity (22, 34). The current study of MHV nsp5 (3CLpro) extends our understanding of RNA virus proteases by demonstrating that multiple nonconserved and structurally distant residues in the CoV nsp5 protease participate in long-distance communication within and between the protease structural domains, function cooperatively to suppress *ts* phenotypes, and are important for nsp5 activity during replicase polyprotein processing. Specifically, we demonstrate that independent and physically distant *ts* alleles in domains 2 and 3 resulted in selection of the same H134Y *ts* suppressor allele, as well as selection of several additional suppressor alleles. With the exception of *ts*V148A, all of the residues are structurally distant from the catalytic, substrate, and dimerization residues in monomers and between nsp5 molecules in solved nsp5 dimer structures. In addition, suppressor mutations arose during reversion analysis in combinations that were required for or augmented restoration of nsp5 processing. Further, artificial combinations of suppressor mutations not seen during biological reversion analysis also showed cooperative suppression of the *ts* phenotype. These results all support the hypothesis that long-distance communication occurs between multiple residue nodes to regulate nsp5 activity.

Since nsp5 must recognize and process 11 closely related cleavage sites in the setting of rapidly changing substrate and cleavage site concentrations from polyproteins and processing intermediates, any explanation for nsp5 intra- or intermolecular communication needs to account for these evolving variables. The observation that a series of residues that are completely conserved across the coronaviruses span the regions between the *ts* and second-site mutations in solved and modeled CoV nsp5 structures (Fig. 5) suggests a possible mechanism for such communication. The potential linkage by conserved residues in the nsp5 tertiary structure would be similar to the model proposed by Ranganathan and coworkers in which networks of distant residues are connected in tertiary structure and are nodes for allosteric communication (22, 42). These networks may further demonstrate coevolution of residues that maintain the protease structure.

Temperature-sensitive mutations and nsp5 activity. Analyses of the MHV nsp5 *ts* mutant viruses demonstrated a defect in nsp5-mediated processing and virus replication at the nonpermissive temperature. However, mature nsp5 cleavage products were still detected for all three *ts* mutant viruses at the nonpermissive temperature, albeit at profoundly reduced levels, indicating that defects in viral growth at the nonpermissive temperature were not the result of complete loss of nsp5 activity. We previously reported that nsp5 activity was not present in *ts*V148A at the nonpermissive temperature, but in that study we did not test nsp5 processing at cleavage sites flanking nsp5. In the present study, we determined that *ts*V148A functions like *ts*S133A and *ts*F219L, exhibiting more impaired processing of nsp8 than nsp5 at the nonpermissive temperature, while retaining residual nsp5 activity at the flanking nsp4-nsp5 and nsp5-nsp6 cleavage sites (40). Overall, our results suggest that the nsp5 *ts* phenotype may be due to an alteration of protease activity at different polyprotein substrate sites. Altered cleavage site specificity has been reported for mutations introduced in the interdomain loop between domains 2 and 3 of the arterivirus equine arteritis virus (EAV) nsp4 protease, a structural orthologue of CoV nsp5 (43). Several of the recovered EAV nsp4 interdomain loop mutants did not abolish nsp4 protease activity but rather altered the substrate specificity. It is possible that the *ts* mutants of nsp5 are acting in a similar manner.

Nonconserved residues in nsp5 protease function. Align-

ment of 130 nonredundant CoV nsp5 sequences demonstrates that the residues that resulted in the nsp5 *ts* mutations are not conserved. There was more conservation of specific residues within the betacoronavirus genus, including MHV and SARS-CoV, specifically at residues V148, F219L, and E285. Finally, the revertants H134Y and S133N selected amino acids that already were present in nsp5 sequences of other alpha- and betacoronaviruses. Among the second-site residues, the H270 position shows the greatest variation with eight different amino acid residues tolerated across the CoVs. The tolerance for a wide variety of residues at this position may explain the acceptance of a duplicated codon resulting in a second histidine residue at the position. While the results have to be interpreted in light of the overall nsp5 sequence variability or conservation, they suggest that the *ts* and revertant residues may represent an evolutionarily adaptive network coevolving with other interacting proteins or cleavage sites. In support of this hypothesis, it was reported that a mutation at the P1 position in the cleavage site between nsp15 and nsp16 of the CoV infectious bronchitis virus resulted in a debilitated virus whose phenotype was compensated for by a mutation in nsp5 (17). Similarly, *ts* mutations in MHV nsp3 and nsp10 resulted in altered nsp5-mediated nsp5 processing (14, 41). A network of nonconserved and mutationally flexible residues could account for the rapid emergence of second-site revertants of *ts* viruses with combinations of more than one second-site mutation that also function in combinations not derived during virus reversion analysis.

Models for testing nsp5 long-distance communication. Coronavirus nsp5 functions in the setting of the largest known RNA virus polyprotein and must orchestrate 11 distinct cleavage events. It is clear from our findings and others that nsp5 activity is affected by changes within the structure of the protease, by changes at nsp5 cleavage sites, and by changes in other replicase proteins. Biochemical studies of the *ts* and revertant nsp5 molecules from this study should determine whether protein alterations and communication are intramolecular or within the nsp5 dimer. Continued mutagenesis of nsp5 in the context of a virus will permit testing for allosteric interactions across the replicase polyprotein that impact specific *ts* and revertant alleles during virus replication. To test for additional communication nodes in nsp5, we will continue to use an iterative approach for mutagenesis at *ts* and revertant alleles with alanine, nonconservative changes, and substitution of residues from other CoV nsp5 sequences. Finally, since some of the MHV revertant alleles such as H134Y are already present in other CoVs, it may be possible to test whether the variability at the revertant residues across different coronaviruses affects whether a particular substitution results in a *ts* phenotype in different nsp5 backgrounds. The application of this data set in combination with structure information and bioinformatic analysis should allow dissection of the extent and mechanism of the communication network that regulates nsp5 activity during coronavirus replication.

ACKNOWLEDGMENTS

We thank Dia Beachboard, Michelle Becker, Megan Culler, Lance Eckerle, Mark Gadlage, and Wayne Hsieh for technical assistance and helpful advice regarding manuscript preparation. We also thank Andrew Mesecar, Aimee Eggler, and Sakshi Tomar for helpful conversations and suggestions in preparing the manuscript and Eric Donaldson for assistance in preparing and using the multiple-sequence alignment for the manuscript.

Support for this work was provided by National Institutes of Health grant R01AI26603 (M.R.D.) from the National Institute of Allergy and Infectious Disease and the Virology Training Grant T32 AI089554 (C.C.S.) through Vanderbilt University. This work was also supported by The Elizabeth B. Lamb Center for Pediatric Research.

REFERENCES

1. **Anand K, et al.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J. **21**:3213–3224.
2. **Anand K, Ziebuhr J, Wadhwani P, Mesters JR, Hilgenfeld R.** 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science **300**:1763–1767.
3. **Ansarah-Sobrinho C, Moss B.** 2004. Role of the I7 protein in proteolytic processing of vaccinia virus membrane and core components. J. Virol. **78**:6335–6343.
4. **Ashkenazy H, Erez E, Martz E, Pupko T, Ben-Tal N.** 2010. ConSurf 2010: calculating evolutionary conservation in sequence and structure of proteins and nucleic acids. Nucleic Acids Res. **38**:W529–W533.
5. **Bacha U, et al.** 2008. Development of broad-spectrum halomethyl ketone inhibitors against coronavirus main protease 3CL(pro). Chem. Biol. Drug Des. **72**:34–49.
6. **Barrette-Ng IH, et al.** 2002. Structure of arterivirus nsp4.The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. J. Biol. Chem. **277**: 39960–39966.
7. **Barrila J, Bacha U, Freire E.** 2006. Long-range cooperative interactions modulate dimerization in SARS 3CLpro. Biochemistry **45**:14908–14916.
8. **Bost AG, Carnahan RH, Lu XT, Denison MR.** 2000. Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. J. Virol. **74**:3379–3387.
9. **Brierley I, Digard P, Inglis SC.** 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell **57**:537–547.
10. **Cavallaro JJ, Monto AS.** 1970. Community-wide outbreak of infection with a 229E-like coronavirus in Tecumseh, Michigan. J. Infect. Dis. **122**: 272–279.
11. **Chen H, et al.** 2006. Only one protomer is active in the dimer of SARS 3C-like proteinase. J. Biol. Chem. **281**:13894–13898.
12. **Chen S, et al.** 2008. Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. J. Biol. Chem. **283**:554–564.
13. **Crooks GE, Hon G, Chandonia JM, Brenner SE.** 2004. WebLogo: a sequence logo generator. Genome Res. **14**:1188–1190.
14. **Donaldson EF, Graham RL, Sims AC, Denison MR, Baric RS.** 2007. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J. Virol. **81**:7086–7098.
15. **Drosten C, et al.** 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N. Engl. J. Med. **348**:1967–1976.
16. **Eswar N, Eramian D, Webb B, Shen MY, Sali A.** 2008. Protein structure modeling with MODELLER. Methods Mol. Biol. **426**:145–159.
17. **Fang S, Shen H, Wang J, Tay FP, Liu DX.** 2010. Functional and genetic studies of the substrate specificity of coronavirus infectious bronchitis virus 3C-like proteinase. J. Virol. **84**:7325–7336.
18. **Gadlage MJ, Denison MR.** 2010. Exchange of the coronavirus replicase polyprotein cleavage sites alters protease specificity and processing. J. Virol. **84**:6894–6898.
19. **Gadlage MJ, et al.** 2010. Murine hepatitis virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. J. Virol. **84**:280–290.
20. **Gorbalenya AE, Koonin EV, Donchenko AP, Blinov VM.** 1989. Coronavirus genome: prediction of putative functional domains in the nonstructural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res. **17**:4847–4861.
21. **Grum-Tokars V, Ratia K, Begaye A, Baker SC, Mesecar AD.** 2008. Evaluating the 3C-like protease activity of SARS-coronavirus: recommendations for standardized assays for drug discovery. Virus Res. **133**:63–73.
22. **Halabi N, Rivoire O, Leibler S, Ranganathan R.** 2009. Protein sectors: evolutionary units of three-dimensional structure. Cell **138**:774–786.
23. **Kaplan AH, Manchester M, Smith T, Yang YL, Swanstrom R.** 1996.

Conditional human immunodeficiency virus type 1 protease mutants show no role for the viral protease early in virus replication. J. Virol. **70**:5840–5844.
24. **Kean KM, Agut H, Fichot O, Wimmer E, Girard M.** 1988. A poliovirus mutant defective for self-cleavage at the COOH-terminus of the 3C protease exhibits secondary processing defects. Virology **163**:330–340.
25. **Lee HJ, et al.** 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology **180**:567–582.
26. **Lu G, et al.** 2011. Enterovirus 71 and coxsackievirus A16 3C proteases: binding to rupintrivir and their substrates and anti-hand, foot, and mouth disease virus drug design. J. Virol. **85**:10319–10331.
27. **Lu X, Lu Y, Denison MR.** 1996. Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. Virology **222**:375–382.
28. **Lu Y, Denison MR.** 1997. Determinants of mouse hepatitis virus 3C-like proteinase activity. Virology **230**:335–342.
29. **Mayuri Geders TW, Smith JL, Kuhn RJ.** 2008. Role for conserved residues of Sindbis virus nonstructural protein 2 methyltransferase-like domain in regulation of minus-strand synthesis and development of cytopathic infection. J. Virol. **82**:7284–7297.
30. **McIntosh K, Becker WB, Chanock RM.** 1967. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. Proc. Natl. Acad. Sci. U. S. A. **58**:2268–2273.
31. **McIntosh K, Dees JH, Becker WB, Kapikian AZ, Chanock RM.** 1967. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. Proc. Natl. Acad. Sci. U. S. A. **57**:933–940.
32. **Nguyen TT, et al.** 2011. Virtual screening identification of novel severe acute respiratory syndrome 3C-like protease inhibitors and in vitro confirmation. Bioorg. Med. Chem. Lett. **21**:3088–3091.
33. **Perlman S, Netland J.** 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nat. Rev. Microbiol. **7**:439–450.
34. **Perryman AL, Lin JH, McCammon JA.** 2004. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci. **13**:1108–1123.
35. **Poordad F, Khungar V.** 2011. Emerging therapeutic options in hepatitis C virus infection. Am. J. Manag. Care. **17**:S123–S130.
36. **Sawicki SG, et al.** 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog. **1**:e39.
37. **Shi J, Sivaraman J, Song J.** 2008. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. J. Virol. **82**:4620–4629.
38. **Shi J, Song J.** 2006. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. FEBS J. **273**:1035–1045.
39. **Sims AC, Ostermann J, Denison MR.** 2000. Mouse hepatitis virus replicase proteins associate with two distinct populations of intracellular membranes. J. Virol. **74**:5647–5654.
40. **Sparks JS, Donaldson EF, Lu X, Baric RS, Denison MR.** 2008. A novel mutation in murine hepatitis virus nsp5, the viral 3C-like proteinase, causes temperature-sensitive defects in viral growth and protein processing. J. Virol. **82**:5999–6008.
41. **Stokes HL, et al.** 2010. A new cistron in the murine hepatitis virus replicase gene. J. Virol. **84**:10148–10158.
42. **Suel GM, Lockless SW, Wall MA, Ranganathan R.** 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nat. Struct. Biol. **10**:59–69.
43. **van Aken D, Snijder EJ, Gorbalenya AE.** 2006. Mutagenesis analysis of the nsp4 main proteinase reveals determinants of arterivirus replicase polyprotein autoprocessing. J. Virol. **80**:3428–3437.
44. **van der Hoek L, et al.** 2004. Identification of a new human coronavirus. Nat. Med. **10**:368–373.
45. **Woo PC, et al.** 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J. Virol. **79**:884–895.
46. **Xue X, et al.** 2007. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J. Mol. Biol. **366**:965–975.
47. **Xue X, et al.** 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J. Virol. **82**:2515–2527.
48. **Yount B, Denison MR, Weiss SR, Baric RS.** 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. J. Virol. **76**:11065–11078.
49. **Zhao Q, et al.** 2008. Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. J. Virol. **82**:8647–8655.
50. **Ziebuhr J, Snijder EJ, Gorbalenya AE.** 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J. Gen. Virol. **81**:853–879.

REFERENCES

Anand, K., Palm, G.J., Mesters, J.R., Siddell, S.G., Ziebuhr, J., Hilgenfeld, R., 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra alpha-helical domain. EMBO J 21, 3213-3224.

Anand, K., Ziebuhr, J., Wadhwani, P., Mesters, J.R., Hilgenfeld, R., 2003. Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. Science 300, 1763-1767.

Anthony, S.J., Ojeda-Flores, R., Rico-Chavez, O., Navarrete-Macias, I., Zambrana-Torrelio, C.M., Rostal, M.K., Epstein, J.H., Tipps, T., Liang, E., Sanchez-Leon, M., Sotomayor-Bonilla, J., Aguirre, A.A., Avila-Flores, R., Medellin, R.A., Goldstein, T., Suzan, G., Daszak, P., Lipkin, W.I., 2013. Coronaviruses in bats from Mexico. J Gen Virol 94, 1028-1038.

Atilgan, A.R., Akan, P., Baysal, C., 2004. Small-world communication of residues and significance for protein dynamics. Biophys J 86, 85-91.

Bacha, U., Barrila, J., Gabelli, S.B., Kiso, Y., Mario Amzel, L., Freire, E., 2008. Development of broad-spectrum halomethyl ketone inhibitors against coronavirus main protease 3CL(pro). Chem Biol Drug Des 72, 34-49.

Baric, R.S., Fu, K., Schaad, M.C., Stohlman, S.A., 1990. Establishing a genetic recombination map for murine coronavirus strain A59 complementation groups. Virology 177, 646-656.

Barrette-Ng, I.H., Ng, K.K., Mark, B.L., Van Aken, D., Cherney, M.M., Garen, C., Kolodenko, Y., Gorbalenya, A.E., Snijder, E.J., James, M.N., 2002. Structure of arterivirus nsp4. The smallest chymotrypsin-like proteinase with an alpha/beta C-terminal extension and alternate conformations of the oxyanion hole. J Biol Chem 277, 39960-39966.

Barrila, J., Bacha, U., Freire, E., 2006. Long-range cooperative interactions modulate dimerization in SARS 3CLpro. Biochemistry 45, 14908-14916.
Barrila, J., Gabelli, S.B., Bacha, U., Amzel, L.M., Freire, E., 2010. Mutation of Asn28 disrupts the dimerization and enzymatic activity of SARS 3CL(pro). Biochemistry 49, 4308-4317.

Becker, M.M., Graham, R.L., Donaldson, E.F., Rockx, B., Sims, A.C., Sheahan, T., Pickles, R.J., Corti, D., Johnston, R.E., Baric, R.S., Denison, M.R., 2008. Synthetic recombinant bat SARS-like coronavirus is infectious in cultured cells and in mice. Proc Natl Acad Sci U S A 105, 19944-19949.

Bost, A.G., Carnahan, R.H., Lu, X.T., Denison, M.R., 2000. Four proteins processed from the replicase gene polyprotein of mouse hepatitis virus colocalize in the cell periphery and adjacent to sites of virion assembly. J Virol 74, 3379-3387.

Brierley, I., Digard, P., Inglis, S.C., 1989. Characterization of an efficient coronavirus ribosomal frameshifting signal: requirement for an RNA pseudoknot. Cell 57, 537-547.
Cavallaro, J.J., Monto, A.S., 1970. Community-wide outbreak of infection with a 229E-like coronavirus in Tecumseh, Michigan. J Infect Dis 122, 272-279.

Chen, H., Wei, P., Huang, C., Tan, L., Liu, Y., Lai, L., 2006. Only one protomer is active in the dimer of SARS 3C-like proteinase. J Biol Chem 281, 13894-13898.

Chen, S., Hu, T., Zhang, J., Chen, J., Chen, K., Ding, J., Jiang, H., Shen, X., 2008. Mutation of Gly-11 on the dimer interface results in the complete crystallographic dimer dissociation of severe acute respiratory syndrome coronavirus 3C-like protease: crystal structure with molecular dynamics simulations. J Biol Chem 283, 554-564.

Chen, S., Jonas, F., Shen, C., Hilgenfeld, R., 2010. Liberation of SARS-CoV main protease from the viral polyprotein: N-terminal autocleavage does not depend on the mature dimerization mode. Protein & cell 1, 59-74.

Cheng, S.C., Chang, G.G., Chou, C.Y., 2010. Mutation of Glu-166 blocks the substrate-induced dimerization of SARS coronavirus main protease. Biophys J 98, 1327-1336.

Chuck, C.P., Chow, H.F., Wan, D.C., Wong, K.B., 2011. Profiling of substrate specificities of 3C-like proteases from group 1, 2a, 2b, and 3 coronaviruses. PloS one 6, e27228.

Crooks, G.E., Hon, G., Chandonia, J.M., Brenner, S.E., 2004. WebLogo: a sequence logo generator. Genome Res 14, 1188-1190.

Cui, L.J., Zhang, C., Zhang, T., Lu, R.J., Xie, Z.D., Zhang, L.L., Liu, C.Y., Zhou, W.M., Ruan, L., Ma, X.J., Tan, W.J., 2011. Human Coronaviruses HCoV-NL63 and HCoV-HKU1 in Hospitalized Children with Acute Respiratory Infections in Beijing, China. Advances in virology 2011, 129134.

Deming, D.J., Graham, R.L., Denison, M.R., Baric, R.S., 2007. Processing of open reading frame 1a replicase proteins nsp7 to nsp10 in murine hepatitis virus strain A59 replication. J Virol 81, 10280-10291.

Dijkman, R., Jebbink, M.F., Gaunt, E., Rossen, J.W., Templeton, K.E., Kuijpers, T.W., van der Hoek, L., 2012. The dominance of human coronavirus OC43 and NL63 infections in infants. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 53, 135-139.

Donaldson, E.F., Graham, R.L., Sims, A.C., Denison, M.R., Baric, R.S., 2007. Analysis of murine hepatitis virus strain A59 temperature-sensitive mutant TS-LA6 suggests that nsp10 plays a critical role in polyprotein processing. J Virol 81, 7086-7098.

Drosten, C., Gunther, S., Preiser, W., van der Werf, S., Brodt, H.R., Becker, S., Rabenau, H., Panning, M., Kolesnikova, L., Fouchier, R.A., Berger, A., Burguiere, A.M., Cinatl, J., Eickmann, M., Escriou, N., Grywna, K., Kramme, S., Manuguerra, J.C., Muller, S., Rickerts, V., Sturmer, M., Vieth, S., Klenk, H.D., Osterhaus, A.D., Schmitz, H., Doerr, H.W., 2003. Identification of a novel coronavirus in patients with severe acute respiratory syndrome. N Engl J Med 348, 1967-1976.

Eswar, N., Eramian, D., Webb, B., Shen, M.Y., Sali, A., 2008. Protein structure modeling with MODELLER. Methods Mol Biol 426, 145-159.

Fan, K., Ma, L., Han, X., Liang, H., Wei, P., Liu, Y., Lai, L., 2005. The substrate specificity of SARS coronavirus 3C-like proteinase. Biochemical and biophysical research communications 329, 934-940.

Fang, S., Shen, H., Wang, J., Tay, F.P., Liu, D.X., 2010. Functional and genetic studies of the substrate specificity of coronavirus infectious bronchitis virus 3C-like proteinase. J Virol 84, 7325-7336.

Felber, L.M., Cloutier, S.M., Kundig, C., Kishi, T., Brossard, V., Jichlinski, P., Leisinger, H.J., Deperthes, D., 2004. Evaluation of the CFP-substrate-YFP system for protease studies: advantages and limitations. BioTechniques 36, 878-885.

Gadlage, M.J., Denison, M.R., 2010. Exchange of the coronavirus replicase polyprotein cleavage sites alters protease specificity and processing. J Virol 84, 6894-6898.

Gadlage, M.J., Sparks, J.S., Beachboard, D.C., Cox, R.G., Doyle, J.D., Stobart, C.C., Denison, M.R., 2010. Murine hepatitis virus nonstructural protein 4 regulates virus-induced membrane modifications and replication complex function. J Virol 84, 280-290.

Goetz, D.H., Choe, Y., Hansell, E., Chen, Y.T., McDowell, M., Jonsson, C.B., Roush, W.R., McKerrow, J., Craik, C.S., 2007. Substrate specificity profiling and identification of a new class of inhibitor for the major protease of the SARS coronavirus. Biochemistry 46, 8744-8752.

Gorbalenya, A.E., Koonin, E.V., Donchenko, A.P., Blinov, V.M., 1989. Coronavirus genome: prediction of putative functional domains in the non-structural polyprotein by comparative amino acid sequence analysis. Nucleic Acids Res 17, 4847-4861.

Graham, R.L., Becker, M.M., Eckerle, L.D., Bolles, M., Denison, M.R., Baric, R.S., 2012. A live, impaired-fidelity coronavirus vaccine protects in an aged, immunocompromised mouse model of lethal disease. Nat Med 18, 1820-1826.

Graham, R.L., Denison, M.R., 2006. Replication of murine hepatitis virus is regulated by papain-like proteinase 1 processing of nonstructural proteins 1, 2, and 3. J Virol 80, 11610-11620.

Grum-Tokars, V., Ratia, K., Begaye, A., Baker, S.C., Mesecar, A.D., 2008. Evaluating the 3C-like protease activity of SARS-Coronavirus: recommendations for standardized assays for drug discovery. Virus Res 133, 63-73.

Hegyi, A., Ziebuhr, J., 2002. Conservation of substrate specificities among coronavirus main proteases. J Gen Virol 83, 595-599.

Hsu, W.C., Chang, H.C., Chou, C.Y., Tsai, P.J., Lin, P.I., Chang, G.G., 2005. Critical assessment of important regions in the subunit association and catalytic action of the severe acute respiratory syndrome coronavirus main protease. J Biol Chem 280, 22741-22748.

Hudson, C.B., Beaudette, F.R., 1932. Infection of the Cloaca with the Virus of Infectious Bronchitis. Science 76, 34.

Jacobs, J., Grum-Tokars, V., Zhou, Y., Turlington, M., Saldanha, S.A., Chase, P., Eggler, A., Dawson, E.S., Baez-Santos, Y.M., Tomar, S., Mielech, A.M., Baker, S.C., Lindsley, C.W., Hodder, P., Mesecar, A., Stauffer, S.R., 2013. Discovery, Synthesis, And Structure-Based Optimization of a Series of N-(tert-Butyl)-2-(N-arylamido)-2-(pyridin-3-yl) Acetamides (ML188) as Potent Noncovalent Small Molecule Inhibitors of the Severe Acute Respiratory Syndrome Coronavirus (SARS-CoV) 3CL Protease. Journal of medicinal chemistry 56, 534-546.

Jung, H., Choi, Y.S., Lee, K.S., Han, D.S., Yu, Y.S., Im, M.Y., Fischer, P., Kim, S.K., 2012. Logic operations based on magnetic-vortex-state networks. ACS nano 6, 3712-3717.

Kindler, E., Jonsdottir, H.R., Muth, D., Hamming, O.J., Hartmann, R., Rodriguez, R., Geffers, R., Fouchier, R.A., Drosten, C., Muller, M.A., Dijkman, R., Thiel, V., 2013. Efficient replication of the novel human betacoronavirus EMC on primary human epithelium highlights its zoonotic potential. mBio 4, e00611-00612.

Knoops, K., Kikkert, M., Worm, S.H., Zevenhoven-Dobbe, J.C., van der Meer, Y., Koster, A.J., Mommaas, A.M., Snijder, E.J., 2008. SARS-coronavirus replication is supported by a reticulovesicular network of modified endoplasmic reticulum. PLoS biology 6, e226.

Kooi, C., Cervin, M., Anderson, R., 1991. Differentiation of acid-pH-dependent and -nondependent entry pathways for mouse hepatitis virus. Virology 180, 108-119.
Lau, S.K., Lee, P., Tsang, A.K., Yip, C.C., Tse, H., Lee, R.A., So, L.Y., Lau, Y.L., Chan, K.H., Woo, P.C., Yuen, K.Y., 2011. Molecular epidemiology of human coronavirus

OC43 reveals evolution of different genotypes over time and recent emergence of a novel genotype due to natural recombination. J Virol 85, 11325-11337.

Lau, S.K., Woo, P.C., Li, K.S., Huang, Y., Tsoi, H.W., Wong, B.H., Wong, S.S., Leung, S.Y., Chan, K.H., Yuen, K.Y., 2005. Severe acute respiratory syndrome coronavirus-like virus in Chinese horseshoe bats. Proc Natl Acad Sci U S A 102, 14040-14045.

Lee, H.J., Shieh, C.K., Gorbalenya, A.E., Koonin, E.V., La Monica, N., Tuler, J., Bagdzhadzhyan, A., Lai, M.M., 1991. The complete sequence (22 kilobases) of murine coronavirus gene 1 encoding the putative proteases and RNA polymerase. Virology 180, 567-582.

Lu, G., Liu, D., 2012. SARS-like virus in the Middle East: a truly bat-related coronavirus causing human diseases. Protein & cell 3, 803-805.

Lu, G., Qi, J., Chen, Z., Xu, X., Gao, F., Lin, D., Qian, W., Liu, H., Jiang, H., Yan, J., Gao, G.F., 2011. Enterovirus 71 and Coxsackievirus A16 3C proteases: binding to Rupintrivir and their substrate, and anti-HFMD drug design. J Virol.

Lu, X., Lu, Y., Denison, M.R., 1996. Intracellular and in vitro-translated 27-kDa proteins contain the 3C-like proteinase activity of the coronavirus MHV-A59. Virology 222, 375-382.

Lu, Y., Denison, M.R., 1997. Determinants of mouse hepatitis virus 3C-like proteinase activity. Virology 230, 335-342.

Mackay, I.M., Arden, K.E., Speicher, D.J., O'Neil, N.T., McErlean, P.K., Greer, R.M., Nissen, M.D., Sloots, T.P., 2012. Co-circulation of four human coronaviruses (HCoVs) in Queensland children with acute respiratory tract illnesses in 2004. Viruses 4, 637-653.
McIntosh, K., Becker, W.B., Chanock, R.M., 1967a. Growth in suckling-mouse brain of "IBV-like" viruses from patients with upper respiratory tract disease. Proc Natl Acad Sci U S A 58, 2268-2273.

McIntosh, K., Dees, J.H., Becker, W.B., Kapikian, A.Z., Chanock, R.M., 1967b. Recovery in tracheal organ cultures of novel viruses from patients with respiratory disease. Proc Natl Acad Sci U S A 57, 933-940.

Mihindukulasuriya, K.A., Wu, G., St Leger, J., Nordhausen, R.W., Wang, D., 2008. Identification of a novel coronavirus from a beluga whale by using a panviral microarray. J Virol 82, 5084-5088.

Muramatsu, T., Kim, Y.T., Nishii, W., Terada, T., Shirouzu, M., Yokoyama, S., 2013. Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CL) from its polyproteins. FEBS J.

Nguyen, T.T., Ryu, H.J., Lee, S.H., Hwang, S., Breton, V., Rhee, J.H., Kim, D., 2011. Virtual screening identification of novel severe acute respiratory syndrome 3C-like protease inhibitors and in vitro confirmation. Bioorg Med Chem Lett 21, 3088-3091.
Normile, D., 2005. Virology. Researchers tie deadly SARS virus to bats. Science 309, 2154-2155.

Okamoto, D.N., Oliveira, L.C., Kondo, M.Y., Cezari, M.H., Szeltner, Z., Juhasz, T., Juliano, M.A., Polgar, L., Juliano, L., Gouvea, I.E., 2010. Increase of SARS-CoV 3CL peptidase activity due to macromolecular crowding effects in the milieu composition. Biol Chem 391, 1461-1468.

Oliner, J.D., Kinzler, K.W., Vogelstein, B., 1993. In vivo cloning of PCR products in E. coli. Nucleic Acids Res 21, 5192-5197.

Pasternak, A.O., Spaan, W.J., Snijder, E.J., 2006. Nidovirus transcription: how to make sense...? J Gen Virol 87, 1403-1421.

Perlman, S., Netland, J., 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nat Rev Microbiol 7, 439-450.

Perlman, S., Zhao, J., 2013. Human coronavirus EMC is not the same as severe acute respiratory syndrome coronavirus. mBio 4.

Perryman, A.L., Lin, J.H., McCammon, J.A., 2004. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. Protein Sci 13, 1108-1123.

Pinon, J.D., Mayreddy, R.R., Turner, J.D., Khan, F.S., Bonilla, P.J., Weiss, S.R., 1997. Efficient autoproteolytic processing of the MHV-A59 3C-like proteinase from the flanking hydrophobic domains requires membranes. Virology 230, 309-322.
Poordad, F., Khungar, V., 2011. Emerging therapeutic options in hepatitis c virus infection. Am J Manag Care 17, S123-130.

Prill, M.M., Iwane, M.K., Edwards, K.M., Williams, J.V., Weinberg, G.A., Staat, M.A., Willby, M.J., Talbot, H.K., Hall, C.B., Szilagyi, P.G., Griffin, M.R., Curns, A.T., Erdman, D.D., 2012. Human coronavirus in young children hospitalized for acute respiratory illness and asymptomatic controls. The Pediatric infectious disease journal 31, 235-240.

Pyrc, K., Sims, A.C., Dijkman, R., Jebbink, M., Long, C., Deming, D., Donaldson, E., Vabret, A., Baric, R., van der Hoek, L., Pickles, R., 2010. Culturing the unculturable: human coronavirus HKU1 infects, replicates, and produces progeny virions in human ciliated airway epithelial cell cultures. J Virol 84, 11255-11263.

Raj, V.S., Mou, H., Smits, S.L., Dekkers, D.H., Muller, M.A., Dijkman, R., Muth, D., Demmers, J.A., Zaki, A., Fouchier, R.A., Thiel, V., Drosten, C., Rottier, P.J., Osterhaus,

A.D., Bosch, B.J., Haagmans, B.L., 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature 495, 251-254.

Rota, P.A., Oberste, M.S., Monroe, S.S., Nix, W.A., Campagnoli, R., Icenogle, J.P., Penaranda, S., Bankamp, B., Maher, K., Chen, M.H., Tong, S., Tamin, A., Lowe, L., Frace, M., DeRisi, J.L., Chen, Q., Wang, D., Erdman, D.D., Peret, T.C., Burns, C., Ksiazek, T.G., Rollin, P.E., Sanchez, A., Liffick, S., Holloway, B., Limor, J., McCaustland, K., Olsen-Rasmussen, M., Fouchier, R., Gunther, S., Osterhaus, A.D., Drosten, C., Pallansch, M.A., Anderson, L.J., Bellini, W.J., 2003. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 300, 1394-1399.

Sawicki, S.G., Sawicki, D.L., 1998. A new model for coronavirus transcription. Adv Exp Med Biol 440, 215-219.

Sawicki, S.G., Sawicki, D.L., Younker, D., Meyer, Y., Thiel, V., Stokes, H., Siddell, S.G., 2005. Functional and genetic analysis of coronavirus replicase-transcriptase proteins. PLoS Pathog 1, e39.

Shi, J., Sivaraman, J., Song, J., 2008. Mechanism for controlling the dimer-monomer switch and coupling dimerization to catalysis of the severe acute respiratory syndrome coronavirus 3C-like protease. J Virol 82, 4620-4629.

Shi, J., Song, J., 2006. The catalysis of the SARS 3C-like protease is under extensive regulation by its extra domain. FEBS J 273, 1035-1045.

Sims, A.C., Ostermann, J., Denison, M.R., 2000. Mouse hepatitis virus replicase proteins associate with two distinct populations of intracellular membranes. J Virol 74, 5647-5654.

Sparks, J.S., Donaldson, E.F., Lu, X., Baric, R.S., Denison, M.R., 2008. A novel mutation in murine hepatitis virus nsp5, the viral 3C-like proteinase, causes temperature-sensitive defects in viral growth and protein processing. J Virol 82, 5999-6008.

Stobart, C.C., Lee, A.S., Lu, X., Denison, M.R., 2012. Temperature-sensitive mutants and revertants in the coronavirus nonstructural protein 5 protease (3CLpro) define residues involved in long-distance communication and regulation of protease activity. J Virol 86, 4801-4810.

Stokes, H.L., Baliji, S., Hui, C.G., Sawicki, S.G., Baker, S.C., Siddell, S.G., 2010. A new cistron in the murine hepatitis virus replicase gene. J Virol 84, 10148-10158.
Tang, S., Liao, J.C., Dunn, A.R., Altman, R.B., Spudich, J.A., Schmidt, J.P., 2007. Predicting allosteric communication in myosin via a pathway of conserved residues. J Mol Biol 373, 1361-1373.

van Aken, D., Snijder, E.J., Gorbalenya, A.E., 2006. Mutagenesis analysis of the nsp4 main proteinase reveals determinants of arterivirus replicase polyprotein autoprocessing. J Virol 80, 3428-3437.

van der Hoek, L., Pyrc, K., Jebbink, M.F., Vermeulen-Oost, W., Berkhout, R.J., Wolthers, K.C., Wertheim-van Dillen, P.M., Kaandorp, J., Spaargaren, J., Berkhout, B., 2004. Identification of a new human coronavirus. Nat Med 10, 368-373.

van Hemert, M.J., van den Worm, S.H., Knoops, K., Mommaas, A.M., Gorbalenya, A.E., Snijder, E.J., 2008. SARS-coronavirus replication/transcription complexes are membrane-protected and need a host factor for activity in vitro. PLoS Pathog 4, e1000054.

Vijayanand, P., Wilkins, E., Woodhead, M., 2004. Severe acute respiratory syndrome (SARS): a review. Clin Med 4, 152-160.

Vijgen, L., Keyaerts, E., Lemey, P., Maes, P., Van Reeth, K., Nauwynck, H., Pensaert, M., Van Ranst, M., 2006. Evolutionary history of the closely related group 2 coronaviruses: porcine hemagglutinating encephalomyelitis virus, bovine coronavirus, and human coronavirus OC43. J Virol 80, 7270-7274.

Vijgen, L., Keyaerts, E., Moes, E., Thoelen, I., Wollants, E., Lemey, P., Vandamme, A.M., Van Ranst, M., 2005. Complete genomic sequence of human coronavirus OC43: molecular clock analysis suggests a relatively recent zoonotic coronavirus transmission event. J Virol 79, 1595-1604.

Whitley, M.J., Lee, A.L., 2009. Frameworks for understanding long-range intra-protein communication. Curr Protein Pept Sci 10, 116-127.

Williams, R.K., Jiang, G.S., Holmes, K.V., 1991. Receptor for mouse hepatitis virus is a member of the carcinoembryonic antigen family of glycoproteins. Proc Natl Acad Sci U S A 88, 5533-5536.

Woo, P.C., Huang, Y., Lau, S.K., Yuen, K.Y., 2010. Coronavirus genomics and bioinformatics analysis. Viruses 2, 1804-1820.

Woo, P.C., Lau, S.K., Chu, C.M., Chan, K.H., Tsoi, H.W., Huang, Y., Wong, B.H., Poon, R.W., Cai, J.J., Luk, W.K., Poon, L.L., Wong, S.S., Guan, Y., Peiris, J.S., Yuen, K.Y., 2005. Characterization and complete genome sequence of a novel coronavirus, coronavirus HKU1, from patients with pneumonia. J Virol 79, 884-895.

Woo, P.C., Lau, S.K., Huang, Y., Yuen, K.Y., 2009. Coronavirus diversity, phylogeny and interspecies jumping. Exp Biol Med (Maywood) 234, 1117-1127.

Woo, P.C., Lau, S.K., Lam, C.S., Lau, C.C., Tsang, A.K., Lau, J.H., Bai, R., Teng, J.L., Tsang, C.C., Wang, M., Zheng, B.J., Chan, K.H., Yuen, K.Y., 2012. Discovery of seven

novel Mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol 86, 3995-4008.

Woo, P.C., Lau, S.K., Yip, C.C., Huang, Y., Tsoi, H.W., Chan, K.H., Yuen, K.Y., 2006. Comparative analysis of 22 coronavirus HKU1 genomes reveals a novel genotype and evidence of natural recombination in coronavirus HKU1. J Virol 80, 7136-7145.

Xue, X., Yang, H., Shen, W., Zhao, Q., Li, J., Yang, K., Chen, C., Jin, Y., Bartlam, M., Rao, Z., 2007. Production of authentic SARS-CoV M(pro) with enhanced activity: application as a novel tag-cleavage endopeptidase for protein overproduction. J Mol Biol 366, 965-975.

Xue, X., Yu, H., Yang, H., Xue, F., Wu, Z., Shen, W., Li, J., Zhou, Z., Ding, Y., Zhao, Q., Zhang, X.C., Liao, M., Bartlam, M., Rao, Z., 2008. Structures of two coronavirus main proteases: implications for substrate binding and antiviral drug design. J Virol 82, 2515-2527.

Yount, B., Denison, M.R., Weiss, S.R., Baric, R.S., 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. J Virol 76, 11065-11078.

Zaki, A.M., van Boheemen, S., Bestebroer, T.M., Osterhaus, A.D., Fouchier, R.A., 2012. Isolation of a novel coronavirus from a man with pneumonia in Saudi Arabia. N Engl J Med 367, 1814-1820.

Zhao, Q., Li, S., Xue, F., Zou, Y., Chen, C., Bartlam, M., Rao, Z., 2008. Structure of the main protease from a global infectious human coronavirus, HCoV-HKU1. J Virol 82, 8647-8655.

Ziebuhr, J., Snijder, E.J., Gorbalenya, A.E., 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. J Gen Virol 81, 853-879.