STRUCTURAL DETERMINANTS OF CORONAVIRUS NSP5 FUNCTION AND

INHIBITION

By

Lindsay Alexandra Maxwell

Thesis

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in fulfillment of the requirements

for the degree of

MASTER OF SCIENCE

in

Microbiology and Immunology

August 2014

Nashville, Tennessee

Approved:

Mark R. Denison

Chris R. Aiken

ACKNOWLEDGEMENTS

TABLE OF CONTENTS

LIST OF TABLES

Table   Page

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACE2, Angiotensin-converting enzyme 2

ARDS, acute respiratory distress syndrome

BHK-MHVR, Baby hamster kidney-MHV receptor cells

CD, Circular dichroism

CEACAM1, Carcinoembyronic antigen adhesion molecule 1

CM, Convoluted membrane

CoV, Coronavirus

CPE, Cytopathic effects

CTD, C-terminal domain

D1, Domain 1 (MHV residues 1 to 100)

D2, Domain 2 (MHV residues 101 to 199)

D3, Domain 3, C-terminal domain (MHV residues 200 to 303)

DBT, Delayed brain tumor cells

DEER, Double electron-electron resonance

DMV, Double membrane vesicle

DPP4, Dipeptidyl peptidase 4

E, Small membrane

EAV, Equine Arteritis Virus

EC_{50}, Effective concentration 50%

EPR, Electron paramagnetic resonance

ER, Endoplasmic reticulum

ERGIC, Endoplasmic reticulum-Golgi intermediate compartment

FFL, Firefly luciferase protein

FRET, Fluorescence resonance energy transfer

H5-MHV, nsp5 from HCoV-HKU1 in MHV backbone

H5-MHV-FFL, FFL fused to the N-terminal portion of nsp2 within H5-MHV

HCoVs, Human coronaviruses

HE, Hemagglutinin esterase

HIV-1, Human immunodeficiency virus type 1

h p.i., Hours post infection

HR, Hinge region

IBV, infectious bronchitis virus

IC_{50}, Inhibitory concentration 50%

IDL, Interdomain loop, C terminal region of domain 2 (MHV residues 184 to 199)

IP, immunoprecipitated

ITC, Isothermal titration calorimetry

L2, Rat lung epithelium

M, Membrane

MERS-CoV, Middle East respiratory syndrome CoV

MHV-FFL, FFL fused to the N-terminal portion of MHV nsp2

MOI, Multiplicity of infection

MTSSL, 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl-methanethiosulfonate

MuCoV, MHV, β-CoV murine coronavirus (murine hepatitis virus)

N, Nucleocapsid

NMR, Nuclear magnetic resonance

Nsp, Non-structural proteins

Nsp5, 3CLpro, Mpro Chymotrypsin-like cysteine protease (non-structural protein 5, 3C

like protease, main protease)

Nsp5-C, Nsp5 C-terminal domain alone

O5-MHV, nsp5 from HCoV-OC43 in MHV backbone

O5-MHV-FFL, FFL fused to the N-terminal portion of nsp2 within O5-MHV

P3, Passage 3

PBS, Phosphate buffered saline

PFU/cell, Plaque forming unit per cell

PLpro, Papain-like protease (nsp3)

P(r), Probability distribution of distances between two spin labels

r_{av}, average distance between two spin labels

RT-PCR, reverse transcription-polymerase chain reaction

RLU, Relative light units

S, Spike

SARS-CoV, Severe acute respiratory syndrome associated CoV

SDSL, Site-directed spin labeling

SEM, Standard error of the mean

TGEV, Transmissible gastroenteritis coronavirus

ts, Temperature sensitive

WT, Wild type

CHAPTER 1 :

BACKGROUND AND LITERATURE REVIEW

Introduction

Coronavirus (CoV) infections cause a number of human diseases ranging in severity from the common cold to severe acute respiratory syndrome (SARS). Middle East respiratory syndrome CoV (MERS-CoV) and severe acute respiratory syndrome CoV (SARS-CoV) demonstrate the continued capacity of zoonotic CoVs to cause severe human respiratory disease. The likelihood of future zoonotic CoV emergence coupled with a lack of effective therapeutics or vaccines emphasizes the need to understand mechanisms of CoV replication. CoVs have a positive-sense RNA genome that is translated into polyproteins that are processed by up to three encoded proteases. The chymotrypsin-like cysteine protease (nsp5, 3CLpro) is responsible for 11 distinct cleavages within the polyprotein. Nsp5 is required for coronavirus replication; thus, it is a key target for structure function studies and for viral inhibition. We propose a series of experiments designed to elucidate the mechanism of nsp5 activity in order to inform *de novo* inhibitor design. Data obtained from this study will (i) define the role of the novel third domain and interdomain loop of nsp5, which connects the catalytic fold to the third domain, during substrate selection and catalytic activity, and (ii) develop novel inhibitors of nsp5 that exploit the mechanism and functional targets of nsp5. Results from these studies will inform antiviral and vaccine

development. This work will may also result in the first experimentally demonstrated model of how nsp5 functions in cell culture infection.

Coronavirus identification, emergence, and disease

Coronaviruses (CoVs) are classified within the subfamily coronavirinae, within the family of coronaviridae and within the order of nidovirales. They are divided into three genera, alpha (α-CoV), beta (β-CoV), and gammacoronaviruses (γ-CoV), which correspond to the previous groupings of 1, 2, and 3, respectively (Figure 1.1) CoVs are positive strand RNA viruses that infect a wide range of species (Figure 1.2). Currently there are 6 known human coronaviruses (HCoVs) (Figure 1.1), which cause a broad range of disease in human hosts from the common cold to acute respiratory distress syndrome (ARDS). HCoV-NL63, an α-CoV, is believed to cause bronchiolitis in children while HCoV-HKU1, a β-CoV, is associated with chronic respiratory disease in the elderly (48). HCoV-OC43 and HCoV-229E, two other HCoVs are known to cause the common cold. While most HCoV symptoms are relatively mild, SARS-CoV and MERS-CoV are emerging zoonotic pathogens associated with severe acute respiratory illness and mortality.

Coronavirus infections pose significant human health risks as well as a huge economic burden. SARS-CoV putatively emerged into the human population from bats in the Guangdong province of China in November of 2002 (57). SARS-CoV spread throughout over twenty countries and caused a pandemic from 2002-2003. Of the roughly 8100 people infected, 774 died, resulting in a 10% case to fatality ratio. More severe illness was observed in elderly patients, with fatality rates reaching near 50% for those

over 50 years of age (47), while children younger than 12 years of age suffered much more mild symptoms (55). It is estimated that the SARS-CoV pandemic cost the United States healthcare system $30 – 100 billion (57), however, the majority of costs related to pandemics occur in non-health sectors, making it difficult to estimate the economic impact. For example, the WHO estimates that international travel fell by 50 - 70% and hotel occupancy dropped by >60% during the SARS-CoV pandemic (57).

MERS-CoV emerged in the human population putatively from close human contact to camels in Saudi Arabia in 2012. The case-fatality ratio has been around 40%, but it has proven much less transmissible than SARS-CoV, with 827 laboratory confirmed cases since emergence in 2012 and at least 287 deaths (41, 44, 46). Recent outbreaks of MERS-CoV infections have caused world health officials concern due to the potential for sustained person-to-person spread coupled with inadequately prepared healthcare systems (41). Further, current US military presence in the Middle East is concerning as military personnel are particularly susceptible to zoonotic infections due to increased stress and naïve immune systems. The mechanisms of CoV emergence are not understood, and no effective treatments or vaccines are available for CoV infections. Thus, is it essential to understand mechanisms of viral pathogenesis and differences in viral genomes that contribute to increased virulence.
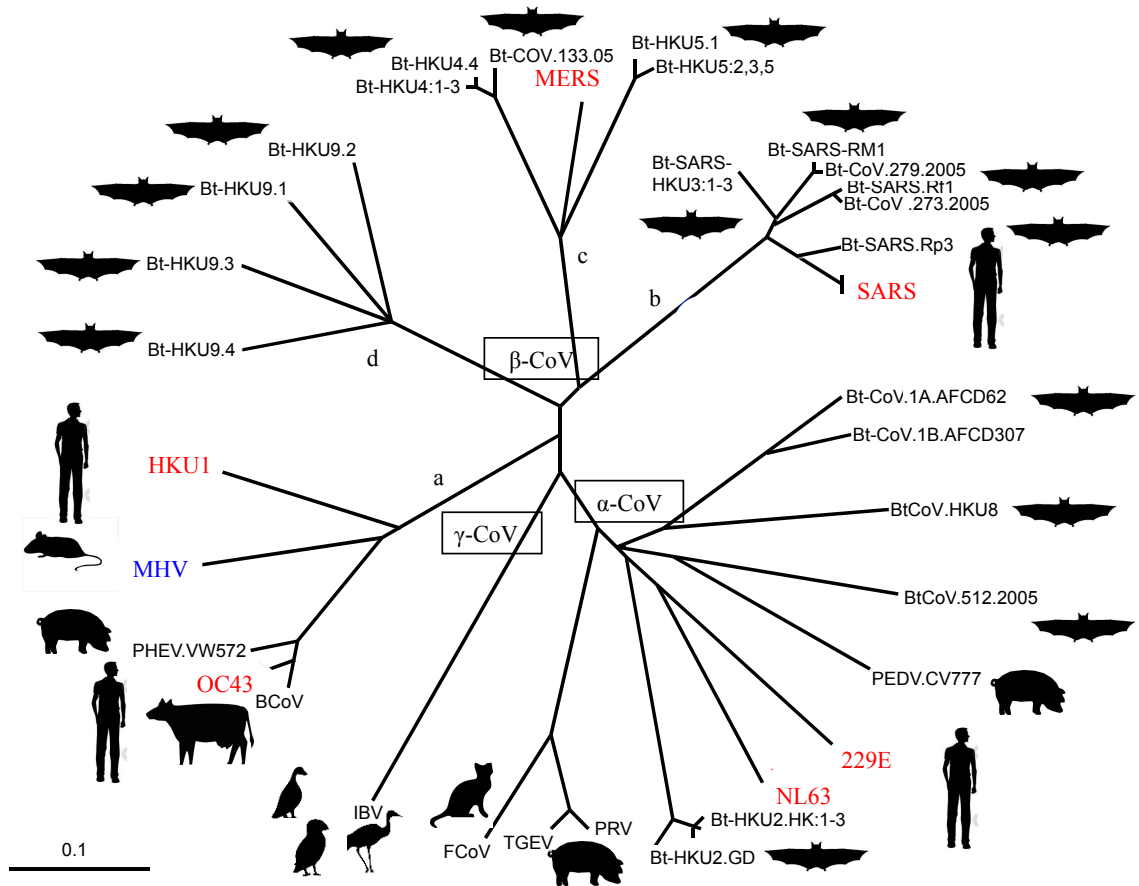
Figure 1.1 Phylogeny of Coronaviridae

The phylogeny of the family Coronaviridae is shown. Coronaviridae falls within the order of Nidovirales and is sub-classified into the genera α-, β-, γ-, and δ-CoV. The 6 known human CoVs are indicated in red, and the BSL-2 model CoV, murine hepatitis virus, is indicated in blue. *Modified from a drawing created by Eric Donaldson.*

CoVs contain the some of the largest known positive strand RNA viral genomes, ranging from 27-32 kb. CoVs are spherical or pleomorphic enveloped particles containing positive strand RNA associated with a helical nucleocapsid. The envelope has club-shaped glycoprotein surface projections, termed spike, which resemble solar coronas (Figure 1.2). Thus, CoVs are named in reference to their characteristic corona morphology seen in electron micrographs (Figure 1.2B and C). CoV genomes are arranged similarly with a 5' replicase gene, 3' structural genes, and 3' accessory genes. The encoded structural genes include hemagglutinin esterase (HE), if present, and all of the following: spike (S), small membrane (E), membrane (M), and nucleocapsid (N). Encoded accessory genes vary between virus strains. The replicase gene is translated as a polyprotein that is cleaved into 16 individual non-structural proteins (nsp) by encoded papain-like protease(s), which are domain(s) within nsp3, and the 3C-like protease (3CLpro, Mpro, nsp5) (Figure 1.2C) (42). The β-CoV murine coronavirus (MuCoV, MHV) has been highly studied as a model system to understand CoV replication, genetics, tropism, and pathogenesis (Figure 1.1). In fact, MHV represents a collection of virus strains that have been used to study pathogenesis the central nervous system, the liver, and the lung. MHV is also noted as one of the few animal models for the study of chronic demyelinating diseases such as multiple sclerosis (67). In the Denison lab, we use a strain of MHV, MHV-A59, as a model to study CoVs. The genome of MHV is indicated in Figure 1.2C.
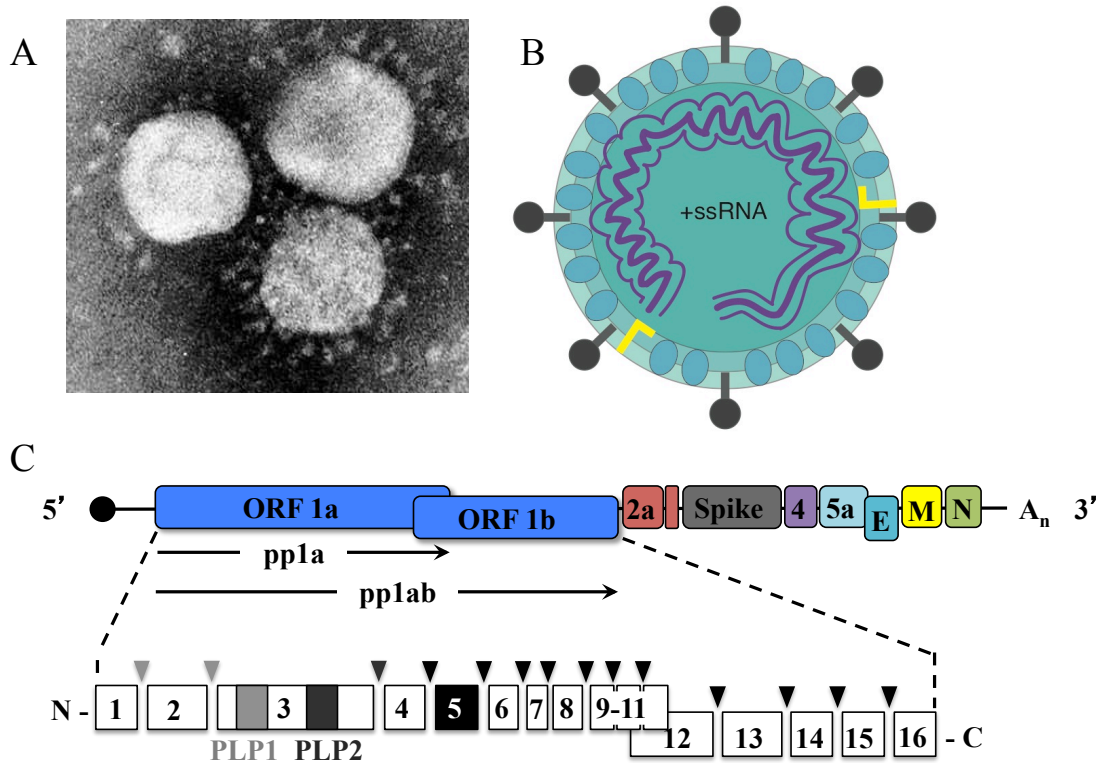
Figure 1.2 Coronavirus Structure and Genome Organization

A) Electron micrograph image of human coronavirus 229E (*Courtesy CDC Public Health Image Library*) showing distinct "corona" of spike proteins on membrane. B) Virion structure with spike protein indicated in grey, membrane protein indicated in light blue, and envelope protein indicated in yellow. *Modified from a drawing created by Megan Culler-Freeman* C) Genomic organization of MHV CoV. Polycistronic transcript is shown with genes indicated. Polyprotein 1a and 1ab (pp1a, pp1ab) translation products are shown with protease cleavage sites indicated by arrowheads.

Coronavirus lifecycle

Coronaviruses initiate cellular infection by receptor-mediated endocytosis. The CoV class I fusion S protein binds cellular receptors. Cellular receptors used by CoVs include carcinoembyronic antigen adhesion molecule 1 (CEACAM1) by MHV, aminopeptidase N by HCoV-229E, angiotensin-converting enzyme 2 (ACE2) by HCoV-NL63 and SARS-CoV, 9-*O*-acetylated sialic acid by HCoV-OC43, and dipeptidyl peptidase 4 (DPP4) by MERS-CoV (47, 49, 68). Once endocytosed, viral nucleocapsid is released in the cytoplasm, and the message sense viral RNA is translated into the replicase polyprotein (pp1a or pp1ab) by host ribosomes (31, 33). Genomic replication occurs within replication complexes formed on modified endoplasmic reticulum (ER) membranes. Nsps 3, 4, and 6 have been demonstrated to modify host ER membranes, resulting in double membrane vesicle (DMV) and convoluted membrane (CM) structures (Figure 1.3B) (3, 30, 66). The genome is replicated through a full-length negative sense RNA intermediate that is used as template for new positive sense progeny genomes.

The 3' end of the genome is replicated through subgenomic mRNAs, which are then translated into structural proteins (Figure 1.2B), HE, S, E, M, and N. The structural proteins are trafficked through the endoplasmic reticulum-Golgi intermediate compartment (ERGIC), where N protein encapsulates progeny genomes and virus assembly occurs. Progeny virions are released via non-lytic exocytosis pathways (Figure 1.3B). Finally, extra S protein is transported to the cell membrane, where it can cause cells to fuse and form multinucleated giant cells, syncytia. This allows the virus to infect new cells while putatively avoiding host immune detection (Figure 1.3) (42).
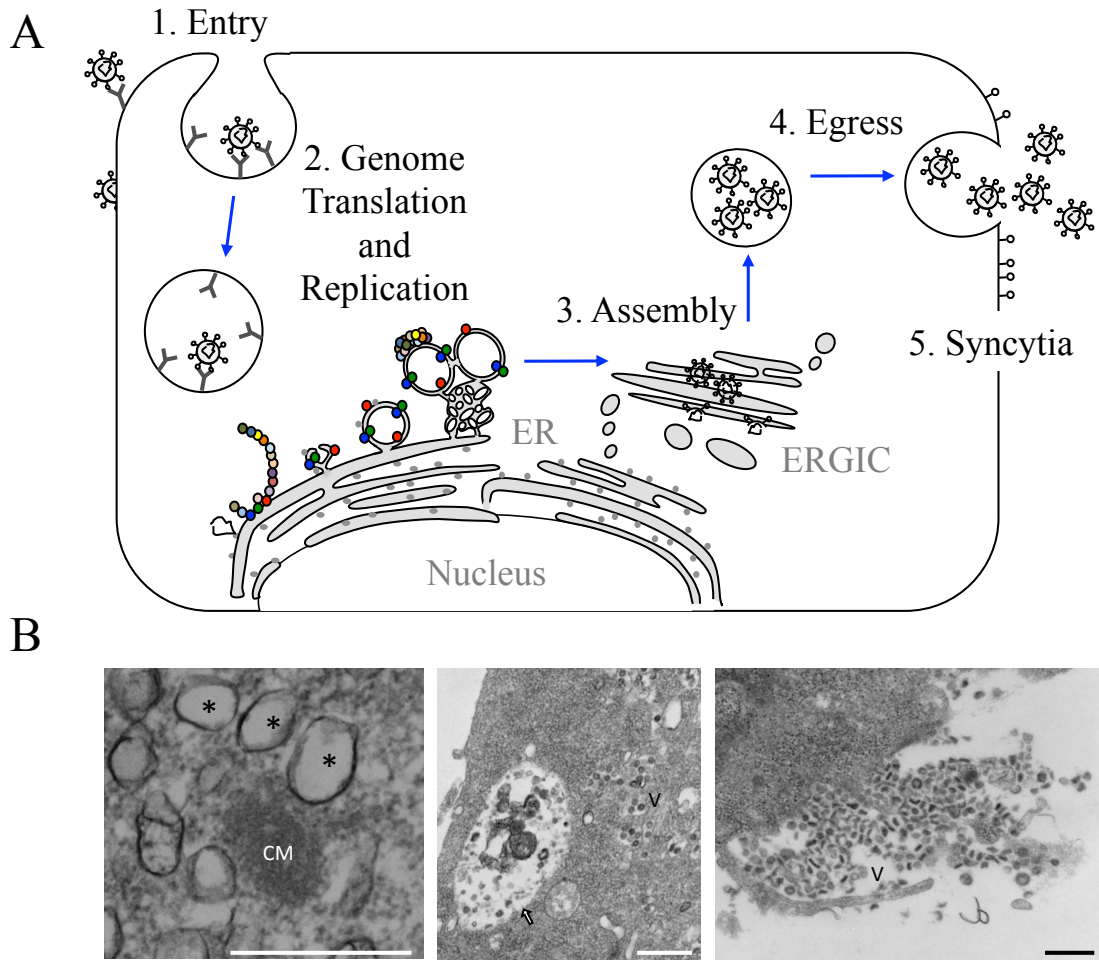
Figure 1.3 Coronavirus Replication

A) Model of coronavirus replication is shown. Briefly, the virus enters by receptor mediated endocytosis, the genome is translated and replication occurs at replication complexes on modified ER membranes, assembly occurs in the ERGIC, new virions emerge via non-lytic pathways, and syncytia formation likely occurs due to extra viral spike protein on the cell surface. Blue, green and red represent nsps 3, 4, and 6. Other colors indicate nsps 1, 2, 5, and 7-16. B) CM and DMV are indicated (left), virions in vesicles prior to egress (middle), and progeny virions exiting the cell (right). Scale bars represent 500 nm. *Electron micrographs by Dia Beachboard, unpublished.*

Coronavirus nsp5 protease structure

The CoV genome is translated into polyproteins pp1a (486 kDa) and pp1ab (790 kDa), which are then cleaved into functional non-structural proteins by encoded proteases (13, 20). One or two encoded papain-like protease(s) (PLpro, nsp3) process the N-terminal portions of pp1a and pp1ab, resulting in autoproteolytic cleavage and mature nsp1, nsp2, and nsp3 (22) (Figure 1.2). PLpro is considered papain-like due to a conserved Cys/His catalytic dyad and α/β fold (76). The C-terminal regions of pp1a and pp1ab are processed by nsp5 at 7 and 11 sites, respectively, resulting in mature nsp4-nsp16 as well as its own autoproteolytic cleavage (11, 24, 34, 45).

Originally reported by Lu *et. al.*, nsp5 is a 27 kDa protein composed of three domains. Domain 1 (MHV residues 1 – 100 [D1]) and domain 2 (MHV residues 101 – 199 [D2]) form a chymotrypsin-like fold with a His41 Cys145 catalytic dyad (Figure 1.4A and B) (38-40). It has been demonstrated that mutation of Cys145 to Ala145 (C145A) abolishes catalytic activity without altering folding (4, 5, 35, 62). CoV nsp5 proteases are among the largest of their type, consisting of 300-306 amino acids, which is about 100 residues more than observed in picornavirus 3C protease homologs (76). The size discrepancy is due to an additional C-terminal domain, domain 3 (MHV residues 200 - 303 [D3]) that is unique to CoVs (Figure 1.4B). This unique domain is an α-helical bundle required for replication (51, 64). It has been proposed that D3 mediates dimerization (Figure 1.4C), but the exact function remains to be elucidated (53, 54, 63). A conserved interdomain loop (IDL) (MHV residues 184 to 199) connects the domains 1/2 to D3.

Cleavage of nsp5 substrates occurs most frequently after a conserved Gln (Q) in the P1 position, similar to picornavirus 3Cpro cleavage sites (Figure 1.4D) (39). The consensus nsp5 cleavage site among known CoVs is P2-LQ†(S/G/A/V)-P1' (Figure 1.4D) (10, 14, 76). Thus, CoV nsp5 proteases have functional and structural homology to picornavirus 3C proteases (51).

The substrate-binding pocket is between D1 and D2, where the P1 residue of the substrate interacts with a conserved Tyr – x – His motif (Figure 1.4D). In MHV nsp5, the substrate-binding pocket is composed of Tyr161, Met162, and His163 (Figure 1.4A). Multiple groups have suggested that the His163 of this motif forms hydrogen bonds with the P1 Gln residue of substrate (19, 70).

Studies suggest that nsp5 acts as a functional dimer, with the two monomers oriented perpendicular to one another (Figure 1.4C) (8, 36, 52, 69, 74). Although dimerization is required for nsp5 activity, studies suggest that N-terminal autoprocessing occurs via a monomeric form (6, 7). Further, *in vitro* studies suggest that only one monomer is catalytically active within the dimer (5). This activity loss has been attributed to the collapse of the oxyanion hole of the S1 substrate-binding site (Figure 1.4D) (61).
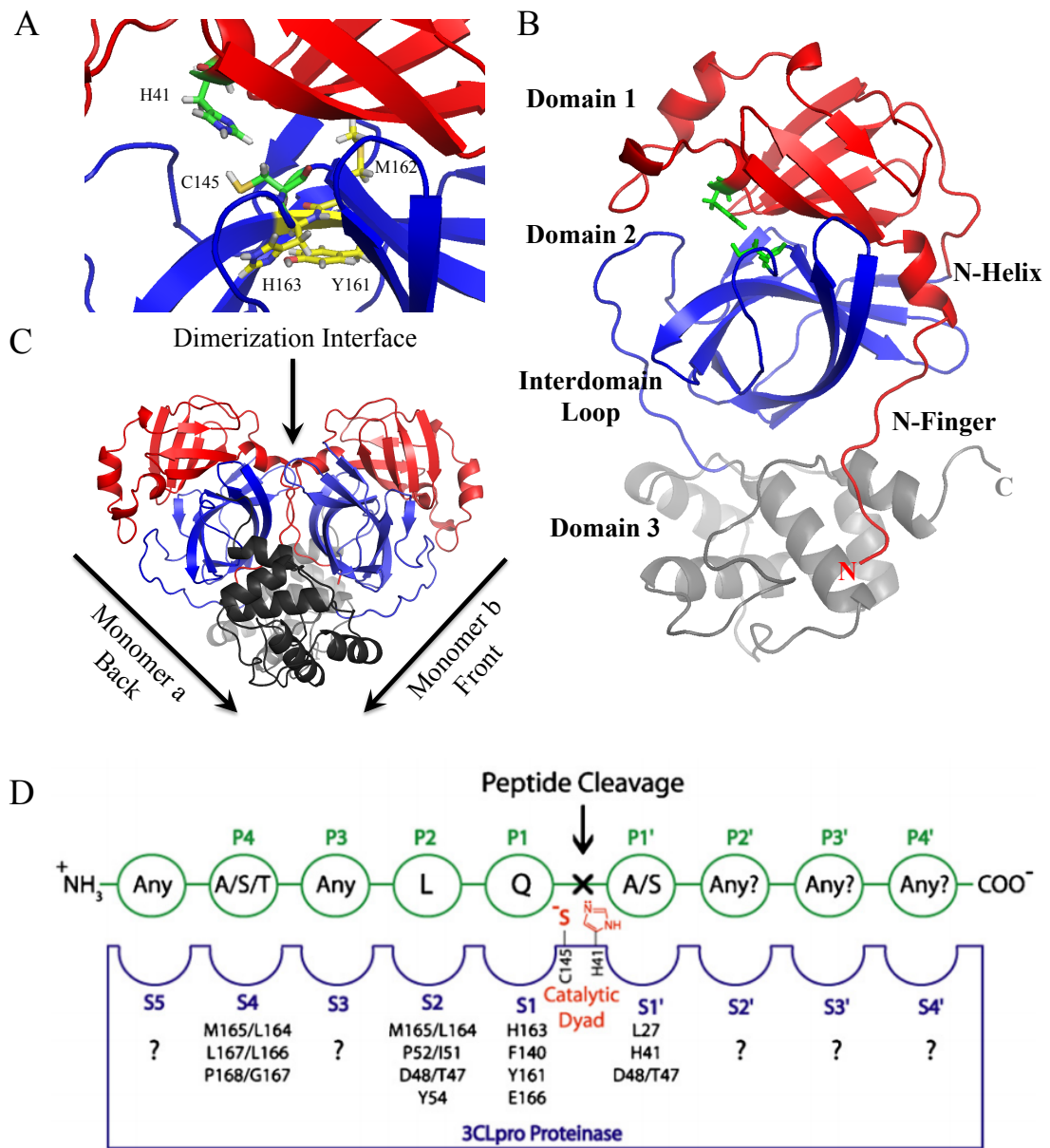
Figure 1.4 Nsp5 Structure and Functional Regions

A) The two catalytic residues, His41 and Cys145, are shown in green. The substrate binding pocket, a conserved Tyr – x – His motif (yellow) binds the P1 residue (Gln) of substrate. B) A model of MHV nsp5 using HKU1 nsp5 structure is shown. Domain 1 (red), domain 2 (blue), domain 3 (grey), and the catalytic dyad (green) are shown. C) Model of nsp5 in putative active dimer form. D) Canonical nsp5 substrate (green) binding to nsp5 protease (blue). Substrate binding pocket residues are indicated with "S" and substrate residues are indicated with "P." Nsp5 catalytic dyad is indicated in red, and cleavage occurs after a Glu in the P1 position. *(D Adapted form Grum-Tokars (2008) (21).)*

Of the 6 known HCoVs, the crystal structure has been solved for nsp5 from SARS-CoV, HCoV-NL63, HCoV-229E, HCoV-HKU1, and MERS-CoV (unpublished) (2, 9, 32, 73). The crystal structure of nsp5 has also been solved for the γ-CoV infectious bronchitis virus (IBV) (70), the α-CoV transmissible gastroenteritis coronavirus (TGEV) (71), and the β-CoV BtHKU4. Despite primary sequence differences, with percent identities as low as 21% for HCoV-NL63 nsp5, the tertiary structure of nsp5 is highly conserved (20, 37, 59). Domains 1 and 2 exhibit high tertiary structure conservation while D3 is much more variable (Figure 1.5). The variability of D3 may be explained by the uniqueness of D3 to CoVs or the fact that D3 does not contain catalytic or substrate binding residues. Our lab has previously demonstrated the limitations and fitness costs of substituting nsp5 proteases from other CoVs into the MHV genome (59). Due to the structural and functional conservation of the CoV nsp5 chymotrypsin-like fold, it is likely that this genetic limitation is associated with the unique domain 3.

Domain 1

Domain 2

Domain 3

■ (black)	MHV*
■ (blue)	SARS-CoV (50%)
■ (red)	HCoV-HKU1 (85%)
■ (orange)	HCoV-229E (45%)
■ (cyan)	HCoV-NL63 (21%)
■ (green)	IBV (41%)

* MHV nsp5 modeled on HCoV-HKU1
nsp5 (PDB 3D23) using Modeller
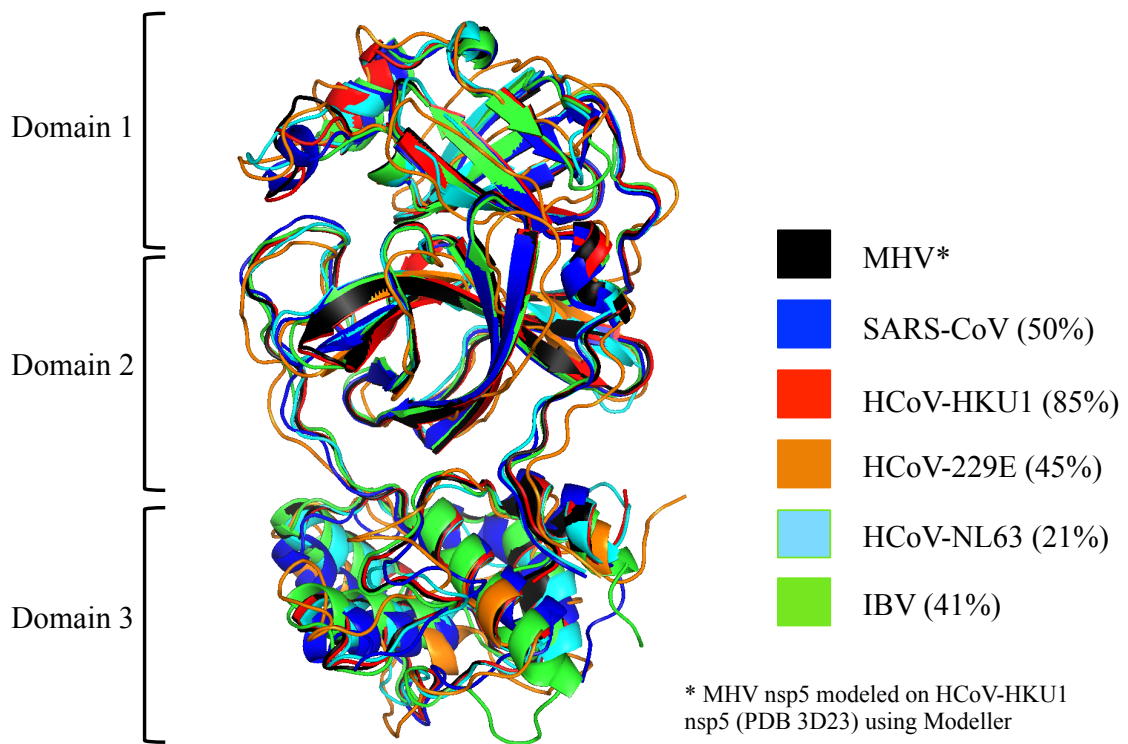
Figure 1.5 Nsp5 is Structurally Conserved

Structural alignment of nsp5 proteases from SARS-CoV (blue), HCoV-HKU1 (red), HCoV-229E (orange), HCoV-NL63 (cyan), and IBV (green). The structure of MHV (black) was modeled on the structure of HCoV-HKU1 using Modeller. The percent identities of CoV nsp5 proteases with respect to MHV are shown in parenthesis.

Proteolytic processing in coronaviruses

Nsp5 dimerization has been shown to be indispensable for trans- cleavage activity. Interestingly, Chen et. al. showed that Glu290 and Arg298, two residues known to contribute to the mature protease active dimer can be mutated to Arg and Glu, respectively. E290R and R298E mutants can both perform N-terminal autoprocessing, while the dimerization of mature protease and trans- cleavage activity were completely abolished (7).

Several other studies have demonstrated the requirement of dimerization for enzymatic activity. For example, Lin et. al. demonstrated that deletion of Gln299 or Arg298 decreased enzymatic activity to 98-99% compared to wild type (WT). Decreased enzymatic activity correlated with an increase in the monomeric form of the enzyme by sedimentation assay. Further, analysis of nearby hydrogen bonding partners of Gln299 and Arg298, Ser123 and Ser139, respectively, revealed a strong correlation between the enzyme activity loss and dimer dissociation (36).

Zhong et. al. further elucidated the dimerization mechanism by solving the structure of the monomeric and dimeric forms of the nsp5 C-terminal domain alone (nsp5-C). They determined that the nsp5-C dimer structure is characterized by 3D domain swapping, in which the first helices of the two protomers are interchanged and each is encircled by four other helices from the other protomer (75).

Numerous groups have reported models of nsp5 folding, dimerization and activity. As discussed above, nsp5 is active as a dimer, with only one monomer active within the dimer (7, 24, 34, 45). The immature form of nsp5 has proteolytic activity, with approximately 8-fold stimulation of activity by N-terminal cleavage, approximately 4-

fold stimulation by C-terminal cleavage, and 23-fold stimulation by the cleavage of both termini. Again, nsp5 autoprocessing is understood to occur in trans at the C-terminus (7, 24, 34, 45).

Chen et. al. proposed a model of nsp5 autoprocessing and dimerization, where two immature, folded nsp5 monomers approach one another and their D3s dimerize. This triggers movement of the N-termini into the substrate-binding pocket of the other monomer. Once N-terminal autoprocessing occurs, the monomers and N-termini assume their final conformation within the dimer. Next, the uncleaved C-terminus of one mature dimer can insert into an active site of another mature dimer. In the final step, the C-terminus is processed in trans, resulting in the final mature nsp5 dimer observed in crystal structures of nsp5 proteases (Figure 1.6) (7). Once folding, stabilization, and dimerization of nsp5 have taken place, the protease processes viral polyproteins within replication complexes (Figure 1.6). This model suggests that we should only be able to find nsp4/5/6 and nsp5/6 polyprotein intermediates, but data from our lab indicate that an additional nsp4/5 intermediate may also exist. Thus, the exact mechanism of nsp5 activity remains to be elucidated.

Despite structural and functional conservation among known CoV nsp5 proteases, there is a strong genetic limitation to the capacity to swap the nsp5 protease between CoVs (59). The existence of this genetic barrier emphasizes the need to further understand structural determinants of nsp5 function. Our studies could uncover a functional determinant that could be targeted for broadly reactive small molecule protease inhibitors.
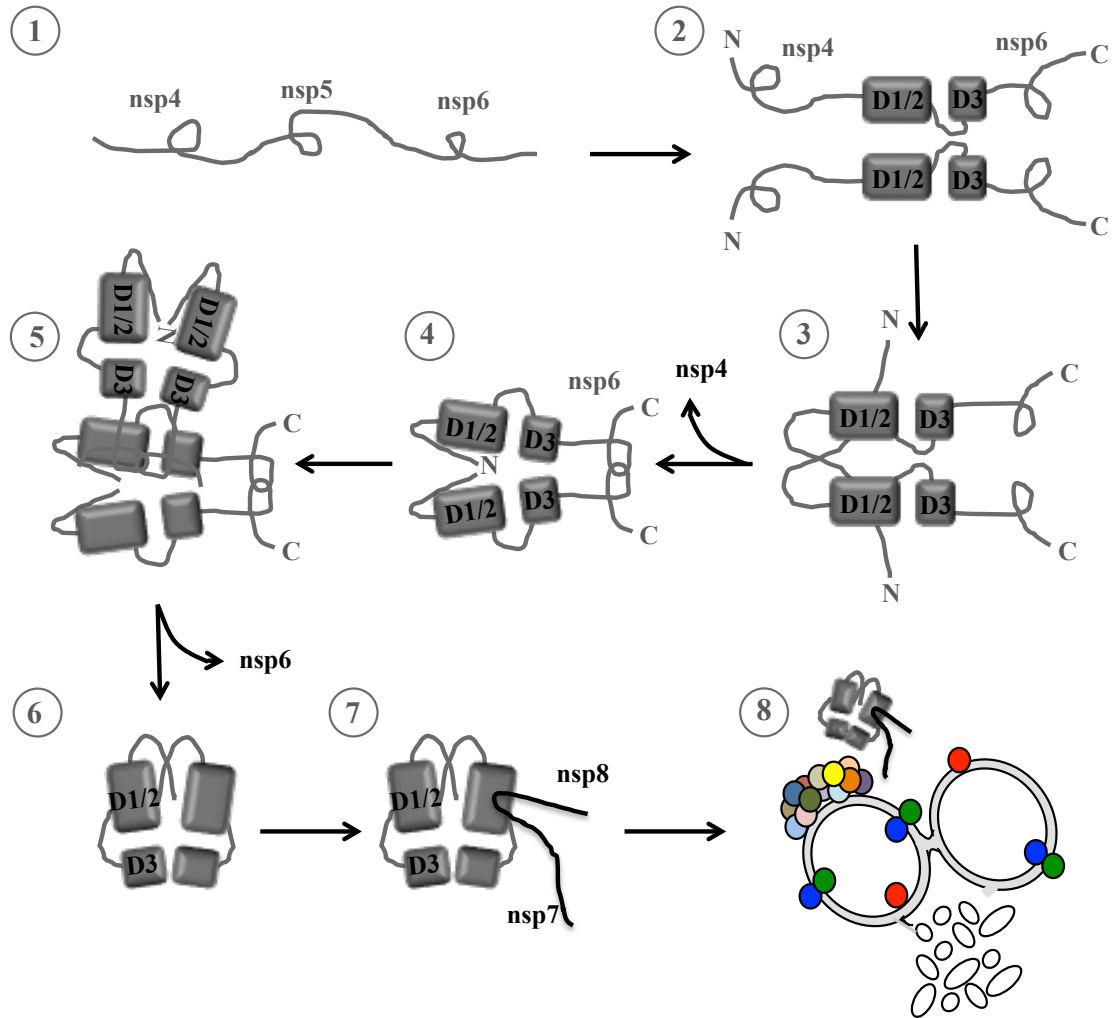
Figure 1.6 Model of Nsp5 Dimerization and Activity

Model of nsp5 autoprocessing, dimerization, activity during infection. 1. CoV genome is translated as a polyprotein. 2. Nsp5 domains fold and associate to form an immature dimer. 3. Immature nsp5 dimer cleaves the nsp4-5 cleavage site in cis, liberating nsp4. 4. N-terminal portion of monomer assume mature positions away from the active site. 5. The nsp5-6 cleavage site is processed in trans, liberating nsp6. 6. Nsp5 dimer is in the final mature, active form. 7. Nsp5 dimer recognizes polyprotein cleavage sites 8. Nsp5 activity occurs on ER membranes at replication complexes in order to liberate viral nsps for replication. *Adapted from Chen et. al (2010) (7).*

Active-site inhibitors of coronavirus nsp5 protease

The lack of efficacious therapeutics or vaccines for CoV infections underscores the need to continue to develop CoV therapeutics. Nsp5 is a promising target for viral inhibition since its activity is required for viral replication. Several groups have reported the identification of nsp5 protease active site inhibitors that prevent viral replication (18, 19, 25, 27, 29, 50). Due to the recent emergence of MERS-CoV, several groups have revisited this work, each suggesting that protease inhibition could lead to CoV therapeutics.

Kilianski et. al. reported an *in vitro* study on a MERS-CoV nsp5 inhibitor screening platform. The authors expressed MERS-CoV nsp5, and they used luciferase-based biosensors to measure protease activity. The authors report the ability to perform both endpoint evaluation and live-cell imaging to profile protease activity. The authors evaluated a chloropyridine ester small molecule inhibitor, CE-5, within their platform and calculated an effective concentration 50% (EC_{50}) of 12.5 μM against MERS-CoV nsp5 (27). Previously, this compound was reported to have an inhibitory concentration 50% (IC_{50}) against SARS-CoV nsp5 of 0.31±0.05 μM and an EC_{50} concentration of 24±0.9 μM against SARS-CoV (18). Further, Agnihothram et. al. reported this compound to be effective at inhibiting Bat HKU-5 CoV and MERS-CoV replication during viral infection of cells by 100-fold at 24 h p.i. at 50 μM (1). These studies indicate the feasibility of developing a class of small molecule inhibitors that is broadly reactive against CoVs. Minimally, this work suggests that broadly reactive inhibitors against each CoV genus (α-, β-, and γ-CoV) could be developed.

Summary

Coronaviruses cause both acute and chronic upper-respiratory disease and lower-respiratory tract illnesses. Both SARS-CoV and MERS-CoV are associated with severe febrile respiratory illness, pneumonia, and mortality. The mechanisms of HCoV emergence and pathogenesis are not understood, making it essential to continue to develop and evaluate broad-spectrum CoV replication inhibitors. Previous work has shown small molecule inhibitors to be effective at inhibiting nsp5. Thus, nsp5 is a proven target for viral inhibition since its activity is required for viral replication. In order to more effectively target nsp5 for viral inhibition, we must understand structural determinants of nsp5 protease function. Our work to understand nsp5 structure-function relationships in order to define novel regions within nsp5 for allosteric inhibition will be critical to developing viral therapeutics.

The CoV nsp5 protease has evolved a unique modification to the chymotrypsin fold. It contains a chymotrypsin fold connected to a novel α-helical domain (D3) by a 16 residue interdomain loop (IDL). Previous work has implicated D3 in substrate binding, *trans*-cleavage activity, and dimerization of protomers to form an active protease. Further, the IDL of equine arterivirus main protease (nsp4) contains putative hinge regions that are important for determining the timing and specificity of processing (65). Thus, we hypothesized that the novel third domain and IDL of CoV nsp5 protease have a unique, undefined function. To test this hypothesis, we probed interactions of residues spanning the Domain 2 (D2)/Domain 3 (D3) interface using our reverse genetics system and evaluation of polyprotein processing. We evaluated putative hinge regions within the IDL via alanine scanning mutagenesis and biochemical evaluation of mutants. This work

combined with future studies may shed light on structural determinants of nsp5 protease function that could be targeted allosterically for protease inhibition.

In addition to structure-function studies of nsp5, we tested and evaluated 9 small molecule nsp5 inhibitors of several classes for the ability to inhibit CoV replication. I evaluated covalent modifying inhibitors, non-covalent inhibitors, and peptide-like inhibitors of nsp5 that were previously biochemically validated. Further, I predict CoV nsp5 protease will mutate to escape inhibition by conserved intermolecular interactions. To test inhibitors and our hypothesis, I used firefly luciferase (FFL) reporter viruses to generate EC_{50} values based on change in reporter luminescence activity. I passaged WT virus in the presence of inhibitors to assess the potential for mutational escape. This work combined with future studies may shed light on CoV evolutionary mechanisms and phylogenetic relationships based on the similarities and differences observed in antiviral escape mutations between different nsp5 proteases. This work will also impact therapeutic design as detailed understanding of nsp5 function and the similarities and differences between various nsp5 proteases will be used to inform therapeutic design.

CHAPTER 2 :

TO DEFINE STRUCTURAL DETERMINANTS WITHIN NSP5 CRITICAL TO

PROTEASE FUNCTION AND DYNAMICS

Introduction

CoV nsp5 protease contains a novel α-helical domain (D3) connected to the

chymotrypsin fold, D1/2, through a 16 residue IDL. Previous work from Stobart et. al.

and others have implicated D3 and the IDL in the functionality of nsp5 (58). This work

identified a number of residues within nsp5, which confer temperature-sensitivity, that

are non-conserved among CoVs and can readily mutate to augment protease activity.

When passaged at non-permissive temperatures, temperature-sensitive (ts) mutants

S133A, V148A, and F219L were found to have developed secondary mutations that

restored nsp5 activity. These secondary mutations included T129M, H134Y, H270HH,

and E285V, in domains 2 and 3 (Figure 2.1) (58). No two of these residues are closer

than 10 Å in distance based on a homology model of the MHV nsp5 structure. It was

concluded that structurally distant, non-conserved residues are connected through tertiary

structure to provide nodes of allosteric communication within the protease (Figure 2.1).

Suel et. al. provide further support for networks of distant residues that confer allosteric

communication within proteins (60).

In equine arteritis virus (EAV), an arterivirus in the order *Nidovirales*, Aken et. al.

described an 8 residue loop connecting the chymotrypsin fold of the nsp4 main protease

to a novel C-terminal domain (CTD, D3). Structurally and functionally, the EAV nsp4

main protease is very similar to the CoV nsp5 main protease. The authors proposed that residues within this loop could be acting as a hinge region (HR) to facilitate the movement of the CTD. Crystal structures of EAV nsp4 have shown the CTD in one of four unique positions. It was concluded that residues within the HR control the spatial orientation of the CTD in order to regulate protease activity and the use of cofactors (65). The CoV nsp5 protease is structurally similar to EAV nsp4 with both a chymotrypsin-like fold and a unique C-terminal domain connected by a loop of amino acid residues. My work provides evidence that the CoV nsp5 IDL may contain determinants of protease function similar to the HR of EAV. Thus, I performed scanning mutagenesis on the CoV nsp5 IDL and domain 3 in order to identify these functional determinants.
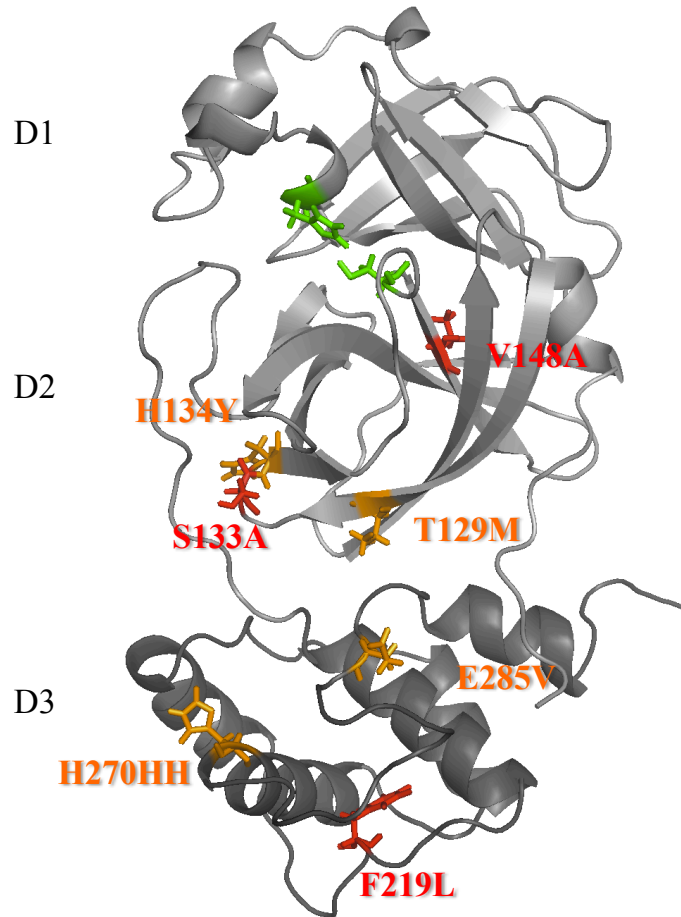
Figure 2.1 Distant Nsp5 Residues are Nodes For Allosteric Communication

MHV nsp5 structure modeled on HKU1 nsp5 is shown with D1/2 in light grey, D3 in dark grey, and active site residues in green. Temperature-sensitive alleles (red) and suppressor alleles (orange) identified by C. Stobart et. al. (2012) are indicated (58).

IDL Scanning Mutagenesis

To gain insight into structural determinants of nsp5 function, I performed scanning

mutagenesis on each of the 16 residues of the IDL. I mutated each residue to an alanine,

or I made more conservative mutations for less variable residues (Figure 2.2). For

example, Asp187 and Gln192 are both 100% conserved across all CoV nsp5 sequences.

Thus, I mutated those residues to glutamic acid and asparagine, respectively. I introduced

additions and truncations at the IDL C-terminus. I expected that viruses with mutations at

100%-conserved residues would not be recoverable. I also expected that mutations in

variable regions of the IDL, which alter amino acid interactions, would result in

debilitated or non-viable virus.

I ordered primers to introduce IDL mutations, additions, and truncations as listed

in Table 5.1. I performed Quick Change PCR to introduce each mutation within MHV C

fragment. Viruses were generated using the established reverse genetics system used by

the Denison lab. Briefly, I ligated the 7 MHV fragments using T4 ligase. I *in vitro*

transcribed ligated DNA and electroporated into baby hamster kidney-MHV receptor

cells (BHK-MHVR). I then overlaid the electroporated BHK-MHVR cells onto delayed

brain tumor cells (DBT) and monitored for signs of syncytia for a minimum of 3 days. If

no cytopathic effects (CPE) were observed after two attempts, I extracted total cellular

RNA in TRIzol reagent, performed reverse transcription-polymerase chain reaction (RT-

PCR) using primers for nsp5, and tested for the presence of virus. I then attempted to

recover virus a third time. I recovered P184A, R186A, A188I, V190I, V191I, L193A,

P194A, Q196A, and Y198A (Figure 2.2A). The nsp5 coding region was sequenced and

no additional mutations were identified. The D187E, T199A, +199, +197-199, or Δ199

viruses were not recovered. I did not attempt to recover Tyr185, Gln189, Val195, and Asp197.

Of the 9 recovered viruses, several contained residues that are between 2.2 Å and 2.7 Å from each other and could potentially interact via electrostatics to stabilize the IDL. This is interesting since electrostatic bonds, hydrogen bonds, and van der Waals forces are all potential bonding interactions that could occur at these distance ranges and such interactions are generally conserved to maintain protein structure. The residue Gln192 is 2.2 Å from Arg186, 2.4 Å from Ala188, 2.7 Å from Val190. Residue Pro184 is 2.6 Å from Val195. And Pro194 is 2.5 Å from Tyr185 (Figure 2.2B – D). Further, Val190, Ala188, and Asp187 all have interacting distances between 3.3 Å and 4.3 Å suggesting that these interactions could be conserved to maintain structure and function. However, the residues themselves are variable among CoV nsp5 proteases. Many of the variable IDL residues are somewhat functionally conserved as hydrophobic/hydrophilic and acidic/basic. Several residues including residue 186 and 188 readily convert between acidic and basic amino acids across all CoV nsp5 sequences.

Among the unrecoverable viruses were truncations and additions at the C-terminal end of nsp5, mutation of the C-terminal threonine, and mutation of 100% conserved residues (Asp187 and Gln192). This suggests that precise positioning of D3 by the IDL either contributes to protein folding, stability, or function. Further, residues Asp187 and Gln192 are putative "Achilles' heels" of the protease since they are /required for viral recovery. Additionally, Brett Case in the Denison lab recently discovered that DBT cells secrete low levels of interferon (personal communication). This may impact the recovery of extremely debilitated viruses since low levels of interferon may be enough to clear the

viral infection. It is possible that some unrecovered viruses may be recoverable if attempted in BHK-MHVR cells in the absence of DBT cells.
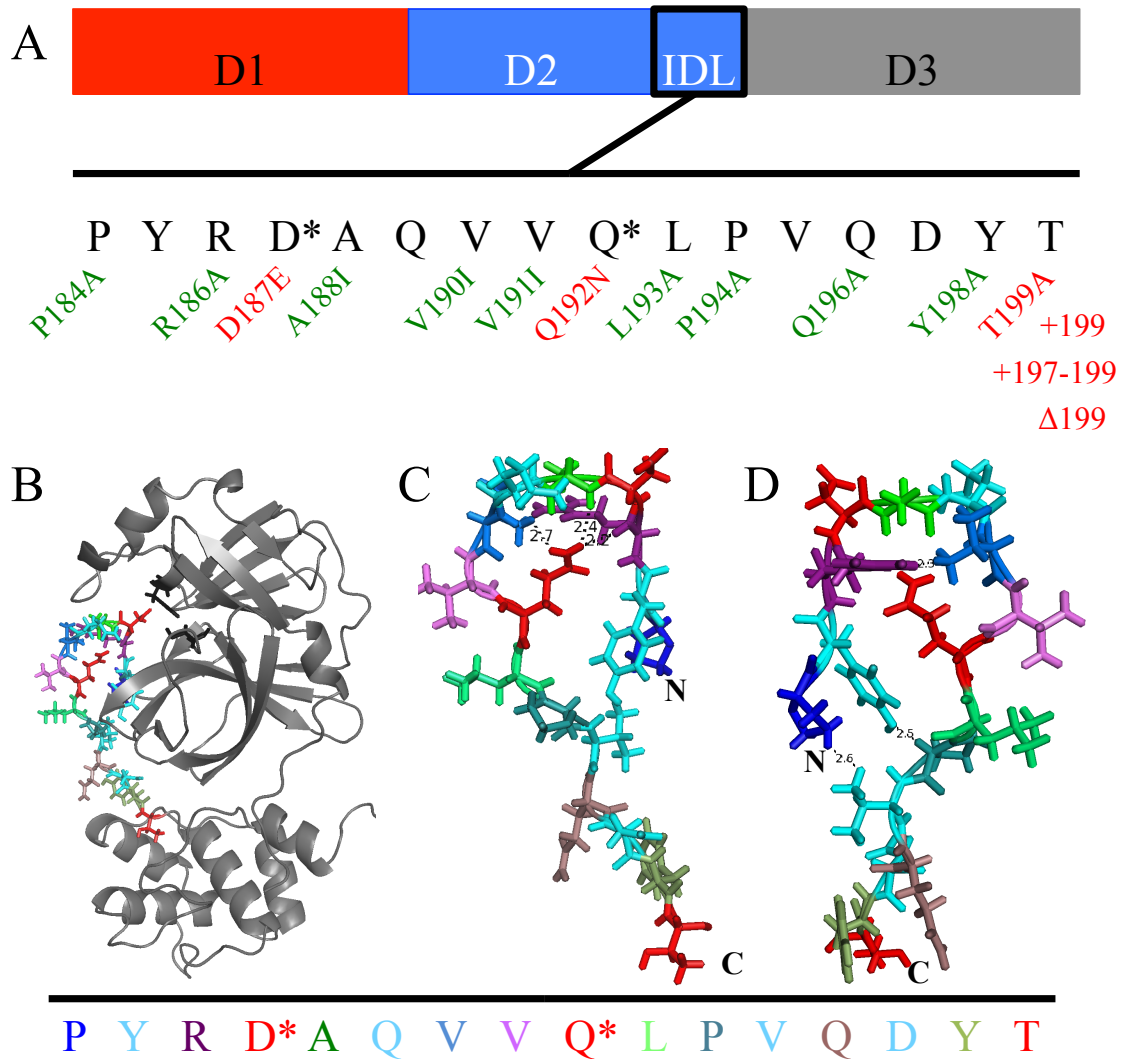
Figure 2.2 IDL Scanning Mutagenesis

Alanine scanning mutagenesis was performed on residues 184-199 of the interdomain loop as indicated using Quick Change protocol (Stratagene). A) Domain 1 is indicated in red, domain 2 including the IDL is indicated in blue, and domain 3 is indicated in grey. The *indicates 100% conserved residue. Green indicates recovered virus. Red indicates unrecoverable virus. B) IDL residues indicated in rainbow scheme. C and D) Putative interacting IDL residues with distance between amino acids indicated in Å.

To determine whether the substitutions alter virus replication, I infected DBT cells with WT-MHV as a control, P184A, V190I, V191I, Q196A, R186A, Y198A, and A188I at a multiplicity of infection (MOI) of 1 plaque-forming unit per cell (PFU/cell). Interestingly, all of the viruses displayed similar replication kinetics to WT-MHV (Figure 2.3). WT-MHV replicated exponentially between 6 and 8 h p.i. and achieved peak titer at 12 h p.i. Each mutant virus tested replicated indistinguishably from WT-MHV. This data indicates that any putative stabilizing interactions between residues across the IDL, as indicated in Figure 2.2, are not required for proper protease function. Further, the capacity of each mutant virus to replicate with kinetics similar to that of WT-MHV suggests that each protease is properly folded, selects substrates appropriately, and catalyzes peptide bond cleavages at a WT-like rate. However, further studies need to be performed to confirm protein folding and activity.
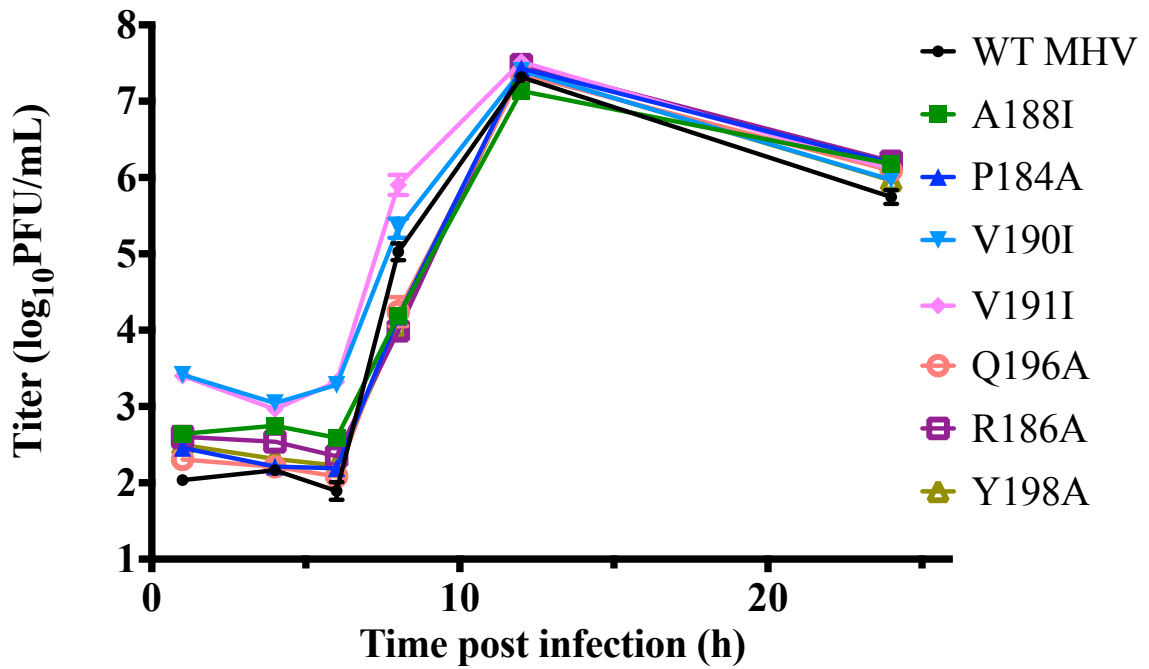
Figure 2.3 Replication Kinetics of IDL Mutant Viruses

DBT cells were infected with indicated viruses at MOI=1 PFU/cell. Aliquots of supernatant were collected at indicated times post infection, and titers were determined by plaque assay. Error bars represent the standard error of the mean of three independent replicates performed in duplicate.

Mutagenesis of Domain 2-Domain 3 Interface Residues

In addition to evaluating residues within the IDL, I wanted to evaluate residues spanning the D2/D3 interface for interactions that modulate activity. As indicated in Figure 2.4A, Arg131 (R131) of D2 and Asp286 (D286) of D3 are 3.5 Å apart while Gly109 (G109) of D2 and Asn203 (N203) of D3 are 2.2 Å apart. These distances are short enough that electrostatic interactions, hydrogen bonding, or van der Waals forces could occur between these residues. Additionally, N203, G109, R131, and D286 are all 100% conserved residues across CoV nsp5 proteases. To assess these putative interactions, I introduced the mutations indicated in Table 2.1. I was unable to recover any of the attempted mutants except for N203C. I recovered N203C with T111A as putative compensatory mutation. I then attempted to plaque purify the virus, but 2 separate plaque purifications resulted in additional, putative compensatory mutations. Plaque 1 contained N203C, T111A, and T129A, while plaque 2 contained N203C and T129A. I then passaged the virus 4 times in DBT cells. Both passage 3 (P3) and passage 4 (P4) contained N203C with T111A (Figure 2.4C).

These results suggest that N203 interacts with either or both T129 and T111 across the D2/D3 interface. Further, these results support work from Stobart et. al. who initially reported allosteric communication networks occurring within nsp5. Interestingly, these residues are within short enough distances from one another to interact directly as opposed to long-distance allosteric interaction. T129M was reported by Stobart et. al. as a compensatory mutation that arose when ts mutant viruses were passaged at non-permissive temperatures (Figure 2.1) (58).
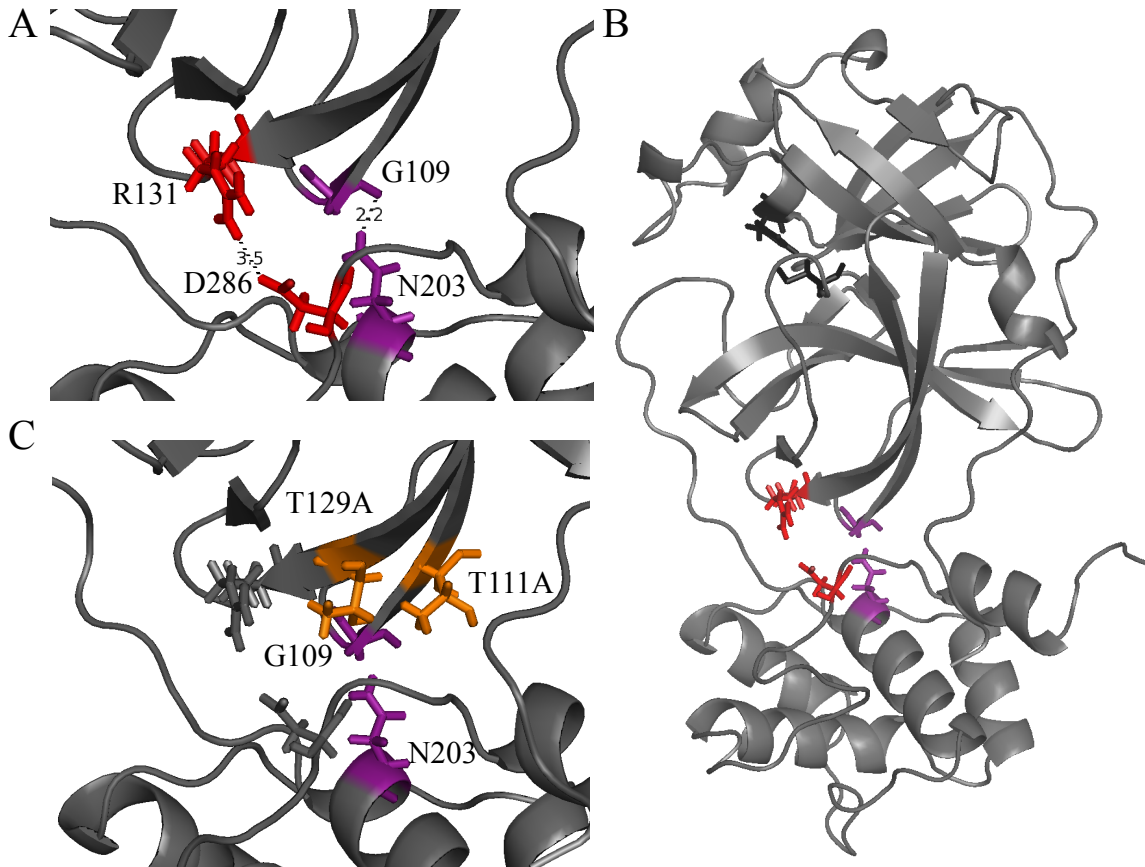
Figure 2.4 Domain 2- Domain 3 Interface Residues

A and B) Putative interacting residue pairs spanning the domain 2 – domain 3 interface are indicated in red and purple. Catalytic residues are indicated in black. C) Compensatory mutations isolated in combination with N203C mutant virus are indicated in orange (T129A and T111A). N203 and G109 are indicated in purple.

Table 2.1 Domain 2 – Domain 3 Interface Residue Mutant Virus Recovery

Mutation	Status
G109L	Not Recovered (1x)
G109C	Not Attempted
R131A	Not Recovered (1x)
R131K	Not Recovered (1x)
R131C	Not Recovered (2x)
N203A	Not Recovered (1x)
N203C	Recovered with T129A and/or T111A
D286A	Not Recovered (1x)
D286C	Not Recovered (1x)

In order to test whether the substitutions and compensatory mutations alter virus replication, DBT cells were infected with WT-MHV as a control, N203C P0 (N203C/T111A), N203C P4 (N203C/T111A), and N203C Plaque 2 (N203C/T129A) at MOI = 1 PFU/cell. Each mutant virus replicated with kinetics indistinguishable from WT-MHV (Figure 2.5). Previous work has shown T129 to be a residue capable of mutating under specific pressures in order to confer normal protease activity. Therefore, the WT-like replication kinetics of each N203C mutant virus suggests that T129A and T111A act as compensatory mutations.
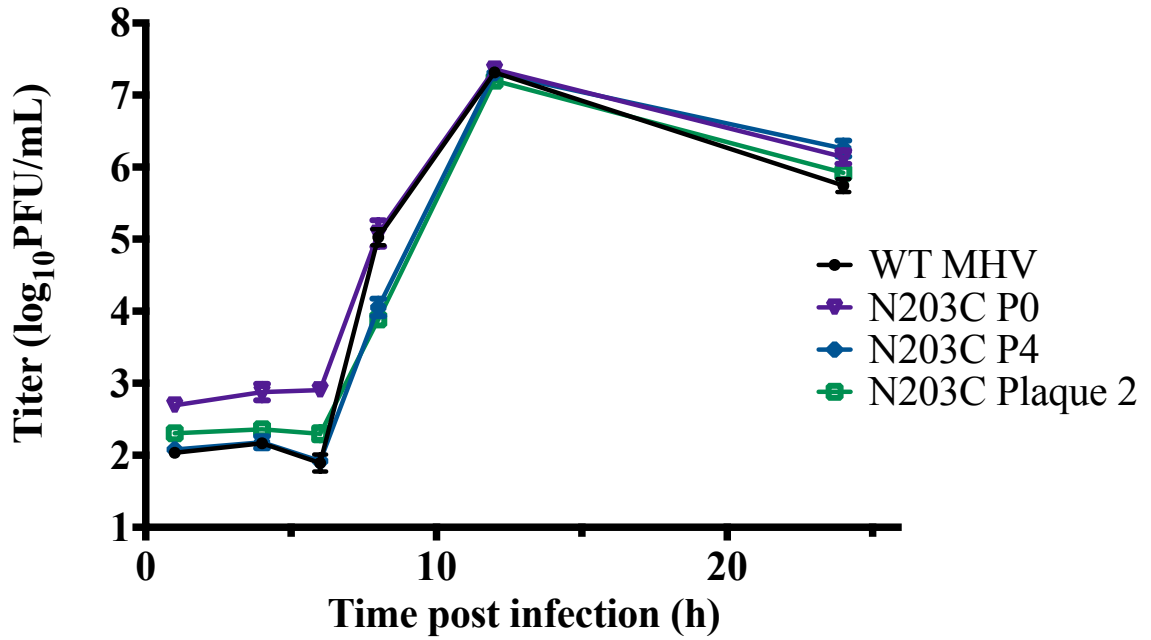
Figure 2.5 Replication Kinetics of N203 Mutant Viruses

DBT cells were infected with indicated viruses at MOI=1 PFU/cell. Aliquots of supernatant were collected at indicated times post infection, and titers were determined by plaque assay. Error bars represent the standard error of the mean of three independent replicates performed in duplicate.

Discussion and Future Directions

Several mutations within the IDL resulted in genomes that could not be recovered as viruses, suggesting that these residues may be particularly important in protease structure, function, or folding. Interestingly, a number of the residues within the IDL were readily mutated resulting in no replication defects including Pro184 and Arg186. These residues were of particular importance since I thought they might contribute to IDL stabilization through inter-residue interactions. Further, the N203C substitution in D3 was only recovered in the presence of additional, putative compensatory mutations that arose in D2. This suggests that N203 may be required for protease function. This work also suggests that residues within the IDL and residues that span the D2/D3 interface communicate either directly or over long distances with residues of other domains and contribute to protease function.

The recovered IDL mutant viruses should be evaluated for altered polyprotein processing by immunoprecipitation. If the IDL does in fact act as a hinge region to mediate selectivity and order of polyprotein processing, it is possible that different polyprotein intermediates could be identified. It would be necessary to compare polyprotein intermediates between mutant viruses and WT-MHV at multiple time points using antibodies to each MHV non-structural protein. For example, Aken et. al. identified altered concentrations of several non-structural proteins and partially processed polyprotein intermediates in their study of EAV nsp4 main protease (65). I would expect to see a similar result with our IDL mutant viruses if any of the engineered mutations alter function and substrate selection. A pulse-chase experiment with chase times between 15 minutes and 300 minutes followed by immunoprecipitation could be useful to

assess protein degradation versus diminished production. It would also be useful to assess any changes in timing of various polyprotein cleavages by monitoring polyprotein intermediates during pulse-chase. Further, competition assays between the mutant viruses and WT-MHV could be performed. If I saw strong differences in viral fitness, I would conclude that intervening, variable IDL residues do modulate protease function. If I saw similar fitness, I would conclude that the virus has evolved to allow large amounts of mutation in the IDL because maintenance of the few conserved residues and interactions are absolutely critical to protease function.

Data obtained from experiments in cultured cells including replication kinetics, polyprotein processing, and pulse-chase radiolabeled translation will allow us to determine the kinetics of viral infection. However, it is important to determine the difference between binding effects vs. unfolding effects in mutant viruses. It will be essential to perform isothermal titration calorimetry (ITC) or fluorometry to determine if peptide substrates are still able to bind the mutant nsp5. ITC is commonly used in biochemistry laboratories to assess both binding affinity and thermodynamic parameters. Fluorometry could be used to evaluate binding using fluorescence resonance energy transfer (FRET), labeled peptide substrates available from Andrew Mesecar's lab (Purdue University), or intrinsic protein fluorescence. Nuclear magnetic resonance (NMR) or a crystal structure would be needed for strong experimental evidence of peptide binding. In terms of folding, intrinsic protein fluorescence or circular dichroism (CD) can be used to determine if mutant nsp5 proteins are folding properly or unfolding. CD could also be used to determine ts effects of mutations if performed at a gradient of temperatures. This work could be performed in collaboration with the Mesecar lab.

Peptide-protein docking could be used to computationally identify or confirm residues mediating important interactions during polyprotein processing. This method involves docking short peptides into the modeled MHV nsp5 structure. The online server FlexPepDock (http://flexpepdock.furmanlab.cs.huji.ac.il/index.php) is an available resource to begin this sort of study, however, the online server is limited in resources/options. We could model each mutation in the IDL or spanning the D2/D3 interface to identify structural perturbations once a peptide is docked in the active site. Biochemical and cellular data will improve these models. This work could be performed in collaboration with Jens Meiler's lab at Vanderbilt University.

In order to determine the how important IDL flexibility is for function, we could engineer several mutations into the IDL simultaneously. We would choose a group of mutations that made the IDL as rigid as possible as well as a group of residue that made the IDL as flexible as possible. We could also try to kink the IDL more by adding in Pro mutations at a number of residues. Finally, we could try to swap the IDL from a different CoV, such as MERS-CoV, into MHV nsp5 and attempt virus recovery. For each of these mutations and swaps within nsp5, we could perform computational, biochemical, and cell culture studies described above.

I have identified residue N203 as an important mediator of nsp5 function since it is 100% conserved and the N203C substitution can only be recovered in the presence of putative compensatory mutations. It would be useful to model N203C, T129A, and T111A into the HKU1 nsp5 structure using Rosetta or Modeller to identify possible changes in interactions. We could reverse engineer each compensatory mutation into WT-MHV and assess replication kinetics, polyproteins processing, and protein

degradation. We could also attempt to recover G109, R131, and D286 mutants by reverse engineering putative compensatory mutations identified through Rosetta/Modeller modeling. Results from these studies and previous studies may suggest that conserved interactions across the D2-D3 interface in addition to interactions mediated by D187, Q193, and T199 maintain nsp5 structure-function.

In addition to activity studies, we also would like to understand the dynamic movement of domains and the IDL during protease function. The Electron Paramagnetic Resonance (EPR) method of site-directed spin labeling (SDSL) with double electron-electron resonance (DEER) spectroscopy is an established biophysical technique that bridges the gap between protein structure and function. Spin labeling in conjunction with EPR spectroscopy can give insight into protein dynamics. EPR relies on SDSL where surface exposed residues along a protein are mutated to a cysteine and then complexed to a spin label (43). The most commonly used spin label is 1-oxyl-2,2,5,5-tetramethylpyrroline-3-methyl-methanethiosulfonate (MTSSL). Quantitative measurements of spin label solvent accessibility and inter-label distances can be made with EPR. Deconvolution of EPR spectra allows determination of protein structure, topology, and packing. DEER is a relatively new, yet established, technique that allows long-range distance measurements (60Å- 80Å) between pairs of spin labels (16, 17, 26). It allows quantitative analysis of equilibrium dynamics and conformational changes. For example, DEER can be used to assess movement of secondary structures, flexible loops, and side chain movement when triggered. The readout from DEER spectroscopy is changes in a distance profile, P(r), which is the probability of distances (r_{av}) between two spin labels (43).

Figure 2.6A (bottom left) shows a scenario where a membrane protein converts from state A to state B. The distance distribution, P(r), clearly changes modes. In Figure 2.6A (bottom right), the membrane protein is in equilibrium between states A and B where condition 1 may indicate the lack of a substrate/cofactor and condition 2 may indicate the presence of a substrate/cofactor that has triggered a shift in the conformational equilibrium. The DEER technique has been successfully employed by Gail Fanucci's lab (University of Florida) in the study of Human immunodeficiency virus type 1 (HIV-1) protease dynamics (26). DEER could be used in the study of CoV nsp5 to understand functional dynamics of the IDL and D3. To perform this study, each solvent exposed cysteine residue on the protease would need to be mutated to alanine without altering protease folding or function. The cysteine residues of MHV nsp5 are indicated in cyan in Figure 2.6B. Once mutated, any remaining or added solvent exposed cysteine residues in the protease would be labeled by SDSL and EPR spectroscopy would be performed. The addition of nsp5 inhibitors or universal substrates would be added to trigger a conformational change in the protein that can be monitored by DEER. This work could be performed in collaboration with Hassane Mchaourab's lab at Vanderbilt University.

Taken together, these studies will allow us to answer questions regarding the role of the CoV nsp5 IDL and D3 during protease function. IDL mutagenesis and polyprotein processing in addition to biochemical and computational studies will help us to identify residues that may act as determinants of protease function. This data can be used to inform which residues to spin label and then probe via DEER experiments. Finally,

DEER spectroscopy will provide us with the necessary link between structural/computational data and functional/cell culture data.
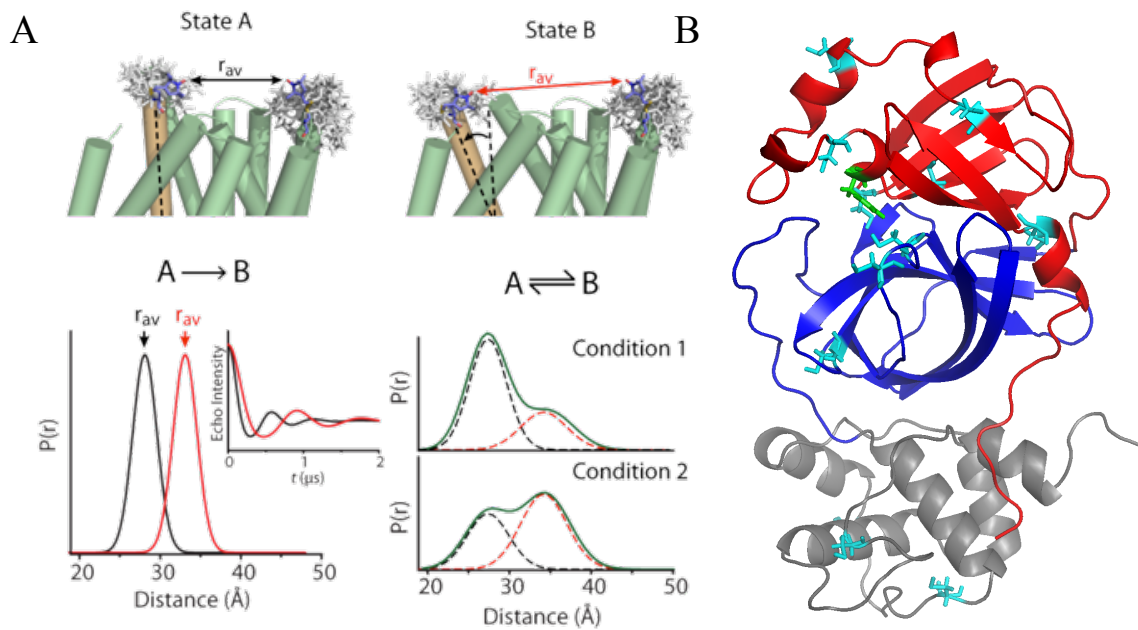
Figure 2.6 Double Electron-Electron Resonance (DEER) Approach

A) DEER readout gives changes in distance profile with or without substrate or inhibitor. $P(r)$ = probability of distance between two spin labels (distance distribution) from r_{av}. Graphs indicate model is sampling either state A or is in equilibrium between sates A and B. *Courtesy of H. Mchaourab (43)*. B) Cysteine residues of nsp5 that may need to be mutated for DEER experiments are indicated in cyan, active site residues are indicated in green.

PLATFORM TO ASSESS EFFICACY OF NSP5 INHIBITORS AND PATHWAYS OF

ANTIVIRAL RESISTANCE

Introduction

Coronaviruses (CoVs) cause a number of human diseases, most notably Severe Acute

Respiratory Syndrome and the more recent Middle East Respiratory Syndrome. These

zoonotic pathogens are of importance since they exhibit pandemic potential and high

mortality rates. The mechanisms of CoV emergence are not understood, making it

essential to develop broadly effective therapeutics to treat future emergent human CoVs.

CoVs have positive-sense RNA genomes that are translated into polyproteins that are

processed into mature proteins by encoded proteases including the main protease (Mpro,

nsp5). Nsp5 is required for CoV replication, and it is responsible for 11 polyprotein

cleavage events. Thus, it is a key target for the design of both active-site inhibitors and

allosteric inhibitors. Mechanisms of protease function and dynamics identified by IDL

and D2/3 interface analysis will guide inhibitor design in addition to concepts such as cell

penetration and activity of side-products. .

Key obstacles to developing nsp5 inhibitors are (1) the inability to culture and test

most CoVs and (2) antiviral resistance. Therefore, I have generated a novel

biotechnological platform for rapid testing of biochemically-validated inhibitors and

screening for new inhibitors. I generated fusions of firefly luciferase protein (FFL) to

nsp2 for each of two previously reported MHV chimeric viruses. These chimeric viruses

contain nsp5 from HCoV-HKU1 (H5-MHV) or HCoV-OC43 (O5-MHV) and have been previously shown to have replication kinetics similar to WT-MHV (59). Our lab has previously demonstrated that FFL-nsp2 fusions do not alter viral replication kinetics, and we have shown that luciferase activity can be used to quantitate protein translation and genome replication (15).

Interestingly, we have been unable to recover chimeric viruses containing nsp5 protease substitutions from viruses outside of subgroup 2a β-CoV (Figure 1.1). Stobart et. al. reported that substitutions into the MHV background of subgroup 2a β-CoV nsp5 proteases containing up to 84% sequence identity (Figure 1.5) to MHV nsp5 resulted in fitness costs (59). Our lab was unable to recover chimeric viruses containing nsp5 proteases from the following CoVs in the MHV background: HCoV-229E, HCoV-NL63, BtCoV-HKU4, and SARS-CoV. In addition, I have been unable to recover MERS-CoV nsp5 protease in an MHV or SARS-CoV background. I propose that the inability to recover chimeric viruses outside of particular subgroups is due to the divergence of function and regulation of protease activity. Stobart et. al. suggested that conservation of intermolecular residue interactions may be most closely associated with members of the same subgroup (59). Thus, it may be most feasible to generate nsp5 inhibitors that are broadly effective across specific CoV genera. For example, the covalent modifying inhibitor, CE-5 (GRL-001-13S) has been shown by several labs to be effective across the β-CoV genus.

The Mesecar lab at Purdue University has developed a series of chloro-pyridyl ester nsp5 active site inhibitors, among which CE-5 (GRL-001-13S) has shown particular promise for broad reactivity. Kilianski et. al. reported an EC_{50} of 12.5 μM against MERS-

CoV nsp5 (27). Previously, this compound was reported to have an IC_{50} against SARS-CoV nsp5 of 0.31 ± 0.05 μM and an EC_{50} concentration of 24 ± 0.9 μM against SARS-CoV (18). Further, Agnihothram et. al. reported this compound to be effective at inhibiting BtCoV-HKU-5 and MERS-CoV 100-fold at 50 μM (1). Further, our studies using chimeric reporter viruses containing nsp5 proteases from HCoV-HKU1 and HCoV-OC43 show EC_{50} values ranging between 1.5 μM – 5 μM. Thus, this inhibitor, CE-5, is broadly effective against the β-CoV genus.

I generated and used chimeric reporter viruses to determine EC_{50} values for nsp5 inhibitors. Preliminary data suggests that inhibitors are specifically inhibiting the nsp5 protease since our assay shows differential sensitivity between chimeric viruses to several tested inhibitors including GRL-002-13S. The only difference between chimeric viruses is the protease itself, so differences in log reductions should reflect protease inhibition. I also passaged viruses in the presence of inhibitors in an effort to identify pathways of antiviral resistance. I anticipate that this system will prove useful for optimizing inhibitor design, resulting in enhanced therapeutic potential.

Experimental Design

In order to test compounds, I needed a platform comprised of i) a reporter molecule and ii) multiple nsp5 proteases. Both were separately available resources in the lab, so I combined them to generate chimeric nsp5 reporter viruses. I generated N-terminal FFL fusions to nsp2 using previously established cloning techniques (15) and our reverse genetics system for MHV (72). Our lab has previously demonstrated that luciferase activity from MHV containing FFL fused to the N-terminal portion of nsp2 (MHV-FFL) can be directly used to quantitate virus replication. Further, virus replication directly depends on nsp5 protease activity. Thus, I used luciferase activity to measure nsp5 inhibition. I fused FFL to the N-terminal portion of nsp2 within H5-MHV and O5-MHV resulting in H5-MHV-FFL and O5-MHV-FFL (Figure 3.1). Each chimeric virus contains normal MHV nsp5 cleavage sites excluding the nsp4-5 and nsp5-6 cleavage sites. The nsp4-5 cleavage site was designed to contain MHV P sites and HKU1/OC43 P' sites, while the nsp5-6 cleavage sites contain HKU1/OC43 P sites and MHV P' sites. Both viruses were readily recovered in DBT cells at $37^{o}C$ with similar CPE to that of WT-MHV. Once recovered, I used TRIzol (Invitrogen) to extract total cellular RNA. I used RT-PCR to generate amplicons that span FFL and nsp5 coding regions. These amplicons were Sanger sequenced and confirmed to contain intact regions with no additional mutations.
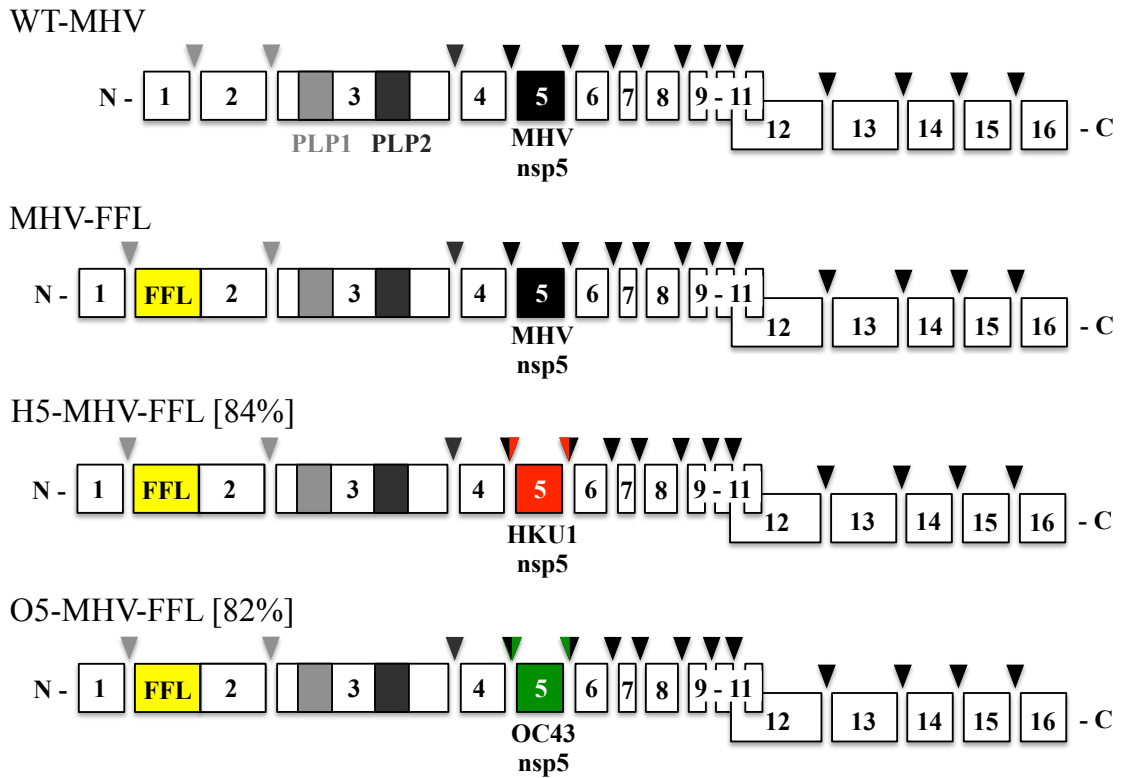
Figure 3.1 Engineered Nsp5 Chimeric Reporter Viruses

Nsps 1-16 are indicated. Protease cleavage sites are indicated by arrowheads above the polyprotein. FFL reporter was engineered as a fusion with nsp2 (yellow). Chimeric nsp5 proteases are indicated in red (HCoV-HKu1) and green (HCoV-OC43).

In order to assess the chimeric viruses for utility in testing nsp5 inhibitors, I first needed to test viral replication and luciferase activity. DBT cells were infected with indicated viruses (Figure 3.2A) at MOI = 1 PFU/cell. At each time point indicated, supernatants were collected and titered by plaque assay. Once supernatants were sampled, cells were harvested and tested for luciferase activity. WT-MHV began exponential replication between 6 – 8 h p.i. and reach peak titer at 12 h p.i. (Figure 3.2A). Each virus replicated indistinguishable from WT-MHV, indicating that both chimeric nsp5 proteases and FFL-nsp2 fusions do not alter the timing or efficiency of viral replication.

To assess luciferase activity, I sampled infected cells at multiple time-points p.i. at the same time I measured viral replication kinetics. WT-MHV lacking an FFL protein was incorporated as a negative control. Briefly, cells infected at MOI = 1 PFU/cell were washed at desired time points with phosphate buffered saline (PBS), and the appropriate amount of reporter lysis buffer (Promega) was added. Cells were frozen at -80°C until the luciferase assay was performed. To measure RLU, cells were warmed to ambient temperature and transferred into a 96-well plate. The luciferase assay reagent (Promega) was suspended according to directions, and the reagent was added and activity read by a Veritas luminometer. MHV-FFL showed 5 logs of luciferase activity measured in relative light units (RLU), and it reached peak luciferase activity at 12 h p.i (Figure 3.2B). Both H5-MHV-FFL and O5-MHV-FFL displayed luciferase activity indistinguishable from MHV-FFL. Additionally, luciferase activity directly correlated with replication kinetics with peak luciferase activity and peak viral titer both occurring at 12 h p.i.

In order to verify the FFL-nsp2 fusion, I performed an immunoprecipitation (IP) experiment on cells infected with each chimeric reporter virus. I infected DBT cells at MOI = 10 PFU/cell, radiolabeled with [^{35}S]-Met/-Cys media, harvested cell lysates, then IP with rabbit polyclonal α-MHV nsp2 antibody. As indicated in Figure 3.2C, WT-MHV, H5-MHV, and O5-MHV each show a band at 65-kDa, indicative of native nsp2, while MHV-FFL, H5-MHV-FFL, and O5-MHV-FFL lacked the 65-kDa band and had an additional band at 100-kDa. This band is consistent in size with the FFL-nsp2 fusion. Together, the data in Figure 3.2 indicates that our chimeric reporter viruses were successfully engineered and recovered with replication kinetics and luciferase activity indistinguishable from WT-MHV (or MHV-FFL).
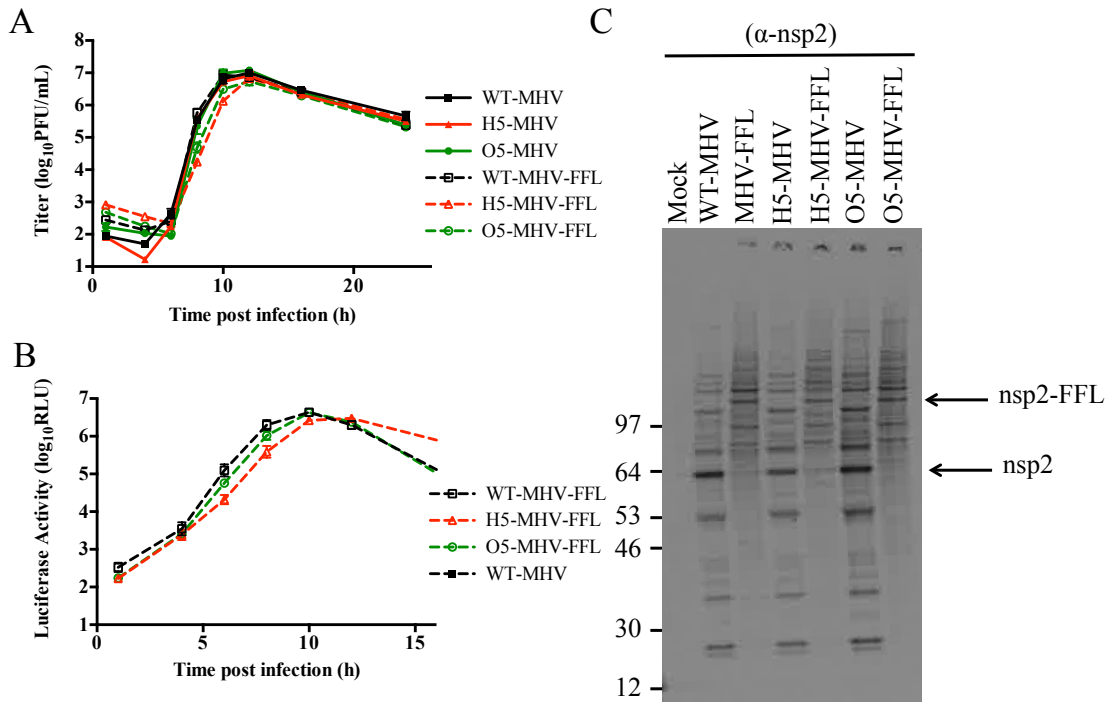
Figure 3.2 Characterization of Chimeric Reporter Viruses

A) Chimeric reporter viruses have same replication kinetics as WT-MHV. DBT cells were infected with indicated viruses at MOI=1 PFU/cell. Aliquots of virus were taken at indicated times p.i. and titers were determined by plaque assay. B) FFL activity of chimeric reporter viruses correlates with replication kinetics. DBT cells were infected with indicated viruses at MOI=1 PFU/cell. Aliquots of virus were taken at indicated times p.i. and RLU were determined using Luciferase Assay (Promega). FFL relative light units (RLU) are indicated. Error bars represent the standard error of the mean of three independent replicates performed in duplicate. C) Chimeric reporter viruses contain nsp2-FFL fusion. DBT cells were infected with indicated viruses at MOI = 10 PFU/cell, and radiolabeled with [^{35}S]-Met/-Cys media. Cell lysates were IP with rabbit polyclonal α-MHV nsp2 antibody, run on an SDS-PAGE gel, and imaged by autoradiography.

Next, I sought to validate our chimeric reporter viruses as a platform to test nsp5 inhibitors. In collaboration with the Mesecar lab, I obtained a series of biochemically validated nsp5 active-site inhibitors. These inhibitors, listed in Table 3.1, include several covalently modifying chloropyridyl ester-derived compounds identified by Ghosh *et. al.* (18), noncovalent acetamide-derived inhibitors (25), and peptide-like inhibitors. I assessed the cellular toxicity of each of these compounds up to 100 μM in rat lung epithelium (L2) cells using a CellTiter-Glo assay (Promega) in a Veritas luminometer. At 100 μM, no inhibitor reduced cell viability by more than 15% (Figure 3.3). I then tested each inhibitor for reduction in RLU, which is indicative of reduction in viral replication (15). Briefly, I incubated cells infected with MHV-FFL, H5-FFL, or O5-FFL with biochemically validated nsp5 active-site inhibitors at concentration ranges from 0 μM - 100 μM. At 10 h p.i., I harvested infected cells, and tested for luciferase activity. GRL-001-13S (also CE-5) reduced RLU and therefore viral replication at low μM concentrations for each viral chimera tested (Figure 3.3). However GRL-002-13S was much less effective at reducing viral replication for each virus. This compound has differential effects on each viral chimera tested, with EC_{50} values ranging from $6 - 200$ μM, demonstrating that the inhibitors specifically act on nsp5 and not the MHV backbone. Estimates of EC_{50} values for all compounds are listed in Table 3.1. Of the compounds tested, GRL-001-13S and RWB exhibited the lowest EC_{50} values for each viral chimera (Table 3.1). This data demonstrates that these viruses can be used to rapidly screen for nsp5 inhibitors against OC43, HKU1, and MHV. This is useful since HKU1 and OC43 are otherwise extremely difficult to cultivate and therefore test for protease inhibition.
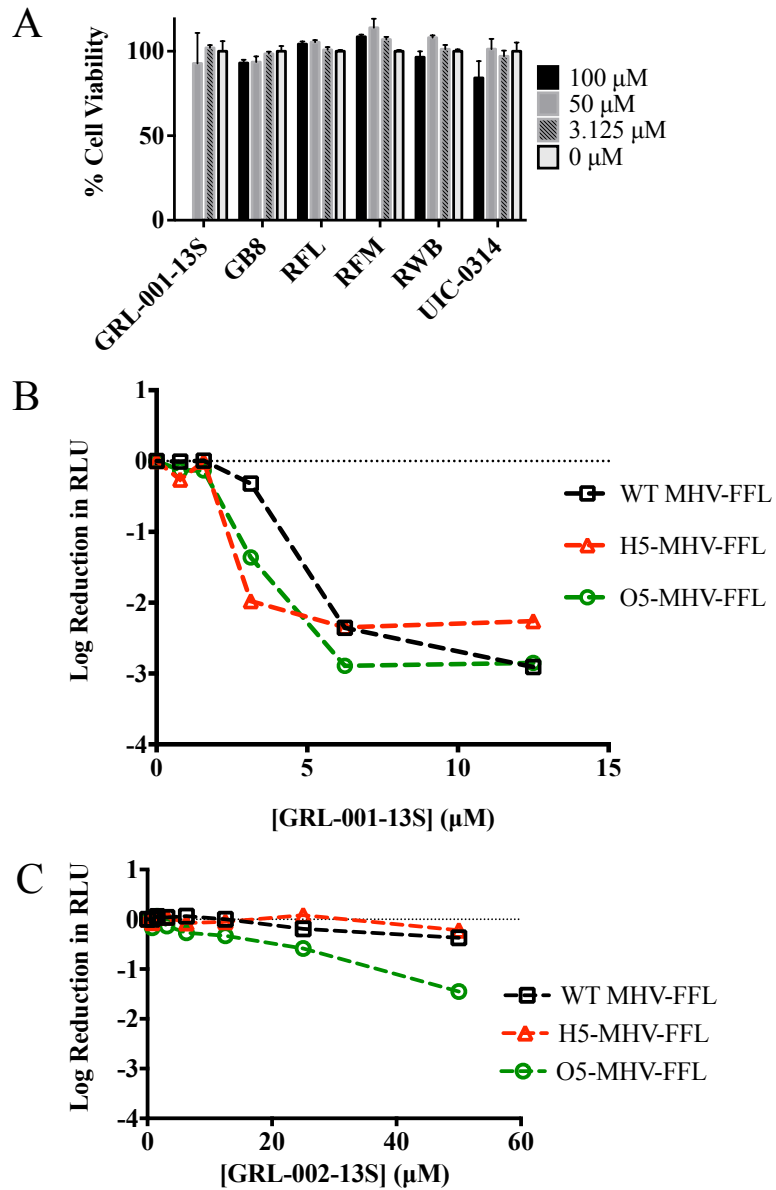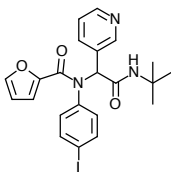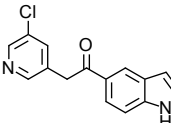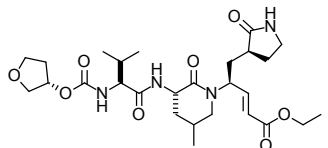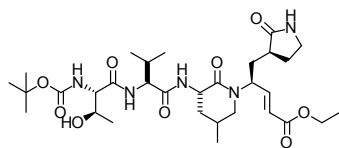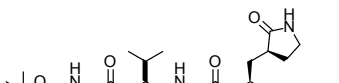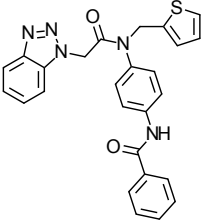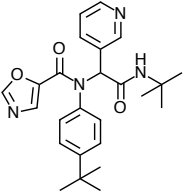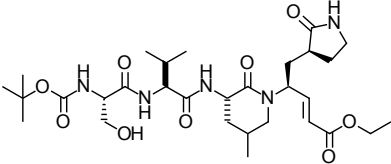
Figure 3.3 Small Molecule Inhibitor Assay

A) L2 cells were treated with indicated concentrations of inhibitors. RLU were determined using Promega CellTiter-Glo Assay. Percent cell viability was calculated relative to DMSO treated cells. Error bars represent the standard error of the mean of three independent replicates. B and C) L2 cells were infected with the indicated FFL viruses at MOI=1 PFU/cell, and treated with indicated concentrations of GRL-001-13S (B) or GRL-002-13S (C). RLU were determined using Promega Luciferase Assay. Fold change in RLU was calculated relative to mock treated infected cells. Points represent average of three independent replicates.

Table 3.1 Compounds Tested for Nsp5 Inhibitory Activity

Compound	Structure	IC_{50}*(μM) (SARS)	IC_{50} (μM) (MHV)	IC_{50} (μM) (HKu1)	IC_{50} (μM) (OC43)	EC_{50} (μM) (MHV)	EC_{50} (μM) (H5)	EC_{50} (μM) (O5)	K_i (μM) (MERS)
GB8		27	10.9	9.6	24	100-200	25-50	100-200	ND
GRL-001-13S		0.31± 0.05	ND	ND	ND	3-5	2-3	1.5-3	ND
GRL-002-13S		ND	ND	ND	ND	25-50	50-200	6-12.5	9.9±2.6
GRL-006-13S		ND	ND	ND	ND	30-70	25-100	70-200	4.7±0.6
GRL-007-13S		ND	ND	ND	ND	15-50	12-50	0.1-50	9.0±2.3

Compound	Structure	IC$_{50}$ (µM) (SARS)	IC$_{50}$ (µM) (MHV)	IC$_{50}$ (µM) (HKu1)	IC$_{50}$ (µM) (OC43)	EC$_{50}$ (µM) (MHV)	EC$_{50}$ (µM) (H5)	EC$_{50}$ (µM) (O5)	K$_i$ (µM) (MERS)
RFM		ND	3.4	19.0	0.49	30-100	3-6.5	25-125	ND
RWB		1.5*	7.4	9.8	8.3	12	4.5	0.35	ND
UIC-0310		ND	ND	ND	ND	100+	100+	70-150	3.6±0.8
UIC-0314		ND	ND	ND	ND	25-50	12.5-30	25-50	ND

* indicates previously published values
** ND = not determined

$$K_i = {IC_{50}} \Big/ {1 + \dfrac{[S]}{K_m}}$$

Discussion and Future Directions

Based on my data, my method of evaluating nsp5 inhibitors represents a platform for rapid EC_{50} determination. My dose response curves show saturability and several logs of luciferase activity. Additionally, my concentration curves show differential effects between various nsp5 proteases. This indicates that reduction in RLU and thus viral replication in the presence of nsp5 inhibitor is directly due to nsp5 inhibition and not an off-target effect. Further, this platform provides us the ability to test inhibitors of nsp5 proteases from CoVs that are not readily cultivatable.

My data and previous work from our lab has demonstrated that nsp5 is likely highly co-evolved with the rest of the genome in largely unidentified ways. Specifically, while we were able to generate chimeric viruses by exchange of nsp5 into the MHV genome from closely related β-CoVs, we could not recover chimeric MHV expressing nsp5 from more distantly related β-CoVs (SARS-CoV, MERS-CoV) or α-CoVs (HCoV-229E, HCoV-NL63). In order to more broadly test inhibitors and investigate the genetic relationships among nsp5 proteases of diverse CoVs, I propose that a series of viruses from each CoV genus subgroup could be used as a backbone to generate nsp5 chimeras. For example, BtCoV-HKU5 could be used as an isogenic background to generate nsp5 chimeras from MERS-CoV and BtCoV-HKU4 (**Error! Reference source not found.**). his would be feasible since our collaborators at the University of North Carolina Chapel Hill in the Ralph Baric lab have recently developed a reverse genetics system for BtCoV-HKU5, thus, it is readily cultivatable (1). HCoV-229E could be used as an isogenic background to put nsp5 proteases from NL63, BtCoV-HKU2, and BtCoV-HKU8 (**Error! eference source not found.**). Volker Thiel's lab of the University of Berne, Switzerland,

has an established reverse genetics system for cultivation and study of HCoV-229E (23). These chimeras would be useful since they would make testing for inhibitors of many uncultivable CoVs feasible. They would also allow us to use our platform to quickly assess for broad reactivity of small molecule inhibitors.

In an effort to assess the potential for antiviral resistance, I passaged WT-MHV four times in the presence of EC_{50} (5 μM) GRL-001-13S. I extracted viral RNA, RT-PCR amplified nsp5, and sequenced across the nsp5 region by Sanger sequencing. No mutations were identified within nsp5. While passaging the virus in the presence of drug, there was a delay in CPE followed by a rapid increase in CPE at 12-16 h p.i. This may indicate that the virus overcame the initial inhibition or that the drug was effective but degraded over time. Further, cells may not take up the drug readily, or it may precipitate out of solution. Future experiments could use temperature dependent ^{1}H NMR to assess molecule stability and degradation. This passage experiment should be repeated using higher concentrations of inhibitor, up to EC_{99}, in order to identify resistance mutations. Once we have isolated anti-viral resistance mutations, we could reverse engineer the mutations into each chimeric reporter virus and perform a luciferase assay to confirm an increase in EC_{50} values. Biochemical studies could also be performed to evaluate inhibitory concentration 50% values (IC_{50}). We expect to find differences in antiviral resistance mutations based on the protease under inhibition. Knowing potential regions for compensation could improve the design of small molecule inhibitors that the virus cannot escape. Additionally, this data could inform methods of protease intramolecular communication as well as modes of functional and genetic divergence.

CHAPTER 4 :

SUMMARY AND IMPLICATIONS

This work has described a series of experiments designed to elucidate the mechanism of nsp5 activity in order to inform *de novo* inhibitor design. Several mutations within the IDL resulted in unrecoverable viruses, suggesting that these residues may be particularly important in protease function or folding. Conversely, several residues within the IDL were readily mutated resulting in no replication defects. Further, I was unable to recover N203C alone, suggesting that the 100% conserved N203 residue in D3 is likely to contribute to protease function. This work has shown that residues within the IDL likely interact with proximal residues as well as engaging in long distance communication across protease domains. Our studies to understand the structure-function relationships of CoV nsp5 proteases are of particular importance since they may reveal regions of the protease critical to function that may be targeted for allosteric inhibition.

In addition to elucidating structure-function relationships of CoV nsp5 proteases, I have developed a platform to test inhibitors during infection as opposed to *in vitro*. This platform is useful since it provides us a way to rapidly quantitate viral replication via luciferase activity. It also provides us the ability to test inhibitors of nsp5 proteases from CoVs that are not readily cultivatable. We hope to develop an even broader platform, containing nsp5 chimeras for each CoV genus subgroup with nsp2-FFL fusions. A platform with this breadth would allow us to test new protease inhibitors for broad reactivity. It would also allow us to quickly evaluate nsp5 protease targeted drugs against

newly emergent zoonotic CoVs. Further, our efforts to generate nsp5 protease chimeras will help us to understand the evolution of nsp5 function and relationships to other replicase proteins. This, in turn, might identify novel residues or functional regions that are conserved across all CoVs, which could be targeted for allosteric inhibition by *de novo* design.

A common problem with antiviral therapy is antiviral resistance. We expect that CoVs will mutate specific sets of variable residues within nsp5 in order to evade protease inhibition and promote viral replication. Although we cannot prevent antiviral resistance mutations from arising during nsp5 active-site inhibition, we can preemptively isolate resistance mutations and study them. Our plan to passage viruses in the presence of nsp5 inhibitors in order to develop resistance mutations will allow us to understand modes of communication within the protease. This will allow us to iteratively develop better active-site inhibitors or even *de novo* design allosteric inhibitors that prevent intermolecular protease communication.

It is critical that we develop several strategies to treat CoV infections as new HCoVs continue to emerge and cause severe human disease and mortality. We do not understand the zoonotic mechanisms by which HCoVs have emerged into the human population, thus, we are unable to predict or prevent future HCoV emergence. However, we can prepare platforms and approaches that are widely applicable to CoV genera in order to be prepared for future emergent disease. Data obtained from this and future studies will (i) define the role of the novel third domain and interdomain loop of nsp5, which connects the catalytic fold to the third domain, during substrate selection and catalytic activity, and (ii) develop novel inhibitors of nsp5 that exploit the functional

regions and activity of nsp5. Results from these studies will inform antiviral and vaccine

development with an emphasis on developing a broadly applicable inhibitor screening

platform and identifying broadly reactive small molecule protease inhibitors.

CHAPTER 5 :

MATERIALS AND METHODS

WT-MHV, Cells and Antibodies

Recombinant MHV A59 (GenBank accession number AY910861) virus was used

as WT-MHV for all experiments. Delayed brain tumor (DBT) cells, baby hamster kidney

(BHK) cells, and rat lung epithelium (L2) cells were grown in Dulbecco's modified Eagle

medium (DMEM; Gibco) supplemented with 10% fetal bovine serum, 1% HEPES, 1%

penicillin/streptomycin and 0.1% Amphotericin B (complete DMEM). Media for the

BHK-MHV receptor (MHVR) cells was supplemented with G418 (Mediatech) at

0.8mg/ml to maintain selection of MHVR. Rabbit polyclonal antibodies were used for

immunoprecipitation directed at the viral nonstructural protein nsp5 (B3-VU6).

Mutagenesis

In order to introduce substitutions into nsp5, the MHV reverse genetics system was used.

Briefly, the MHV genome is divided into seven plasmids, and nsp5 is located in plasmid

pCR-XL-pSMART C. Nucleotides 9556-11510 are in fragment C. Substitutions were

introduced into the C fragment using Quick Change (Stratagene) PCR mutagenesis, and

the primers listed in Table 5.1. Changes to the manufacturer's protocol include the use of

PFU turbo and the following PCR conditions: initial denaturation at 95°C for 2 min,

followed by 16 cycles of denaturation at 95°C for 30 sec, annealing at temperatures

dependent on the primers for 1 min, and extension for 10 min at 72°C. All C fragment

plasmids containing mutations in nsp5 were sequenced to ensure that PCR amplification did not introduce additional changes in the coding region.

Table 5.1: Mutagenesis Primers

Primer Name	Sequence	Purpose
P184A sense	5'-GAACTTTTATGGTGCATATAGAGATGCGAAG-3'	Mutagenesis for P184A
P184A antisense	5'-CTTCGCATCTCTATATGCACCATAAAAGTTC-3'	Mutagenesis for P184A
Y185A sense	5'-TATGGTCCCGCGAGAGATGCG-3'	Mutagenesis for Y185A
Y185A antisense	5'-CGCATCTCTCGCGGGACCATA-3'	Mutagenesis for Y185A
R186A sense	5'-GTACAACTTGCGCATCAGCATAGGGACCATAAAAG-3'	Mutagenesis for R186A
R186A antisense	5'- CTTTTATGGTCCCTATGCTGATGCGCAAGTTGTAC -3'	Mutagenesis for R186A
D187E sense	5'-CCCTATAGAGAGGCGCAAGTT-3'	Mutagenesis for D187E
D187E antisense	5'-AACTTGCGCCTCTCTATAGGG-3'	Mutagenesis for D187E
A188I sense	5'- GCAATTGTACAACTTGAATATCTCTATAGGGAC -3'	Mutagenesis for A188I
A188I antisense	5'- GTCCCTATAGAGATATTCAAGTTGTACAATTGC -3'	Mutagenesis for A188I
Q189A sense	5'-CAGGCAATTGTACAACGGCCGCATCTCTATAGGG-3'	Mutagenesis for Q189A
Q189A antisense	5'-CCCTATAGAGATGCGGCCGTTGTACAATTGCCTG-3'	Mutagenesis for Q189A
V190I sense	5'- GAACAGGCAATTGTACGATTTGCGCATCTCTATAGG -3'	Mutagenesis for V190I
V190I antisense	5'- CCT ATA GAG ATG CGC AAA TCG TAC AAT TGC CTG TTC -3'	Mutagenesis for V190I
V191I sense	5'- CTGAACAGGCAATTGGATAACTTGCGCATCTC -3'	Mutagenesis for V191I
V191I antisense	5'- GAGATGCGCAAGTTATCCAATTGCCTGTTCAG -3'	Mutagenesis for V191I
Q192N sense	5'-GATGCGCAAGTTGTAAACTTGCCTGT-3'	Mutagenesis for Q192N
Q192N antisense	5'-GCGTATAATCCTGAACAGGCAAGTTT-3'	Mutagenesis for Q192N
L193A sense	5'- GTATAATCCTGAACAGGGGCTTGTACAACTTGCGC -3'	Mutagenesis for L193A
L193A antisense	5'- GCGCAAGTTGTACAAGCCCCTGTTCAGGATTATAC -3'	Mutagenesis for L193A
P194A sense	5'-GTTGTACAATTGGCGGTTCAGGATGATTATAC-3'	Mutagenesis for P194A
P194A antisense	5'-GTATAATCATCCTGAACCGCCAATTGTACAAC-3'	Mutagenesis for P194A
V195I sense	5'-CTGCGTATAATCCTGGATAGGCAATTGTACAAC-3'	Mutagenesis for V195I
V195I antisense	5'-GTTGTACAATTGCCTATCCAGGATTATACGCAG-3'	Mutagenesis for V195I
Q196A sense	5'-CAATTGCCTGTTGCAGATTATACGCAG-3'	Mutagenesis for Q196A
Q196A antisense	5'-CTGCGTATAATCTGCAACAGGCAATTG-3'	Mutagenesis for Q196A
D197A sense	5'-CAGTCTGCGTATATGCCTGAACAGGCAATTG-3'	Mutagenesis for D197A
D197A antisense	5'-CAATTGCCTGTTCAGGCATATACGCAGACTG -3'	Mutagenesis for D197A
Y198A sense	5'-CCTGTTCAGGATGCGACGCAGACTGT-3'	Mutagenesis for Y198A
Y198A antisense	5'-AACAGTCTGCGTCGCATCCTGAACAG-3'	Mutagenesis for Y198A
T199A sense	5'-CCTGTTCAGGATTATGCACAGACTGT-3'	Mutagenesis for T199A
T199A antisense	5'-AACAGTCTGTGCATAATCCTGAACAG-3'	Mutagenesis for T199A

Primer Name	Sequence	Purpose
Δ199 sense	5'- GCTACAACATTAACAGTCTGATAATCCTGAACAGGCAATTGTAC AAC -3'	Mutagenesis for Δ199
Δ199 antisense	5'- GTTGTACAATTGCCTGTTCAGGATTATCAGACTGTTAATGTTGTA GC -3'	Mutagenesis for Δ199
+197-199 sense	5'- GCTACAACATTAACAGTCTGCGTATAATCCGTATAATCCTGAAC AGGCAATTG -3'	Mutagenesis for +197-199
+197-199 antisense	5'- CAATTGCCTGTTCAGGATTATACGGATTATACGCAGACTGTTAAT GTTGTAGC -3'	Mutagenesis for +197-199
+199 sense	5'- GCTACAACATTAACAGTCTGCGTCGTATAATCCTGAACAGGCAA TTG -3'	Mutagenesis for +199
+199 antisense	5'- CAATTGCCTGTTCAGGATTATACGACGCAGACTGTTAATGTTGTA GC -3'	Mutagenesis for +199
R131A sense	5'-CTTTATGGTATGGCTACTCGCAAGCGTAACATGGAAG-3'	Mutagenesis for R131A
R131A antisense	5'- CTTCCATGTTACGCTTGCGAGTAGCCATACCATAAAG -3'	Mutagenesis for R131A
R131K sense	5'-CTTTATGGTATGGCTACTTTTAAGCGTAACATGGAAG-3'	Mutagenesis for R131K
R131K antisense	5'- CTTCCATGTTACGCTTAAAAGTAGCCATACCATAAAG -3'	Mutagenesis for R131K
R131C sense	5'-CTTTATGGTATGGCTACTGCAAAGCGTAACATGGAAG-3'	Mutagenesis for R131C
R131C antisense	5'- CTTCCATGTTACGCTTTGCAGTAGCCATACCATAAAG -3'	Mutagenesis for R131C
D286A sense	5'- CTTGGTGTCAGCTCTGCTTCAAGCACACAAC -3'	Mutagenesis for D286A
D286A antisense	5'- GTTGTGTGCTTGAAGCAGAGCTGACACCAAG -3'	Mutagenesis for D286A
D286C sense	5'-CTTGGTGTCAGCTCGCATTCAAGCACACAAC-3'	Mutagenesis for D286C
D286C antisense	5'- GTTGTGTGCTTGAA GCGAGCTGACACCAAG -3'	Mutagenesis for D286C
G109L sense	5'-GTACAGTAAATGTCTCTAGAGGCTTAACAACACC-3'	Mutagenesis for G109L
G109L antisense	5'- GGTGTTGTTAAGCCTCTAGAGACATTTACTGTAC -3'	Mutagenesis for G109L
G109C sense	5'-GTACAGTAAATGTCTCGCAAGGCTTAACAACACC-3'	Mutagenesis for G109C
G109C antisense	5'- GGTGTTGTTAAGCCTTGCGAGACATTTACTGTAC -3'	Mutagenesis for G109C
N203A sense	5'-CATAAAGCCAAGCTACAACTGCAACAGTCTGCGTATAATC-3'	Mutagenesis for N203A
N203A antisense	5'- GATTATACGCAGACTGTTGCAGTTGTAGCTTGGCTTTATG -3'	Mutagenesis for N203A
N203C sense	5'-CATAAAGCCAAGCTACAACGCAAACAGTCTGCGTATAATC-3'	Mutagenesis for N203C
N203C antisense	5'- GATTATACGCAGACTGTTTGCGTTGTAGCTTGGCTTTATG -3'	Mutagenesis for N203C

Virus Recovery

Viruses containing the nsp5 mutations were generated using the reverse genetics system for MHV-A59 described by Yount *et al*. (72) and modified by Denison *et al*. (12) and Sparks *et al* (56). Briefly, the MHV-A59 genome is divided into seven cDNA fragments, which were digested using the appropriate restriction enzymes. These digested fragments were then ligated at 16 °C overnight before the DNA was purified, *in vitro* transcribed and electroporated into BHK-MHVR cells along with N gene transcripts. Electroporated cells were co-cultured with DBT cells and incubated at 37 °C until cytopathic effects were observed. The cytopathic effect see in MHV-infected cells is the formation of multinucleated giant cells (syncytia formation). The virus produced from electroporated cells (passage 0, P0) was passaged onto uninfected DBT cells to generate a P1 stock virus that was used for all experiments. The P0 virus was sequenced across the nsp5 gene to ensure that no additional mutations were present. If the *in vitro* transcribed genome did not produce virus on the first attempt, virus assembly was attempted at least two additional times.

RT-PCR and Sequencing

Total intracellular RNA was isolated using TRIzol (Invitrogen) and the manufacturer's protocol. Viral RNA was then reverse transcribed using Superscript III (Invitrogen) and random hexamers (Roche). The nsp5 coding sequence was amplified by PCR using primers complementary to 10158 to 10177 (sense) and 11783 to 11799 (antisense). The nsp5 amplicons generated were directly Sanger sequenced to analyze retention of the engineered mutations and absence of additional mutations.

Viral replication Assay

DBT cells were infected with either WT or nsp4 mutant viruses at an MOI of 1 PFU/cell and absorbed for 30 min. Cells were then washed twice with PBS, media was replaced, and cells were incubated at 37°C. Supernatants were sampled from 1 hour to 24 h post infection (p.i.) and viral titers were determined by plaque assay, as previously described (28).

Radiolabel of viral proteins and protein immunoprecipitation

DBT cells were infected at an MOI of 10 PFU/cell or mock infected. At 4 h p.i., the medium was replaced with DMEM that lacks cysteine and methionine and supplemented with 20 μg/ml actinomycin D (ActD; Sigma). At 5 h p.i., [^{35}S]-Methionine/-Cysteine ([^{35}S]-Met/-Cys) was added to the cells and monitored for CPE. When cells were 90-100% involved in CPE, lysates were harvested in lysis buffer (1% NP-40, 0.5% sodium deoxycholate, 150 mM NaCl, and 50 mM Tris, pH 8.0). Lysates were centrifuged at 6,000X g for 3 min to remove nuclei. Viral proteins were immunoprecipitated in a total volume of 1ml with 0.6 mg of Protein A-Sepharose beads (Sigma), 200ul lysate, 1:250 dilution of polyclonal nsp5 antisera, and a protease inhibitor cocktail (Roche) in lysis buffer. After an overnight incubation, immunoprecipitated proteins were pelleted by centrifugation at 12,000 x g for 1 min. Protein-bead conjugates were then washed in low salt lysis buffer (lysis buffer with 150mM NaCl), high salt lysis buffer (Lysis buffer with 1M NaCl) and low salt lysis buffer. Samples were then resuspended in 2x SDS protein sample buffer, heated to 70°C for 10 min. Proteins were

resolved by SDS-PAGE gel electrophoresis on 4-12% Bis-Tris polyacrylamide gels (Invitrogen) and imaged by autoradiography. A full range rainbow ladder (GE Healthcare) and ^{14}C ladder (PerkinElmer LAS) were used as molecular weight standards.

Luciferase assay for nsp5 inhibition

DBT cells were infected at MOI = 1 PFU/cell and absorbed for 30 minutes in 96-well plates. Nsp5 inhibitors were diluted in 1:2 serial dilutions at 50X concentration. Drug was added to wells with cells and complete DMEM (3 replicates each) resulting in 1X concentration of drug, 2% DMSO, and 2% fetal bovine serum. At 10 h p.i., cells were washed with PBS, and 100 μL of reporter lysis buffer (Promega) was added. Cells were frozen at -80°C until the luciferase assay was performed. To measure RLU, cells were warmed to ambient temperature. The luciferase assay reagent (Promega) was suspended according to directions, and the reagent was added and activity read by a Veritas luminometer. RLU raw data was normalized to untreated, infected wells and log transformed. Error bars represent standard error of the mean (SEM).

REFERENCES

1. **Agnihothram, S., B. L. Yount, Jr., E. F. Donaldson, J. Huynh, V. D. Menachery, L. E. Gralinski, R. L. Graham, M. M. Becker, S. Tomar, T. D. Scobey, H. L. Osswald, A. Whitmore, R. Gopal, A. K. Ghosh, A. Mesecar, M. Zambon, M. Heise, M. R. Denison, and R. S. Baric.** 2014. A mouse model for Betacoronavirus subgroup 2c using a bat coronavirus strain HKU5 variant. mBio **5:**e00047-00014.

2. **Anand, K., G. J. Palm, J. R. Mesters, S. G. Siddell, J. Ziebuhr, and R. Hilgenfeld.** 2002. Structure of coronavirus main proteinase reveals combination of a chymotrypsin fold with an extra α‒helical domain. The EMBO Journal **21:**3213-3224.

3. **Angelini, M. M., M. Akhlaghpour, B. W. Neuman, and M. J. Buchmeier.** 2013. Severe acute respiratory syndrome coronavirus nonstructural proteins 3, 4, and 6 induce double-membrane vesicles. mBio **4**.

4. **Chang, H. P., C. Y. Chou, and G. G. Chang.** 2007. Reversible unfolding of the severe acute respiratory syndrome coronavirus main protease in guanidinium chloride. Biophysical journal **92:**1374-1383.

5. **Chen, H., P. Wei, C. Huang, L. Tan, Y. Liu, and L. Lai.** 2006. Only one protomer is active in the dimer of SARS 3C-like proteinase. The Journal of biological chemistry **281:**13894-13898.

6. **Chen, S., L. Chen, J. Tan, J. Chen, L. Du, T. Sun, J. Shen, K. Chen, H. Jiang, and X. Shen.** 2005. Severe acute respiratory syndrome coronavirus 3C-like proteinase N terminus is indispensable for proteolytic activity but not for enzyme dimerization. Biochemical and thermodynamic investigation in conjunction with molecular dynamics simulations. The Journal of biological chemistry **280:**164-173.

7. **Chen, S., F. Jonas, C. Shen, and R. Hilgenfeld.** 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.

8. **Chen, S., J. Zhang, T. Hu, K. Chen, H. Jiang, and X. Shen.** 2008. Residues on the Dimer Interface of SARS Coronavirus 3C-like Protease: Dimer Stability Characterization and Enzyme Catalytic Activity Analysis. Journal of biochemistry **143:**525-536.

9. **Chuck, C. P., C. Chen, Z. Ke, D. C. Wan, H. F. Chow, and K. B. Wong.** 2013. Design, synthesis and crystallographic analysis of nitrile-based broad-spectrum peptidomimetic inhibitors for coronavirus 3C-like proteases. European journal of medicinal chemistry **59:**1-6.

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

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

12. **Denison, M. R., B. Yount, S. M. Brockway, R. L. Graham, A. C. Sims, X. Lu, and R. S. Baric.** 2004. Cleavage between replicase proteins p28 and p65 of mouse hepatitis virus is not required for virus replication. Journal of virology **78:**5957-5965.

13. **Denison, M. R., P. W. Zoltick, S. A. Hughes, B. Giangreco, A. L. Olson, S. Perlman, J. L. Leibowitz, and S. R. Weiss.** 1992. Intracellular processing of the N-terminal ORF 1a proteins of the coronavirus MHV-A59 requires multiple proteolytic events. Virology **189:**274-284.

14. **Fan, K., P. Wei, Q. Feng, S. Chen, C. Huang, L. Ma, B. Lai, J. Pei, Y. Liu, J. Chen, and L. Lai.** 2004. Biosynthesis, Purification, and Substrate Specificity of Severe Acute Respiratory Syndrome Coronavirus 3C-like Proteinase. Journal of Biological Chemistry **279:**1637-1642.

15. **Freeman, M. C., R. L. Graham, X. Lu, C. T. Peek, and M. R. Denison.** 2014. Coronavirus replicase-reporter fusions provide quantitative analysis of replication and replication complex formation. Journal of virology **88:**5319-5327.

16. **Galiano, L., M. Bonora, and G. E. Fanucci.** 2007. Interflap distances in HIV-1 protease determined by pulsed EPR measurements. J Am Chem Soc **129:**11004-11005.

17. **Galiano, L., F. Ding, A. M. Veloro, M. E. Blackburn, C. Simmerling, and G. E. Fanucci.** 2009. Drug Pressure Selected Mutations in HIV-1 Protease Alter Flap Conformations. J Am Chem Soc **131:**430-+.

18. **Ghosh, A. K., G. Gong, V. Grum-Tokars, D. C. Mulhearn, S. C. Baker, M. Coughlin, B. S. Prabhakar, K. Sleeman, M. E. Johnson, and A. D. Mesecar.** 2008. Design, synthesis and antiviral efficacy of a series of potent chloropyridyl ester-derived SARS-CoV 3CLpro inhibitors. Bioorganic & medicinal chemistry letters **18:**5684-5688.

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

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

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

22. **Harcourt, B. H., D. Jukneliene, A. Kanjanahaluethai, J. Bechill, K. M. Severson, C. M. Smith, P. A. Rota, and S. C. Baker.** 2004. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. Journal of virology **78:**13600-13612.

23. **Herold, J., V. Thiel, and S. G. Siddell.** 1998. A strategy for the generation of infectious RNAs and autonomously replicating RNAs based on the HCV 229E genome. Advances in experimental medicine and biology **440:**265-268.

24. **Hsu, M.-F., C.-J. Kuo, K.-T. Chang, H.-C. Chang, C.-C. Chou, T.-P. Ko, H.-L. Shr, G.-G. Chang, A. H.-J. Wang, and P.-H. Liang.** 2005. Mechanism of the Maturation Process of SARS-CoV 3CL Protease. Journal of Biological Chemistry **280:**31257-31266.

25. **Jacobs, J., V. Grum-Tokars, Y. Zhou, M. Turlington, S. A. Saldanha, P. Chase, A. Eggler, E. S. Dawson, Y. M. Baez-Santos, S. Tomar, A. M. Mielech, S. C. Baker, C. W. Lindsley, P. Hodder, A. Mesecar, and S. R. Stauffer.** 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.

26. **Kear, J. L., M. E. Blackburn, A. M. Veloro, B. M. Dunn, and G. E. Fanucci.** 2009. Subtype polymorphisms among HIV-1 protease variants confer altered flap conformations and flexibility. J Am Chem Soc **131:**14650-14651.

27. **Kilianski, A., A. M. Mielech, X. Deng, and S. C. Baker.** 2013. Assessing activity and inhibition of Middle East respiratory syndrome coronavirus papain-like and 3C-like proteases using luciferase-based biosensors. Journal of virology **87:**11955-11962.

28. **Kim, J. C., R. A. Spence, P. F. Currier, X. Lu, and M. R. Denison.** 1995. Coronavirus protein processing and RNA synthesis is inhibited by the cysteine proteinase inhibitor E64d. Virology **208:**1-8.

29. **Kim, Y., S. Lovell, K. C. Tiew, S. R. Mandadapu, K. R. Alliston, K. P. Battaile, W. C. Groutas, and K. O. Chang.** 2012. Broad-spectrum antivirals against 3C or 3C-like proteases of picornaviruses, noroviruses, and coronaviruses. Journal of virology **86:**11754-11762.

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

31. **Kooi, C., M. Cervin, and R. Anderson.** 1991. Differentiation of acid-pH-dependent and -nondependent entry pathways for mouse hepatitis virus. Virology **180:**108-119.

32. **Lee, C. C., C. J. Kuo, T. P. Ko, M. F. Hsu, Y. C. Tsui, S. C. Chang, S. Yang, S. J. Chen, H. C. Chen, M. C. Hsu, S. R. Shih, P. H. Liang, and A. H. Wang.** 2009. Structural basis of inhibition specificities of 3C and 3C-like proteases by zinc-coordinating and peptidomimetic compounds. The Journal of biological chemistry **284:**7646-7655.

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

34. **Li, C., Y. Qi, X. Teng, Z. Yang, P. Wei, C. Zhang, L. Tan, L. Zhou, Y. Liu, and L. Lai.** 2010. Maturation Mechanism of Severe Acute Respiratory Syndrome (SARS) Coronavirus 3C-like Proteinase. Journal of Biological Chemistry **285:**28134-28140.

35. **Liao, H. H., Y. C. Wang, M. C. Chen, H. Y. Tsai, J. Lin, S. T. Chen, G. J. Tsay, and S. L. Cheng.** 2011. Down-regulation of granulocyte-macrophage colony-stimulating factor by 3C-like proteinase in transfected A549 human lung carcinoma cells. BMC immunology **12:**16.

36. **Lin, P. Y., C. Y. Chou, H. C. Chang, W. C. Hsu, and G. G. Chang.** 2008. Correlation between dissociation and catalysis of SARS-CoV main protease. Archives of biochemistry and biophysics **472:**34-42.

37. **Lu, J. H., D. M. Zhang, G. L. Wang, Z. M. Guo, J. Li, B. Y. Tan, L. P. Ou-Yang, W. H. Ling, X. B. Yu, and N. S. Zhong.** 2005. Sequence analysis and structural prediction of the severe acute respiratory syndrome coronavirus nsp5. Acta biochimica et biophysica Sinica **37:**473-479.

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

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

40. **Lu, Y., X. Lu, and M. R. Denison.** 1995. Identification and characterization of a serine-like proteinase of the murine coronavirus MHV-A59. Journal of virology **69:**3554-3559.

41. **Malani, P.** 2014. First US MERS-CoV cases underscore need for preparedness. JAMA : the journal of the American Medical Association **311:**2160-2161.

42. **Masters, P. S.** 2006. The molecular biology of coronaviruses. Advances in virus research **66:**193-292.

43. **McHaourab, H. S., P. R. Steed, and K. Kazmier.** 2011. Toward the fourth dimension of membrane protein structure: insight into dynamics from spin-labeling EPR spectroscopy. Structure (London, England : 1993) **19:**1549-1561.

44. **Milne-Price, S., K. L. Miazgowicz, and V. J. Munster.** 2014. The emergence of the Middle East Respiratory Syndrome coronavirus (MERS-CoV). Pathogens and disease.

45. **Muramatsu, T., Y. T. Kim, W. Nishii, T. Terada, M. Shirouzu, and S. Yokoyama.** 2013. Autoprocessing mechanism of severe acute respiratory syndrome coronavirus 3C-like protease (SARS-CoV 3CL(pro)) from its polyproteins. The FEBS journal **280:**2002-2013.

46. **Organization, W. H.** July 4, 2014 2014, posting date. Middle East respiratory syndrome coronavirus (MERS-CoV) - update. [Online.]

47. **Perlman, S., and J. Netland.** 2009. Coronaviruses post-SARS: update on replication and pathogenesis. Nature reviews. Microbiology **7:**439-450.

48. **Pyrc, K., B. Berkhout, and L. van der Hoek.** 2007. Identification of new human coronaviruses. Expert review of anti-infective therapy **5:**245-253.

49. **Raj, V. S., H. Mou, S. L. Smits, D. H. Dekkers, M. A. Muller, R. Dijkman, D. Muth, J. A. Demmers, A. Zaki, R. A. Fouchier, V. Thiel, C. Drosten, P. J. Rottier, A. D. Osterhaus, B. J. Bosch, and B. L. Haagmans.** 2013. Dipeptidyl peptidase 4 is a functional receptor for the emerging human coronavirus-EMC. Nature **495:**251-254.

50. **Ren, Z., L. Yan, N. Zhang, Y. Guo, C. Yang, Z. Lou, and Z. Rao.** 2013. The newly emerged SARS-like coronavirus HCoV-EMC also has an "Achilles' heel": current effective inhibitor targeting a 3C-like protease. Protein & cell **4:**248-250.

51. **Seybert, A., J. Ziebuhr, and S. G. Siddell.** 1997. Expression and characterization of a recombinant murine coronavirus 3C-like proteinase. The Journal of general virology **78 (Pt 1):**71-75.

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

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

54. **Shi, J., Z. Wei, and J. Song.** 2004. Dissection study on the severe acute respiratory syndrome 3C-like protease reveals the critical role of the extra domain in dimerization of the enzyme: defining the extra domain as a new target for design of highly specific protease inhibitors. The Journal of biological chemistry **279:**24765-24773.

55. **SØRensen, M. D., B. SØRensen, R. Gonzalez-Dosal, C. J. Melchjorsen, J. Weibel, J. Wang, C. W. Jun, Y. Huanming, and P. Kristensen.** 2006. Severe Acute Respiratory Syndrome (SARS). Annals of the New York Academy of Sciences **1067:**500-505.

56. **Sparks, J. S., X. Lu, and M. R. Denison.** 2007. Genetic analysis of Murine hepatitis virus nsp4 in virus replication. Journal of virology **81:**12554-12563.

57. **Stacey Knobler, A. M., Stanley Lemon, Alison Mack, Laura Sivitz, and Katherine Oberholtzer,.** 2004. Learning from SARS: Preparing for the Next Disease Outbreak. National Academies Press.

58. **Stobart, C. C., A. S. Lee, X. Lu, and M. R. Denison.** 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. Journal of virology **86:**4801-4810.

59. **Stobart, C. C., N. R. Sexton, H. Munjal, X. Lu, K. L. Molland, S. Tomar, A. D. Mesecar, and M. R. Denison.** 2013. Chimeric exchange of coronavirus nsp5 proteases (3CLpro) identifies common and divergent regulatory determinants of protease activity. Journal of virology **87:**12611-12618.

60. **Suel, G. M., S. W. Lockless, M. A. Wall, and R. Ranganathan.** 2003. Evolutionarily conserved networks of residues mediate allosteric communication in proteins. Nature structural biology **10:**59-69.

61. **Tan, J., K. H. Verschueren, K. Anand, J. Shen, M. Yang, Y. Xu, Z. Rao, J. Bigalke, B. Heisen, J. R. Mesters, K. Chen, X. Shen, H. Jiang, and R. Hilgenfeld.** 2005. pH-dependent conformational flexibility of the SARS-CoV main proteinase (M(pro)) dimer: molecular dynamics simulations and multiple X-ray structure analyses. Journal of molecular biology **354:**25-40.

62. **Tibbles, K. W., D. Cavanagh, and T. D. Brown.** 1999. Activity of a purified His-tagged 3C-like proteinase from the coronavirus infectious bronchitis virus. Virus research **60:**137-145.

63. **Tsai, M.-Y., W.-H. Chang, J.-Y. Liang, L.-L. Lin, G.-G. Chang, and H.-P. Chang.** 2010. Essential covalent linkage between the chymotrypsin-like domain and the extra domain of the SARS-CoV main protease. Journal of biochemistry **148:**349-358.

64. **Tsai, M. Y., W. H. Chang, J. Y. Liang, L. L. Lin, G. G. Chang, and H. P. Chang.** 2010. Essential covalent linkage between the chymotrypsin-like domain and the extra domain of the SARS-CoV main protease. Journal of biochemistry **148:**349-358.

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

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

67. **Weiss, S. R., and J. L. Leibowitz.** 2011. Coronavirus pathogenesis. Advances in virus research **81:**85-164.

68. **Williams, R. K., G. S. Jiang, and K. V. Holmes.** 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.

69. **Wu, C. G., S. C. Cheng, S. C. Chen, J. Y. Li, Y. H. Fang, Y. H. Chen, and C. Y. Chou.** 2013. Mechanism for controlling the monomer-dimer conversion of SARS coronavirus main protease. Acta crystallographica. Section D, Biological crystallography **69:**747-755.

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

71. **Yang, H., W. Xie, X. Xue, K. Yang, J. Ma, W. Liang, Q. Zhao, Z. Zhou, D. Pei, J. Ziebuhr, R. Hilgenfeld, K. Y. Yuen, L. Wong, G. Gao, S. Chen, Z. Chen, D. Ma, M. Bartlam, and Z. Rao.** 2005. Design of wide-spectrum inhibitors targeting coronavirus main proteases. PLoS biology **3:**e324.

72. **Yount, B., M. R. Denison, S. R. Weiss, and R. S. Baric.** 2002. Systematic assembly of a full-length infectious cDNA of mouse hepatitis virus strain A59. Journal of virology **76:**11065-11078.

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

74. **Zheng, K., G. Ma, J. Zhou, M. Zen, W. Zhao, Y. Jiang, Q. Yu, and J. Feng.** 2007. Insight into the activity of SARS main protease: Molecular dynamics study of dimeric and monomeric form of enzyme. Proteins **66:**467-479.

75. **Zhong, N., S. Zhang, F. Xue, X. Kang, P. Zou, J. Chen, C. Liang, Z. Rao, C. Jin, Z. Lou, and B. Xia.** 2009. C-terminal domain of SARS-CoV main protease can form a 3D domain-swapped dimer. Protein science : a publication of the Protein Society **18:**839-844.

76. **Ziebuhr, J., E. J. Snijder, and A. E. Gorbalenya.** 2000. Virus-encoded proteinases and proteolytic processing in the Nidovirales. The Journal of general virology **81:**853-879.