

University of Massachusetts Medical School

eScholarship@UMMS

---

GSBS Dissertations and Theses

Graduate School of Biomedical Sciences

---

2012-08-31

## Hepatitis C Virus Non-Structural Protein 3/4A: A Tale of Two Domains: A Dissertation

Cihan Aydin

*University of Massachusetts Medical School*

Let us know how access to this document benefits you.

Follow this and additional works at: [https://escholarship.umassmed.edu/gsbs\\_diss](https://escholarship.umassmed.edu/gsbs_diss)



Part of the [Amino Acids, Peptides, and Proteins Commons](#), [Biochemistry, Biophysics, and Structural Biology Commons](#), [Digestive System Diseases Commons](#), [Enzymes and Coenzymes Commons](#), [Nucleic Acids, Nucleotides, and Nucleosides Commons](#), [Pharmaceutical Preparations Commons](#), [Therapeutics Commons](#), [Virus Diseases Commons](#), and the [Viruses Commons](#)

---

### Repository Citation

Aydin C. (2012). Hepatitis C Virus Non-Structural Protein 3/4A: A Tale of Two Domains: A Dissertation. GSBS Dissertations and Theses. <https://doi.org/10.13028/9pmf-a841>. Retrieved from [https://escholarship.umassmed.edu/gsbs\\_diss/626](https://escholarship.umassmed.edu/gsbs_diss/626)

This material is brought to you by eScholarship@UMMS. It has been accepted for inclusion in GSBS Dissertations and Theses by an authorized administrator of eScholarship@UMMS. For more information, please contact [Lisa.Palmer@umassmed.edu](mailto:Lisa.Palmer@umassmed.edu).

**HEPATITIS C VIRUS NON-STRUCTURAL PROTEIN 3/4A**

**A TALE OF TWO DOMAINS**

A Dissertation Presented

By

**Cihan AYDIN**

Submitted to the Faculty of the  
University of Massachusetts Graduate School of Biomedical Sciences, Worcester

In partial fulfillment of the requirements for the degree of

**DOCTOR OF PHILOSOPHY**

August 31, 2012

**Biochemistry and Molecular Pharmacology**

**HEPATITIS C VIRUS NON-STRUCTURAL PROTEIN 3/4A  
A TALE OF TWO DOMAINS**

A Dissertation Presented  
By  
**Cihan AYDIN**

The signatures of the Dissertation Defense Committee signify completion and approval as to style and content of the Dissertation.

---

**Celia A. Schiffer, Ph.D., Thesis Advisor**

---

**Osman Bilsel, Ph.D., Member of Committee**

---

**Andrei Korostelev, Ph.D., Member of Committee**

---

**David G. Lambright, Ph.D., Member of Committee**

---

**Juswinder Singh, Ph.D., Member of Committee**

The signature of the Chair of the Committee signifies that the written dissertation meets the requirements of the Dissertation Committee.

---

**William E. Royer, Ph.D., Chair of Committee**

The signature of the Dean of the Graduate School of Biomedical Sciences signifies that the student has met all graduation requirements of the school.

---

**Anthony Carruthers, Ph.D.**  
**Dean of the Graduate School of Biomedical Sciences**  
**Biochemistry and Molecular Pharmacology Program**

August 31, 2012

*This dissertation is dedicated to my father (1947-2009).*

***PhD***

*You wanted it,*

*You couldn't get it,*

*I will have it,*

*You aren't here to see it.*

## **Acknowledgements**

Through my life, I have crossed paths with many people – family, friends, colleagues, people passing by... – whom had positive and negative impacts on my life. However, if I am here today, it is because of the long and thin road of life I am walking on, from cradle to the grave, and the people that walked with me on this way. I would like to thank you altogether, without separating one of you from the other. You are the reason that I am typing these words here today.

Contrary to the tradition, I am not going to acknowledge anyone specifically here, because I don't think a paragraph here about a person or some people would not do justice – I would have needed many pages, that would overshadow the thesis work. However, you should know that every one of you has a special part in my soul.

Looking at my five years as a non-resident alien PhD student in US, I probably have one vital advice that I will carry out with me throughout my life from a very special person (and this may count as an exception to the above) and I am beholden to you for it – "It is not the technique, it is the question."

## **Abstract**

Two decades after the discovery of the Hepatitis C Virus (HCV), Hepatitis C infection still persists to be a global health problem. With the recent approval of the first set of directly acting antivirals (DAAs), the rate of sustained viral response for HCV-infected patients increased significantly. However, a complete cure has not been found yet. Drug development efforts primarily target NS3/4A protease, bifunctional serine protease – RNA helicase of HCV. HCV NS3/4A is critical in viral function; protease domain processes the viral polyprotein and helicase domain aids replication of HCV genome by unwinding double stranded RNA transcripts produced by NS5B, RNA-dependent RNA polymerase of HCV. Protease and helicase domains can be isolated, expressed and purified separately while retaining function. Isolated domains of HCV NS3/4A have been extensively used in biochemical and biophysical studies for scientific and therapeutic purposes to evaluate functional capability and mechanism. However, these domains are highly interdependent and modulate the activities of each other bidirectionally. Interdomain dependence was demonstrated in comparative studies where activities of isolated domains versus the full length protein were evaluated. Nevertheless, specific factors affecting interdependence have not been thoroughly studied.

Chapter II investigates the domain-domain interface formed between protease and helicase domains as a determinant in interdependence. Molecular dynamics simulations performed on single chain NS3/4A constructs demonstrated the importance of interface in the coupled dynamics of the two domains. The role of the interface in interdomain communication was experimentally probed by disrupting the domain-domain interface

through Ala-scanning mutations in selected residues in the interface with significant buried surface areas. These interface mutants were assayed for both helicase and protease related activities. Instead of downregulating the activities of either domain, interface mutants caused enhancement of protease and helicase activities. In addition, the interface had minimal effect in RNA unwinding activity of the helicase domain, the mere presence of the protease domain was the main protagonist in elevated RNA unwinding activity. In conclusion, I suspect that the interface formed between the domains is transient in nature and plays a regulatory role more than a functional role. In addition, I found results supporting the suggestion that an alternate domain-domain arrangement other than what is observed in crystal structures is the active, biologically relevant conformation for both the helicase and the protease.

Chapter III investigates structural features of HCV NS3/4A protease inhibitors in relation to effects on inhibitor potency, susceptibility to drug resistance and modulation of potency by the helicase domain. Nearly all NS3/4A protease inhibitors share common features, with major differences only in bulky P2 extension groups and macrocyclization statuses. Enzymatic inhibition profiles of different drugs were analyzed for wildtype isolated protease domain and single chain NS3/4A helicase – protease construct, their multi drug resistant variants, and additional helicase mutants. Inhibitor potency was mainly influenced by macrocyclization, where macrocyclic drugs were significantly more potent compared to acyclic variants. Potency loss with respect to resistance mutations primarily depended on the P2 extension, while macrocyclization had minimal effect except for P2-P4 macrocyclic compounds which were up to an order of magnitude more

susceptible to mutations A156T and, in lesser extent, D168A. Modulation by helicase domain was also dependent on P2 extension, although opposite trends were observed for danoprevir analogs versus others. In conclusion, this study provides a basis for future inhibitor development in both avoiding drug resistance and exploitation of the helicase domain for additional efficacy.

In this thesis, I have provided evidence further supporting and revealing the details of domain-domain dependency in HCV NS3/4A. Lessons learned here will aid future research for dissecting the interdependency to gain a better understanding of HCV NS3/4A function, which can possibly be extended to all *Flaviviridae* NS3 protease – helicase complexes. In addition, interdomain dependence can be exploited in future drug development efforts to create better drugs that will pave the way to an effective cure.



<b>Table of Contents</b>	
<b>Title Page</b>	i
<b>Signature Page</b>	ii
<b>Dedication</b>	iii
<b>Acknowledgements</b>	iv
<b>Abstract</b>	v
<b>Table of Contents</b>	viii
<b>List of Tables</b>	xiii
<b>List of Figures</b>	xiv
<b>List of Third Party Copyrighted Material</b>	xvi
<b>List of Abbreviations</b>	xvii
<b>List of Prefixes and Suffixes</b>	xix
<b>Preface</b>	xx
<b><u>Chapter I – Introduction</u></b>	1
<b>Epidemiology of Hepatitis C Virus</b>	2
<b>Hepatitis C Virus Life Cycle</b>	6
<b>Treatment of Hepatitis C and Anti Hepatitis C Drug Development</b>	
<b>Effort</b>	15
<b>Non-Structural Protein 3/4A – “Swiss Army Knife” of Hepatitis C</b>	
<b>Virus</b>	17

	ix
<i>Protease Domain</i>	17
<i>Helicase Domain</i>	18
<i>NS4A</i>	23
<b>Interdomain Dependency in Non-Structural Protein 3/4A</b>	24
<b>Structures of Single-Chain NS3 Protease – Helicase Complexes from Hepatitis C Virus and <i>Flaviviruses</i></b>	26
<b>Inhibitors of Non-Structural Protein 3/4A Protease</b>	35
<b>Thesis Summary</b>	39
<b>References</b>	40
<b><u>Chapter II – Destabilization of Protease – Helicase Interface</u></b>	
<b><u>Enhances Both Protease and Helicase Activities of Hepatitis C Virus Non- Structural Protein NS3-NS4A Complex</u></b>	
<b>Author Contributions</b>	
<b>Abstract</b>	61
<b>Introduction</b>	62
<b>Materials and Methods</b>	65
<i>Molecular Dynamics Simulations</i>	65
<i>Nucleic Acid Substrates</i>	65
<i>Protein Expression Plasmids</i>	66
<i>Protein Expression and Purification</i>	66
<i>Protease Assays</i>	68

<i>DNA Unwinding Assays</i>	71
<i>RNA Unwinding Assays</i>	72
<i>DNA Binding Assays</i>	75
<i>Poly(U)-Stimulated ATPase Assays</i>	76
<b>Results</b>	77
<i>Structural Analysis of Domain-Domain Interface</i>	77
<i>Dynamic Interaction between the Domains</i>	81
<i>Purification of scNS3/4A with Amino Acid Substitutions Designed to Disrupt Protease – Helicase Interface</i>	90
<i>Protease Cleavage Efficiencies</i>	93
<i>Helicase Unwinding Rates for DNA and RNA Substrates</i>	93
<i>Helicase Single-Stranded DNA Affinities</i>	103
<i>Helicase RNA-Stimulated ATP Hydrolysis Rates</i>	104
<b>Discussion</b>	107
<i>Protease and Helicase Domains of HCV NS3 are Dynamically Coupled In Silico</i>	107
<i>The Protease is Restricted by the Helicase</i>	108
<i>The Helicase – Protease Interface Influences NS3 Helicase Action</i>	110
<i>ATP Hydrolysis is Independent of Protease – Helicase Association</i>	111
<i>HCV NS3/4A Possibly Functions in an Extended Domain Arrangement</i>	112
<b>Acknowledgements</b>	118

<b>Supplementary Methods</b>	119
<i>Molecular Dynamics Simulations</i>	119
<b>References</b>	125

### **Chapter III – P2 Moiety in the Hepatitis C Virus Protease Inhibitors**

#### **is the Major Determinant in Drug Resistance and Modulation of Inhibitor**

<b><u>Potency by the Helicase Domain</u></b>	133
<b>Author Contributions</b>	134
<b>Abstract</b>	135
<b>Introduction</b>	136
<b>Materials and Methods</b>	141
<i>Synthesis of Inhibitors</i>	141
<i>Expression and Purification of NS3/4A Protease Constructs</i>	141
<i>Protease Cleavage Assays</i>	143
<i>Enzymatic Inhibition Assays</i>	144
<b>Results</b>	145
<i>Protease Inhibitors Used in this Study</i>	145
<i>Macrocyclization Status is the Major Determinant in Inhibitor</i>	
<i>Potency</i>	145
<i>Identity of the P2 Moiety is the Major Determinant in Drug</i>	
<i>Resistance</i>	146
<i>P2 Moiety is also the Major Determinant in Modulation of</i>	

<i>Inhibitor Potency by the Helicase Domain</i>	149
<b>Discussion</b>	157
<b>Acknowledgements</b>	161
<b>Supplementary Methods</b>	162
<i>Inhibitor Design and Synthesis</i>	162
<b>References</b>	165
<b><u>Chapter IV – Discussion</u></b>	171
<b>Discussion</b>	172
<i>Effect of the Domain-Domain Interface in Interdomain</i>	
<i>Dependence and Function</i>	172
<i>Effect of the Protease Domain in Helicase RNA Preference</i>	176
<i>Domain-Domain Arrangement in Vivo</i>	178
<i>Effect of Helicase Domain in Protease Inhibition</i>	180
<b>Conclusion</b>	183
<b>References</b>	184

**List of Tables**

<b>Table I.1</b>	Primary cleavage sites for NS3/4A protease for genotype 1a.	19
<b>Table II.1</b>	Parameters for correlation distributions for <i>in silico</i> NS3 constructs and visualization cutoffs.	85
<b>Table II.2</b>	Catalytic efficiencies of RET-S1 cleavage for different NS3 protease constructs.	97
<b>Table II.3</b>	Specific activities of different NS3 helicase constructs in DNA and RNA unwinding.	101
<b>Table II.4</b>	ssDNA binding constants, RNA-stimulated ATP hydrolysis catalytic rate constants and the extent of RNA stimulation of ATP hydrolysis for different NS3 helicase constructs.	105

**List of Figures**

<b>Figure I.1</b>	Prevalence of different genotypes of HCV around the globe and variance across genotypes and subtypes.	3
<b>Figure I.2</b>	Structure of HCV.	7
<b>Figure I.3</b>	HCV life cycle and entry of HCV into host cells.	9
<b>Figure I.4</b>	HCV genome, translation and polyprotein processing.	11
<b>Figure I.5</b>	Structure of single chain HCV NS3/4A protease – helicase complex and individual domains.	21
<b>Figure I.6</b>	Domain-domain interface in scNS3/4A.	27
<b>Figure I.7</b>	Comparison of single chain NS3 protease – helicase structures from HCV and DV.	31
<b>Figure I.8</b>	Model of membrane integration for full length HCV NS3/4A.	33
<b>Figure I.9</b>	Inhibitors of HCV NS3/4A protease.	37
<b>Figure II.1</b>	Protease cleavage assay.	69
<b>Figure II.2</b>	DNA and RNA unwinding assays.	73
<b>Figure II.3</b>	Structure of scNS3/4A and the domain-domain interface.	79
<b>Figure II.4</b>	Correlation maps for <i>in silico</i> deletion constructs and distributions of total correlations.	83
<b>Figure II.5</b>	Interdomain correlations for <i>in silico</i> deletion constructs.	87

<b>Figure II.6</b>	Distance distributions around the interfaces.	91
<b>Figure II.7</b>	Proteolysis catalytic efficiencies for interface mutants.	95
<b>Figure II.8</b>	Kinetic parameters for the multiple activities of the NS3hel constructs.	99
<b>Figure II.9</b>	A model for NS3/4A function consistent with <i>in vitro</i> results.	115
<b>Figure II.10</b>	Stability of individual trajectories.	121
<b>Figure III.1</b>	Inhibitors used in this study.	139
<b>Figure III.2</b>	Potencies of protease inhibitors used in this study.	147
<b>Figure III.3</b>	Drug resistance profiles of the inhibitors for multi drug resistant variants.	151
<b>Figure III.4</b>	The modulation of drug activities by the helicase domain in multi drug resistance mutants.	153
<b>Figure III.5</b>	The effect of helicase mutations on inhibitor potency.	155



### List of Third Party Copyrighted Material

The following figures were reproduced with permission.

<b>Figure</b>	<b>Publisher</b>	<b>License Number</b>
<b>Figure I.3a</b>	Nature Publishing Group	2947740948043
<b>Figure I.3b</b>	Nature Publishing Group	2947740948043
<b>Figure I.4</b>	Nature Publishing Group	2947740948043
<b>Figure I.8</b>	National Academy of Sciences	N/A

The following figures were reproduced. No permission required.

<b>Figure</b>	<b>Source</b>
<b>Figure I.1a</b>	World Health Organization ( <a href="http://www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html">http://www.who.int/vaccine_research/diseases/viral_cancers/en/index2.html</a> )
<b>Figure I.2</b>	Toronto Western Hospital Francis Family Liver Clinic ( <a href="http://www.torontoliver.ca/content/hepatitisc.html">http://www.torontoliver.ca/content/hepatitisc.html</a> )

**List of Abbreviations**

<b>AIDS</b>	Acquired Immuno Deficiency Syndrome
<b>ARFP</b>	Alternate reading frame protein
<b>BSA</b>	Bovine serum albumin
<b>C</b>	Core protein
<b>CD81</b>	Tetraspanin
<b>Cy5</b>	Cyanine 5
<b>DAA</b>	Directly acting antiviral
<b>DMSO</b>	Dimethyl sulphoxide
<b>DTT</b>	Dithiothreitol
<b>DV</b>	Dengue Virus
<b>E1</b>	Envelope protein 1
<b>E2</b>	Envelope protein 2
<b>FRET</b>	Förster resonance energy transfer
<b>GAG</b>	Glycosaminoglycan
<b>HCV</b>	Hepatitis C Virus
<b>HEPES</b>	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
<b>HIV</b>	Human Immunodeficiency Virus
<b>IBRQ</b>	Iowa Black RQ
<b>IPTG</b>	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<b>IRES</b>	Internal ribosome entry site

<b>LDAO</b>	Lauryldimethylamine-oxide
<b>LDL</b>	Low-density lipoprotein
<b>MAVS</b>	Mitochondrial antiviral signaling protein
<b>MDR</b>	Multi drug resistant
<b>MES</b>	2-(N-morpholino)ethanesulfonic acid
<b>MOPS</b>	3-(N-morpholino)propanesulfonic acid
<b>ORF</b>	Open reading frame
<b>O<math>\beta</math>G</b>	n-octyl-beta-glucopyranoside
<b>PEG-IF</b>	Pegylated interferon
<b>RBV</b>	Ribavirin
<b>RIG-I</b>	Retinoic acid-inducible gene I
<b>RMSD</b>	Root-mean-square deviation
<b>SAR</b>	Structure-activity relationship
<b>SR-BI</b>	Scavenger Receptor Class B Type I
<b>SVR</b>	Sustained viral response
<b>TCEP</b>	Tris(2-carboxyethyl)phosphine
<b>TRIF</b>	Toll interleukin-1 receptor domain-containing adapter-inducing interferon- $\beta$
<b>Tris</b>	Tris(hydroxymethyl)aminomethane
<b>TRL3</b>	Toll-like receptor 3
<b>UTR</b>	Untranslated region
<b><math>\beta</math>ME</b>	Beta-mercapto ethanol

**List of Prefixes and Suffixes**

<b>d-</b>	Deoxyribo
<b>ds-</b>	Double stranded
<b>NS-</b>	Non-structural protein
<b>P-</b>	Product position
<b>r-</b>	Ribo
<b>sc-</b>	Single chain
<b>SD-</b>	Subdomain
<b>ss-</b>	Single stranded
<b>Δ-</b>	Deletion
<b>-hel</b>	Helicase
<b>-pro</b>	Protease
<b>-SD</b>	Subdomain

## **Preface**

Chapter II will be submitted for publication as “Destabilization of Protease – Helicase Interface Enhances Both Protease and Helicase Activities of Hepatitis C Virus Non-Structural Protein NS3-NS4A Complex” by “Cihan Aydin (first author), Sourav Mukherjee, Alicia M. Hanson, David N. Frick and Celia Schiffer”.

A modified version of Chapter III will be submitted for publication as “Drug Resistance against Hepatitis C Virus NS3/4A Protease Inhibitors is Mainly Caused by the P2 Groups” by “Akbar Ali, Cihan Aydin (co-first author) and Celia A. Schiffer”. Part of the data presented in Chapter III was published in the paper “Romano, K. P., A. Ali, C. Aydin, et al. (2012). "The Molecular Basis of Drug Resistance against Hepatitis C Virus NS3/4A Protease Inhibitors." *PLoS Pathog* **8**(7): e1002832.”, which is available online as of July 27<sup>th</sup>, 2012.

**CHAPTER I**

**INTRODUCTION**

## **Epidemiology of Hepatitis C Virus**

Hepatitis C Virus (HCV) was initially discovered in 1989 as the causative agent of non-A non-B viral hepatitis [1]. Approximately 150 million people around the globe are chronically infected with HCV with 3-4 million new infections happening annually, and an annual death toll of 350,000 people [2]. HCV is primarily transmitted via exposure to blood of an infected individual and, to a lesser extent, through sexual contact. High risk individuals for HCV infection include recipients of blood products or organ transplants prior to discovery of HCV and HCV detection systems, intravenous drug users, long-term hemodialysis patients, healthcare workers and individuals who are infected with Human Immunodeficiency Virus (HIV) [3]. Acute HCV infection is often asymptomatic or produces mild symptoms such as abdominal pain and flu-like symptoms, thus Hepatitis C is usually not diagnosed in acute phase [4, 5]. HCV infection can be cleared by host immune system, however, majority of acute HCV infections progress into chronic viral hepatitis. Around 20% of chronically infected patients develop cirrhosis within two decades, which further progresses into hepatocellular carcinoma and liver failure [6]. Chronic HCV infection is the major cause of liver transplantation in the United States [7, 8].

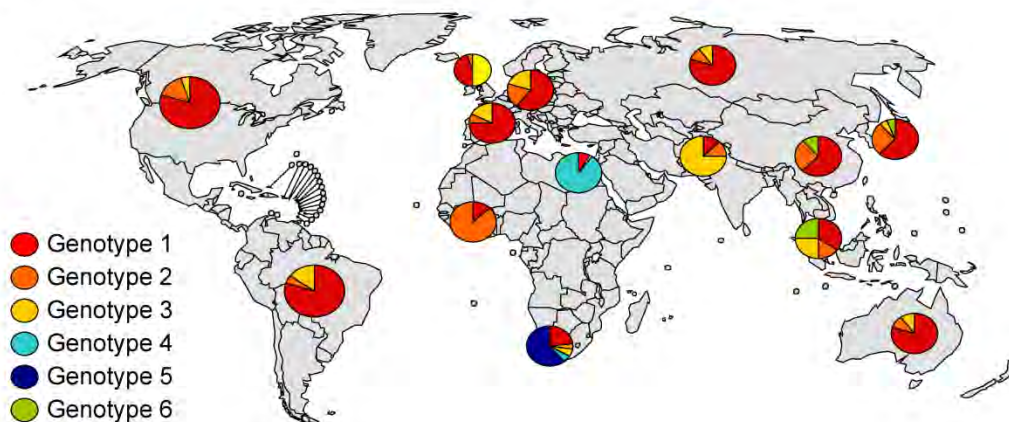
HCV is genotypically diverse – six genotypes were classified (1-6) and subcategorized further into different subtypes (a, b, c, etc.) [9]. Depending on geography, distinct patterns of genotypical distribution is present (Figure I.1a). Genotype 1 is the major genotype in the world, especially in continental America and Europe. Genotype 3, the second predominant genotype, is mostly observed in Southeastern Asia and countries

**Figure I.1 – Prevalence of different genotypes of HCV around the globe and variance across genotypes and subtypes.** (a) Genotypical distribution of HCV exhibits distinct patterns with respect to geographical location. Genotype 1 is the most dominant across the world, with the others following (Copyright 2009, World Health Organization)

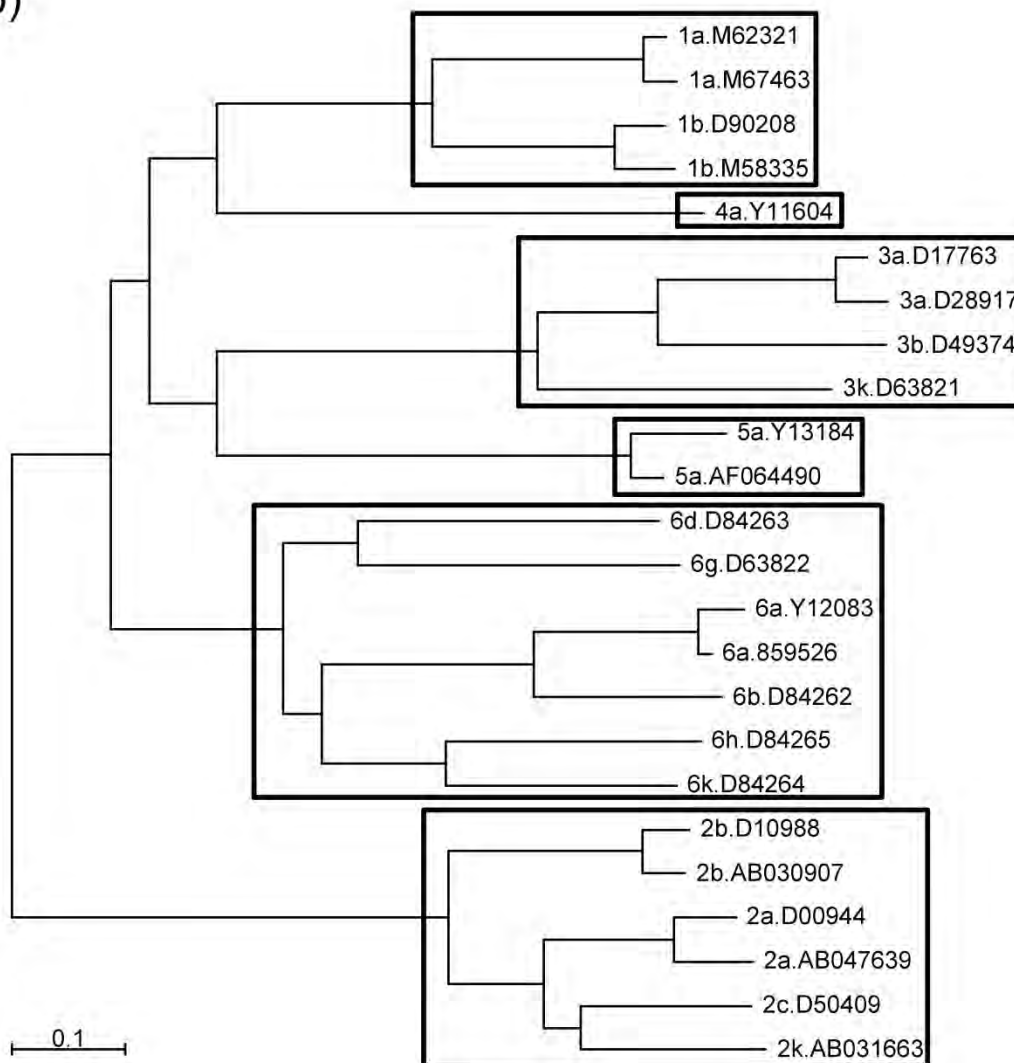
(b) Confirmed genotypes and subtypes of HCV are genetically diverse. Genomic sequences for confirmed HCV subtypes were obtained from GenBank [10] using Simmonds *et al* classification [9]. Alignment of sequences and tree generation was performed in SeaView [11] using MUSCLE algorithm [12] for sequence alignment and PhyML [13] for generating phylogenetic trees. Texts next to the branches denote genotype, subtype and accession numbers for confirmed genomic sequences. Scale bar represents the average number of substitutions per site.



(a)



(b)



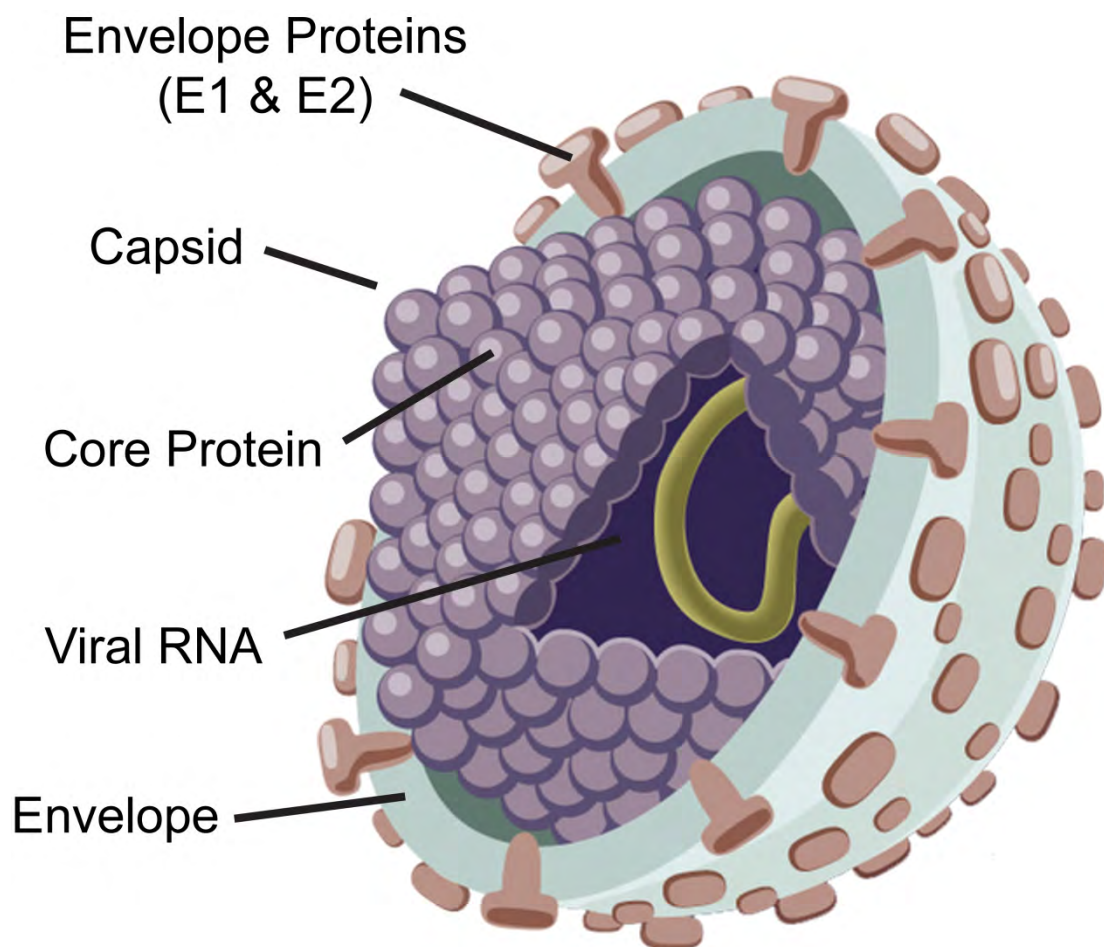
like Russia, England and Australia. Other genotypes are relatively minor, but distinct geographical patterns of spread has been observed – genotype 2 is more dominant in sub-saharan Africa, genotype 4 in Middle East, genotype 5 in South Africa and genotype 6 in Asia [9, 14]. These genotypes and subtypes are distinct both genetically and epidemiologically, relative nucleotide differences are around 30% at genotype level and around 25% at subtype level [14].

## Hepatitis C Virus Life Cycle

HCV is a member of the family *Flaviviridae* genus *Hepacivirus*, enveloped, single stranded positive-sense RNA viruses. The genome of HCV is encased in an icosahedral capsid formed by the Core proteins and further enveloped in host-derived membrane, embedded with viral envelope glycoproteins E1 and E2 (Figure I.2). The life cycle of HCV consists of entry, uncoating, translation, replication, packaging and exit phases (Figure I.3a). HCV circulates in the host organism in various ways – either in association with low density lipoproteins (LDLs), immunoglobulins or as a free virus particle [15]. Association of the envelope protein E2 with Tetraspanin (CD81) and Scavenger Receptor Class B Type I (SR-BI), in combination with LDL receptors (LDL-R), glycosaminoglycans (GAG), Claudin-I and Occludin, facilitate fusion with human hepatocytes [16-23]. The virion is then taken into the cell via clathrin-mediated endocytosis [24]. Acidification of the endosomes facilitates fusion with endosomal membranes and subsequent release of HCV RNA into cytoplasm [25, 26] (Figure I.3b).

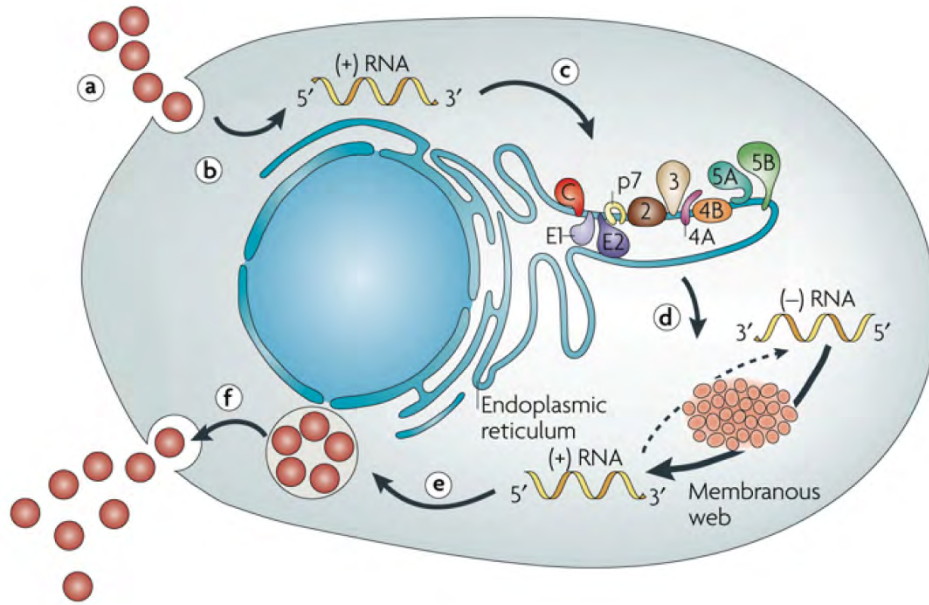
The 9.6 kb genome of HCV encodes an open reading frame (ORF) flanked by 5' and 3' highly structured untranslated regions (UTRs) (Figure I.4a). The 5' UTR contains the internal ribosome entry site (IRES) which is essential for the translation of the viral genome [27]. Additionally, composition of the 3' UTR is important for viral propagation *in vivo* [28]. HCV positive sense RNA is translated by the host replication machinery in a cap-independent manner, producing a single polyprotein composed of structural proteins (Core, E1, E2), p7 and non-structural proteins (NS2, NS3, NS4A, NS4B, NS5A, NS5B),

**Figure I.2 – Structure of HCV.** RNA genome of the HCV is encased in an icosahedral capsid formed by core proteins and further enveloped in host-derived membrane embedded with envelope glycoproteins E1 and E2 (Toronto Western Hospital Liver Centre).

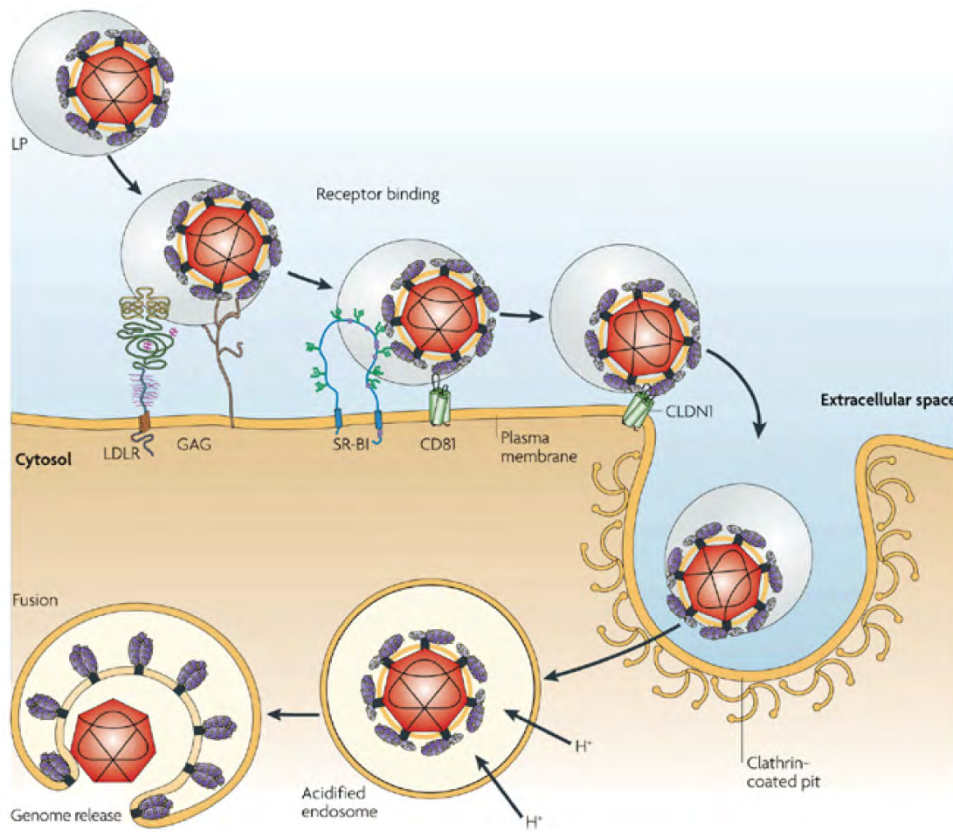


**Figure I.3 – HCV life cycle and entry of HCV into host cells.** (a) Life cycle of HCV starts with the entry of viral particles into host cells and subsequent release of viral positive sense RNA, which is then translated into structural and non-structural proteins in endoplasmic reticulum. Processing of the viral polyprotein is performed by host and viral proteases. At some point, negative sense RNA is generated, which will serve as a template for positive sense RNA that is packed inside mature virions. Replication of viral genome commences in specialized structures called "membranous web". Replicated viral RNA is then packaged into virus particles and exocytosed (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology, Moradpour, D., F. Penin, et al. (2007). "Replication of hepatitis C virus." *Nat Rev Microbiol* **5**(6): 453-463., Copyright 2007). (b) HCV possibly associates with LDLs while travelling inside the host organism. LDL receptors and GAGs are identified as early factors in HCV entry into the cells. Additional factors such as SR-BI and CD81 is required for tethering viral particles to the cell membrane. Claudin I and Occludin are late entry factors that possibly translocate viral particles to the tight junctions of polarized hepatocytes. HCV is then endocytosed in clathrin-coated vesicles. Acidification of endosomes results in the uncoating of viral genome and injection of virus RNA into host cell cytoplasm (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology, Moradpour, D., F. Penin, et al. (2007). "Replication of hepatitis C virus." *Nat Rev Microbiol* **5**(6): 453-463., Copyright 2007).

(a)

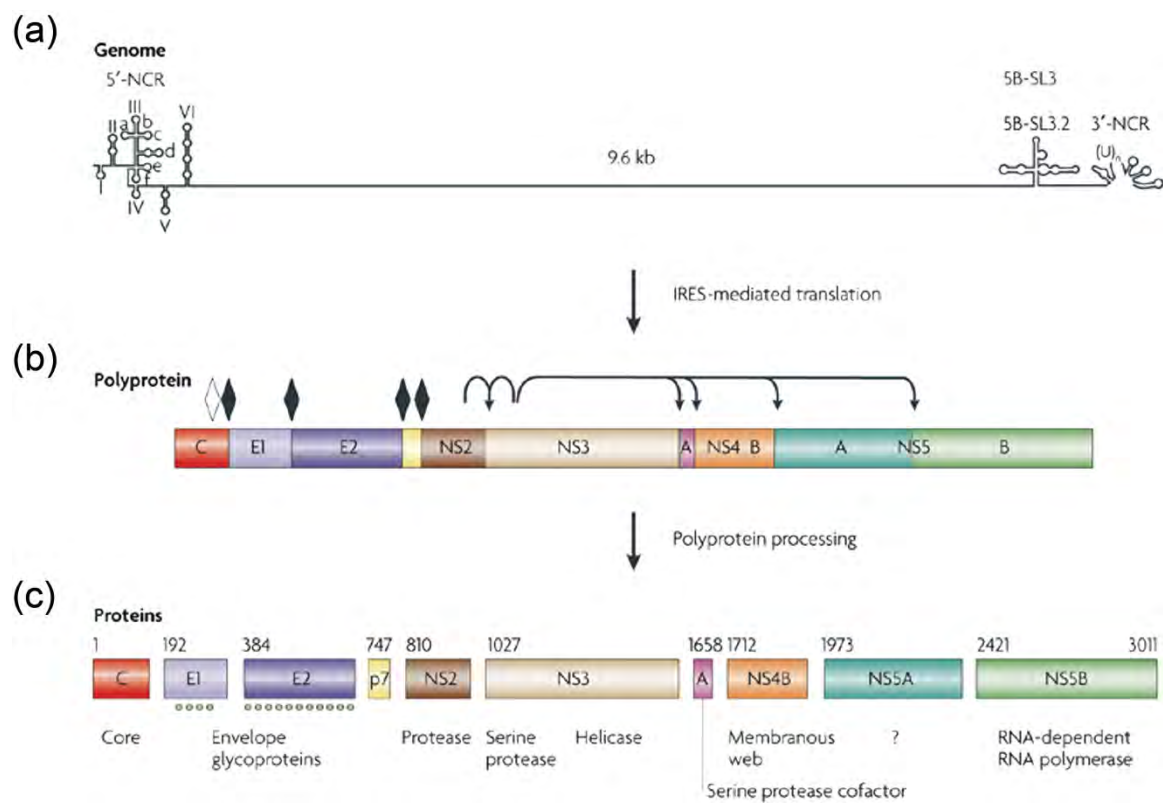


(b)



**Figure I.4 – HCV genome, translation and polyprotein processing.** (a) The 9.6 kb genome contains 5' IRES and 3' structured motifs which are conserved. (b) HCV genome is translated into a polyprotein which is processed co- and post-translationally. Structural proteins and p7 is processed by ER signal peptidase (solid diamond) and the core protein is further matured by signal peptide peptidase (open diamond). Non-structural proteins are processed by two viral proteases – NS2/3 cysteine protease cleaves the NS2-NS3 junction and NS3/4A serine protease cleaves NS3-NS4A junction in *cis* and others in *trans*. (c) Processing of viral proteins result in 3 structural proteins (C, E1, E2), p7 viroporin protein and 5 non-structural proteins (NS2, NS3/4A, NS4B, NS5A and NS5B) (Reprinted by permission from Macmillan Publishers Ltd: Nature Reviews Microbiology, Moradpour, D., F. Penin, et al. (2007). "Replication of hepatitis C virus." Nat Rev Microbiol 5(6): 453-463., Copyright 2007).





Structural proteins and p7 are liberated from the viral polyprotein by host cell signalases and host signal peptide peptidases inside ER lumen [29-31]. Subsequent cleavage of non-structural proteins are performed by viral proteases denoted as NS2/3 [32] and NS3/4A [33] (Figure I.4c).

Both structural and non-structural proteins are critical for HCV propagation. Most of these proteins perform multiple functions required for viral propagation, host immune suppression and pathogenesis. Core and envelope proteins are an integral part of the virus particle, forming the viral capsid, facilitating virus – host cell interaction and infection. Core protein forms the viral capsid that binds the viral RNA [34, 35] and also alters the host cell lipid mechanism, which is implied in steatosis [36]. Highly glycosylated envelope proteins E1 and E2 form a heterodimeric complex that facilitates host – virus interactions [37, 38]. The p7, a member of viroporins, has been shown to form oligomeric ion channels *in vitro* and to be critical for infectivity *in vivo* [39-41]. Non-structural proteins form the viral replication machinery and are also responsible for packaging of infectious virus particles. NS2 is implicated in virus assembly and polyprotein processing [42, 43]. In combination with N-terminal domain of NS3 protease, NS2/3 cysteine protease cleaves the NS2-NS3 junction [32]. NS3/4A is a multifunctional protease – helicase complex that processes the remaining viral polyprotein, aids in replication by unwinding RNA in ATP-dependent fashion and suppresses immune response [44-47]. NS4A cofactor, an integral part of the NS3/4A protease, also anchors NS3/4A to the membrane and aids in the processivity and translocation efficiency of NS3 helicase [48-51]. NS4B is implicated in the formation of "membranous webs" where viral replication

happens [52, 53]. NS5A is an essential phosphoprotein that binds to RNA and implicated in replication, virus assembly and regulation of host processes [54]. NS5B is the RNA-dependent RNA polymerase that replicates the viral genome [55-57]. In addition, an alternate reading frame protein (ARFP) overlapping with the core protein on the +1 frame was also discovered, but a specific function for this protein has not been attributed [58, 59].

Viral replication commences in a region of ER where the membrane is altered to form vesicular structures, called "membranous web", which contains non-structural proteins and viral RNA in production [60, 61]. Replicated viral RNA is then packaged into viral particles and exocytosed through the secretory pathway. However, the specific mechanisms associated with viral replication, packaging and release have not been fully explored.

### **Treatment of Hepatitis C and Anti Hepatitis C Drug Discovery Effort**

Hepatitis C, like Acquired Immuno Deficiency Syndrome (AIDS), is a major health problem across the globe; but, unlike AIDS, HCV infection can be cured in majority of cases. Until mid-2011, the treatment regime against Hepatitis C involved pegylated interferon- $\alpha$  (PEG-IFN) and ribavirin (RBV) [62, 63]. Interferons are cytokines that are produced by the body in response to viral infection and activate host immune system [64-66]. Pegylation increases the half life of interferons, so PEG-IFN is stable in bloodstream for extended periods of time [67, 68]. Ribavirin is a nucleotide analog, similar to ribo-purines, that gets incorporated into the viral genome during replication and induces hypermutations that are harmful for viral viability [69]. Duration and effectiveness of this dual therapy heavily depends on host and viral factors [70, 71]. Sustained viral responses (SVRs) were observed for 50% of the patients who exhibit genotype 1 virus and up to 90% for genotypes 2 and 3 [72, 73]. However, PEG-IFN – RBV combination therapy is costly, lengthy (up to 72 weeks) and side effects such as anemia, flu-like symptoms and depression is frequently observed in patients undergoing therapy [74, 75]. Thus, pharmaceutical industry is more focused on creating directly acting antiviral (DAA) agents against crucial viral machinery and, potentially, vaccines to create specific drugs to combat Hepatitis C Virus.

Non-structural proteins constitute critical machinery essential for viral propagation, hence potential DAA development targets. Drug candidates against non-structural proteins are being developed actively – especially for NS3/4A protease (reviewed below in depth in “*Inhibitors of Non-Structural Protein 3/4A Protease*”),

NS5A (e.g. BMS-790052 (Bristol-Myers Squibb) [76]), nucleoside inhibitors (e.g. RG-7128 (Roche/Pharmasset) [77], PSI-7977 (Pharmasset) [78]) and non-nucleoside inhibitors (e.g. filibuvir (Pfizer) [79], BI-207127 (Boehringer Ingelheim) [80], GS-9190 (Gilead) [81]) for NS5B [82]. Two protease inhibitors have recently been approved by the Food and Drug Administration (FDA) and are currently being utilized in triple combination therapies including PEG-IFN, RBV and either telaprevir (Vertex Pharmaceuticals – Incivek®) or bocepravir (Merck – Victrelis®) [83, 84]. Up to 20% improvement in SVR rates for treatment-naïve and non-responsive patients have been observed in the triple therapy regime over the old PEG-IFN/RBV regime [85-87]. Additional inhibitors along the development pipeline exhibit good activities against respective targets and are progressing in various stages of clinical trials. Eventually, near-term goals for drug development is to combine new generation DAAs with the current triple combination PEG-IFN/RBV/PI therapy to improve SVR rates and in long-term, use highly effective DAAs in mono- or combinatory therapies without PEG-IFN/RBV.

### **Non-Structural Protein 3/4A – "Swiss Army Knife" of Hepatitis C Virus**

NS3/4A, nicknamed the "Swiss Army Knife" of HCV by Kenichi Morikawa and colleagues [88], is a 685 residue multi functional protein formed by a serine protease domain and an RNA helicase domain connected by a flexible linker, with the residues 12-25 of NS4A integrated into the protease non-covalently (Figure I.5a). Protease domain plays a critical role in viral protein processing (Figure I.4) and overcoming host immune response. Helicase domain unwinds dsDNA and dsRNA in ATP-dependent manner *in vitro* and likely assists in the replication of the viral genome. NS4A serves as a cofactor for both domains and anchors the protein to membranes. Both unwinding and proteolysis functionalities are essential in a mutually exclusive manner; hence NS3/4A is a prime target for drug development.

#### *Protease Domain*

HCV NS3 protease comprise the N-terminal 181 residues of the protein [44] (Figure I.5b). Protease domain folds into two  $\beta$ -barrels, a common motif observed in chymotrypsin-like serine proteases. The distorted N-terminal  $\beta$ -fold is stabilized by the contributing  $\beta$ -strand from the cofactor NS4A. The active site is formed by three residues in the cleft between the two subdomains, H57 and D81 from the N-terminal subdomain, and S139 from the C-terminal subdomain. Crystal structures of the protease domain in isolation show that protease active site is shallow and solvent exposed [48].

NS3 protease is necessary for cleaving NS3-4A, NS4A-4B, NS4B-5A and NS5A-5B junctions, the first being in *cis* and the remaining in *trans* [33, 89, 90]. In addition, HCV protease circumvents host immune response by cleaving Toll interleukin-1 receptor

domain-containing adapter-inducing interferon- $\beta$  (TRIF) and mitochondrial antiviral signaling protein (MAVS), adapter proteins for Toll-like receptor 3 (TLR3) and Retinoic acid-inducible gene I (RIG-I), hence disrupting viral RNA sensors in host cells [91-94]. Viral and host cleavage sites show little sequence homology – only P6, P1 and P1' residues show a distinct trend (D/E-X-X-X-X-C/T||S/A/H-X-X-X) [44, 95] (Table I.1). However, even with limited sequence complementarity, all protease substrates occupy the protease active site in a distinct, conserved fashion [95, 96]. The common volume occupied by bound substrates, called the "substrate envelope", has been demonstrated and extensively studied for HIV-1 and HCV NS3/4A proteases [95, 97].

#### *Helicase Domain*

HCV NS3 helicase is a member of the DEXH/D-box family of RNA helicases from helicase superfamily II, which are characterized by at least twelve conserved motifs [98] (Figure I.5c). Most prominent of these motifs are the Walker A motif (G-X-X-X-X-G-K-T/S), which interacts with the  $\beta$ - and  $\gamma$ - phosphates of NTPs and the Walker B motif (DEAH/D-box), which chelates divalent cations [99-101]. Similar to other DEAD-box RNA helicases, the helicase core consists of two conserved RecA-like subdomains 1 and 2 [100, 102]. In addition, a C-terminal accessory domain with high  $\alpha$ -helical content is also present for *Flaviviridae* NS3 helicases, possibly conferring translocation polarity [103].

NS3 helicase exhibits both RNA and DNA unwinding activity *in vitro* [45, 104]. Helicase domain associates with the single stranded region of duplexes, accommodating the nucleic acid in the cleft formed between subdomains 1-2 and 3 [105, 106]. Bound

**Table I.1 – Primary cleavage sites for NS3/4A protease for genotype 1a.**

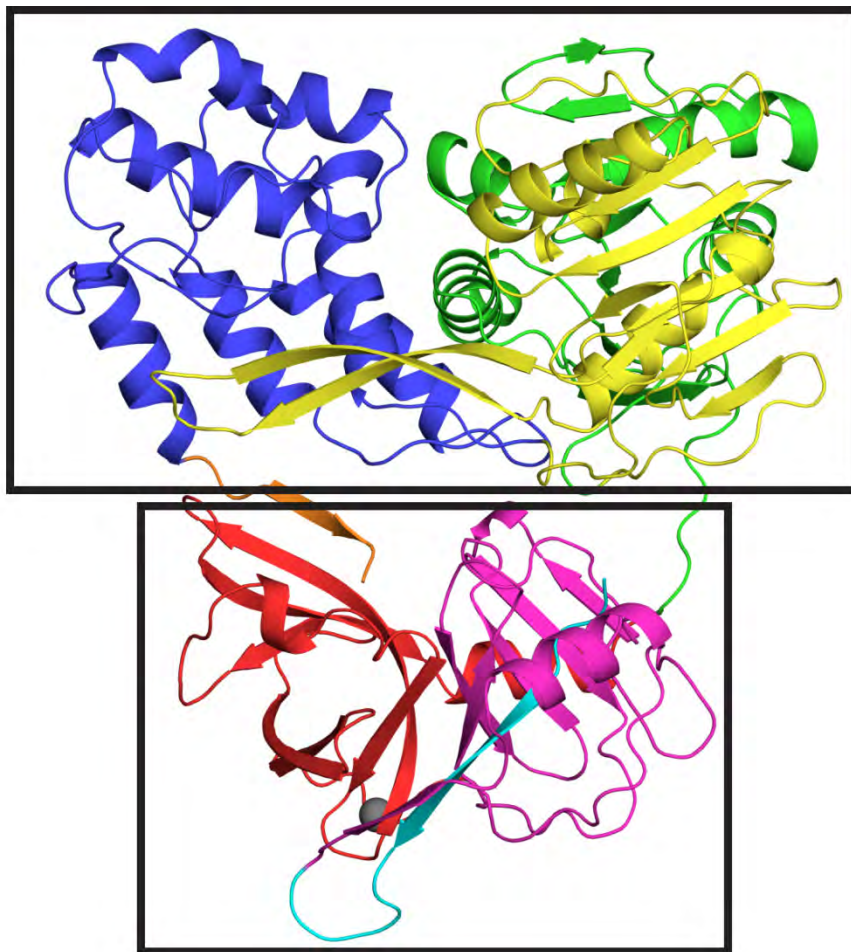
Amino acid sequence of polyprotein processing and immune suppression sites cleaved by NS3/4A. Cleavage sequences are diverse except an acidic residue at P6, a cysteine at P1 and a serine at P1' positions. Similar amino acids are highlighted.



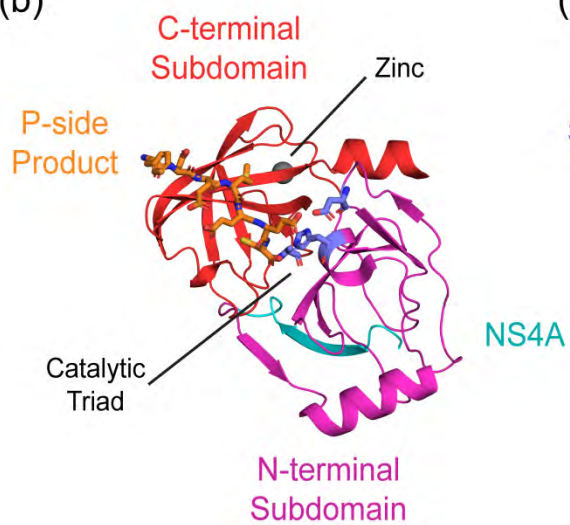
<b>Substrate</b>	<b>P6</b>	<b>P5</b>	<b>P4</b>	<b>P3</b>	<b>P2</b>	<b>P1</b>	<b>P1'</b>	<b>P2'</b>	<b>P3'</b>	<b>P4'</b>
<b>3-4A</b>	D	L	E	V	V	T	S	T	W	V
<b>4A-4B</b>	D	E	M	E	E	C	S	Q	H	L
<b>4B-5A</b>	E	C	T	T	P	C	S	G	S	W
<b>5A-5B</b>	E	D	V	V	C	C	S	M	S	Y
<b>TRIF</b>	P(8)	S	S	T	P	C	S	A	H	L
<b>MAVS</b>	E	R	E	V	P	C	H	R	P	S

**Figure I.5 – Structure of single chain HCV NS3/4A protease – helicase complex and individual domains.** (a) Crystal structure of scNS3/4A protease (lower box) – helicase (upper box) complex (PDB-ID: 1CU1) in apo form, where NS4A protease cofactor is covalently linked to the N-terminus of the protein, solved by Yao *et al* [107]. In this structure, helicase domain covers otherwise shallow protease active site. C-terminus of the protein, the P-side product of NS3-NS4A junction, occupies protease active site. (b) Crystal structure of scNS3/4A protease domain bound to NS4A-NS4B P-side product (PDB-ID: 3M5M) [95]. Protease is composed of two subdomains with  $\beta$ -barrel folds. NS4A cofactor integrates into the N-terminal subdomain  $\beta$ -barrel fold, stabilizing the protease structure. A zinc ion is chelated by three cysteines in C-terminal subdomain and is structural. Catalytic triad lies between two subdomains and the active side is fairly shallow and solvent exposed. (c) Crystal structure of NS3 helicase domain bound to ssDNA (PDB-ID: 3KQK) [108]. NS3 helicase is formed by two recA-like subdomains 1 and 2, and an  $\alpha$ -helical subdomain 3. DNA binding occurs in the groove formed between subdomains 1-2 and 3, and ATP hydrolysis occurs in the groove between subdomains 1 and 2.

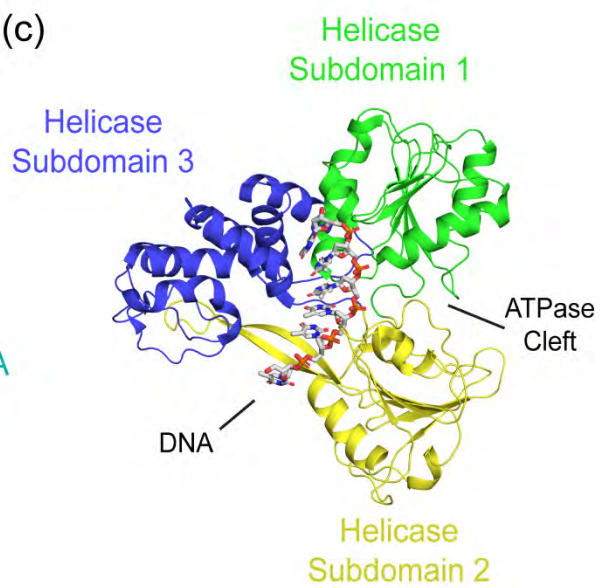
(a)



(b)



(c)



nucleic acid translocates from 3' to 5' direction in ATP-dependent manner [109]. This translocation activity is likely the driving force behind unwinding and removal of nucleic acid bound proteins [109]. Exact mechanism of unwinding is still unknown, but several mechanisms that accounts for biochemical and structural observations have been proposed [98, 110]. NS3 helicase can also form oligomers, NS3 helicase was demonstrated to unwind nucleic acid substrates in monomeric, dimeric or oligomeric forms [111-115]. Unwinding activity of NS3 helicase exhibits cooperative behavior depending on substrate concentration, which may indicate productive association between individual helicase domains [113]. However, whether observed oligomeric form is relevant *in vivo* has not been elucidated.

#### *NS4A*

NS4A consists of three regions that are critical in HCV function – N-terminal transmembrane helix (amino acids 1-20), hydrophobic  $\beta$ -strand (amino acids 21-32) and C-terminal acidic region (amino acids 40-54). N-terminal transmembrane helix anchors NS3/4A to membranes [116]. Hydrophobic  $\beta$ -strand is integrated into the  $\beta$ -barrel fold of the protease domain, thus stabilizing protease fold [48, 89]. Without NS4A integration, protease activity is reduced significantly. C-terminal acidic region is implicated in helicase function, which will be explained in the next section. Additionally, cleavage between NS3 and NS4A happen in the unimolecular NS3-NS4A complex in *cis* fashion [33, 90]. P1-site in the NS3-NS4A cleavage junction is a threonine unlike other cleavage sites, which, in combination with *cis* cleavage, is likely to be important in post-translational integration of NS4A inside NS3 [117].

### **Interdomain Dependency in Non-Structural Protein 3/4A**

Although NS3/4A incorporates two domains with distinct functionalities into one single chain protein complex, most of the studies performed on NS3/4A involve truncated domains isolated independently. Isolated domains are easier to overexpress and purify and both domains are active in isolation [33, 44, 46, 47, 118]. In addition, structural discrepancies between individual domains and the full length complex is minimal [107]. Isolated domains have been utilized extensively in biochemical and structural characterization, and drug discovery. However, studies performed on protease – helicase constructs exhibit possible interplay between the two domains [50, 114, 115, 119, 120].

Effect of domain-domain interdependence on protease function remains a controversial topic. On one side, comparable activities between isolated protease and full length protein were demonstrated [45, 121]. On the opposing side, significant enhancement in protease activity [120] and alterations in protease inhibitor potencies [122, 123] were observed when the helicase domain was present. Direct comparisons between these studies are difficult due to differences in assay conditions and constructs. Three different protease constructs were utilized. The first construct involves expression of NS3 protease (or NS3 protease – helicase complex) without NS4A, where the NS4A protease cofactor peptide is added externally in saturating amounts [124, 125]. In the second construct, NS4A protease cofactor region is ligated to the N-terminus of NS3 protease (or NS3 protease – helicase complex) with a flexible linker (usually GSGS), forming single chain NS3/4A (scNS3/4A) [126, 127]. The third construct is the closest to

the actual NS3/4A protein *in vivo*, since NS3 and non-truncated NS4A is expressed as a single protein, which is autoprocessed to obtain full length NS3/4A [118, 119]. scNS3/4A constructs are popular in the field, since they overexpress relatively well and can be purified with better yield. Each construct has advantages and disadvantages, and the discrepancies between the constructs are possibly the reason behind the observed controversy.

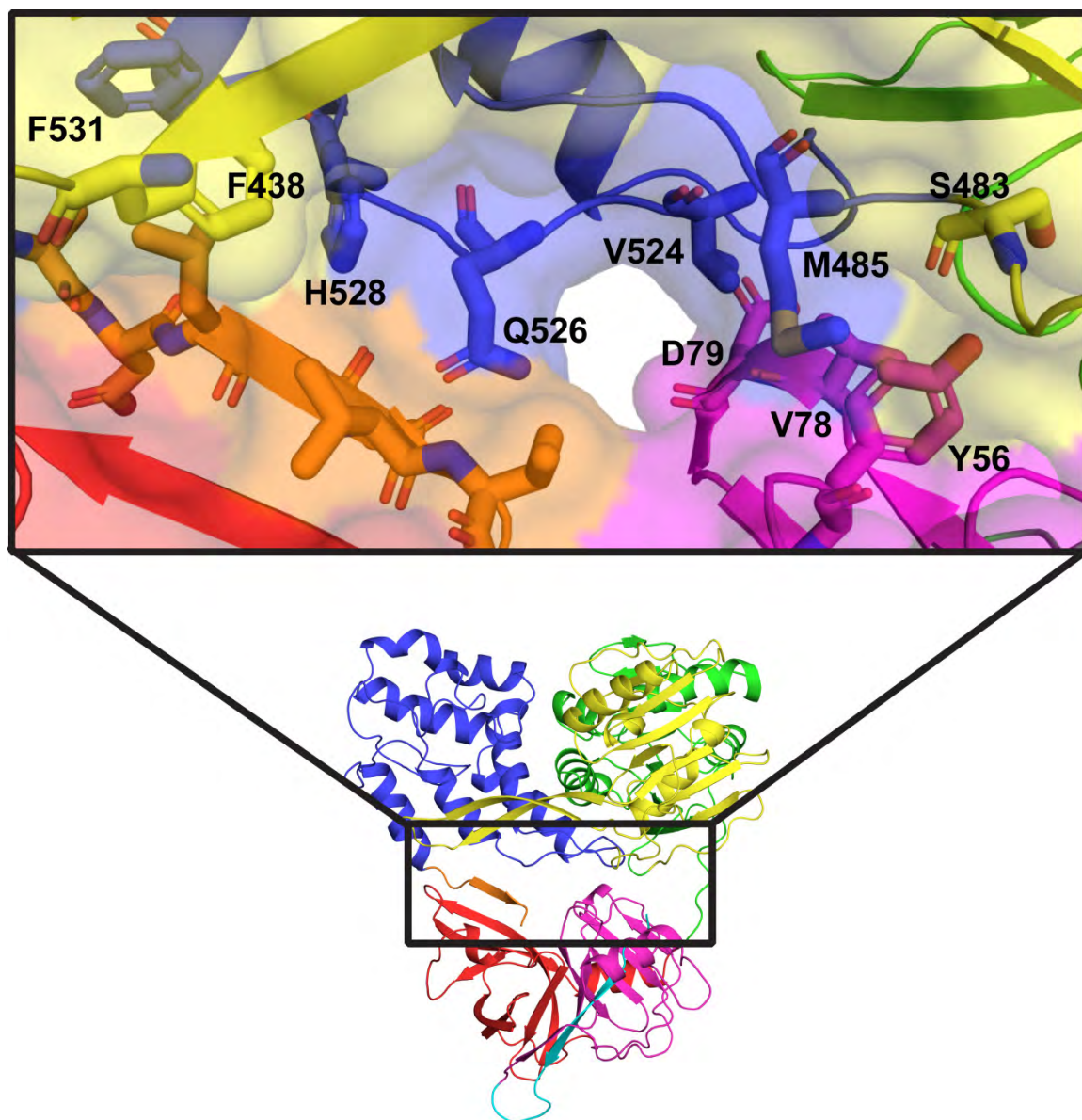
Influence of protease domain on helicase activity is better characterized and more consistent across different protease – helicase constructs. NS3 helicase exhibits both DNA and RNA unwinding activity *in vitro*. An initial study performed on the RNA unwinding activity of helicase domain exhibited comparable activities in the absence and presence of the protease domain [45]. However, enhancements in helicase activities were demonstrated later when protease domain was present [114, 119, 128, 129]. Protease domain specifically enhances RNA unwinding activity of the helicase domain – minimal modulations have been observed either for RNA binding or ATPase activities [114], a fact also observed in West Nile Virus NS3/2B [130]. This effect is more pronounced in full length NS3/4A construct [119], meaning that in addition to a correctly folded active protease domain, NS4A may serve a further purpose in helicase function. Recently, C-terminal acidic region of the NS4A protein was shown to be important in rate of ATP hydrolysis and ATP-dependent RNA binding, both for HCV [50] and *Flavivirus* NS3 helicase [51, 131]. In summary, protease domain – NS4A complex may serve as a cofactor for NS3 helicase to achieve efficient and processive unwinding activity.

## **Structures of Single Chain NS3 Protease – Helicase Constructs from Hepatitis C Virus and Flaviviruses**

Extensive structural studies have been performed on the individual domains and deposited in databases. However, only three sets of structures are available for the full length NS3/4A. Crystallographic studies performed using single chain NS3/4A protein yielded information on the apo form [102], various stages of RNA translocation [109] and inhibitor binding in the presence of helicase domain [132]. In these structures, helicase domain covers the protease active site (Figure I.5a). In addition, the domains contact each other in two distinct regions, separated by the protease active site (Figure I.6). Nearly 900 Å<sup>2</sup> is buried in this domain-domain interface. The first contact region is formed between the N-terminal half of the protease domain and helicase subdomains 2 and 3. Residues Y56, V78 and D79 from protease domain, and S483, M485 and V524 from helicase domain form the core of this region, significant portion of these residues are buried in this 400 Å<sup>2</sup> interface. The second contact region is formed between the C-terminus of the protein, also the P-side product of NS3-NS4A *cis*-proteolysis reaction, and protease active site. P-side products of NS3/4A protease strongly associate with protease active site, exhibiting micromolar affinities [95, 125, 133]. Apart from the linker region, which connects the two domains covalently, this domain-domain interface is the only apparent contact region between the two domains. However, probably due to the small size and composition compared to biologically functional interfaces [134, 135], no investigation has been performed on this interface for functional relevance and possible role in domain-domain communication.

**Figure I.6 – Domain-domain interface in scNS3/4A.** 900 Å<sup>2</sup> interface between the domains is divided by the protease active site into two regions – direct interface (right) between helicase subdomains 2-3 (yellow and blue) and N-terminal subdomain of protease (magenta), and substrate interface (left) between the C-terminus (orange) and C-terminal subdomain of protease (red). Annotated residues were mutated to alanine for functional studies (detailed in Chapter II).



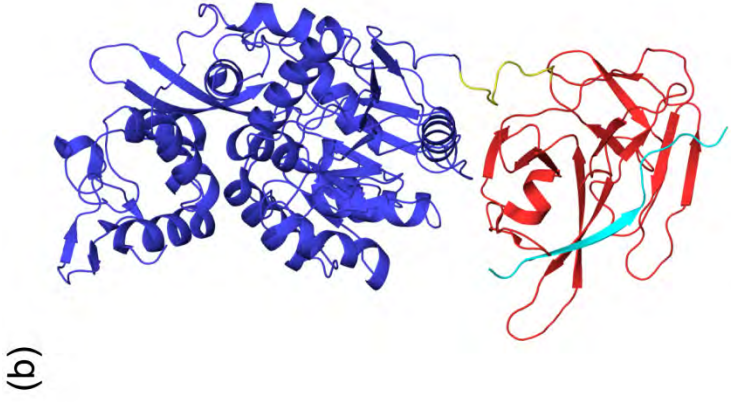


The domain-domain arrangement observed in HCV NS3 crystal structures differs significantly between various *Flaviviruses*. NS3/2B from Dengue Virus (DV), a member of the *Flaviviruses*, is similar to HCV NS3/4A both in composition and function [136]. In addition, helicase dependency of protease activity was also observed for full length DV NS3/2B constructs *in vitro* [130]. Two major differences is observed between *Flavivirus* and HCV NS3 complexes – NS2B being a cofactor instead of NS4A and the missing C-terminal helix in helicase domain, which renders direct contact between C-terminus and protease active site unfavorable. Several crystal and solution structures of single chain NS3/2B constructs are available [137-139] (Figure I.7). In these structures, the protease is rotated away from the helicase domain, resulting in an extended conformation in two alternative arrangements. A smaller interface compared to HCV NS3/4A was formed between protease domain and helicase subdomains 1 and 2 – in this interface protease domain is in direct contact with the ATPase cleft, and possibly functional, minor alterations in ATPase activity was observed with mutations in protease domain [138]. This brings in the question whether the domain-domain arrangement observed in HCV NS3/4A crystal structure is relevant *in vivo*, or an artifact due to crystallization and strong intramolecular contacts between protease active site and C-terminus. Deletion of the C-terminus of HCV NS3/4A was shown to reduce the incubation period required before reaction initialization for helicase domain to become fully active [115], hence the kinetic barrier of transformation from crystallographic conformation to an extended conformation, similar to *Flavivirus* NS3/2B, is possibly lowered by inhibiting the strong interactive force between the two domains.

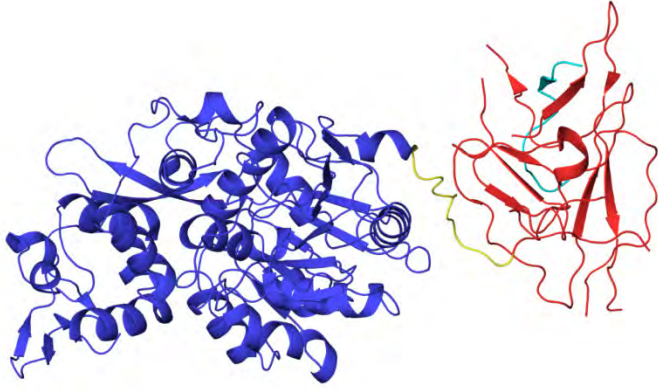
Results from studies investigating the membrane attachment of HCV NS3/4A also suggest an alternative structural arrangement compared to published structures.

Membrane attachment of NS3/4A is critical in HCV life cycle, and two factors – NS4A transmembrane helix and NS3 protease helix  $\alpha_0$  – perform anchorage function [116, 140, 141]. Alterations in either factor result in incorrect localization of the protein, disruption of viral replication and viral polyprotein maturation [116, 141]. However, membrane bound models of NS3/4A suggest that helicase domain has to move away from the protease in order not to clash with the membrane [116] (Figure I.8). Still, the transiency of membrane attachment is not yet fully explored. Cytoplasmic NS3/4A expressed in replicon-based systems cleave membrane-unbound polyprotein junctions whereas membrane-integrated NS3/4A was required for the correct cleavage of membrane-bound substrates [140], which suggests that cycling between membrane-bound and unbound stages, in addition to compact and extended conformations is possibly important in inherent protease and helicase activities.

**Figure I.7 – Comparison of single chain NS3 protease – helicase structures from HCV and DV.** (a) scNS3/4A crystal structure of HCV (PDB-ID:1CU1) [108] and (b) two crystal structures of DV scNS3/2B with alternative domain arrangements (Conformation I – PDB-ID: 2VBC, Conformation II – PDB-ID: 2WHX) [137, 138]. Protease domains are colored in blue, helicase in red, linker in yellow, NS4A or NS2B cofactor in cyan, and the C-terminal  $\alpha$ -helix of HCV scNS3/4A in orange. Unlike HCV scNS3/4A, scNS3/2B structures from DV exhibit extended, rod-shaped conformations.

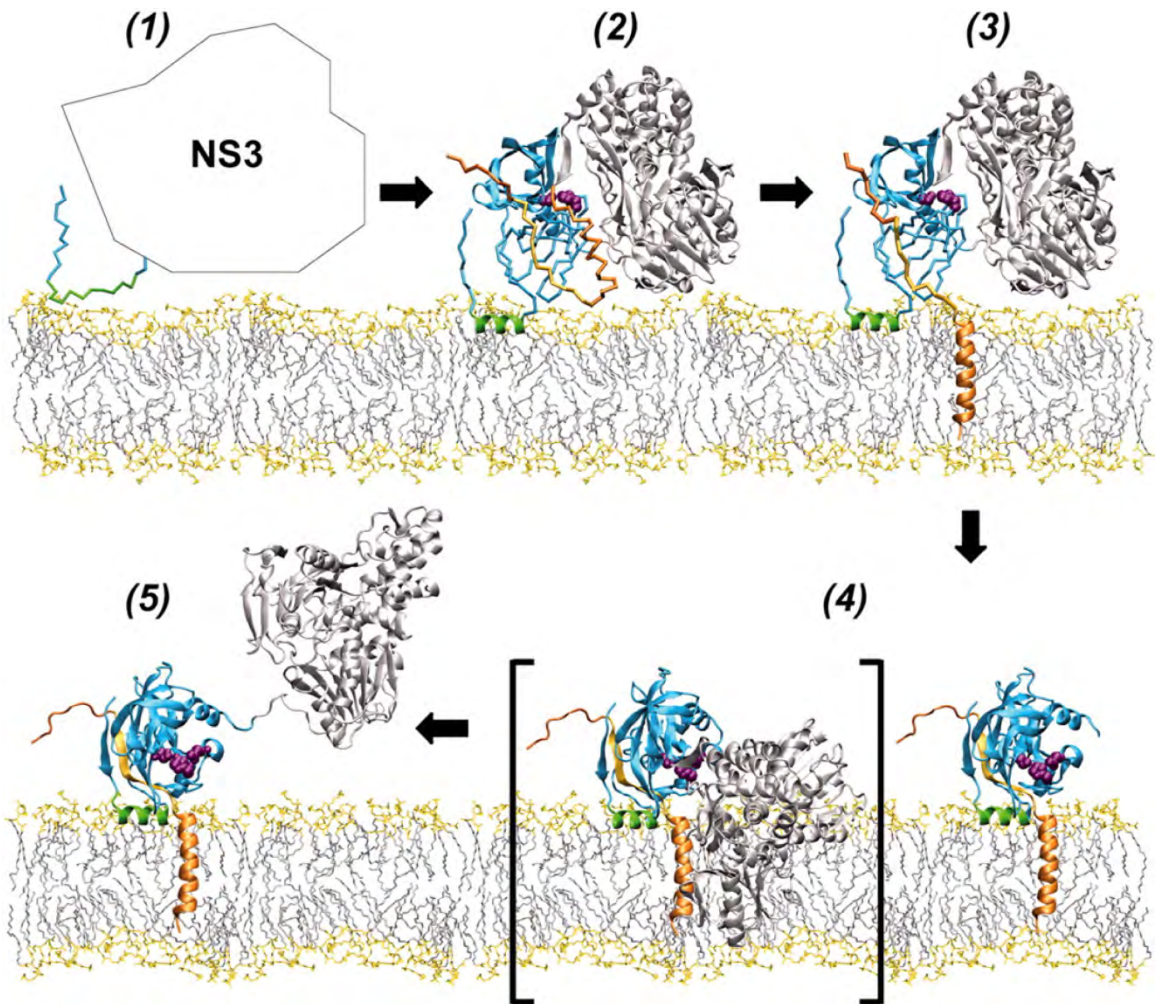


Conformation I



Conformation II

**Figure I.8 – Model of membrane integration for full length HCV NS3/4A.** A model for cotranslational membrane integration of HCV NS3/4A is presented. Protease domain is colored in blue, helicase in gray, NS4A in orange, protease helix  $\alpha_0$  in green and protease catalytic triad is represented as magenta spheres. NS4A and protease helix  $\alpha_0$  are both important in membrane integration. Note that in the crystallographic conformation (in brackets), helicase domain significantly clashes with the membrane when both membrane attachment factors are integrated. (Brass, V., J. M. Berke, et al. (2008). "Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex." *Proc Natl Acad Sci U.S.A.* **105**(38): 14545-14550, Copyright 2009 National Academy of Sciences, U.S.A.).





### **Inhibitors of Non-Structural Protein 3/4A Protease**

Initial proof-of-concept inhibition of HCV NS3/4A protease was discovered in early kinetic studies. N-terminal products of protease substrates were shown to inhibit protease function with micromolar inhibition constants [142]. Thus, similarities between N-terminal products have been exploited by the pharmaceutical industry in structure-based drug design to create NS3/4A protease inhibitors. The first inhibitor from product-based structure-activity relationship (SAR) studies that exhibited high potency in clinical trials was ciluprevir (Boehringer-Ingelheim) [143, 144]. Although ciluprevir was later dropped from development due to cardiotoxicity [145], proof of inhibition for product-based inhibitors paved the way for further drug development.

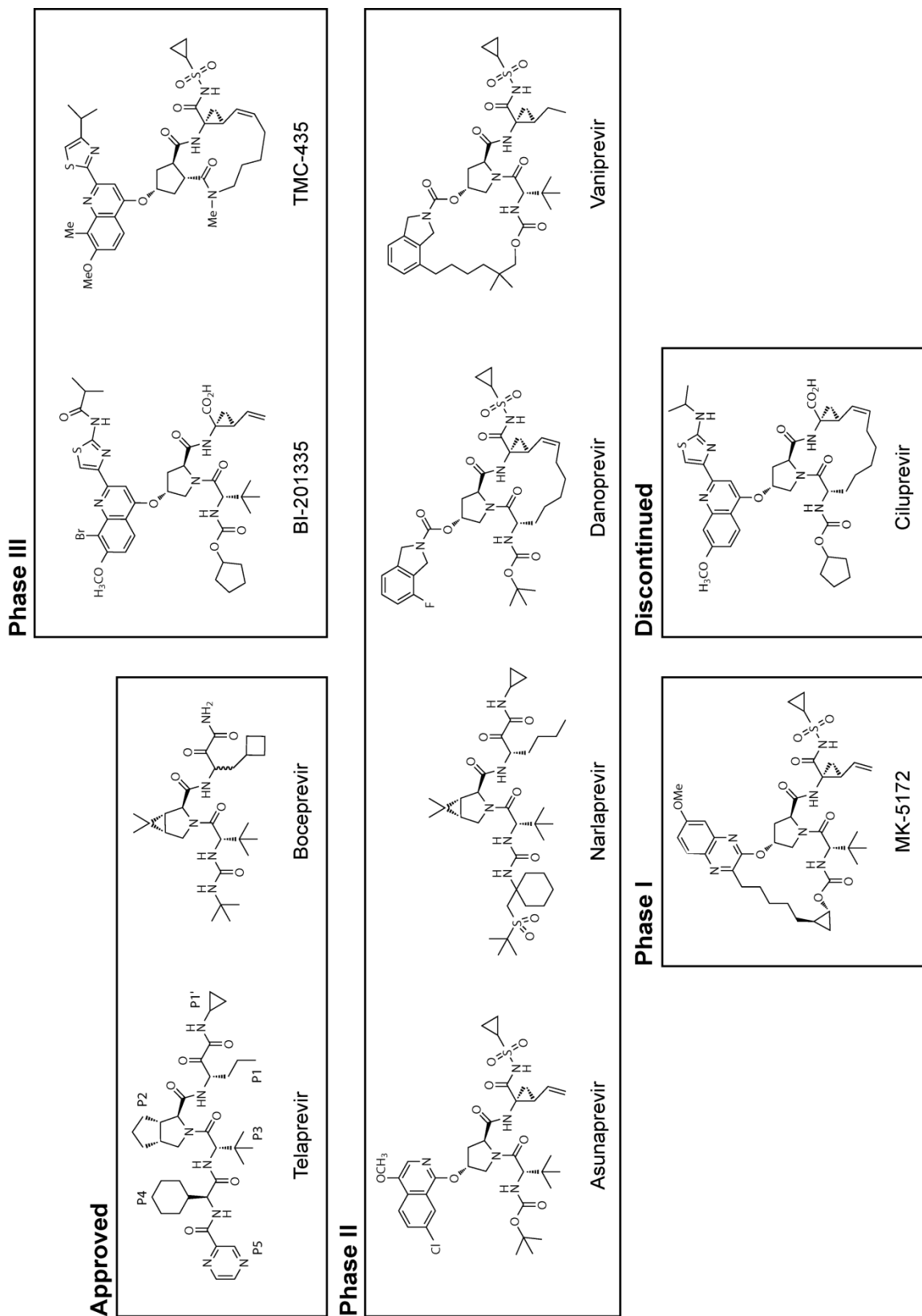
Over the past decade, significant drug development has been performed on peptidomimetic inhibitors by multiple pharmaceutical companies, which resulted in generation of various linear and macrocyclic inhibitors [146, 147] (Figure I.9). Structural features of these drugs are similar – all mimic the P4-P1' region of protease substrates. These drugs contain a ketoamide or sulfonamide warhead in P1' position, a cysteine-like moiety at P1 with terminal olefin that integrates into protease catalytic triad, and a proline-like moiety at P2 with a bulky extension. *Trans* amide-bond geometry between P2 and P3 [148] is stabilized by either macrocyclization or bulky side chain at P3 position. Recently approved inhibitors telaprevir and boceprevir are both linear ketoamides. Some other inhibitors that are currently in clinical trials are the macrocyclic sulfonamide TMC-435 (Tibotec) [149] and linear carboxylate BI-201335 (Boehringer-Ingelheim) [150] in phase 3; macrocyclic sulfonamide drugs danoprevir (Roche) [151]



and vaniprevir (Merck) [152], the linear sulfonamide asunaprevir (Bristol-Myers Squibb) [153] and linear ketoamide narlaprevir (Schering Plough) [154] in phase 2; and macrocyclic sulfonamide MK-5172 (Merck) [155] in phase 1 clinical trials. There are many more inhibitors undergoing clinical investigation and pre-clinical development, but chemical structures for most of these drugs are not publicly available.

However, drug resistance mutants in NS3/4A protease are rapidly selected in both pre-clinical experiments and clinical trials under drug pressure [156]. HCV NS5B polymerase lacks proof reading activity; hence misincorporation of bases into the viral genome causes high degree of mutations in viral proteins [157-159]. These mutations potentially alter the viability of viral production, but under drug pressure, less viable species containing drug resistance mutations thrive. Especially, mutations in the protease domain residues R155, A156 and D168, which contact the P2 extension of protease inhibitors, confer resistance to a multitude of drugs [96, 156, 160-162]. These mutations significantly alter inhibitor – protease contacts without disrupting substrate – protease interactions and cleavage kinetics. Structural basis of drug resistance for HCV NS3/4A protease inhibitors was demonstrated extensively [96], thus new generation drugs should be developed with drug resistance and drug-protein interactions in consideration.

**Figure I.9 – Inhibitors of HCV NS3/4A protease.** HCV NS3/4A protease inhibitors with disclosed chemical structures. These drugs are grouped (rectangles) with respect to developmental stages.



## Thesis Summary

Although individual domains have been studied extensively, interdependence of protease and helicase domains of HCV NS3/4A and underlying mechanisms of domain-domain communication has not been elucidated. This thesis work attempts to evaluate this interdependency by investigating the role of domain-domain interface in protease and helicase function, and by investigating the role of helicase domain in protease inhibitor recognition. Chapter II demonstrates that the interface between the two domains is transient and regulatory *in silico* and *in vitro*, both protease and helicase activities were enhanced when the interface was disrupted. In turn, an alternative domain arrangement relevant for *in vivo* function is proposed, which is in agreement with recent reports covering both *in vitro* and *in vivo* observations. Chapter III demonstrates that the helicase domain plays a limited but consistent role in protease inhibitor potency modulation. Overall, this thesis expands the knowledge on interdomain dependence in the bifunctional HCV NS3/4A enzyme, and provides information for potential drug development exploiting this dependence.

## References

1. Choo, Q.L., et al., *Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome*. Science, 1989. **244**(4902): p. 359-62.
2. WHO. *Fact Sheet No 164*. 2012; Available from: <http://www.who.int/mediacentre/factsheets/fs164/en/index.html>.
3. Alter, M.J., *HCV routes of transmission: what goes around comes around*. Semin Liver Dis, 2011. **31**(4): p. 340-6.
4. Seeff, L.B., *Natural history of chronic hepatitis C*. Hepatology, 2002. **36**(5 Suppl 1): p. S35-46.
5. Hoofnagle, J.H., *Course and outcome of hepatitis C*. Hepatology, 2002. **36**(5 Suppl 1): p. S21-9.
6. Thein, H.H., et al., *Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression*. Hepatology, 2008. **48**(2): p. 418-31.
7. Kim, W.R., et al., *Burden of liver disease in the United States: summary of a workshop*. Hepatology, 2002. **36**(1): p. 227-42.
8. Davis, G.L., et al., *Projecting future complications of chronic hepatitis C in the United States*. Liver Transpl, 2003. **9**(4): p. 331-8.
9. Simmonds, P., et al., *Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes*. Hepatology, 2005. **42**(4): p. 962-73.
10. Benson, D.A., et al., *GenBank*. Nucleic Acids Res, 2011. **39**(Database issue): p. D32-7.

11. Gouy, M., S. Guindon, and O. Gascuel, *SeaView version 4: A multiplatform graphical user interface for sequence alignment and phylogenetic tree building*. Mol Biol Evol, 2010. **27**(2): p. 221-4.
12. Edgar, R.C., *MUSCLE: multiple sequence alignment with high accuracy and high throughput*. Nucleic Acids Res, 2004. **32**(5): p. 1792-7.
13. Guindon, S., et al., *New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0*. Syst Biol, 2010. **59**(3): p. 307-21.
14. Kuiken, C. and P. Simmonds, *Nomenclature and numbering of the hepatitis C virus*. Methods Mol Biol, 2009. **510**: p. 33-53.
15. Andre, P., et al., *Hepatitis C virus particles and lipoprotein metabolism*. Semin Liver Dis, 2005. **25**(1): p. 93-104.
16. Bartosch, B., et al., *Cell entry of hepatitis C virus requires a set of co-receptors that include the CD81 tetraspanin and the SR-B1 scavenger receptor*. J Biol Chem, 2003. **278**(43): p. 41624-30.
17. Scarselli, E., et al., *The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus*. EMBO J, 2002. **21**(19): p. 5017-25.
18. Pileri, P., et al., *Binding of hepatitis C virus to CD81*. Science, 1998. **282**(5390): p. 938-41.
19. Ploss, A. and M.J. Evans, *Hepatitis C virus host cell entry*. Curr Opin Virol, 2012. **2**(1): p. 14-9.
20. Ploss, A., et al., *Human occludin is a hepatitis C virus entry factor required for infection of mouse cells*. Nature, 2009. **457**(7231): p. 882-6.

21. Owen, D.M., et al., *Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor*. *Virology*, 2009. **394**(1): p. 99-108.
22. Barth, H., et al., *Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate*. *J Biol Chem*, 2003. **278**(42): p. 41003-12.
23. Koutsoudakis, G., et al., *The level of CD81 cell surface expression is a key determinant for productive entry of hepatitis C virus into host cells*. *J Virol*, 2007. **81**(2): p. 588-98.
24. Blanchard, E., et al., *Hepatitis C virus entry depends on clathrin-mediated endocytosis*. *J Virol*, 2006. **80**(14): p. 6964-72.
25. Koutsoudakis, G., et al., *Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses*. *J Virol*, 2006. **80**(11): p. 5308-20.
26. Tscherne, D.M., et al., *Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry*. *J Virol*, 2006. **80**(4): p. 1734-41.
27. Friebe, P., et al., *Sequences in the 5' nontranslated region of hepatitis C virus required for RNA replication*. *J Virol*, 2001. **75**(24): p. 12047-57.
28. Kolykhalov, A.A., et al., *Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo*. *J Virol*, 2000. **74**(4): p. 2046-51.
29. Mizushima, H., et al., *Two hepatitis C virus glycoprotein E2 products with different C termini*. *J Virol*, 1994. **68**(10): p. 6215-22.

30. Mizushima, H., et al., *Analysis of N-terminal processing of hepatitis C virus nonstructural protein 2*. J Virol, 1994. **68**(4): p. 2731-4.
31. Lin, C., et al., *Processing in the hepatitis C virus E2-NS2 region: identification of p7 and two distinct E2-specific products with different C termini*. J Virol, 1994. **68**(8): p. 5063-73.
32. Reed, K.E., A. Grakoui, and C.M. Rice, *Hepatitis C virus-encoded NS2-3 protease: cleavage-site mutagenesis and requirements for bimolecular cleavage*. J Virol, 1995. **69**(7): p. 4127-36.
33. Bartenschlager, R., et al., *Kinetic and structural analyses of hepatitis C virus polyprotein processing*. J Virol, 1994. **68**(8): p. 5045-55.
34. Santolini, E., G. Migliaccio, and N. La Monica, *Biosynthesis and biochemical properties of the hepatitis C virus core protein*. J Virol, 1994. **68**(6): p. 3631-41.
35. Hwang, S.B., et al., *Detection of Cellular Proteins and Viral Core Protein Interacting with the 5' Untranslated Region of Hepatitis C Virus RNA*. J Biomed Sci, 1995. **2**(3): p. 227-236.
36. Perlemuter, G., et al., *Hepatitis C virus core protein inhibits microsomal triglyceride transfer protein activity and very low density lipoprotein secretion: a model of viral-related steatosis*. FASEB J, 2002. **16**(2): p. 185-94.
37. Hijikata, M., et al., *Hypervariable regions in the putative glycoprotein of hepatitis C virus*. Biochem Biophys Res Commun, 1991. **175**(1): p. 220-8.
38. Goffard, A. and J. Dubuisson, *Glycosylation of hepatitis C virus envelope proteins*. Biochimie, 2003. **85**(3-4): p. 295-301.



39. Carrere-Kremer, S., et al., *Subcellular localization and topology of the p7 polypeptide of hepatitis C virus*. J Virol, 2002. **76**(8): p. 3720-30.
40. Griffin, S.D., et al., *The p7 protein of hepatitis C virus forms an ion channel that is blocked by the antiviral drug, Amantadine*. FEBS Lett, 2003. **535**(1-3): p. 34-8.
41. Pavlovic, D., et al., *The hepatitis C virus p7 protein forms an ion channel that is inhibited by long-alkyl-chain iminosugar derivatives*. Proc Natl Acad Sci U S A, 2003. **100**(10): p. 6104-8.
42. Grakoui, A., et al., *A second hepatitis C virus-encoded proteinase*. Proc Natl Acad Sci U S A, 1993. **90**(22): p. 10583-7.
43. Lorenz, I.C., et al., *Structure of the catalytic domain of the hepatitis C virus NS2-3 protease*. Nature, 2006. **442**(7104): p. 831-5.
44. Grakoui, A., et al., *Characterization of the hepatitis C virus-encoded serine proteinase: determination of proteinase-dependent polyprotein cleavage sites*. J Virol, 1993. **67**(5): p. 2832-43.
45. Gallinari, P., et al., *Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus*. J Virol, 1998. **72**(8): p. 6758-69.
46. Kanai, A., K. Tanabe, and M. Kohara, *Poly(U) binding activity of hepatitis C virus NS3 protein, a putative RNA helicase*. FEBS Lett, 1995. **376**(3): p. 221-4.
47. Kim, D.W., et al., *C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity*. Biochem Biophys Res Commun, 1995. **215**(1): p. 160-6.

48. Kim, J.L., et al., *Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide*. Cell, 1996. **87**(2): p. 343-55.
49. Sali, D.L., et al., *Serine protease of hepatitis C virus expressed in insect cells as the NS3/4A complex*. Biochemistry, 1998. **37**(10): p. 3392-401.
50. Beran, R.K., B.D. Lindenbach, and A.M. Pyle, *The NS4A protein of hepatitis C virus promotes RNA-coupled ATP hydrolysis by the NS3 helicase*. J Virol, 2009. **83**(7): p. 3268-75.
51. Shiryayev, S.A., et al., *The acidic sequence of the NS4A cofactor regulates ATP hydrolysis by the HCV NS3 helicase*. Arch Virol, 2011. **156**(2): p. 313-8.
52. Gouttenoire, J., et al., *Identification of a novel determinant for membrane association in hepatitis C virus nonstructural protein 4B*. J Virol, 2009. **83**(12): p. 6257-68.
53. Gouttenoire, J., F. Penin, and D. Moradpour, *Hepatitis C virus nonstructural protein 4B: a journey into unexplored territory*. Rev Med Virol, 2010. **20**(2): p. 117-29.
54. Huang, Y., et al., *Phosphorylation of hepatitis C virus NS5A nonstructural protein: a new paradigm for phosphorylation-dependent viral RNA replication?* Virology, 2007. **364**(1): p. 1-9.
55. Chang, M., et al., *Dynamics of hepatitis C virus replication in human liver*. Am J Pathol, 2003. **163**(2): p. 433-44.
56. Quinkert, D., R. Bartenschlager, and V. Lohmann, *Quantitative analysis of the hepatitis C virus replication complex*. J Virol, 2005. **79**(21): p. 13594-605.

57. Targett-Adams, P., S. Boulant, and J. McLauchlan, *Visualization of double-stranded RNA in cells supporting hepatitis C virus RNA replication*. J Virol, 2008. **82**(5): p. 2182-95.
58. Xu, Z., et al., *Synthesis of a novel hepatitis C virus protein by ribosomal frameshift*. EMBO J, 2001. **20**(14): p. 3840-8.
59. Vassilaki, N. and P. Mavromara, *The HCV ARFP/F/core+1 protein: production and functional analysis of an unconventional viral product*. IUBMB Life, 2009. **61**(7): p. 739-52.
60. Egger, D., et al., *Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex*. J Virol, 2002. **76**(12): p. 5974-84.
61. Gosert, R., et al., *RNA replication of mouse hepatitis virus takes place at double-membrane vesicles*. J Virol, 2002. **76**(8): p. 3697-708.
62. Neumann, A.U., et al., *Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy*. Science, 1998. **282**(5386): p. 103-7.
63. Strader, D.B., et al., *Diagnosis, management, and treatment of hepatitis C*. Hepatology, 2004. **39**(4): p. 1147-71.
64. Lindenmann, J., D.C. Burke, and A. Isaacs, *Studies on the production, mode of action and properties of interferon*. Br J Exp Pathol, 1957. **38**(5): p. 551-62.
65. Isaacs, A., J. Lindenmann, and R.C. Valentine, *Virus interference. II. Some properties of interferon*. Proc R Soc Lond B Biol Sci, 1957. **147**(927): p. 268-73.

66. Isaacs, A. and J. Lindenmann, *Virus interference. I. The interferon*. Proc R Soc Lond B Biol Sci, 1957. **147**(927): p. 258-67.
67. Abuchowski, A., et al., *Alteration of immunological properties of bovine serum albumin by covalent attachment of polyethylene glycol*. J Biol Chem, 1977. **252**(11): p. 3578-81.
68. Abuchowski, A., et al., *Effect of covalent attachment of polyethylene glycol on immunogenicity and circulating life of bovine liver catalase*. J Biol Chem, 1977. **252**(11): p. 3582-6.
69. Sidwell, R.W., et al., *Broad-spectrum antiviral activity of Virazole: 1-beta-D-ribofuranosyl-1,2,4-triazole-3-carboxamide*. Science, 1972. **177**(4050): p. 705-6.
70. Zeuzem, S. and F. Poordad, *Pegylated-interferon plus ribavirin therapy in the treatment of CHC: individualization of treatment duration according to on-treatment virologic response*. Curr Med Res Opin, 2010. **26**(7): p. 1733-43.
71. Zeuzem, S., et al., *Management of hepatitis C virus genotype 2 or 3 infection: treatment optimization on the basis of virological response*. Antivir Ther, 2009. **14**(2): p. 143-54.
72. Manns, M.P., et al., *Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial*. Lancet, 2001. **358**(9286): p. 958-65.
73. Fried, M.W., et al., *Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection*. N Engl J Med, 2002. **347**(13): p. 975-82.

74. Fried, M.W., *Side effects of therapy of hepatitis C and their management*. Hepatology, 2002. **36**(5 Suppl 1): p. S237-44.
75. Butt, A.A., et al., *Hepatitis C treatment completion rates in routine clinical care*. Liver Int, 2010. **30**(2): p. 240-50.
76. Gao, M., et al., *Chemical genetics strategy identifies an HCV NS5A inhibitor with a potent clinical effect*. Nature, 2010. **465**(7294): p. 96-100.
77. Chow, J., et al., *Isolation and identification of ester impurities in RG7128, an HCV polymerase inhibitor*. J Pharm Biomed Anal, 2010. **53**(3): p. 710-6.
78. Sofia, M.J., et al., *Discovery of a beta-d-2'-deoxy-2'-alpha-fluoro-2'-beta-C-methyluridine nucleotide prodrug (PSI-7977) for the treatment of hepatitis C virus*. J Med Chem, 2010. **53**(19): p. 7202-18.
79. Li, H., et al., *Discovery of (R)-6-cyclopentyl-6-(2-(2,6-diethylpyridin-4-yl)ethyl)-3-((5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-yl)methyl)-4-hydroxy-5,6-dihydropyran-2-one (PF-00868554) as a potent and orally available hepatitis C virus polymerase inhibitor*. J Med Chem, 2009. **52**(5): p. 1255-8.
80. Zeuzem, S., et al., *Efficacy of the protease inhibitor BI 201335, polymerase inhibitor BI 207127, and ribavirin in patients with chronic HCV infection*. Gastroenterology, 2011. **141**(6): p. 2047-55; quiz e14.
81. Shih, I.H., et al., *Mechanistic characterization of GS-9190 (Tegobuvir), a novel nonnucleoside inhibitor of hepatitis C virus NS5B polymerase*. Antimicrob Agents Chemother, 2011. **55**(9): p. 4196-203.

82. Membreno, F.E. and E.J. Lawitz, *The HCV NS5B nucleoside and non-nucleoside inhibitors*. Clin Liver Dis, 2011. **15**(3): p. 611-26.
83. Hofmann, W.P. and S. Zeuzem, *Hepatitis C in 2011: A new standard of care and the race towards IFN-free therapy*. Nat Rev Gastroenterol Hepatol, 2012. **9**(2): p. 67-8.
84. Butt, A.A. and F. Kanwal, *Boceprevir and telaprevir in the management of hepatitis C virus-infected patients*. Clin Infect Dis, 2012. **54**(1): p. 96-104.
85. Kwo, P.Y., *Phase III results in Genotype 1 naive patients: predictors of response with boceprevir and telaprevir combined with pegylated interferon and ribavirin*. Liver Int, 2012. **32 Suppl 1**: p. 39-43.
86. Forestier, N. and S. Zeuzem, *Triple therapy with telaprevir: results in hepatitis C virus-genotype 1 infected relapsers and non-responders*. Liver Int, 2012. **32 Suppl 1**: p. 44-50.
87. Bacon, B.R. and O. Khalid, *Triple therapy with boceprevir for HCV genotype 1 infection: phase III results in relapsers and nonresponders*. Liver Int, 2012. **32 Suppl 1**: p. 51-3.
88. Morikawa, K., et al., *Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus*. J Viral Hepat, 2011. **18**(5): p. 305-15.
89. Lin, C., et al., *Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics*. J Virol, 1994. **68**(12): p. 8147-57.
90. Bartenschlager, R., et al., *Substrate determinants for cleavage in cis and in trans by the hepatitis C virus NS3 proteinase*. J Virol, 1995. **69**(1): p. 198-205.

91. Li, K., et al., *Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF*. Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2992-7.
92. Foy, E., et al., *Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease*. Science, 2003. **300**(5622): p. 1145-8.
93. Yoneyama, M., et al., *Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity*. J Immunol, 2005. **175**(5): p. 2851-8.
94. Foy, E., et al., *Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-1 signaling*. Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2986-91.
95. Romano, K.P., et al., *Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease*. J Virol, 2011. **85**(13): p. 6106-16.
96. Romano, K.P., et al., *Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding*. Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20986-91.
97. Prabu-Jeyabalan, M., E. Nalivaika, and C.A. Schiffer, *Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes*. Structure, 2002. **10**(3): p. 369-81.
98. Frick, D.N., *The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target*. Curr Issues Mol Biol, 2007. **9**(1): p. 1-20.

99. Walker, J.E., et al., *Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold*. EMBO J, 1982. **1**(8): p. 945-51.
100. Fairman-Williams, M.E., U.P. Guenther, and E. Jankowsky, *SF1 and SF2 helicases: family matters*. Curr Opin Struct Biol, 2010. **20**(3): p. 313-24.
101. Pyle, A.M., *RNA helicases and remodeling proteins*. Curr Opin Chem Biol, 2011. **15**(5): p. 636-42.
102. Yao, N., et al., *Structure of the hepatitis C virus RNA helicase domain*. Nat Struct Biol, 1997. **4**(6): p. 463-7.
103. Pyle, A.M., *Translocation and unwinding mechanisms of RNA and DNA helicases*. Annu Rev Biophys, 2008. **37**: p. 317-36.
104. Paolini, C., R. De Francesco, and P. Gallinari, *Enzymatic properties of hepatitis C virus NS3-associated helicase*. J Gen Virol, 2000. **81**(Pt 5): p. 1335-45.
105. Kim, J.L., et al., *Hepatitis C virus NS3 RNA helicase domain with a bound oligonucleotide: the crystal structure provides insights into the mode of unwinding*. Structure, 1998. **6**(1): p. 89-100.
106. Mackintosh, S.G., et al., *Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus*. J Biol Chem, 2006. **281**(6): p. 3528-35.
107. Yao, N., et al., *Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase*. Structure, 1999. **7**(11): p. 1353-63.



108. Gu, M. and C.M. Rice, *Three conformational snapshots of the hepatitis C virus NS3 helicase reveal a ratchet translocation mechanism*. Proc Natl Acad Sci U S A, 2010. **107**(2): p. 521-8.
109. Appleby, T.C., et al., *Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV*. J Mol Biol, 2011. **405**(5): p. 1139-53.
110. Dumont, S., et al., *RNA translocation and unwinding mechanism of HCV NS3 helicase and its coordination by ATP*. Nature, 2006. **439**(7072): p. 105-8.
111. Levin, M.K. and S.S. Patel, *The helicase from hepatitis C virus is active as an oligomer*. J Biol Chem, 1999. **274**(45): p. 31839-46.
112. Sikora, B., et al., *Hepatitis C virus NS3 helicase forms oligomeric structures that exhibit optimal DNA unwinding activity in vitro*. J Biol Chem, 2008. **283**(17): p. 11516-25.
113. Jennings, T.A., et al., *NS3 helicase from the hepatitis C virus can function as a monomer or oligomer depending on enzyme and substrate concentrations*. J Biol Chem, 2009. **284**(8): p. 4806-14.
114. Frick, D.N., et al., *The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex RNA efficiently*. J Biol Chem, 2004. **279**(2): p. 1269-80.
115. Ding, S.C., A.S. Kohlway, and A.M. Pyle, *Unmasking the active helicase conformation of nonstructural protein 3 from hepatitis C virus*. J Virol, 2011. **85**(9): p. 4343-53.

116. Brass, V., et al., *Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex*. Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14545-50.
117. Wang, W., et al., *Conserved C-terminal threonine of hepatitis C virus NS3 regulates autoproteolysis and prevents product inhibition*. J Virol, 2004. **78**(2): p. 700-9.
118. Morgenstern, K.A., et al., *Polynucleotide modulation of the protease, nucleoside triphosphatase, and helicase activities of a hepatitis C virus NS3-NS4A complex isolated from transfected COS cells*. J Virol, 1997. **71**(5): p. 3767-75.
119. Beran, R.K., V. Serebrov, and A.M. Pyle, *The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate*. J Biol Chem, 2007. **282**(48): p. 34913-20.
120. Beran, R.K. and A.M. Pyle, *Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase*. J Biol Chem, 2008. **283**(44): p. 29929-37.
121. Thibeault, D., et al., *Use of the fused NS4A peptide-NS3 protease domain to study the importance of the helicase domain for protease inhibitor binding to hepatitis C virus NS3-NS4A*. Biochemistry, 2009. **48**(4): p. 744-53.
122. Dahl, G., et al., *Effects on protease inhibition by modifying of helicase residues in hepatitis C virus nonstructural protein 3*. FEBS J, 2007. **274**(22): p. 5979-86.
123. Dahl, G., O.G. Arenas, and U.H. Danielson, *Hepatitis C virus NS3 protease is activated by low concentrations of protease inhibitors*. Biochemistry, 2009. **48**(48): p. 11592-602.

124. Taliani, M., et al., *A continuous assay of hepatitis C virus protease based on resonance energy transfer depsi-peptide substrates*. *Anal Biochem*, 1996. **240**(1): p. 60-7.
125. Urbani, A., et al., *Substrate specificity of the hepatitis C virus serine protease NS3*. *J Biol Chem*, 1997. **272**(14): p. 9204-9.
126. Howe, A.Y., et al., *A novel recombinant single-chain hepatitis C virus NS3-NS4A protein with improved helicase activity*. *Protein Sci*, 1999. **8**(6): p. 1332-41.
127. Taremi, S.S., et al., *Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease*. *Protein Sci*, 1998. **7**(10): p. 2143-9.
128. Pang, P.S., et al., *The hepatitis C viral NS3 protein is a processive DNA helicase with cofactor enhanced RNA unwinding*. *EMBO J*, 2002. **21**(5): p. 1168-76.
129. Zhang, C., et al., *Stimulation of hepatitis C virus (HCV) nonstructural protein 3 (NS3) helicase activity by the NS3 protease domain and by HCV RNA-dependent RNA polymerase*. *J Virol*, 2005. **79**(14): p. 8687-97.
130. Chernov, A.V., et al., *The two-component NS2B-NS3 proteinase represses DNA unwinding activity of the West Nile virus NS3 helicase*. *J Biol Chem*, 2008. **283**(25): p. 17270-8.
131. Shiryayev, S.A., et al., *NS4A regulates the ATPase activity of the NS3 helicase: a novel cofactor role of the non-structural protein NS4A from West Nile virus*. *J Gen Virol*, 2009. **90**(Pt 9): p. 2081-5.

132. Schiering, N., et al., *A macrocyclic HCV NS3/4A protease inhibitor interacts with protease and helicase residues in the complex with its full-length target*. Proc Natl Acad Sci U S A, 2011. **108**(52): p. 21052-6.
133. Zhang, R., et al., *Probing the substrate specificity of hepatitis C virus NS3 serine protease by using synthetic peptides*. J Virol, 1997. **71**(8): p. 6208-13.
134. Wodak, S.J. and J. Janin, *Structural basis of macromolecular recognition*. Adv Protein Chem, 2002. **61**: p. 9-73.
135. Nooren, I.M. and J.M. Thornton, *Structural characterisation and functional significance of transient protein-protein interactions*. J Mol Biol, 2003. **325**(5): p. 991-1018.
136. Bollati, M., et al., *Structure and functionality in flavivirus NS-proteins: perspectives for drug design*. Antiviral Res, 2010. **87**(2): p. 125-48.
137. Luo, D., et al., *Crystal structure of the NS3 protease-helicase from dengue virus*. J Virol, 2008. **82**(1): p. 173-83.
138. Luo, D., et al., *Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications*. J Biol Chem, 2010. **285**(24): p. 18817-27.
139. Mastrangelo, E., et al., *Crystal structure and activity of Kunjin virus NS3 helicase; protease and helicase domain assembly in the full length NS3 protein*. J Mol Biol, 2007. **372**(2): p. 444-55.

140. Horner, S.M., H.S. Park, and M. Gale, Jr., *Control of innate immune signaling and membrane targeting by the Hepatitis C virus NS3/4A protease are governed by the NS3 helix alpha0*. J Virol, 2012. **86**(6): p. 3112-20.
141. He, Y., et al., *The N-terminal helix alpha(0) of hepatitis C virus NS3 protein dictates the subcellular localization and stability of NS3/NS4A complex*. Virology, 2012. **422**(2): p. 214-23.
142. Steinkuhler, C., et al., *Product inhibition of the hepatitis C virus NS3 protease*. Biochemistry, 1998. **37**(25): p. 8899-905.
143. Lamarre, D., et al., *An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus*. Nature, 2003. **426**(6963): p. 186-9.
144. Hinrichsen, H., et al., *Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients*. Gastroenterology, 2004. **127**(5): p. 1347-55.
145. Vanwollegem, T., et al., *Ultra-rapid cardiotoxicity of the hepatitis C virus protease inhibitor BILN 2061 in the urokinase-type plasminogen activator mouse*. Gastroenterology, 2007. **133**(4): p. 1144-55.
146. Tsantrizos, Y.S., et al., *Macrocyclic inhibitors of the NS3 protease as potential therapeutic agents of hepatitis C virus infection*. Angew Chem Int Ed Engl, 2003. **42**(12): p. 1356-60.
147. Tsantrizos, Y.S., *Peptidomimetic therapeutic agents targeting the protease enzyme of the human immunodeficiency virus and hepatitis C virus*. Acc Chem Res, 2008. **41**(10): p. 1252-63.

148. LaPlante, S.R., et al., *Solution structure of substrate-based ligands when bound to hepatitis C virus NS3 protease domain*. J Biol Chem, 1999. **274**(26): p. 18618-24.
149. Raboisson, P., et al., *Structure-activity relationship study on a novel series of cyclopentane-containing macrocyclic inhibitors of the hepatitis C virus NS3/4A protease leading to the discovery of TMC435350*. Bioorg Med Chem Lett, 2008. **18**(17): p. 4853-8.
150. Llinas-Brunet, M., et al., *Discovery of a potent and selective noncovalent linear inhibitor of the hepatitis C virus NS3 protease (BI 201335)*. J Med Chem, 2010. **53**(17): p. 6466-76.
151. Seiwert, S.D., et al., *Discovery and Development of the HCV NS3/4A Protease Inhibitor Danoprevir (ITMN-191/RG7227)*, in *Antiviral Drugs*. 2011, John Wiley & Sons, Inc. p. 257-271.
152. McCauley, J.A., et al., *Discovery of vaniprevir (MK-7009), a macrocyclic hepatitis C virus NS3/4a protease inhibitor*. J Med Chem, 2010. **53**(6): p. 2443-63.
153. McPhee, F., *Identification and preclinical profile of the novel HCV NS3 protease inhibitor BMS-650032*. J Hepatol, 2010. **52**: p. S296.
154. Arasappan, A., et al., *Discovery of Narlaprevir (SCH 900518): A Potent, Second Generation HCV NS3 Serine Protease Inhibitor*. ACS Medicinal Chemistry Letters, 2010. **1**(2): p. 64-69.
155. Harper, S., et al., *Discovery of MK-5172, a Macrocyclic Hepatitis C Virus NS3/4a Protease Inhibitor*. ACS Medicinal Chemistry Letters, 2012: p. 120302153334001.

156. Sarrazin, C. and S. Zeuzem, *Resistance to direct antiviral agents in patients with hepatitis C virus infection*. Gastroenterology, 2010. **138**(2): p. 447-62.
157. Qureshi, S.A., *Hepatitis C virus--biology, host evasion strategies, and promising new therapies on the horizon*. Med Res Rev, 2007. **27**(3): p. 353-73.
158. Cabot, B., et al., *Longitudinal evaluation of the structure of replicating and circulating hepatitis C virus quasispecies in nonprogressive chronic hepatitis C patients*. J Virol, 2001. **75**(24): p. 12005-13.
159. Farci, P., et al., *Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome*. Proc Natl Acad Sci U S A, 2002. **99**(5): p. 3081-6.
160. Tong, X., et al., *Identification and analysis of fitness of resistance mutations against the HCV protease inhibitor SCH 503034*. Antiviral Res, 2006. **70**(2): p. 28-38.
161. Rong, L., et al., *Rapid emergence of protease inhibitor resistance in hepatitis C virus*. Sci Transl Med, 2010. **2**(30): p. 30ra32.
162. Kieffer, T.L., A.D. Kwong, and G.R. Picchio, *Viral resistance to specifically targeted antiviral therapies for hepatitis C (STAT-Cs)*. J Antimicrob Chemother, 2010. **65**(2): p. 202-12.

**CHAPTER II**

**DESTABILIZATION OF PROTEASE – HELICASE INTERFACE ENHANCES  
BOTH PROTEASE AND HELICASE ACTIVITIES OF HEPATITIS C VIRUS  
NON-STRUCTURAL PROTEIN NS3-NS4A COMPLEX**



**Author Contributions**

This study was performed in collaboration with Sourav Mukherjee (University of Wisconsin, Milwaukee – UWM), who performed the DNA binding assays, and Alicia M Hanson (UWM), who performed DNA unwinding assays, with all the remaining work performed by me. The paper was written by me and edited by Nese Kurt Yilmaz, Akbar Ali, Sourav Mukherjee (UWM), David Frick (UWM) and Celia Schiffer.

## Abstract

Hepatitis C Virus (HCV) non-structural proteins 3 (NS3) and 4A (NS4A) form a heterodimer (NS3/4A) that is both a serine protease and RNA helicase, both of which are essential for viral propagation. Recombinant DNA can be used to express the NS3 helicase and protease domains separately in model organisms, but the isolated domains behave differently from full length NS3 in assays, suggesting that “interdomain communication” occurs. However, the mechanism of this communication and the subsequent biological implications is unclear. Here, we investigated interdomain dependence in the context of the domain-domain interface formed in the “compact” conformation observed in the HCV NS3/4A crystal structures, using a combination of computational and biochemical methods. The two domains of the NS3/4A were dynamically coupled through the interfaces *in silico*. However, amino acid substitutions designed to disrupt this interface did not hinder the catalytic activities of either domain; in fact, many enhanced the activity. These findings suggest that in solution the protein might assume the “extended” conformation where the protease is rotated away from the helicase, which is observed in the crystal structures of homologous proteins from *Flaviviruses* and proposed to play a biological role by facilitating cleavage of proteins *in trans*. Based on these results, the disruptions likely caused in the interface by the introduced substitutions lowers the energetic barrier between the compact and the extended conformations, thus the protein is rendered more active. The dynamic interplay between the two enzymatic domains of HCV NS3/4A likely modulates the protease and helicase activities *in vitro* and may very well have implications *in vivo*.

## Introduction

Hepatitis C Virus (HCV) is the causative agent of non-A non-B viral hepatitis that infects 3% of the world population [1]. HCV infection, if not cleared by the host, eventually progresses into chronic hepatitis [2]. Upon infection of the host, the single strand positive sense RNA genome is translated into a 3,000 amino acid polyprotein which is then processed by a combination of host and viral proteases NS2/3 and NS3/4A to yield structural and non-structural proteins [3].

HCV NS3/4A is a protein complex that contains both a protease and an ATP-dependent RNA helicase. The N-terminal region of the NS3 protein forms a chymotrypsin-like serine protease that is only active when bound to the hydrophobic  $\beta$ -sheet of the NS4A protein [4]. NS3/4A protease is required to cleave the junction between NS3 and NS4A in *cis* and the downstream viral polyprotein cleavage sites between NS4A and NS4B, NS4B and NS5A, and NS5A and NS5B [5], all of which are necessary for viral maturation. The protease domain also cleaves the host mitochondrial antiviral signaling protein (MAVS) and TIR-domain-containing adapter-inducing interferon- $\beta$  (TRIF), which are required in the innate immune response that cells mount against HCV [6, 7]. In addition, the C-terminal region of NS3 forms an RNA helicase that is classified as a superfamily 2 DExD-box helicase. NS3 helicase incorporates both polynucleotide stimulated NTPase activity and 3' to 5' unwinding activity for duplex RNA and DNA [8-10]. The NS3 helicase likely aids the viral replication machinery by unwinding the RNA duplexes formed during and/or after replication [11]. Both functions of the HCV NS3/4A are essential for viral propagation, and they have been studied

extensively. Many of these studies were performed using NS3 fragments containing only protease or helicase domains because such fragments can be more easily isolated as recombinant proteins and are generally more stable *in vitro* than NS3/4A. However, studies comparing the full-length protein to the protease fragment or helicase fragment have shown the bidirectional influence of each domain on the functional capability of the other. The influence of the helicase domain on NS3 protease function is still controversial [12-14]. In most studies, helicase domain did not exert an effect on protease function [12, 13, 15], but in one study, the full-length NS3-NS4A complex was significantly more active than NS3 protease [14]. The effect of the protease on the NS3 helicase function is better established – the RNA to DNA preference and RNA unwinding activity of the full-length protein are both higher compared to the isolated helicase domain (NS3hel) [16, 17]. Similar results have been obtained from studies on Flavivirus NS3/NS2B, which is an ortholog of HCV NS3/NS4A complex [18-20]. In summary, prior studies suggest the presence of domain-domain dependence in NS3 protein, but fail to elucidate how these two domains communicate.

Several published crystal structures of recombinant proteins where the portion of NS4A needed to activate NS3 is covalently tethered to the NS3 N-terminus, forming a single chain NS3/4A protease – helicase construct (scNS3/4A) show a proximal arrangement of the two domains, where the helicase domain covers the protease active site [21-23]. In this “compact” conformation, the helicase and the protease domains form extensive interdomain contacts. In this study, we analyzed this interface as a possible medium for domain-domain communication. Molecular dynamics (MD) studies were

performed on scNS3/4A using the apo structure seen in PDB file 1CU1 [21] and *in silico* variants with deletions in the C-terminus and linker regions. MD results show dynamic coupling through the interface. In order to experimentally verify the suggested domain-domain interaction, functional studies were performed to compare protease and helicase activities of scNS3/4A containing amino acid substitutions in the protease-helicase interface with those of wildtype scNS3/4A and isolated domains (*i.e.* single chain NS3/4A protease (scNS3/4Apro) and NS3 helicase (NS3hel)). Most amino acid substitutions in the interface destabilized the interface and enhanced protease and helicase activities. These findings suggest that the HCV NS3/NS4A complex might assume another conformation [24, 25] – possibly similar to the extended conformation seen in crystal structures of homologous proteins from related viruses [26, 27] – and that this alternate conformation not yet seen in scNS3/4A crystal structures is the catalytically active and biologically relevant arrangement.

## Materials and Methods

### *Molecular Dynamics Simulations*

Molecular dynamics (MD) simulations were performed on scNS3/4A constructs. Briefly, the structure of the scNS3/4A was obtained from Protein Data Bank (PDB ID: 1CU1) [21] and additional *in silico* deletion constructs  $\Delta$ C-term ( $\Delta$ 626-631 – C-terminus deletion),  $\Delta$ Lin ( $\Delta$ 184-186 – linker deletion) and  $\Delta$ Double ( $\Delta$ 184-186, 626-631 – C-terminus and linker deletion) were created. MD simulations were performed on quadruplicate on each construct using AMBER [28]. Resulting trajectories were analyzed for cross-correlations between the domains and distance calculations were performed between selected residues across the domains. Details of the simulation setup and analytical calculations are explained in Supplementary Methods.

### *Nucleic Acid Substrates*

Single stranded oligonucleotides used to construct helicase substrates were obtained from Integrated DNA Technologies. Preparation of the DNA molecular beacon substrate for DNA unwinding assay was explained in detail previously [11]. Briefly, the 50  $\mu$ M each of the molecular beacon top strand 5'-Cy5 (Cyanine 5)-GCTCCCAATCGATGAACGGGGAGC-IBRQ (Iowa Black RQ)-3' and the bottom strand 5'-GCTCCCCGTTTCATCGATTGGGGAGC-T<sub>19</sub>-3' were heated to 95°C, slowly cooled to room temperature and further purified from native polyacrylamide gel electrophoresis using a crush and soak method. The RNA-DNA hybrid substrate for RNA unwinding assay was constructed by heating up the 1.1:1.2:1 mixture of top strands 5'-GCUGUAUCGUCAAGGCACUAGUGC-Cy5-3', 5'-IBRQ-

CCTACGCCACCAGCTCCGTAGG-3' and the bottom strand 5'-GGAGCUGGUGGCGUAGGCAAGAGUGCCUUGACGAUACAGCUUUUU-U<sub>15</sub>-3' to 95°C and slowly cooling down to room temperature. No further purification was performed for the RNA-DNA substrate.

#### *Protein Expression Plasmids*

The constructs for the isolated protease and helicase domains were defined elsewhere [29, 30]. The single chain NS3/4A (scNS3/4A) construct was generated by ligating the codon optimized genotype 1a helicase construct (H77c) downstream of the Bristol-Myers-Squibb patented scNS3/4A protease sequence (synthesized by GenScript) [31] and cloned into a pET28a expression vector (Novagen). Mutant constructs Y56A, V78A, D79A, S483A, M485A, V524A, Triple Prot (Y56A-V78A-D79A), Triple Hel (S483A-M485A-V524A), F438A, Q526A, H528A and F531A have been generated using the QuikChange II site-directed mutagenesis kit (Stratagene) with suitable mutagenesis primers generated by PrimerX (<http://bioinformatics.org/primerx>). The construct Hexa (Y56A-V78A-D79A-S483A-M485A-V524A) was constructed by ligating the protease domain of Triple Prot and the helicase domain of Triple Hel. Each mutant construct was sequenced to ensure correctness and full coverage.

#### *Protein Expression and Purification*

The expression and purification of scNS3/4A<sub>pro</sub> and NS3<sub>hel</sub> were detailed elsewhere [11, 30]. For the full-length scNS3/4A constructs, expression vectors containing scNS3/4A full-length inserts were transformed into BL21(DE3) *E. coli* cells. 2 liters of lysogeny broth (LB) culture of transformed cells, supplemented with 20 µg/mL

kanamycin, was grown to an  $A_{600}$  of 0.6 with shaking at 30°C. The cultures were then transferred to 20°C, supplemented with 0.5 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) and incubated for 4 hours with shaking. After expression, the cells were harvested by centrifuging at 4,000 g for 30 min, suspended in 40 mL ice-cold 1X phosphate-buffered saline (PBS), the precipitate was collected by centrifuging at 4,000 g for 30 min and stored at -80°C.

All the subsequent steps were performed at 4°C. The frozen pellet was resuspended in 20 mL Buffer HT (25 mM HEPES, 500 mM NaCl, 10% glycerol, 0.1% octyl- $\beta$ -D-glucopyranoside (O $\beta$ G), 2 mM tris(2-carboxylethyl)phosphine (TCEP) and 20 mM imidazole, pH 8.0). The suspended pellet was supplemented with 10 mM MgCl<sub>2</sub>, 4 mM CaCl<sub>2</sub>, 1 mM MnCl<sub>2</sub> and 200 U DNase I (Roche) and lysed with a cell disruptor (Micro Fluidics) in two passes. The lysate was incubated for 2 hours with gentle mixing and centrifuged at 40,000 g for 30 min. The clarified lysate was first passed through a 0.22  $\mu$ m syringe filter (Millipore) and then applied to a 1 mL HisTrap HP (GE Lifesciences) column, equilibrated in Buffer HT, using an AKTA FPLC (GE Lifesciences) with a flow rate of 0.4 mL/min. The column was washed with 10 column volumes (CV) of Buffer HT, 10 CV of Buffer HT with 40 mM imidazole with a flowrate of 0.25 mL/min. Finally, the protein was eluted with 5 column volumes of Buffer HT with 250 mM imidazole. The eluate was dialyzed against Buffer S (25 mM MES, 150 mM NaCl, 10% glycerol, 0.1% O $\beta$ G, 2 mM TCEP, pH 6.0) overnight.

The dialyzed protein suspension was first centrifuged at 20,000 g for 30 min and the supernatant was applied to an 8 mL Mono S column (GE Lifesciences), equilibrated



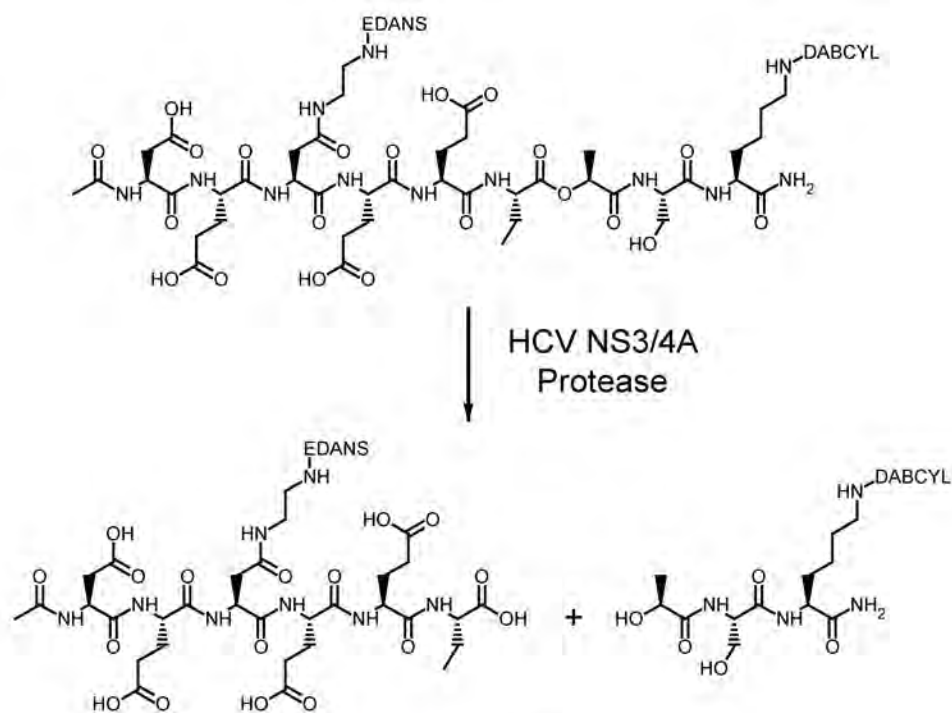
in Buffer S, using an AKTA FPLC with a flowrate of 1 mL/min. A linear gradient from Buffer S to Buffer S with 1 M NaCl was applied to the column in 30 CV, and the protein eluted at ~60% Buffer S with 1 M NaCl. The eluate contained full-length scNS3/4A >90% pure, as judged by denaturing polyacrylamide gel electrophoresis and Western blotting. The eluate was concentrated to ~500  $\mu$ L, flash frozen in 20  $\mu$ L aliquots in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ . The same procedure was followed for all the full-length scNS3/4A constructs and no significant difference in expression and purification profiles had been observed. In addition, purified protein exhibited no discernible DNase and RNase contamination, within the timescale of helicase assays (data not shown).

#### *Protease Assays*

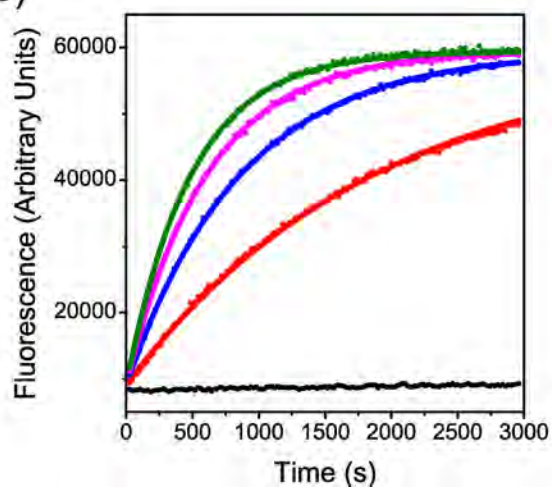
Depsipeptide cleavage assays were performed in a final volume of 60  $\mu$ L containing protease assay buffer (50 mM Tris, 2.5% glycerol, 0.1% O $\beta$ G, 5 mM TCEP, 1% DMSO, pH 7.5) and 1-20 nM scNS3/4A in black 96-well flat bottom non-binding surface half-area plates (Corning) at room temperature. The protein was pre-incubated in the reaction mixture for 30 min and the reaction was initiated by injecting RET-S1 substrate (Anaspec) (Figure II.1a) [32] with a final concentration of 250 nM in the reaction mixture. The fluorescence output was measured kinetically for at least an hour using an EnVision plate reader (Perkin Elmer) with excitation at 340 nm and emission at 492 nm. Each dilution series was repeated five times. Active site titrations with the high affinity protease inhibitor ITMN-191 [33] were performed for each experiment to ensure accurate active enzyme concentration.

**Figure II.1 – Protease cleavage assay.** (a) Depsipeptide substrate RET-S1 (Anaspec), based on NS4A-NS4B cleavage junction. This substrate is labeled with the FRET pair EDANS and DABCYL and contains an ester bond instead of an amide bond between P1 and P1' moieties to facilitate faster cleavage. (b) Upon cleavage, the FRET pair is liberated and the increase in fluorescence intensity is recorded kinetically (black – no protein control, red to magenta – increasing protein concentration). Progress curves were fit to first order rate equations. (c) Obtained  $k_s$  values are plotted against enzyme concentration and catalytic efficiencies were calculated from the concatenate linear fit.

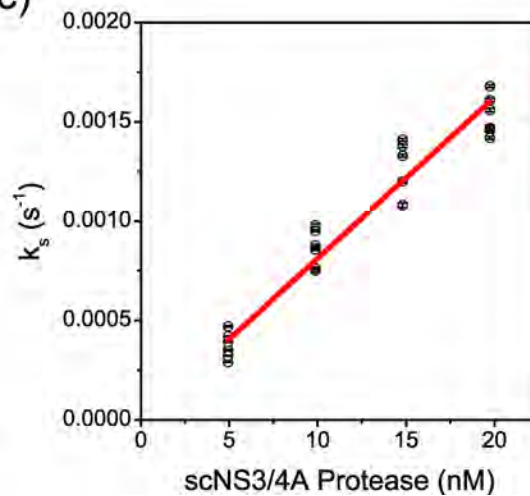
(a)



(b)



(c)



The resulting progress curves were processed using OriginPro 8 (Origin Labs). Progress curves for each construct were fitted globally to Equation (1) (Figure II.1b).

$$F = F_f - Ae^{-k_s t} \quad (1)$$

In Equation (1),  $F$  is the fluorescence,  $F_f$  is the final fluorescence,  $A$  is the amplitude of the reaction which is shared between different datasets and  $k_s$  is the specificity constant [34]. The obtained  $k_s$  values were plotted against enzyme concentration and fitted linearly using all the data from different experiments (*i.e.* concatenate fitting) to Equation (2) (Figure II.1c) [34-36].

$$k_s = \left( \frac{k_{cat}}{K_M} \right) [E]_T \quad (2)$$

In Equation (2),  $k_{cat}/K_M$  is the catalytic efficiency and  $[E]_T$  is the total enzyme concentration.

#### *DNA Unwinding Assay*

The details of the DNA unwinding assay were previously described (Figure II.2a) [37, 38]. Briefly, the assay was performed in a final volume of 60  $\mu$ L containing helicase assay buffer (25 mM MOPS, 1.25 mM  $MgCl_2$ , 0.05 % Tween-20, 0.005 mg/mL BSA, 2.5 mM TCEP, pH 6.5), 12.5 nM dsDNA substrate and 2-200 nM of the NS3 helicase constructs in white flat bottom 96-well half area plates (Corning) at 25°C. The reaction was initiated by injecting a final concentration of 1 mM ATP in multiple turnover conditions or the mixture of 1 mM ATP and 1  $\mu$ M dT<sub>15</sub> DNA trap. The decrease in Fluorescence intensity was monitored using a Varioskan plate reader (Thermo Scientific) with excitation at 643 nm and emission at 667 nm for multiple turnover conditions, and a

FLUOstar Omega plate reader (BMG Labtech) with excitation at 640 nm and emission at 665 nm.

Resulting progress curves were processed using Prism 5 (GraphPad). The progress curves for each construct were fitted individually to Equation (3) (Figure II.2c).

$$F = (F_0 - A) + Ae^{-k_{obs}t} \quad (3)$$

In Equation (3),  $F$  is the fluorescence,  $F_0$  is the initial fluorescence,  $A$  is the amplitude of the reaction and  $k_{obs}$  is the observed first-order rate constant. The rates of the reactions were obtained from Equation (4) (Figure II.2e).

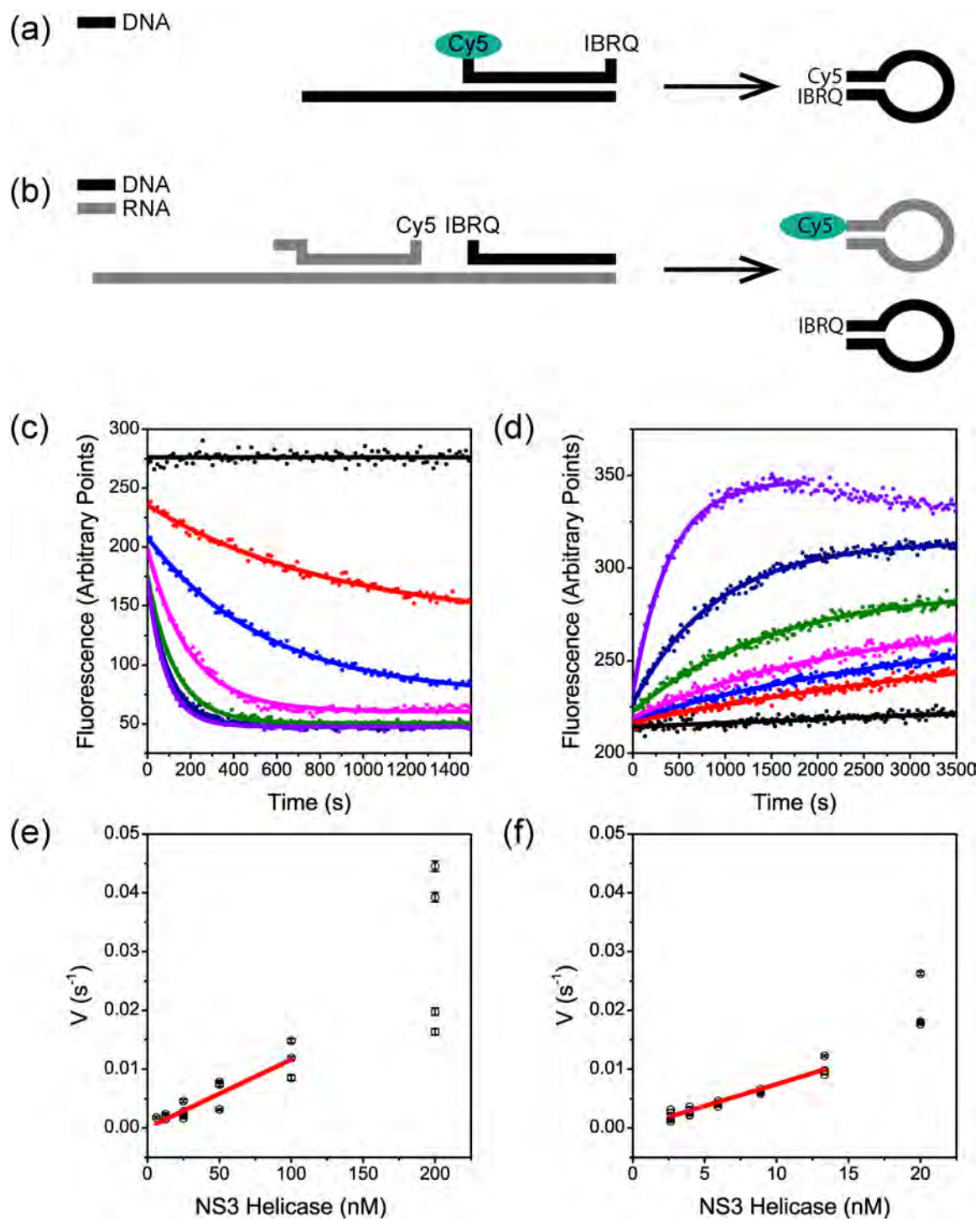
$$V = k_{obs} \left( \frac{A}{F_0 - F_\infty} \right) [S]_T \quad (4)$$

In Equation (4),  $F_\infty$  is the fluorescence when all the substrate is unwound and  $[S]_T$  is the total substrate concentration. Calculated velocities from individual experiments for a given construct were plotted against enzyme concentration and linear portions of these curves were fitted linearly with concatenate fitting (Figure II.2e). Slopes of these lines were reported as specific activities ( $V/E$ ).

#### *RNA Unwinding Assay*

The details of RNA unwinding assay were previously described (Figure II.2b) [38]. Briefly, RNA unwinding assays were performed in a final volume of 60  $\mu$ L containing helicase assay buffer, 25 nM RNA-DNA hybrid substrate and 1-20 nM of the scNS3/4A constructs or 20-150 nM of NS3hel in white flat bottom 96-well half area plates (Corning) at 25°C. The reaction mix was preincubated for 30 min and subsequently initiated by injecting a final concentration of 1 mM ATP. The increase in

**Figure II.2 – DNA and RNA unwinding assays.** (a) dsDNA substrate for DNA unwinding assays consists of a top strand labeled with Cyanine 5 (Cy5) – Iowa Black RQ (IBRQ) FRET pair in both ends and a longer bottom strand. Upon strand separation, the top strand self anneals and Cy5 fluorescence is quenched. (b) dsRNA-DNA hybrid substrate for RNA unwinding assays consists of an RNA Cy5-labeled top strand, a DNA IBRQ-labeled top strand and a long RNA bottom strand. Upon unwinding, both top strands are liberated and self-anneal, resulting in Cy5 fluorescence. (c) Loss of fluorescence in DNA unwinding reaction and (d) gain of fluorescence in RNA unwinding reaction are recorded kinetically and fit to Equations (3) and (5) respectively (black – no protein control, red to magenta – increasing protein concentration). Linear portions of (e) DNA and (f) RNA unwinding reactions are fit concatenately, using data from all independent datasets to obtain specific velocities (*i.e.*  $V/E$ ).



fluorescence was monitored using a Varioskan plate reader (Thermo Scientific) with excitation at 643 nm and emission at 667 nm.

Resulting progress curves were processed using Prism 5 (GraphPad). The fitting scheme is the same as the DNA unwinding assay with the following differences. The progress curves were fit to Equation (5) (Figure II.2d).

$$F = (F_0 + A) - Ae^{-k_{obs}t} \quad (5)$$

The velocities were calculated using Equation (6) (Figure II.2f).

$$V = k_{obs} \left( \frac{A}{F_{\infty} - F_0} \right) [S]_T \quad (6)$$

#### *DNA Binding Assay*

The details of fluorescence anisotropy-based DNA binding assays were previously described [39]. Briefly, DNA binding assays were performed in a final volume of 60  $\mu$ L containing helicase assay buffer, 5 nM Cy5-dT<sub>15</sub> DNA and 1-100 nM NS3 helicase constructs in black flat bottom 96-well half area plates (Corning) at 25°C. The protein-DNA mixture was incubated for 30 min and the fluorescence polarization was measured using an Infinite M1000 plate reader (Tecan) with excitation at 635 nm and emission at 667 nm.

The resulting anisotropies were plotted against enzyme concentrations to obtain binding isotherms which, were subsequently fitted globally using Prism 5 to Equation (7).

$$R = R_0 + \frac{R_{max}}{2[S]_T} \left( \left( K_d + \frac{[E]}{n} + [S]_T \right) - \sqrt{\left( K_d + \frac{[E]}{n} + [S]_T \right)^2 - 4 \frac{[E]}{n} [S]_T} \right) \quad (7)$$



In Equation (7),  $R$  is the anisotropy,  $R_0$  is the background anisotropy,  $R_{max}$  is the maximum anisotropy,  $n$  is the stoichiometry of binding,  $[E]$  is the enzyme concentration,  $[S]_T$  is the total substrate concentration and  $K_d$  is the dissociation constant for DNA, which is shared globally.

#### *Poly(U)-Stimulated ATPase Assay*

The details of the modified malachite green assay were described previously [11]. Briefly, ATPase assays were performed in a final volume of 50  $\mu\text{L}$  containing helicase assay buffer, 0-10 mg/mL poly(U) RNA and 5 nM NS3 helicase constructs at 25°C. The reaction was initiated by rapid mixing of 1 mM ATP (final concentration) and incubated for 10 min. 25  $\mu\text{L}$  of this reaction was quenched in 200  $\mu\text{L}$  of malachide green reagent (3 vol of 0.045% (w/v) malachite green, 1 vol of 4.2% ammonium molybdate in 4 N HCl, 0.01 vol of 10% Tween-20) in clear flat bottom 96-well plates (Corning), immediately followed by the addition of 25  $\mu\text{L}$  of 34% sodium citrate. Color was allowed to develop for 15 minutes and the absorbance was read using a VarioSkan plate reader (Thermo Scientific) at 630 nm. The absorbance values were converted to the inorganic phosphate concentrations using a phosphate standard curve and converted to rates by dividing by incubation times.

Resulting rates were plotted against uracyl concentrations and fitted globally using Prism 5 to Equation (8).

$$\frac{V}{[E]_T} = \frac{k_{cat}[U]}{K_{RNA} + [U]} \quad (8)$$

In Equation (8),  $V$  is the rate of ATP hydrolysis,  $[E]_T$  is the total enzyme concentration,  $[U]$  is the uracyl concentration and  $K_{RNA}$  is the concentration of uracyl needed to stimulate the ATP hydrolysis to half of the maximum velocity, which is shared globally.

## Results

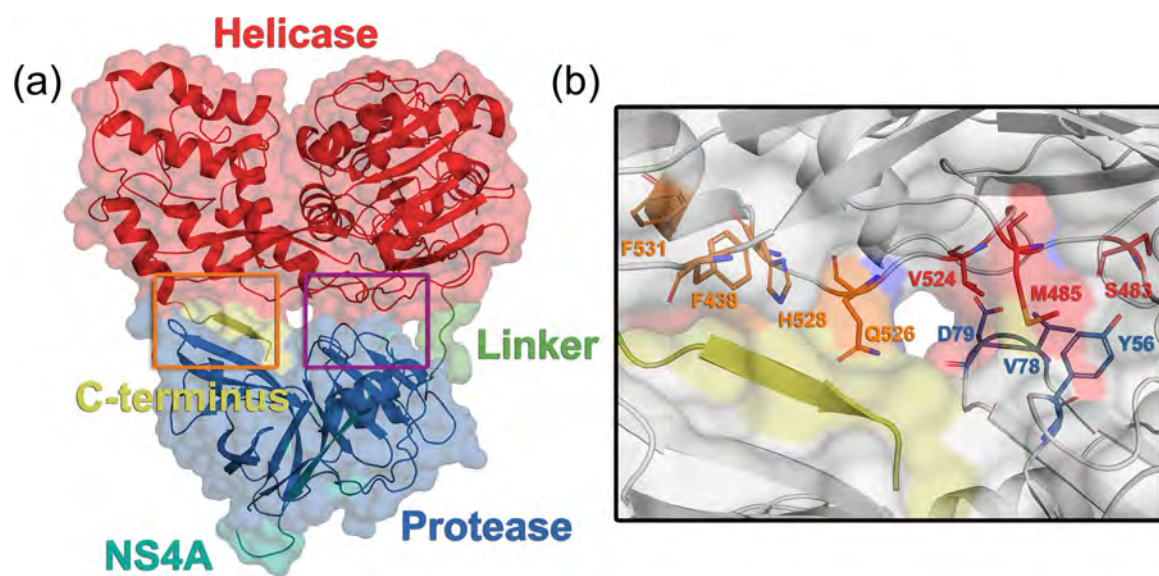
### *Structural Analysis of the Domain-Domain Interfaces*

The single chain NS3/4A (scNS3/4A) protein that is utilized in this study is comprised of two functional domains. The portion of NS3 that forms the N-terminal protease folds into two subdomains with  $\beta$ -barrel folds – N-terminal (N-SD) and C-terminal (C-SD) subdomains of NS3 protease (scNS3/4A<sub>pro</sub>). The serine protease active site is in a cleft between protease subdomains. The C-terminal helicase portion of NS3 (NS3<sub>hel</sub>) is made of three subdomains, two N-terminal recA-like motor domains (subdomains SD1 and SD2) and a third  $\alpha$ -helical domain (subdomain SD3). In order to analyze the extent of non-covalent contacts between the helicase and protease regions, the composition of the buried surface area in the domain-domain interface was evaluated. The crystal structures of scNS3/4A show extensive contacts between the protease and helicase domains (Figure II.3a) [21-23]. Approximately 900 Å<sup>2</sup> is buried between two domains in two discrete interfaces, separated by the protease active site.

The first interface is formed by the scNS3/4A<sub>pro</sub> C-SD and will be referred to here as the “substrate interface” (Figure II.3a, orange box) because this interface is formed with the C-terminus of NS3. The NS3 C-terminus is the *cis*-cleavage product of the NS3-NS4A junction, and remains bound in the P-side of the protease active site. The NS3 C-terminus – protease active site interactions have been studied before [30, 40], so we focused on analyzing contacts between the protease active site cleft and the helicase domains. Residues Q526 and H528 in NS3<sub>hel</sub> SD3 interact with the NS3 C-terminus directly, whereas residues F438 and F531 form a groove to accommodate one of the C-

**Figure II.3 – Structure of scNS3/4A and the domain-domain interface. (a)**

Cartoon representation of HCV scNS3/4A (PDB ID: 1CU1) [21]. Protease domain, helicase domain, linker, C-terminus and NS4A beta helix is colored in blue, red, green, yellow and cyan respectively. C-terminus – protein “substrate” (orange box) and protease – helicase “direct” (magenta box) interfaces are encased in rectangles. In this structure, helicase domain covers the solvent exposed active site of the protease and the C-terminus of the protein occupies the P-side of protease active site. (b) The domain-domain interface formed between the protease and the helicase is magnified. The residues represented in ball-and-stick figures and annotated were used in mutagenesis and biochemical characterization of the interface.



terminal residues. (Figure II.3b – left).

The second interface, which will be referred to here as the “direct interface”, (Figure II.3a, magenta box) is formed between scNS3/4Apro N-SD and NS3hel SD2 and SD3. Residues from helicase and the protease domains form direct contacts, and  $\sim 400 \text{ \AA}^2$  is buried in this interface mainly between Y56, V78, D79 from the protease and S483, M485 and V524 from the helicase. This interface is mostly hydrophobic, with Y56 in the core covered by M485 and V78, and in favorable hydrogen bonding distance with S483. These contacts are further complemented by V524, which forms hydrophobic contacts with V78, and D79 (Figure II.3b – right).

#### *The Dynamic Interaction between the Domains*

To investigate the dynamic coupling between the two domains and to evaluate the role of the interfaces, nanosecond-scale molecular dynamics (MD) simulations were performed on scNS3/4A (PDBID: 1CU1) [21] and three *in silico* deletion constructs, which are  $\Delta$ Lin ( $\Delta$ 184-186 – linker deletion construct),  $\Delta$ C-term ( $\Delta$ 626-631, C-terminus deletion construct) and  $\Delta$ Double ( $\Delta$ 184-186, 626-631 – deletion construct with both linker and C-terminus deletions). Linker deletion was made to break the covalent linkage between the helicase and the protease domains, and C-terminus deletion to disrupt the association of the protease and helicase through the strong substrate product-protease association.

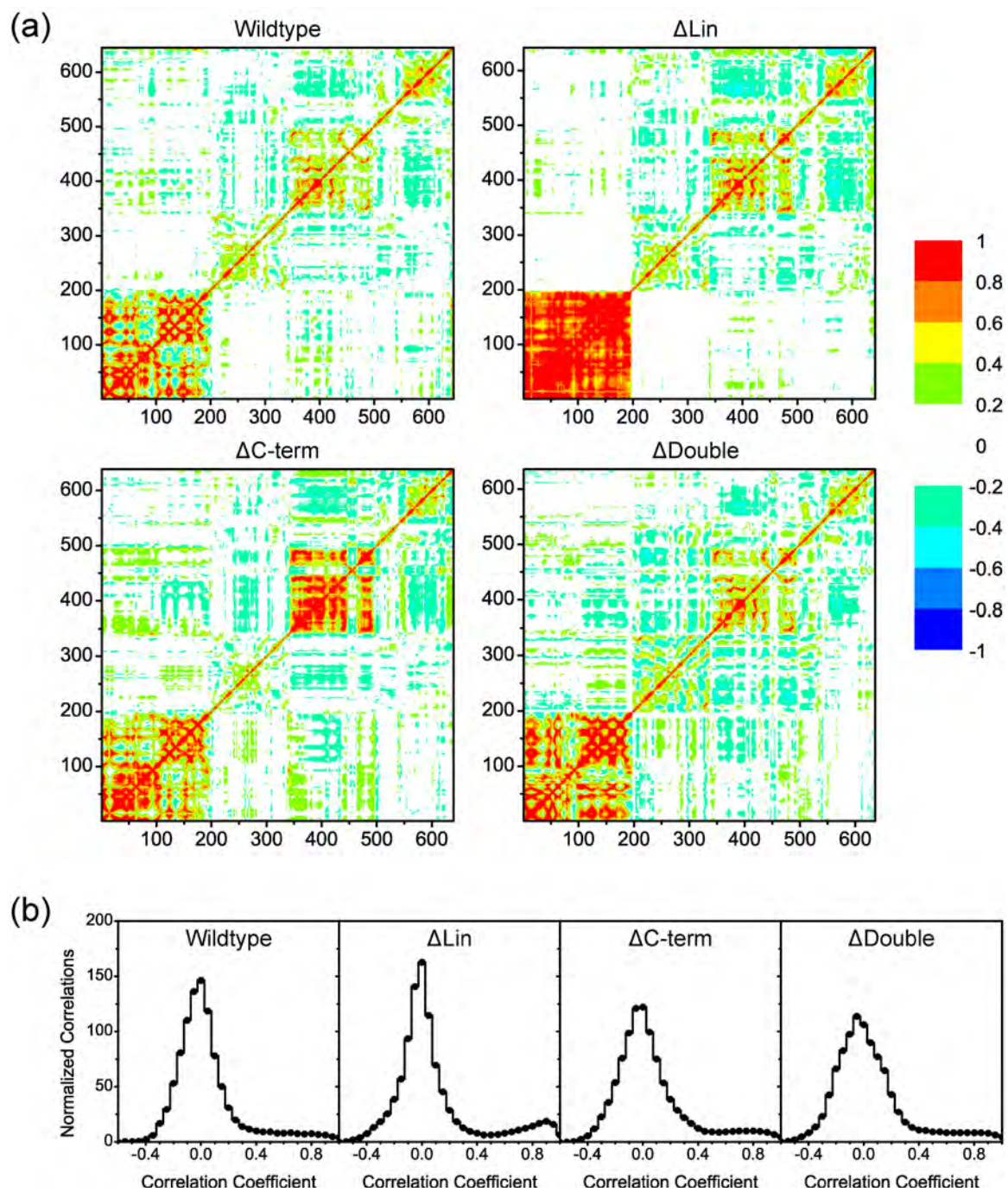
Dynamic cross-correlations were calculated from resulting trajectories for each deletion construct (Figure II.4a). Dynamic cross-correlations between two atoms indicate how these atoms move in the Cartesian space relative to each other and range between -1

(perfect anti-correlation) and 1 (perfect correlation). These correlations also signify the extent of interactions between atom pairs and regions of the protein in general. For all the constructs, the correlations were mainly distributed around 0 (*i.e.* no correlation) symmetrically, and correlations larger than 0.5 (*i.e.* the “shoulder”) were mostly *intra*-domain correlations (Figure II.4b). Remaining interdomain correlations were investigated further to evaluate the significance of the interface and the modulation of domain-domain contacts due to deletions.

Correlation distributions were generated and fitted to Gaussian function for residue pairs across domains (Figure II.5 – bottom, see Table II.1 for fitting parameters) excluding linker and C-terminus, and significantly correlating residue pairs were visualized (Figure II.5 – top, see Supplemental Methods for selection criteria, and Table II.1 for cutoff values). The width of the distributions (Table II.1) indicates the extent of correlation – wider distributions exhibit stronger interdomain correlations and *vice versa*. Compared to the wildtype, the distributions of interdomain correlations were narrower in  $\Delta$ Lin and wider in  $\Delta$ C-term and  $\Delta$ Double, suggesting that the two domains were less correlated for  $\Delta$ Lin and more for the latter two. Changes in correlation distributions were also evident in correlation maps where more residue pairs exhibited correlations larger than 0.2 and smaller than  $-0.2$  (Figure II.4a). Most of the significant correlations clustered around the direct interface. Interdomain correlations were stronger and more focused around the direct interface for the C-terminus deletion constructs ( $\Delta$ C-term and  $\Delta$ Double), whereas correlated motion between distal residues was more dominant for the

**Figure II.4 – Correlation maps for *in silico* deletion constructs and distributions of total correlations.** (a)  $C_{\alpha}$ - $C_{\alpha}$  correlation maps for each construct. Axes represent residue indexes and coloring represents strength of correlations. (b) Distributions for total correlations. Correlations were distributed around 0 (*i.e.* no correlation) symmetrically with “shoulder-like” regions for correlations larger than 0.5, which are *intra*-domain correlations (correlations within domains, either between or within individual subdomains of protease and helicase).



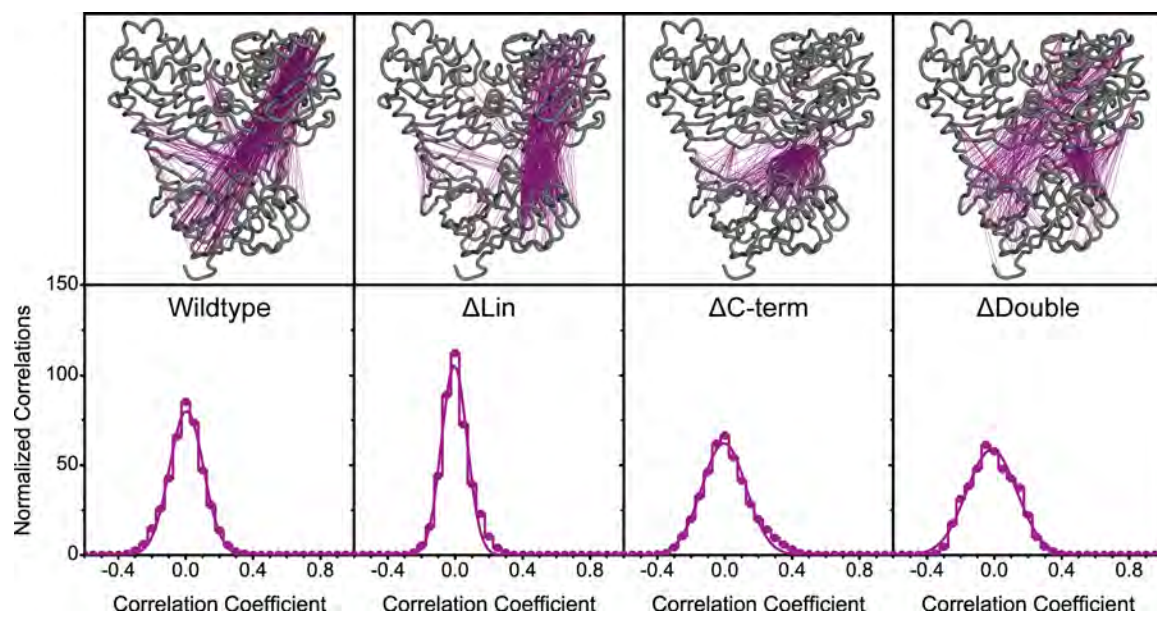


**Table II.1** – Parameters for correlation distributions for *in silico* NS3/4A constructs and visualization cutoffs.

<b>Construct</b>	<b><math>\mu</math> (Mean)</b>	<b><math>\sigma</math> (SD)<sup>a</sup></b>
	<b>Cutoff<sup>b</sup> (Number of Pairs<sup>c</sup>)</b>	
<b>Wildtype</b>	0.005±0.002	0.101±0.001
	<b>±0.33 (375)</b>	
<b><math>\Delta</math>Lin</b>	-0.005±0.002	0.076±0.002
	<b>±0.28 (340)</b>	
<b><math>\Delta</math>C-term</b>	-0.005±0.003	0.132±0.003
	<b>±0.44 (326)</b>	
<b><math>\Delta</math>Double</b>	-0.015±0.003	0.148±0.003
	<b>±0.35 (362)</b>	

<sup>a</sup> Standard deviation of the Gaussian fit to distributions  
“Cutoff” denotes the visualization cutoffs used to create Figure 2 and “±” means  
<sup>b</sup> negative correlations smaller than the cutoff and positive correlations larger than the cutoff were visualized  
<sup>c</sup> “Number of Pairs” denotes the total count of residue pairs that were selected for visualization

**Figure II.5 – Interdomain correlations for *in silico* deletion constructs.** Visual representation and distribution of interdomain correlations. For each of the *in silico* constructs, selected significant protease – helicase correlations were visualized (top panels) and the correlation distributions were plotted (bottom panels).



others, which can be explained as the propagation of the intra and inter subdomain correlation network through the interfaces.

Additionally, distance difference distributions between  $C_{\alpha}$  atoms of selected residues (Figure II.6a) across protease and helicase domains were generated (Figure II.6b). These distributions indicate the change in  $C_{\alpha}$ - $C_{\alpha}$  distances of a given pair across simulation time, reflecting the relative freedom of movement between the domains. For the direct interface, the distributions for buried residue pairs (Y56-S483, Y56-M485) are similar. However, the movement of outer residue pairs (V78-V524, D79-V524) is more constricted for  $\Delta C$ -term and wider for linker deletion constructs. The wildtype construct exhibited intermediary spread. These results correlate with the cross-correlation distributions. For the substrate interface, all distance distributions except C159-F438 were constricted for wildtype and  $\Delta Lin$ , and wider for the other two constructs.

Taken together, removal of C-terminus caused stronger correlations through the direct interface, but removal of the linker caused the exact opposite effect. Removal of both the linker and the C-terminal caused the protein to behave similar to C-terminus deletion constructs, suggesting coupling through the direct interface dominates in the complex. The wild-type construct exhibited the combined characteristics of both deletions. In either case, the deletions affect the relative dynamics of the two domains acting through the interfaces. The C-terminus deletion confers freedom of movement between the domains at the substrate interface while strengthening the association through the direct interface and vice versa, indicating that both interfaces aid in the dynamic coupling of the domains, but work in an anti-coordinated manner.

*Design of scNS3/4A Amino Acid Substitutions to Disrupt the Protease/Helicase Interface*

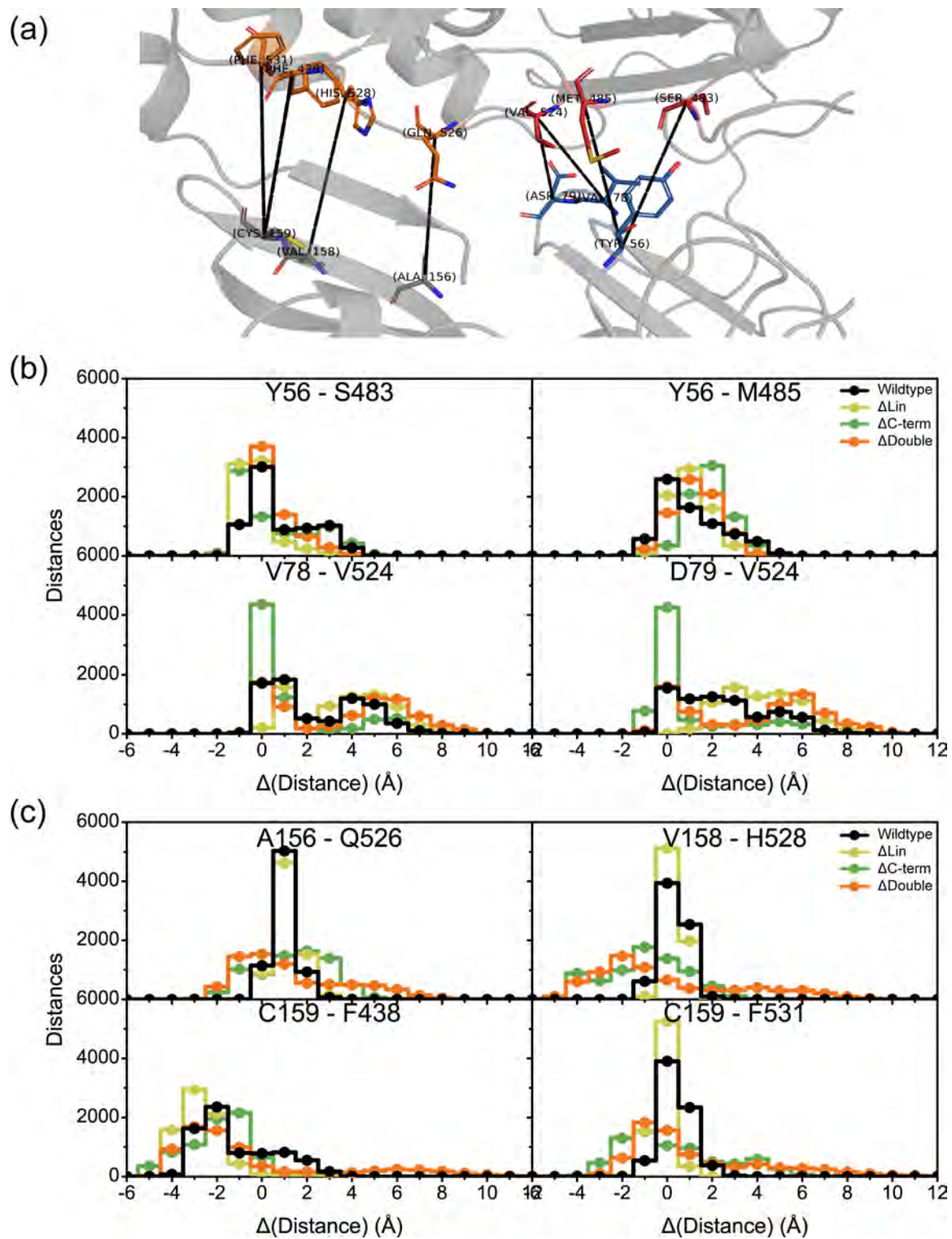
In order to evaluate the functional relevance of the interfaces, a total of ten residues from either interface were mutated to alanine in the plasmid expressing scNS3/4A to yield two sets of proteins with amino acid substitution in domain interfaces.

The aim of the first set of substitutions was to disrupt the “direct” interface. Four proteins in this set had alterations in the protease region. Three of these “protease mutants” had one of three single point mutations Y56A, V78A, D79A, and one had all three substitutions in scNS3/4Apro N-SD (Y56A, V78A and D79A *i.e.* Triple Prot) (Figure II.3b – blue). Four proteins had mutations in NS3hel SD2 and SD3. Three of these “helicase mutants” had one of the following point mutations – S483A, M485A, V524A, and one protein had all three substitutions (S483A, M485A, and V524A, *i.e.* Triple Hel) (Figure II.3b – red). One protein (called “Hexa”) had all six amino acid substitutions in the interface (Y56A, V78A, D79A, S483A, M485A, V524A).

The aim of the second set of altered scNS3/4A proteins was to disrupt the “substrate” interface formed between scNS3/4A C-SD, NS3hel SD3 and the NS3 C-terminus, with a goal to further focus on the role of helicase in substrate recognition. The proteins in this second set had one of the following amino acid substitutions – F438A, Q526A, H528A and F531A (Figure II.3b – orange). All mutant proteins were expressed, purified and biochemically characterized for protease, DNA/RNA unwinding, DNA binding and RNA-stimulated ATPase activities; in comparison with the scNS3/4A and the NS3 fragments lacking either the protease or helicase regions (scNS3/4Apro and, NS3hel).

**Figure II.6 – Distance distributions around the interfaces.** (a) Residues used for distance calculations. Each residue pair consists of one residue from the protease and one residue from the helicase. (b) Distance distributions for the residue pairs in the direct interface. (c) Distance distributions for the residue pairs either side of the C-terminus.





### *Protease Cleavage Efficiencies*

The catalytic efficiencies ( $k_{cat} / K_M$ ) of scNS3/4A and its site directed mutants were determined using a FRET-based cleavage assay (Figure II.7, Table II.2). All proteins harboring substitutions in the direct interface had higher catalytic efficiency than scNS3/4A except Y56A. However, this activation was nullified in the Triple Hel. Y56A containing mutants (Y56A, Triple Prot, Hexa) were less efficient in cleavage assays; Triple Prot performing the worst followed by Hexa and Y56A. Y56 is near the protease active site, so this residue may play a direct role in proteolysis. In addition, Y56 is also a reported drug resistance site [41], which may explain loss of activity for Y56A. The catalytic efficiencies of the proteins with substitutions in the substrate interface were similar to scNS3/4A, except H528A, which was more active. The truncated complex lacking the helicase domain (*i.e.* scNS3/4A<sub>pro</sub>) had an enhanced activity, similar to the interface mutants.

### *Helicase Unwinding Rates for DNA and RNA Substrates*

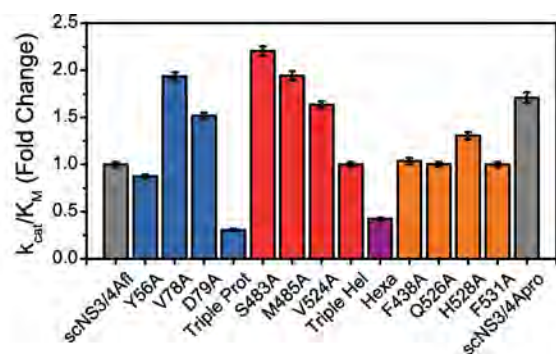
The ability of each of the above proteins to unwind DNA was assessed by measuring initial rates of helicase catalyzed DNA unwinding upon ATP addition at eight different protein concentrations. In order to compare the specific activity of each protein, initial rates were divided by protein concentration and averaged. These specific activities (V/E) were determined using a DNA molecular beacon based unwinding assay performed both in the presence (Figure II.8a, dark bars) and absence of an enzyme trap (Figure II.8a, light bar) under multiple and single turnover conditions (Table II.3). Under multiple turnover conditions, the same amount of NS3hel unwound DNA about 6 times faster

compared to wildtype scNS3/4A. All scNS3/4A proteins harboring amino acid substitutions designed to disrupt the helicase-protease interface unwound DNA with rates between those seen with scNS3/4A and NS3hel. Y56A, V78A and M485A were each significantly more active (more than 3 fold) than scNS3/4A. In addition, Triple Hel performed 2 fold better compared to NS3hel, unwinding DNA about 10 times faster than wildtype scNS3/4A. Similar to the substitutions that disrupt the direct interface, the unwinding rates catalyzed by proteins harboring substitutions designed to disrupt the substrate interface (*i.e.* those in the NS3hel SD3) were also between those seen with the full-length protein and the isolated helicase.

When the ability of different scNS3/4A constructs to unwind DNA in the presence of an enzyme trap was compared, rates catalyzed under single turnover conditions were generally similar to those seen under multiple turnover conditions (*i.e.* in the absence of trap) with two noteworthy exceptions. Both F438A and F531A did not show discernible unwinding activity in the presence of a trap. Both of these residues are conserved, and F438 was previously identified as a critical residue for unwinding activity in NS3hel [42].

In addition to DNA unwinding rates, RNA unwinding rates and specific activities (V/E) were determined for all NS3 helicase constructs using a duplex RNA-DNA hybrid substrate in which the helicase must separate an RNA duplex to cause an increase in substrate fluorescence (Figure II.8a – bottom panel, Table II.3) [38]. Compared to NS3hel, scNS3/4A unwound RNA about 10 times more rapidly than NS3hel, confirming the prior observation that full-length NS3/4A unwinds RNA faster than DNA, while

**Figure II.7 – Proteolysis catalytic efficiencies for interface mutants.** Protease cleavage efficiencies of different scNS3/4A protease constructs. The catalytic efficiencies were normalized against full-length scNS3/4A (abbreviated as scNS3/4Afl here). Coloring represents different regions mutated (blue – protease mutations in direct interface, red – helicase mutations in direct interface, magenta – Hexa, orange – helicase mutations in substrate interface, gray – wildtype protease and protease – helicase complex). Errors reported are the standard errors of the mean (n=6).



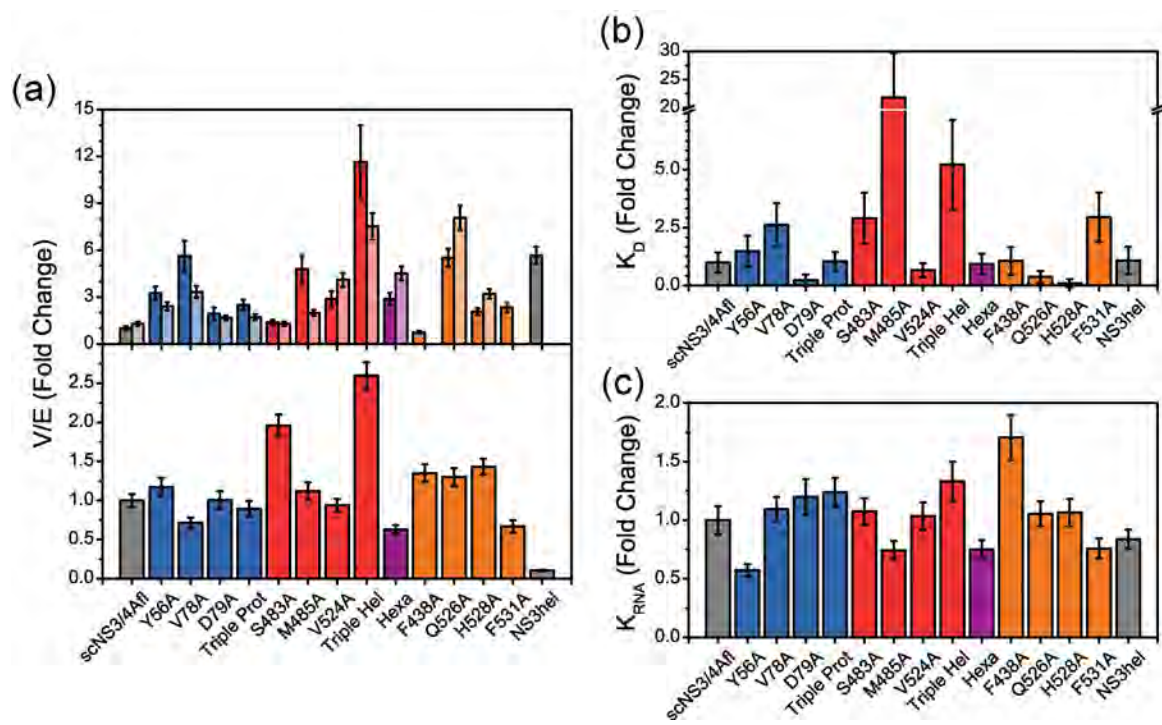
**Table II.2** – Catalytic efficiencies of RET-S1 cleavage for different NS3/4A protease constructs. Errors reported are the standard errors of the mean (n=6).

<b>Construct</b>	<b><math>k_{\text{cat}}/K_M</math> (<math>\text{s}^{-1}\mu\text{M}^{-1}</math>)<sup>a</sup></b>
<b>scNS3/4Afl</b>	$0.082 \pm 0.002$ ( $1.00 \pm 0.03$ )
<b>Y56A</b>	$0.071 \pm 0.001$ ( $0.87 \pm 0.02$ )
<b>V78A</b>	$0.158 \pm 0.001$ ( $1.94 \pm 0.04$ )
<b>D79A</b>	$0.124 \pm 0.001$ ( $1.52 \pm 0.03$ )
<b>Triple Prot</b>	$0.025 \pm 0.001$ ( $0.31 \pm 0.01$ )
<b>S483A</b>	$0.180 \pm 0.002$ ( $2.21 \pm 0.05$ )
<b>M485A</b>	$0.159 \pm 0.002$ ( $1.94 \pm 0.05$ )
<b>V524A</b>	$0.133 \pm 0.001$ ( $1.64 \pm 0.04$ )
<b>Triple Hel</b>	$0.082 \pm 0.001$ ( $1.00 \pm 0.03$ )
<b>Hexa</b>	$0.035 \pm 0.001$ ( $0.43 \pm 0.01$ )
<b>F438A</b>	$0.085 \pm 0.002$ ( $1.04 \pm 0.03$ )
<b>Q526A</b>	$0.082 \pm 0.001$ ( $1.00 \pm 0.03$ )
<b>H528A</b>	$0.107 \pm 0.002$ ( $1.31 \pm 0.04$ )
<b>F531A</b>	$0.082 \pm 0.002$ ( $1.00 \pm 0.03$ )
<b>scNS3/4Apro</b>	$0.140 \pm 0.004$ ( $1.71 \pm 0.06$ )

<sup>a</sup> Values denoted in parenthesis are fold changes with respect to scNS3/4Afl.

**Figure II.8 – Kinetic parameters for the multiple activities of the NS3hel constructs.** (a) Multiple turnover DNA (top panel – dark bars), single turnover DNA (top panel – light bars) and multiple turnover RNA (bottom panel) unwinding activities; (b) ssDNA binding affinities and (c) extent of ATPase stimulation by poly(U) RNA are represented here. All activities were normalized against scNS3/4A (abbreviated as scNS3/4Afl here). Coloring represents different regions mutated (blue – protease mutations in direct interface, red – helicase mutations in direct interface, magenta – Hexa, orange – helicase mutations in substrate interface, gray – wildtype protease and protease – helicase complex). Errors reported are the standard errors of the mean. (n=3 for DNA unwinding and DNA binding assays, n $\geq$ 4 for RNA unwinding assays, and n $\geq$ 3 for ATPase assays).





**Table II.3** – Specific activities of unwinding dsDNA and dsRNA beacons for different NS3 helicase constructs. Errors reported are the standard errors of the mean. (n=3 for DNA unwinding assays, n $\geq$ 4 for RNA unwinding assays).

Construct	V/E (*1000 s <sup>-1</sup> ) <sup>a</sup>		
	DNA Unwinding (Multiple Turnover)	DNA Unwinding (Single Turnover)	RNA Unwinding
<b>scNS3/4Afl</b>	0.28 ± 0.02 (1.00 ± 0.12)	0.37 ± 0.02 (1.29 ± 0.13)	0.74 ± 0.04 (1.00 ± 0.08)
<b>Y56A</b>	0.92 ± 0.09 (3.25 ± 0.44)	0.68 ± 0.05 (2.41 ± 0.27)	0.87 ± 0.07 (1.17 ± 0.12)
<b>V78A</b>	1.60 ± 0.24 (5.63 ± 0.98)	0.95 ± 0.07 (3.34 ± 0.38)	0.53 ± 0.04 (0.71 ± 0.07)
<b>D79A</b>	0.55 ± 0.10 (1.93 ± 0.39)	0.47 ± 0.02 (1.67 ± 0.16)	0.75 ± 0.07 (1.01 ± 0.11)
<b>Triple Prot</b>	0.81 ± 0.12 (2.49 ± 0.34)	0.49 ± 0.03 (1.73 ± 0.19)	0.70 ± 0.04 (0.90 ± 0.10)
<b>S483A</b>	0.70 ± 0.08 (1.38 ± 0.17)	0.36 ± 0.03 (1.27 ± 0.14)	0.66 ± 0.06 (1.96 ± 0.14)
<b>M485A</b>	0.39 ± 0.03 (4.78 ± 0.92)	0.56 ± 0.02 (1.99 ± 0.19)	1.45 ± 0.07 (1.12 ± 0.11)
<b>V524A</b>	1.35 ± 0.23 (2.87 ± 0.50)	1.16 ± 0.07 (4.11 ± 0.44)	0.83 ± 0.07 (0.94 ± 0.08)
<b>Triple Hel</b>	3.30 ± 0.60 (11.64 ± 2.35)	2.14 ± 0.15 (7.53 ± 0.84)	1.92 ± 0.06 (2.59 ± 0.17)
<b>Hexa</b>	0.82 ± 0.09 (2.88 ± 0.39)	1.28 ± 0.06 (4.52 ± 0.45)	0.47 ± 0.03 (0.63 ± 0.05)
<b>F438A</b>	0.21 ± 0.02 (0.75 ± 0.09)	0 <sup>b</sup>	1.00 ± 0.06 (1.35 ± 0.11)
<b>Q526A</b>	1.56 ± 0.08 (5.51 ± 0.56)	2.29 ± 0.10 (8.08 ± 0.79)	0.96 ± 0.07 (1.30 ± 0.12)
<b>H528A</b>	0.58 ± 0.04 (2.05 ± 0.22)	0.91 ± 0.04 (3.20 ± 0.31)	1.06 ± 0.04 (1.43 ± 0.10)
<b>F531A</b>	0.66 ± 0.06 (2.32 ± 0.29)	0 <sup>b</sup>	0.49 ± 0.05 (0.67 ± 0.08)
<b>NS3hel</b>	1.61 ± 0.07 (5.67 ± 0.55)	N/A <sup>c</sup>	0.08 ± 0.01 (0.10 ± 0.01)

<sup>a</sup> Values denoted in parenthesis are fold changes compared to scNS3/4Afl MHBA multiple turnover rate for both MBHAs and scNS3/4Afl SBHA rate for SBHA

<sup>b</sup> No detectable unwinding activity was observed.

<sup>c</sup> Single turnover MBHA has not been performed for this construct.

truncated NS3 lacking the protease prefers DNA to RNA [16, 17]. Proteins harboring mutations in the helicase-protease interface unwound RNA with rates that were more similar to each other than those seen in DNA unwinding assays. RNA unwinding activities for most of the proteins with direct interface mutations were similar to those seen with of scNS3/4A (*i.e.* within a 0.75 – 1.25 fold range), with S483A and Triple Hel being significantly more active and Hexa slightly less active. All proteins harboring substitutions designed to disrupt the substrate interface were slightly more active compared to scNS3/4A, except F531A. Hence, presence of the protease domain enhanced the RNA helicase activity, and destabilizing the interface did not affect RNA unwinding significantly. As discussed later, we interpret these results to mean that the RNA unwinding activity of the helicase is modulated by the protease more directly than DNA unwinding, which depends more on interface contacts.

#### *Helicase Single-Stranded DNA Affinities*

In order to act as a helicase, NS3 must first load on its substrate by tightly binding single stranded nucleic acid. Fluorescence anisotropy based binding assays were performed to assess the ability of scNS3/4A, its various mutants, and NS3hel to load on substrates, which gave dissociation constants ( $K_d$ ) for ssDNA binding (Figure II.8b, Table II.4). In this assay, scN3/4A and NS3hel both exhibited similar affinities to ssDNA. Differences between the various proteins were greater than those seen in either protease or helicase assays, and dissociation constants were spread over a wide range (~100 pM – ~10 nM). Compared to scNS3/4A; V78A, S483A, M485A, Triple Hel and

F531A bound ssDNA more weakly while D79A, Q526A and H528A bound ssDNA more tightly.

#### *Helicase RNA-Stimulated ATP Hydrolysis Rates*

HCV NS3 helicase utilizes ATP to unwind nucleic acids, and in turn, nucleic acid stimulates NS3 catalyzed ATP hydrolysis. In other words, in the presence of nucleic acids, NS3 helicase hydrolyzes ATP faster. The extent of ATPase activation by nucleic acids follows a behavior similar to Michaelis-Menten kinetics, and kinetic parameters for nucleic acid based activation can be obtained. For this purpose, a concentration gradient of poly(U) RNA was used to stimulate NS3 ATPase, and the extent of ATP hydrolysis was measured (Figure II.8c, Table S4). As was seen with the RNA unwinding results, the extent of ATP hydrolysis stimulation by poly(U) RNA ( $K_{RNA}$ ) for most of the constructs was constricted to a narrow range (0.75 – 1.25 fold); with Y56A exhibiting slightly lower  $K_{RNA}$  and, Triple Hel and Hexa slightly higher. The catalytic turnover rate constants ( $k_{cat}$ ) were also similar, varying in a constricted range (0.9 – 1.6 fold, Table S4). Thus, the ATPase activity is largely independent of the presence of the protease domain and the interface.

**Table II.4** – ssDNA binding constants, RNA-stimulated ATP hydrolysis catalytic rate constants and the extent of RNA stimulation of ATP hydrolysis for different NS3 helicase constructs. Errors reported are the standard errors of the mean (n=3 for DNA binding assays and  $n \geq 3$  for ATPase assays).

Construct	Cy5-dT <sub>15</sub> Binding	RNA-Stimulated ATP Hydrolysis	
	K <sub>d</sub> (nM) <sup>a</sup>	k <sub>cat</sub> (s <sup>-1</sup> ) <sup>a</sup>	K <sub>RNA</sub> (ng poly(U)/ml) <sup>a</sup>
<b>scNS3/4Afl</b>	0.66 ± 0.20 (1.00 ± 0.43)	37.79 ± 4.46 (1.00 ± 0.17)	75.53 ± 6.37 (1.00 ± 0.12)
<b>Y56A</b>	0.98 ± 0.33 (1.48 ± 0.68)	36.12 ± 5.00 (0.96 ± 0.17)	43.29 ± 1.53 (0.57 ± 0.05)
<b>V78A</b>	1.73 ± 0.31 (2.63 ± 0.93)	45.89 ± 4.14 (1.21 ± 0.18)	82.64 ± 3.63 (1.09 ± 0.10)
<b>D79A</b>	0.15 ± 0.16 (0.23 ± 0.25)	46.94 ± 5.92 (1.24 ± 0.21)	90.55 ± 8.51 (1.20 ± 0.15)
<b>Triple Prot</b>	0.69 ± 0.15 (1.05 ± 0.39)	46.71 ± 7.75 (1.24 ± 0.25)	78.26 ± 5.83 (1.24 ± 0.13)
<b>S483A</b>	1.91 ± 0.42 (2.91 ± 1.10)	35.94 ± 3.66 (0.95 ± 0.15)	93.39 ± 5.23 (1.08 ± 0.11)
<b>M485A</b>	14.36 ± 2.73 (21.82 ± 7.86)	60.88 ± 10.47 (1.61 ± 0.34)	81.32 ± 4.99 (0.74 ± 0.08)
<b>V524A</b>	0.44 ± 0.15 (0.67 ± 0.30)	37.04 ± 5.44 (0.98 ± 0.18)	56.14 ± 3.44 (1.04 ± 0.12)
<b>Triple Hel</b>	3.43 ± 0.72 (5.22 ± 1.93)	61.23 ± 13.96 (1.62 ± 0.42)	100.65 ± 9.48 (1.33 ± 0.17)
<b>Hexa</b>	0.61 ± 0.22 (0.93 ± 0.44)	55.07 ± 9.90 (1.46 ± 0.31)	56.73 ± 3.52 (0.75 ± 0.08)
<b>F438A</b>	0.70 ± 0.33 (1.06 ± 0.60)	62.87 ± 9.18 (1.66 ± 0.31)	128.82 ± 9.49 (1.71 ± 0.19)
<b>Q526A</b>	0.25 ± 0.15 (0.38 ± 0.26)	55.50 ± 6.32 (1.47 ± 0.24)	79.71 ± 4.38 (1.06 ± 0.11)
<b>H528A</b>	0.07 ± 0.12 (0.11 ± 0.18)	49.58 ± 4.56 (1.31 ± 0.20)	80.51 ± 5.62 (1.07 ± 0.12)
<b>F531A</b>	1.94 ± 0.37 (2.95 ± 1.06)	33.50 ± 2.25 (0.89 ± 0.12)	57.26 ± 4.38 (0.76 ± 0.09)
<b>NS3hel</b>	0.71 ± 0.32 (1.08 ± 0.59)	43.00 ± 7.31 (1.14 ± 0.24)	63.36 ± 2.93 (0.84 ± 0.08)

<sup>a</sup> Values denoted in paranthesis are fold changes compared to scNS3/4Afl.

## Discussion

In this study, the role of the protease-helicase interface in HCV NS3-NS4A complex was examined using a single chain NS3-NS4A protein (scNS3/4A) whose structure was previously determined at atomic resolution [21]. Dynamic coupling between the helicase and protease portions of scNS3/4A was observed *in silico* in the compact conformation the protein assumes in crystal structures. We experimentally showed that amino acid substitutions that destabilize the helicase-protease interface also affect NS3 catalyzed proteolysis, DNA and RNA unwinding, and the affinity of the protein for DNA. Most substitutions in the interface render the protein more active, presumably because of the liberation of domain-domain attachment. The protease has complex effects on NS3 catalyzed DNA and RNA unwinding, with the protease inhibiting DNA unwinding and stimulating RNA unwinding. Residues in the interface modulate DNA unwinding more than RNA unwinding. In contrast, the extent of ATPase stimulation by RNA was less dependent on the protease domain. Taken together, the data suggest that the interface disruptions engineered between the helicase and protease in the compact conformation caused the protein to achieve maximal activity. These results support the presence of an active conformer other than the compact state – the extended conformation, as has been suggested by others [21, 24, 25], and observed in structures of related proteins from similar viruses [26, 27, 43].

### *Protease and Helicase Domains of HCV NS3 are Dynamically Coupled in Silico*

Even though the interface formed between the protease and helicase domains of HCV NS3 is not as extensive as other reported biologically functional interfaces, where



the buried surface area is around  $1500 \text{ \AA}^2$  and mostly hydrophobic in content [44], significant coupling between protease and helicase domains through the interfaces *in silico* was observed, albeit not as strong as *intra*-domain (*i.e.* within a domain, can be within or between subdomains) couplings. The extent and strength of interdomain correlations is changed when either the protease-helicase linker (residues 184-186) and/or the C-terminus of scNS3/4A (residues 626-631) is deleted, but still, upon deletion of these regions, the majority of the interdomain correlations remain within the neighborhood of the interfaces and extend through the *inter* and *intra*-domain correlation network. The interface formed between the two domains is most probably a transient, dynamic interface. The linker and the C-terminus are strong linkage points that keep the two domains in proximity; hence the interdomain interfaces are strongly associated through these regions. With removal of these linkages, the interfaces close to the deleted regions become “unlinked” and free to explore the otherwise constricted conformational space that is available. However, these deletions do not completely uncouple the domains; they only modulate the strength of correlated motion between domains. In order to elucidate the importance of the interfaces in protease and helicase activity, functional characterizations were performed on scNS3/4A mutants harboring amino acid substitutions designed to disrupt the helicase-protease interfaces.

#### *The Protease is Restricted by the Helicase*

The helicase domain acts as a regulatory mechanism that limits the catalytic efficiency of the protease *in vitro*. For all single amino acid substitutions in the direct interface (except Y56A mutants), increased cleavage activity was observed, whereas the

C-terminus – helicase interface mutants showed similar activity as the full-length protein. The combinatorial mutants nullified the effects of single mutations. These direct interface mutants may destabilize the direct interface, which, in turn, can alter the dynamics of protease – helicase interaction and ease the access of the protease substrate to the protease active site. In the combinatorial mutants, the interface was flattened out; this may be an indication of interface restabilization.

Consistent with interface mutants, the isolated protease domain, which is not hindered by the helicase domain, is more active. Our protease results correlate with earlier results [12, 13]; which, in turn, disagree with a more recent report by Beran & Pyle suggesting that the helicase domain enhances NS3/4A protease activity [14]. Unlike other studies, Beran & Pyle compared an NS3-NS4A complex (which includes the complete NS4A that is processed *in cis* by the NS3/4A protease) with a complex formed by NS3 protease (NS3pro) combined with full-length NS4A, after the two peptides were independently expressed and combined *in vitro*. The lower activity observed for the NS4A-NS3pro used by Beran & Pyle could be explained by product inhibition by the C-terminal of NS4A, which is known to act as a competitive inhibitor of the NS3 protease [30, 45].

Our results suggest that the protease and helicase rotate around their flexible linker in order to achieve maximum protease activity. Such a mechanism was first proposed by Yao *et al* [21], that the helicase should move away from the protease domain so that the protease active site is liberated from the C-terminus of the protein and becomes accessible to process substrates *in trans*. Mutations in the direct interface in this

case destabilize this transient interface so that the transition from the “closed” state to the “open” state is easier, thus the protein is rendered more active.

*Helicase – Protease Interface Influences NS3 Helicase Action*

Parallel with the effect of helicase domain on protease function, the protease domain also restricts the helicase domain when the helicase is processing DNA. NS3hel is significantly more efficient in unwinding DNA than scNS3/4A, and the interface mutants exhibited activities between scNS3/4A and NS3hel. Therefore, the restrictive role of the domains is bidirectional – meaning protease and helicase should be liberated from each other in order to achieve full activity. The protease domain also influences the helicase by conferring specificity to RNA and increasing RNA unwinding activity. Unlike in DNA unwinding, interface mutations did not influence RNA unwinding rates significantly. The flatness of the activity profile of our panel of mutants in RNA-based assays suggests that the interface does not affect RNA unwinding. One reason for this difference may be due to the fact that, helicase domain alone unwinds RNA 10-times slower than scNS3/4A, suggesting that RNA might induce a conformational change in scNS3/4A that DNA does not. This is consistent with a direct role of the protease domain in dsRNA unwinding. How can the protease domain aid in the specificity for RNA versus DNA? The interaction of viral RNA with the serine protease 3C from the human rhinovirus has been well established [46]. Similarly, the direct binding of RNA to the HCV protease has been recently demonstrated [47]. In addition, the protease domain can specifically bind to HCV genomic RNA [48]. Taken together, the rate enhancement of RNA processing observed in the full-length protein may be the result of a specific

interaction of the protease domain with RNA; hence the protease domain acts as a cofactor of the helicase domain.

Although there is no direct correlation between ssDNA binding affinities of interface mutants, the differences between multiple and single turnover rates for DNA unwinding for individual constructs can be explained in light of these results. For the constructs that were less active under single turnover conditions compared to multiple turnover rates, the affinity for binding ssDNA were lower and vice versa. This trend was most prominent for V78A, M485A and Triple Prot, where single turnover rates were considerably lower and, conversely, for V524A, Q526A and H528A, where they were higher. Our results indicate that disruptions in the domain-domain interface modulate the DNA unwinding activity of the helicase domain, similar to the protease domain. The ssDNA binding affinity is also modulated by the interface, albeit in a different fashion. In addition, the difference between multiple and single turnover unwinding directly correlates with ssDNA binding affinity.

Curiously, the influence of interface mutations on ssDNA binding affinities does not correlate with their effect on either protease or helicase activities. Amino acid substitutions on both protease and helicase domains significantly alter ssDNA affinities, and these are reflected on the ability of an oligonucleotide to act as an enzyme trap in “single-turnover” helicase assays. These variations are more apparent in the helicase mutants. The mutations at the interface may alter DNA binding and unwinding through a combination of *intra* and *inter*-domain effects. Furthermore, binding of DNA is not likely to induce the liberation of two domains. This is supported by the published full-length

structures of scNS3/4A with bound helicase ligands [22]. Nucleic acid and ATP binding induces structural changes in the helicase domain but these changes were not reflected in the relative arrangement of the helicase and protease domains. Therefore, possibly, the full-length protein cycles through two stages – a dependent stage where two domains interact through the interfaces, and an independent stage where the domains are liberated. Nucleic acid binding likely occurs in the first state, while unwinding and ATP hydrolysis in the second.

*ATP Hydrolysis is Independent of Protease – Helicase Association*

The ATPase activity of the helicase domain is independent of the protease domain. The rates of ATP hydrolysis and the extent of stimulation of ATPase activity with poly(U) was similar in all the constructs tested. These results are not surprising, as previously tested helicase mutants which hamper the unwinding capability of the helicase still showed similar ATPase activities. However, our full-length constructs do not contain the acidic terminal helix of NS4A, which was shown to influence ATPase activity [20, 49, 50]. Therefore, even if the protease does not directly affect ATPase, NS4A may modulate ATP hydrolysis indirectly.

*HCV NS3/4A Possibly Functions in an Extended Domain Arrangement*

All of the crystal structures of full-length NS3/4A [21-23] exhibit a compact domain arrangement. However, recent results from RNA unwinding studies [25] suggest another conformation – the “extended” conformation – similar to the NS3/2B protein [27], which is the functional analog of NS3/4A in Flaviviruses. Upon deletion of the C-terminus, the HCV NS3 helicase domain formed functional complexes faster due to the

domains being more independent and able to sample more conformations [25]. In addition, the extended conformation is the only possible way the full-length protein can be modeled on a membrane. The amphipathic helix  $\alpha 0$  of the protease domain and the transmembrane helix of NS4A, the factors that anchor NS3/4A to the membrane, were found to be functionally important [24, 51, 52].

In the compact conformation, the protease active site is occupied by the C-terminus of the molecule. For the protease to function, the C-terminus must be dislodged from the active site so that the substrate can bind to the protease and get cleaved. Competition of the substrate and the C-terminus may be the key to the initial disruption of C-terminus – protease interactions. Following the association of protease and substrate, the direct interface interactions would also be disrupted, leading to the change to the extended conformation. After the cleavage, the P' side of the substrate dissociates from the protease active site, leaving the P side still occupying the active site, as indicated by the product complex crystal structures [30]. The protease can finally either be recycled to the compact conformation by the C-terminus, or an uncleaved substrate peptide competes out the product for another round of cleavage (Figure II.9a).

Our results also provide clues to the regulation of helicase activity in the closed and open states. The unwinding reaction involves multiple steps (binding, ATP hydrolysis, nucleic acid translocation), and in either of these steps a transformation between the two different states is possible. When the C-terminus of the protein is deleted, the helicase forms functional complexes faster [25], suggesting the conformational shift from closed conformation to the extended conformation potentially involves an interaction between

the protein and RNA. However, our binding results and published crystal structures indicate that binding of the nucleic acid is not enough to induce a conformational change that separates the domains. Therefore, one or more of the following may be possible. First, the three-way interaction between RNA, protease and the helicase can induce the conformational change – this actually explains why the RNA unwinding rates were similar for the various proteases unlike DNA unwinding rates. Second, protein oligomerization may be a trigger for the conformational switch. HCV NS3 helicase operates as an oligomer when unwinding nucleic acids [53-55]. The crystal structure of an NS3hel oligomer bound to DNA, obtained by using a longer ssDNA fragment, exhibits an association between different helicase domains, which would be possible only if the protease rotates away from the helicase [56]. Hence, transient association of different molecules on a nucleic acid track may induce a conformational change for NS3 protomers, which eventually associate to form oligomers. A third possibility is the combination of the first two. The association of RNA with the protein destabilizes the compact conformation and shifts the equilibrium between the closed and open states to favor the latter. Association and oligomerization of these RNA-protein complexes stabilizes the open state. This third model is the one most consistent with our results, since this model explains both the interface effect in DNA binding and unwinding, and the protease domain effect in RNA unwinding. Consistent with our results, we propose the following scheme for helicase action (Figure II.9b). The unwinding cycle first starts with binding to the single stranded overhang of RNA. This step is followed by the oligomerization and domain-domain rearrangement – in this stage, the protease also

**Figure II.9 – A model for NS3/4A function consistent with *in vitro* results. (a)**

Upon binding substrate, the protein switches from compact to extended conformation.

Cleavage reaction proceeds in the extended conformation and the P' side of the substrate

leaves upon cleavage. An uncleaved substrate can reattach for another round of

proteolysis or the C-terminus can compete out the remaining product and return to the

compact conformation. (b) The single stranded overhang of the dsRNA binds to the

helicase in compact conformation. With the oligomerization and the protease-RNA

association, the RNA-protein complex is ready to unwind RNA efficiently. The binding

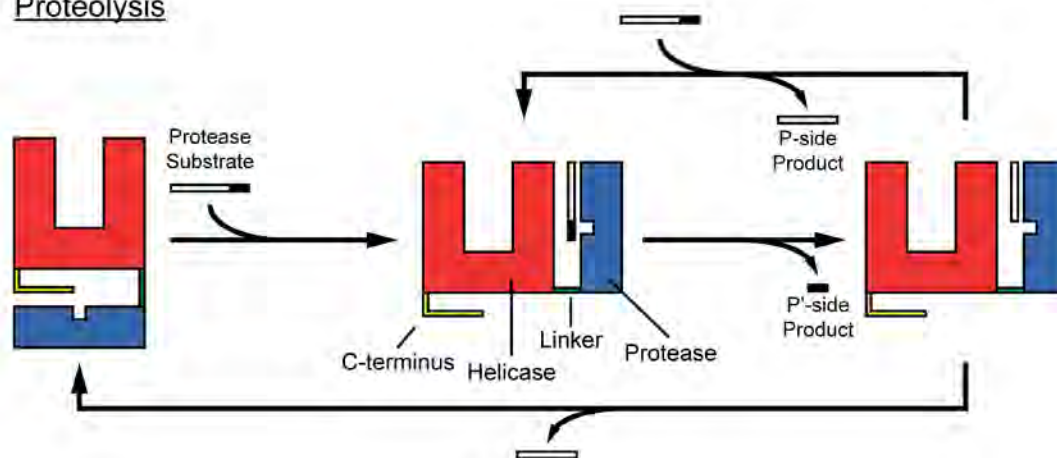
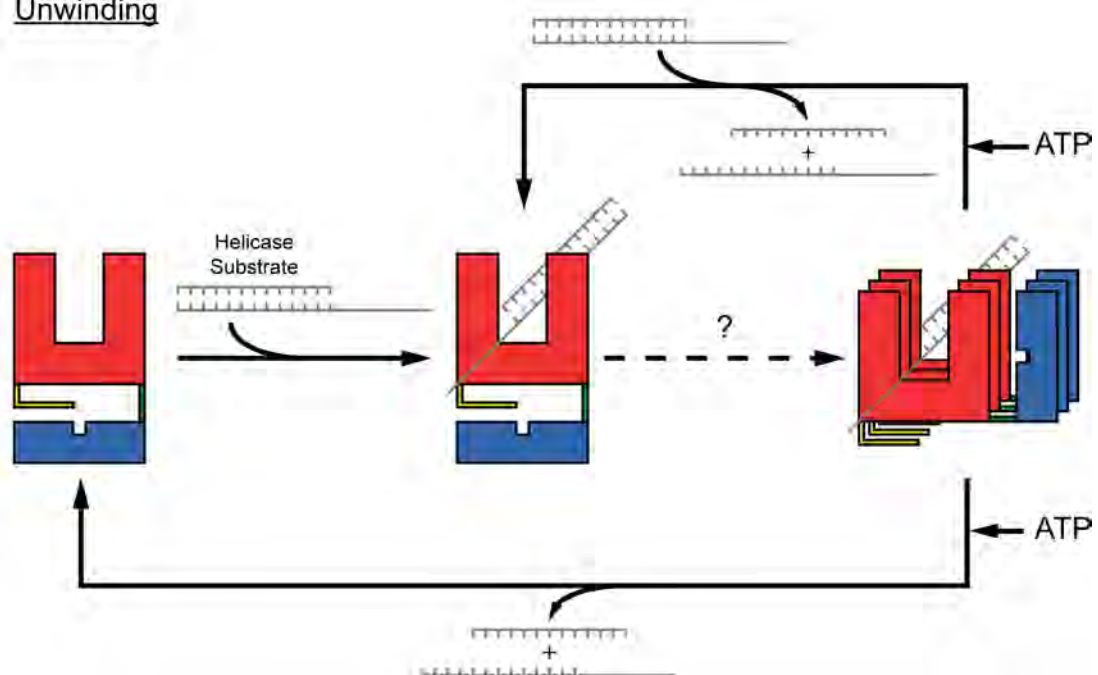
and hydrolysis of ATP fuels the unwinding reaction and the strands get separated. The

protein then can either dissociate from RNA and other subunits of the oligomer, and

return to the closed conformation or another dsRNA can attach to the protein in the

oligomeric state.



(a) Proteolysis(b) Unwinding

associates with the RNA. ATP binding and hydrolysis fuels the movements needed for the helicase to translocate and separate the RNA strands. After unwinding, the protein can either bind a new RNA in the oligomeric state or complete the cycle by returning to the monomeric, apo state.

The protease – helicase interplay in the HCV NS3-NS4A complex is likely even more complex *in vivo*, because NS3/4A attaches itself to the host ER membrane via two regions in the protein – the protease helix  $\alpha_0$  and the NS4A transmembrane helix [24, 51]. Important unanswered questions concern when the protein associates with the membrane, and whether or not membrane association drives the above conformational change *in vivo*. Alternatively, the activation of the protein may lead to the conformational change and induce membrane integration. In either case, attachment to the membrane is required for the efficient cleavage of membrane-anchored cleavage junctions and continuation of HCV life cycle [51, 52]. Still, single chain NS3/4A, similar to the one used in this study, can cleave both membrane bound and free substrates *in vivo* [52], indicating that the protein is able to cycle between cytoplasmic and membrane bound states. Therefore, domain-domain interface may still play a regulatory role in controlling the catalytic activities of both domains *in vivo*.

### **Acknowledgements**

We would like to thank Hong Cao for valuable assistance in enzymatic assay setup; Akbar Ali for synthesis of protease inhibitors; John Hernandez (University of Wisconsin, Milwaukee) for help with preparing RNA substrate; Akbar Ali, Keith Romano, Djade Soumana and Kelly Thayer for valuable discussions; and Nese Kurt Yilmaz for editorial help.

This work was supported by NIH grants R01 AI085051 and RO1 AI088001, and a RGI grant from the UWM research foundation.

## Supplementary Methods

### *Molecular Dynamics Simulations*

The apo crystal structure of single chain HCV NS3/4A was obtained from Protein Data Bank (PDB ID: 1CU1) [21]. Three additional *in silico* deletion constructs were created, where the C-terminus ( $\Delta$ 626-631), the linker region ( $\Delta$ 184-186) and both the C-terminus and the linker ( $\Delta$ 184-186, 626-631) were deleted. These constructs were used as initial coordinates in molecular dynamics (MD) simulations using AMBER10 [28] with the *ff03* force field [57].

The generated constructs were prepared for simulation using the *tLEaP* module of AMBER. First, the missing hydrogen atoms were added. Subsequently, the constructs were immersed in a truncated octahedral periodic water box with TIP3P waters [58], keeping the crystallographic waters. The edges of the box were at least 8 Å away from the nearest protein atom. The net charge of the system was calculated and if necessary, neutralized by adding either Na<sup>+</sup> or Cl<sup>-</sup> as counter ions. The tetrahedral coordinated structural Zn<sup>2+</sup> was parameterized using cationic dummy atom model [59] with the coordinating water and cysteines ionized to hydroxyl and anionic cysteines respectively. The resulting systems were composed of ~96,500 atoms including ~26,500 TIP3P water molecules with box dimension of ~(125 Å)x(125 Å)x(125 Å).

The prepared systems were simulated using the *sander* module of AMBER. Four independent replicas per system were simulated with different initial velocities resulting in a total of sixteen trajectories. Initially, the conformation of each system was relaxed with a series of restrained minimizations. Stepwise gradient method was used in these

restrained minimizations, removing harmonic restraints sequentially with each successive minimization step. A final conjugate gradient minimization was performed on the unrestrained system to fully minimize the energy of the system.

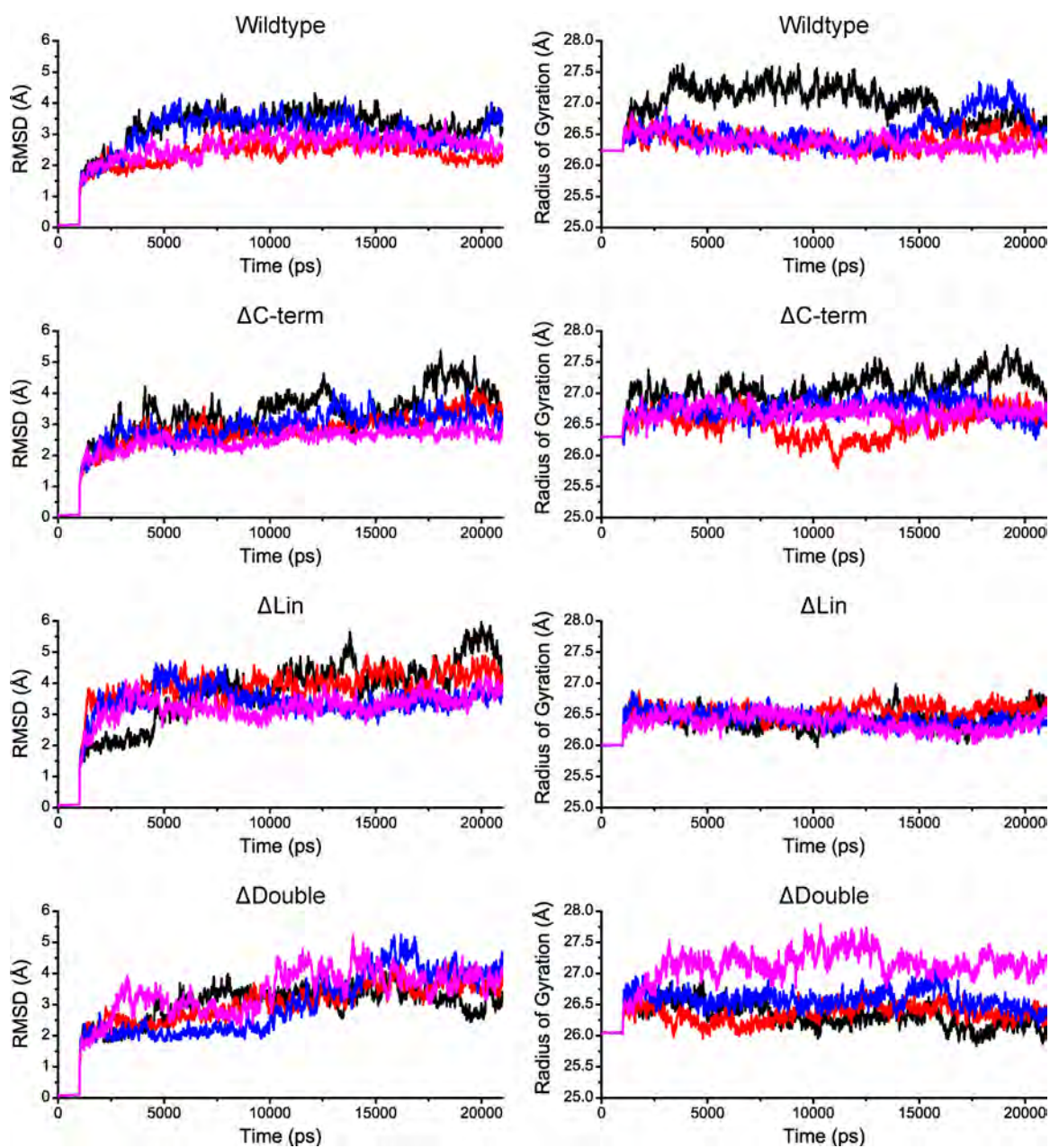
The resulting systems were simulated with periodic boundary conditions, and Particle Mesh Ewald summation method [60] was used for long-range electrostatic interactions with atoms further than 8 Å. The simulations were performed in a NTP (constant pressure – temperature) ensemble with the temperature coupled to a Berendsen thermostat [61]. First, the systems were gradually heated to 300 K with weak positional restraints in 100 ps, followed by the equilibration of waters for 250 ps while keeping the positional restraints on the protein atoms. Finally, all the restraints were removed and 500 ps of equilibration was performed. Production simulations were performed employing SHAKE algorithm [62] with 2 fs time step. 20 ns trajectories were collected for each simulation at a speed of ~400 ps/day using 8 nodes on an AMD Opteron CPU.

#### *Trajectory Analyses*

We have performed four independent replicates for each construct with different initial velocities. Prior to the analyses of resulting trajectories, stability of the simulations were ensured by evaluating root mean square deviations (RMSDs) and radii of gyration over simulation time (Figure II.10). For all constructs, RMSDs and radii of gyration stabilized ~3 ns starting from the initial restrained equilibration, so the data for the first 3 ns were discarded from post simulation analyses. Remaining trajectories for each construct were concatenated, resulting in 72 ns trajectories per construct.

Prior to analyses, concatenated trajectories were aligned using *theseus* [63]. The

**Figure II.10 – Stability of individual trajectories.** Root-mean square deviations (left) and radii of gyration (right) are represented for all the replicas of each construct. Stabilization of parameters was observed after 3 ns of simulation.



*ptraj* module of AMBER was used to analyze the aligned trajectories. Dynamic cross-correlation matrices for all C<sub>α</sub> atom pairs (Figure II.4a) were generated using Equation (9)

$$c_{ij} = \frac{\langle \Delta \vec{r}_i \cdot \Delta \vec{r}_j \rangle}{\sqrt{\langle \Delta r_i^2 \rangle} \sqrt{\langle \Delta r_j^2 \rangle}} \quad (9)$$

In Equation (9),  $c_{ij}$  is the correlation coefficient between atoms  $i$  and  $j$ ,  $\Delta \vec{r}_i$  is the positional fluctuation vector of atom  $i$  with respect to its average position,  $\langle \rangle$  denotes an ensemble average and  $\cdot$  denotes a dot product. Correlation coefficients larger than 0.5 (the “shoulders” in distributions for total correlations) are mostly the correlations within the subdomains (Figure II.4b). Remaining correlations are of interest because they encompass interdomain correlations, which account for roughly 40% of the total correlations.

Histograms of these correlations were constructed with increments of 0.05, and were fit to Equation (10).

$$y = \frac{A}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (10)$$

In Equation (10)  $\mu$  is the mean,  $\sigma$  is the standard deviation and  $A$  is the area of the distribution curve. Distribution parameters were calculated for each construct (Table I.1) to evaluate the extent of correlations.

For visualization purposes, ~300 residue pairs were selected (Table 1, Figure II.5). These numbers roughly correspond to 0.5% of interdomain correlations, excluding linker – protein and C-terminus – protein correlations. The cutoff values indicated are the



correlation coefficient lower limit (or higher limit in the case of anti-correlations) for visualization. This selection procedure served to ease visualization of correlations by optimally representing the interdomain correlations while avoiding clutter in the visual representations.

Additionally, distances between the residues from the protease and the helicase domains were calculated over the simulation time (Figure II.6). For the direct interface, interdomain distances were calculated for the residues constituting the interface (Y56-S483, Y56-M485, V78-V524, D79-V524) and for the substrate interface, residues from both domains covering the C-terminus (A156-Q526, V158-H528, C159-F438, C159-F531). Histograms were generated from these distances with increments of 1 Å.

## References

1. Lavanchy, D., *The global burden of hepatitis C*. Liver Int, 2009. **29 Suppl 1**: p. 74-81.
2. Liang, T.J. and T. Heller, *Pathogenesis of hepatitis C-associated hepatocellular carcinoma*. Gastroenterology, 2004. **127**(5 Suppl 1): p. S62-71.
3. Bartenschlager, R., M. Frese, and T. Pietschmann, *Novel insights into hepatitis C virus replication and persistence*. Adv Virus Res, 2004. **63**: p. 71-180.
4. Morikawa, K., et al., *Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus*. J Viral Hepat, 2011. **18**(5): p. 305-15.
5. Bartenschlager, R., et al., *Substrate determinants for cleavage in cis and in trans by the hepatitis C virus NS3 proteinase*. J Virol, 1995. **69**(1): p. 198-205.
6. Foy, E., et al., *Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease*. Science, 2003. **300**(5622): p. 1145-8.
7. Li, K., et al., *Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF*. Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2992-7.
8. Kanai, A., K. Tanabe, and M. Kohara, *Poly(U) binding activity of hepatitis C virus NS3 protein, a putative RNA helicase*. FEBS Lett, 1995. **376**(3): p. 221-4.
9. Korolev, S., et al., *Comparisons between the structures of HCV and Rep helicases reveal structural similarities between SF1 and SF2 super-families of helicases*. Protein Sci, 1998. **7**(3): p. 605-10.

10. Kim, D.W., et al., *C-terminal domain of the hepatitis C virus NS3 protein contains an RNA helicase activity*. *Biochem Biophys Res Commun*, 1995. **215**(1): p. 160-6.
11. Belon, C.A. and D.N. Frick, *Fuel specificity of the hepatitis C virus NS3 helicase*. *J Mol Biol*, 2009. **388**(4): p. 851-64.
12. Gallinari, P., et al., *Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus*. *J Virol*, 1998. **72**(8): p. 6758-69.
13. Thibeault, D., et al., *Use of the fused NS4A peptide-NS3 protease domain to study the importance of the helicase domain for protease inhibitor binding to hepatitis C virus NS3-NS4A*. *Biochemistry*, 2009. **48**(4): p. 744-53.
14. Beran, R.K. and A.M. Pyle, *Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase*. *J Biol Chem*, 2008. **283**(44): p. 29929-37.
15. Dahl, G., et al., *Effects on protease inhibition by modifying of helicase residues in hepatitis C virus nonstructural protein 3*. *FEBS J*, 2007. **274**(22): p. 5979-86.
16. Frick, D.N., et al., *The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex RNA efficiently*. *J Biol Chem*, 2004. **279**(2): p. 1269-80.
17. Beran, R.K., V. Serebrov, and A.M. Pyle, *The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate*. *J Biol Chem*, 2007. **282**(48): p. 34913-20.
18. Yon, C., et al., *Modulation of the nucleoside triphosphatase/RNA helicase and 5'-RNA triphosphatase activities of Dengue virus type 2 nonstructural protein 3*

- (NS3) by interaction with NS5, the RNA-dependent RNA polymerase. *J Biol Chem*, 2005. **280**(29): p. 27412-9.
19. Chernov, A.V., et al., *The two-component NS2B-NS3 proteinase represses DNA unwinding activity of the West Nile virus NS3 helicase*. *J Biol Chem*, 2008. **283**(25): p. 17270-8.
  20. Shiryayev, S.A., et al., *The acidic sequence of the NS4A cofactor regulates ATP hydrolysis by the HCV NS3 helicase*. *Arch Virol*, 2011. **156**(2): p. 313-8.
  21. Yao, N., et al., *Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase*. *Structure*, 1999. **7**(11): p. 1353-63.
  22. Appleby, T.C., et al., *Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV*. *J Mol Biol*, 2011. **405**(5): p. 1139-53.
  23. Schiering, N., et al., *A macrocyclic HCV NS3/4A protease inhibitor interacts with protease and helicase residues in the complex with its full-length target*. *Proc Natl Acad Sci U S A*, 2011. **108**(52): p. 21052-6.
  24. Brass, V., et al., *Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex*. *Proc Natl Acad Sci U S A*, 2008. **105**(38): p. 14545-50.
  25. Ding, S.C., A.S. Kohlway, and A.M. Pyle, *Unmasking the active helicase conformation of nonstructural protein 3 from hepatitis C virus*. *J Virol*, 2011. **85**(9): p. 4343-53.

26. Mastrangelo, E., et al., *Crystal structure and activity of Kunjin virus NS3 helicase; protease and helicase domain assembly in the full length NS3 protein*. J Mol Biol, 2007. **372**(2): p. 444-55.
27. Luo, D., et al., *Crystal structure of the NS3 protease-helicase from dengue virus*. J Virol, 2008. **82**(1): p. 173-83.
28. Case, D.A., et al., *The Amber biomolecular simulation programs*. J Comput Chem, 2005. **26**(16): p. 1668-88.
29. Belon, C.A. and D.N. Frick, *Helicase inhibitors as specifically targeted antiviral therapy for hepatitis C*. Future Virol, 2009. **4**(3): p. 277-293.
30. Romano, K.P., et al., *Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding*. Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20986-91.
31. Wittekind, M., et al., *Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes*, in *United States Patent Applications Publication*2002: United States of America. p. 26.
32. Taliani, M., et al., *A continuous assay of hepatitis C virus protease based on resonance energy transfer depsipeptide substrates*. Anal Biochem, 1996. **240**(1): p. 60-7.
33. Seiwert, S.D., et al., *Preclinical characteristics of the hepatitis C virus NS3/4A protease inhibitor ITMN-191 (R7227)*. Antimicrob Agents Chemother, 2008. **52**(12): p. 4432-41.

34. Palmier, M.O. and S.R. Van Doren, *Rapid determination of enzyme kinetics from fluorescence: overcoming the inner filter effect*. *Anal Biochem*, 2007. **371**(1): p. 43-51.
35. Niedzwiecki, L., et al., *Substrate specificity of the human matrix metalloproteinase stromelysin and the development of continuous fluorometric assays*. *Biochemistry*, 1992. **31**(50): p. 12618-23.
36. Neumann, U., et al., *Characterization of Mca-Lys-Pro-Leu-Gly-Leu-Dpa-Ala-Arg-NH<sub>2</sub>, a fluorogenic substrate with increased specificity constants for collagenases and tumor necrosis factor converting enzyme*. *Anal Biochem*, 2004. **328**(2): p. 166-73.
37. Belon, C.A. and D.N. Frick, *Monitoring helicase activity with molecular beacons*. *BioTechniques*, 2008. **45**(4): p. 433-40, 442.
38. Hanson, A.M., et al., *Identification and analysis of inhibitors targeting the hepatitis C virus NS3 helicase*. *Methods Enzymol*, 2012. **511**: p. 463-83.
39. Mukherjee, S., et al., *Identification and analysis of hepatitis C virus NS3 helicase inhibitors using nucleic acid binding assays*. *Nucleic Acids Res*, 2012.
40. Romano, K.P., et al., *Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease*. *J Virol*, 2011. **85**(13): p. 6106-16.
41. Vallet, S., et al., *Genetic heterogeneity of the NS3 protease gene in hepatitis C virus genotype 1 from untreated infected patients*. *J Med Virol*, 2005. **75**(4): p. 528-37.

42. Lam, A.M., D. Keeney, and D.N. Frick, *Two novel conserved motifs in the hepatitis C virus NS3 protein critical for helicase action*. J Biol Chem, 2003. **278**(45): p. 44514-24.
43. Luo, D., et al., *Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications*. J Biol Chem, 2010. **285**(24): p. 18817-27.
44. Nooren, I.M. and J.M. Thornton, *Structural characterisation and functional significance of transient protein-protein interactions*. J Mol Biol, 2003. **325**(5): p. 991-1018.
45. Steinkuhler, C., et al., *Product inhibition of the hepatitis C virus NS3 protease*. Biochemistry, 1998. **37**(25): p. 8899-905.
46. Matthews, D.A., et al., *Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein*. Cell, 1994. **77**(5): p. 761-71.
47. Fukuda, K., et al., *Isolation and characterization of RNA aptamers specific for the hepatitis C virus nonstructural protein 3 protease*. Eur J Biochem, 2000. **267**(12): p. 3685-94.
48. Ray, U. and S. Das, *Interplay between NS3 protease and human La protein regulates translation-replication switch of Hepatitis C virus*. Sci Rep, 2011. **1**: p. 1.

49. Beran, R.K., B.D. Lindenbach, and A.M. Pyle, *The NS4A protein of hepatitis C virus promotes RNA-coupled ATP hydrolysis by the NS3 helicase*. J Virol, 2009. **83**(7): p. 3268-75.
50. Shiryayev, S.A., et al., *NS4A regulates the ATPase activity of the NS3 helicase: a novel cofactor role of the non-structural protein NS4A from West Nile virus*. J Gen Virol, 2009. **90**(Pt 9): p. 2081-5.
51. He, Y., et al., *The N-terminal helix alpha(0) of hepatitis C virus NS3 protein dictates the subcellular localization and stability of NS3/NS4A complex*. Virology, 2012. **422**(2): p. 214-23.
52. Horner, S.M., H.S. Park, and M. Gale, Jr., *Control of innate immune signaling and membrane targeting by the Hepatitis C virus NS3/4A protease are governed by the NS3 helix alpha0*. J Virol, 2012. **86**(6): p. 3112-20.
53. Levin, M.K. and S.S. Patel, *The helicase from hepatitis C virus is active as an oligomer*. J Biol Chem, 1999. **274**(45): p. 31839-46.
54. Sikora, B., et al., *Hepatitis C virus NS3 helicase forms oligomeric structures that exhibit optimal DNA unwinding activity in vitro*. J Biol Chem, 2008. **283**(17): p. 11516-25.
55. Jennings, T.A., et al., *NS3 helicase from the hepatitis C virus can function as a monomer or oligomer depending on enzyme and substrate concentrations*. J Biol Chem, 2009. **284**(8): p. 4806-14.



56. Mackintosh, S.G., et al., *Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus*. J Biol Chem, 2006. **281**(6): p. 3528-35.
57. Duan, Y., et al., *A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations*. J Comput Chem, 2003. **24**(16): p. 1999-2012.
58. Jorgensen, W.L., et al., *Comparison of Simple Potential Functions for Simulating Liquid Water*. Journal of Chemical Physics, 1983. **79**(2): p. 926-935.
59. Pang, Y.P., et al., *Successful molecular dynamics simulation of the zinc-bound farnesyltransferase using the cationic dummy atom approach*. Protein Sci, 2000. **9**(10): p. 1857-65.
60. Essmann, U., et al., *A smooth particle mesh Ewald method*. The Journal of Chemical Physics, 1995. **103**(19): p. 8577.
61. Berendsen, H.J.C., et al., *Molecular dynamics with coupling to an external bath*. The Journal of Chemical Physics, 1984. **81**(8): p. 3684.
62. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes*. Journal of Computational Physics, 1977. **23**(3): p. 327-341.
63. Theobald, D.L. and D.S. Wuttke, *THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures*. Bioinformatics, 2006. **22**(17): p. 2171-2.

**CHAPTER III**

**P2 MOIETY IN THE HEPATITIS C VIRUS PROTEASE INHIBITORS IS THE  
MAJOR DETERMINANT IN DRUG RESISTANCE AND MODULATION OF  
INHIBITOR POTENCY BY THE HELICASE DOMAIN**

### **Author Contributions**

This study was performed in collaboration with Akbar Ali, who synthesized the protease inhibitors, and all the data was collected by me. The paper was written by me with Akbar Ali contributing “Inhibitor Design and Synthesis” in Supplementary Methods and edited by Akbar Ali, Nese Kurt Yilmaz and Celia Schiffer.

## **Abstract**

Hepatitis C is a global health problem that plagues about 3% of the world's population. Despite being curable, the current treatment with pegylated interferon and ribavirin is lengthy, not completely effective, and generally intolerable. Thus, research and development of directly acting antiviral drugs are being performed actively against critical viral machinery, such as NS3/4A protease, the primary and indispensable proteolysis machinery for polyprotein maturation and immune suppression. Drug development efforts on NS3/4A protease currently produced two drugs on the market – boceprevir and telaprevir – with promising new protease inhibitors in clinical trials. However, drug resistance emerges quickly under drug pressure and resistance patterns arise as early as in clinical trials. In addition, most of these drug development attempts disregard the helicase domain, which is an essential part of the Hepatitis C virus machinery and covalently linked to the protease. Furthermore, the protease and the helicase domains are interdependent functionally. However, there are conflicting results on the role of helicase domain in protease inhibitor recognition. In this study, four protease inhibitors and their macrocyclic analogs were investigated for susceptibility to drug resistance and modulation of potency by the helicase domain. Potency is mainly affected by macrocyclization. Conversely, for both potency loss due to drug resistance mutations and potency modulation by the presence of the helicase domain, the P2 group is identified as the major determinant. The results signify the importance of the P2 group in inhibitor resistance and assign a possible role for the P2 group in association with the helicase.

## Introduction

Hepatitis C Virus (HCV), a member of the *Flaviviridae*, is the causative agent of non-A non-B viral hepatitis that infects 170 million people worldwide [1, 2]. HCV is also genetically diverse; six major genotypes (1-6) were classified with the genotypes 1 and 3 most prevalent around the globe [3, 4]. Acute HCV infection is asymptomatic, but persistent viral infection progresses into chronic viral hepatitis which leads to cirrhosis, liver failure and liver cancer; HCV is the main cause of liver transplantation in the United States [5]. Current state-of-art treatment regime includes the combination of either telaprevir or boceprevir, recently approved direct acting antivirals (DAAs) targeting HCV non-structural protein 3/4A (NS3/4A) protease, pegylated interferon- $\alpha$  and ribavirin [6]. This triple combination therapy led to higher sustained virologic response (SVR) rates, which is an indicator of cure, for both treatment-naive patients and patients who failed the previous dual therapy regime (with interferon and ribavirin alone) [7-11]. However, SVR strongly depends on host and viral factors, even with the new triple therapy regime [8, 12]. Also, serious side effects have been observed in these treatment regimes – most commonly anemia, flu-like symptoms, eczematous skin rash, depression and gastrointestinal symptoms [13]. Thus, the current HCV treatment is still not completely effective and not well tolerated by many patients and specific DAAs are necessary for an effective and better quality treatment.

Pharmaceutical development effort for DAAs against HCV primarily focuses on the NS3/4A protein. HCV NS3/4A is the bifunctional protease – helicase complex that is pivotal in polyprotein processing and replication [14]. NS3/4A protease plays an

additional role in circumventing host immune response by cleaving host factors mitochondrial antiviral signaling protein (MAVS) and TIR-domain-containing adapter-inducing interferon- $\beta$  (TRIF) [15-17]. Drug development effort on the NS3/4A protease produced two drugs that are currently in market – telaprevir and boceprevir – which are linear ketoamides [18, 19]. In addition, many more compounds are in various stages of clinical development, both acyclic compounds (asunaprevir [20], BI-201335 [21]) and macrocyclic compounds including P1-P3 macrocycles (danoprevir [22], TMC-435 [23]) or P2-P4 macrocycles (vaniprevir [24], MK-5172 [25]). However this drug development effort primarily focuses on the protease domain alone. NS3 helicase is also required in HCV replication, and essential for viral propagation [26, 27]. Furthermore, both domains are interdependent and influence their activities bidirectionally [28-30]. Although there is no concrete evidence for a direct interaction between the helicase domain and protease inhibitors [31], helicase may play an accessory role [32, 33].

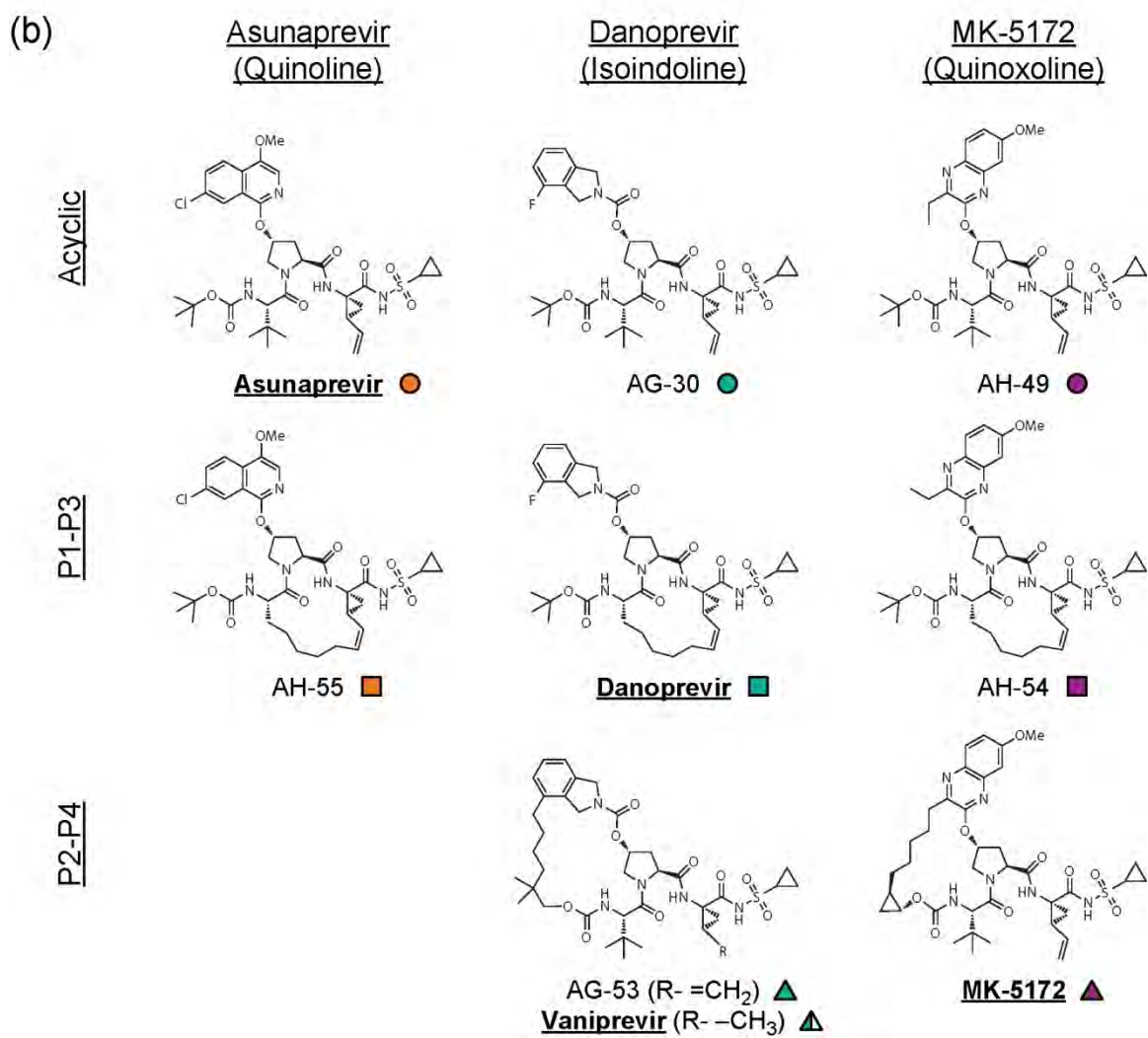
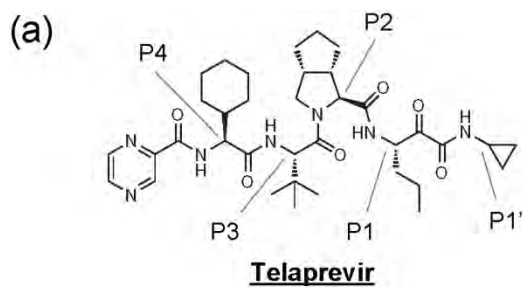
HCV replicates rapidly in host liver, producing around a trillion virus particles per day [34]. HCV NS5B, the RNA-dependent RNA polymerase, is inherently error-prone – misincorporation of bases into the HCV genome produces a low frequency of genetically diverse viral particles [35]. Under drug pressure, populations that contain mutations which confer resistance are selected rapidly. Most strikingly, mutations at the protease residues R155, A156 and D168 confer resistance to a multitude of protease inhibitors [36, 37]. The different mutation patterns due to drug resistance are influenced by many factors such as genotypic diversity, genetic barrier for mutations and viral fitness. Emergence of

inhibitor resistance in HCV NS3/4A protease is thus the major drawback in cure of Hepatitis C using a monotherapy with a single protease inhibitor.

NS3/4A protease inhibitors mostly share the same architecture (Figure III.1). The common features of these drugs in clinical development are either a ketoamide or an acylsulphonamide warhead at P1' position, a cysteine-like P1 moiety with a cyclopropane and a terminal olefin, a proline-like P2 moiety with a bulky extension and capped at the P4 moiety [38]. On the other hand, these drugs differ in two major ways. First, the P2 extensions are different – they all contain bulky, hydrophobic moieties but their identities are different. Second, these drugs are either macrocyclized or acyclic. Macrocyclization happens either between the P1 terminus and the P3, or the P2 extension and the P4 terminus. All of these drugs have different potencies and different resistance profiles. In this study, four protease inhibitors with different P2 extensions and macrocyclization statuses were selected and their acyclic and/or macrocyclic analogs were generated to evaluate the effect of the different structural features of these inhibitors. Macrocyclization was observed to be the major determinant in inhibitor potency, and the P2 extension to be the major determinant in susceptibility to drug resistance and modulation by the helicase domain. These results explain the impact of different structural features in potency and susceptibility to drug resistance, and will form a basis for future HCV NS3/4A protease inhibitor development.

**Figure III.1 – Inhibitors used in this study.** (a) Telaprevir and (b) other protease inhibitors used in this study are shown. The inhibitors are classified according to macrocyclization status and identity of the P2 group. Macrocyclization class separates the compounds depending on the presence and position of the macrocycle (acyclic, P1-P3 macrocycle, P2-P4 macrocycle). P2 class separates the compounds depending on the P2 group (quinoline, isoindoline, quinoxaline). Inhibitors are colored according to the identity of the P2 group (asunaprevir – orange, danoprevir – teal, MK-5172 – magenta), and annotated according to the macrocyclization status (acyclic – circle, P1-P3 – square, P2-P4 – triangle).





## **Materials and Methods**

### *Synthesis of Inhibitors*

HCV NS3/4A inhibitors telaprevir, asunaprevir, danoprevir, vaniprevir, MK-5172 and their acyclic and macrocyclic analogs were synthesized in lab. Chemical syntheses of these inhibitors are explained in detail in the Supplementary Methods.

### *NS3/4A Protease Constructs*

HCV NS3/4A genotype 1a protease domains for both the full length and the isolated protease constructs were derived from the Bristol-Myers-Squibb patent [39] and cloned into a pET28a bacterial expression vector. Both of these constructs contain the NS4A cofactor peptide ligated to the N-terminus. In the full length construct, codon optimized genotype 1a NS3 helicase (H77) were cloned downstream. Multi drug resistant mutants R155K, A156T and D168A for both isolated protease and full length constructs and additional helicase mutations Triple Hel (S483A-M485A-V524A), Q526A and H528A on full length construct were generated by site-directed mutagenesis kit QuikChange II (Stratagene), using primer pairs generated by Primer X (<http://www.bioinformatics.org/primerx/>).

### *Expression and Purification of NS3/4A Protease Constructs*

The expression and purification scheme for the isolated protease domain is detailed elsewhere [40]. Briefly, transformed BL21(DE3) *E. coli* expression cells were grown to an  $A_{600}$  of 0.6, induced with 1 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) and incubated with shaking for 5 hours at 37°C. Cells expressing the protein were

harvested by centrifugation and stored at  $-80^{\circ}\text{C}$ . Frozen pellets were resuspended in resuspension buffer (50 mM phosphate buffer, 500 mM NaCl, 10% glycerol, 2 mM beta mercapto ethanol ( $\beta\text{ME}$ ), pH 7.5), lysed and centrifuged to pellet the cell debris and resulting supernatant was applied to a nickel-affinity column (Qiagen). The column was washed with resuspension buffer and the protein was eluted with resuspension buffer supplemented with 200 mM imidazole. The eluate was treated with thrombin and dialyzed overnight to cleave the His-tag and remove the imidazole. The purified protein was flash frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$ .

The expression and purification scheme for full length protein is detailed in Chapter II. Briefly, transformed BL21(DE3) *E. coli* expression cells were grown to an  $A_{600}$  of 0.6 at  $37^{\circ}\text{C}$ , transferred to  $20^{\circ}\text{C}$ , induced by 0.5 mM IPTG and incubated with shaking for 4 hours. Cells were then harvested via centrifugation; cell pellets were washed with 1X phosphate-buffered saline (PBS), repelleted and stored at  $-80^{\circ}\text{C}$ . Frozen pellets were resuspended in Buffer HT (25 mM HEPES, 500 mM NaCl, 10% glycerol, 0.1% O $\beta$ G, 2 mM TCEP and 20 mM imidazole, pH 8.0) supplemented with DNase I (Roche) and homogenized using a cell disruptor (Micro Fluidics). Lysed cells were centrifuged to clear the cell debris and applied to a 1 mL HisTrap HP column (GE Lifesciences) using an AKTA Purifier (GE Lifesciences). The protein was washed with Buffer HT supplemented with 40 mM imidazole and eluted with Buffer HT supplemented with 250 mM imidazole. The eluate was dialysed overnight against Buffer S (25 mM MES, 150 mM NaCl, 10% glycerol, 0.1% O $\beta$ G, 2 mM TCEP, pH 6.0). Dialyzed protein was applied to a Mono S Column (GE Lifesciences) and eluted with a

linear gradient of NaCl up to 1M. The eluate was judged > 90% pure by polyacrylamide gel electrophoresis, concentrated, flash frozen and stored at -80°C.

#### *Protease Cleavage Assays*

Protease cleavage assays were performed in a final volume of 60  $\mu$ L containing protease assay buffer (50 mM Tris, 2.5% glycerol, 0.1% O $\beta$ G, 5 mM TCEP, 1% DMSO, pH 7.5) and up to 20  $\mu$ M HCV NS3/4A protease substrate Ac-DE-Dap(QXL-520)-EE-Abu- $\psi$ -[COO]AS-C(5-FAMsp)-NH<sub>2</sub> (Anaspec) in black 96-well flat bottom non-binding surface half-area plates (Corning) at room temperature. The reaction was initiated by the rapid addition of 10  $\mu$ L of HCV NS3/4A protease to a final concentration of 20 nM. The fluorescence output was measured kinetically for at least an hour using an EnVision plate reader (Perkin Elmer) with excitation wavelength at 485 nm and emission at 530 nm. End point readings were also taken after 6 hours. For each construct, three independent experiments were performed.

Inner filter effect in higher concentrations of substrate was corrected using the following scheme. End point readings were subtracted from the background fluorescence values to obtain reaction amplitudes and these amplitudes were plotted against substrate concentration. Initial linear portion of this curve is fitted with linear regression and extrapolated to maximum substrate concentration, to obtain the ideal amplitude line. Corresponding amplitude values in the ideal amplitude line were divided by the actual amplitudes to obtain inner filter effect correction parameters which are then multiplied by initial velocities to obtain corrected velocities. Corrected velocities were plotted against substrate concentrations and all the replicates were fitted to Michaelis-Menten equation

globally to obtain Michaelis-Menten constants ( $K_M$ ), which was shared between different replicates.

#### *Enzymatic Inhibition Assays*

All enzymatic inhibition assays were performed in non binding surface 96-well black half-area plates (Corning) in a reaction volume of 60  $\mu$ L. ~1 nM NS3/4A protease was preincubated with increasing concentration of drugs in protease reaction buffer for an hour. The reaction was initiated by the rapid addition of 5  $\mu$ L of HCV NS3/4A protease substrate to a final concentration of 200 nM and kinetically monitored using a Perkin Elmer EnVision plate reader (excitation 485 nm – emission 530 nm) for an hour with 30 s between data points. Initial velocities were obtained from the progress curves and plotted against inhibitor concentrations to get inhibition curves. Resulting curves were fitted to Morrison's equation:

$$\frac{V_i}{V_0} = 1 - \frac{([E]_T + [I]_T + K_i^{app}) - \sqrt{([E]_T + [I]_T + K_i^{app})^2 - 4[E]_T[I]_T}}{2[E]_T}$$

$$K_i^{app} = K_i \left( 1 + \frac{[S]}{K_M} \right)$$

Where  $[E]_T$  is the total enzyme concentration,  $[I]_T$  is the total inhibitor concentration,  $[S]$  is the substrate concentration,  $K_M$  is the Michaelis-Menten constant,  $V_0$  is the initial velocity when there is no drug,  $V_i$  is the initial velocity at  $[I]_T$ ,  $K_i^{app}$  is the apparent inhibition constant and  $K_i$  is the inhibition constant. Non-linear regression analyses were performed where independent replicates from different protease constructs with different drugs were fitted globally, sharing  $K_i$ .

## Results

### *Protease Inhibitors Used in this Study*

In this study, four drugs that are currently in clinical trials were selected with different macrocyclization statuses and P2 extensions. These drugs are asunaprevir (BMS-650032), which is an acyclic compound with a quinoline in the P2 extension; danoprevir (ITMN-191), which is a P1-P3 macrocyclic compound with an isoindoline; and MK-5172, which is a P2-P4 macrocyclic compound with a quinoxaline. In addition, macrocyclization analogs of these drugs were generated – namely AH-55 (asunaprevir P1-P3 macrocyclic), AG-30 (danoprevir acyclic), AG-53 (danoprevir P2-P4 macrocyclic), AH-49 (MK-5172 acyclic) and AH-54 (MK-5172 P1-P3 macrocyclic). Vaniprevir (MK-7009) was also synthesized because vaniprevir is an inhibitor also being investigated in clinical trials; and similar to danoprevir and nearly the same as AG-53 with the only difference P1 terminus being saturated. With the addition of telaprevir (Figure III.1a), a total of ten protease inhibitors (Figure III.1b) were investigated in this study.

### *Macrocyclization Status is the Major Determinant in Inhibitor Potency*

To determine the effect of the P2 moiety and different macrocyclization patterns, *in vitro* potencies of HCV NS3/4A protease inhibitors were determined in inhibition assays for full length protein, isolated protease domain and their variants. Initially,  $K_M$  values for each protein construct were determined (Table 1).  $K_M$ 's for all the constructs, except full length A156T mutant, were similar, which signify that the substrate

recognition was not altered significantly with respect to the different multi drug resistance mutations and helicase mutations.

Telaprevir and boceprevir are the only two drugs that are approved by FDA and currently being utilized in clinical treatment regimes in combination with ribavirin and interferon [41]. However, compared to the other drugs, telaprevir was significantly less active against both wildtype and drug resistant mutants, except D168A (Figure III.2a). Other acyclic drugs (asunaprevir, AG-30 and AH-49) are more active compared to telaprevir, with their potencies ranging from lower nanomolar for the wildtype up to lower micromolar for drug resistant mutants (Figure III.2b).

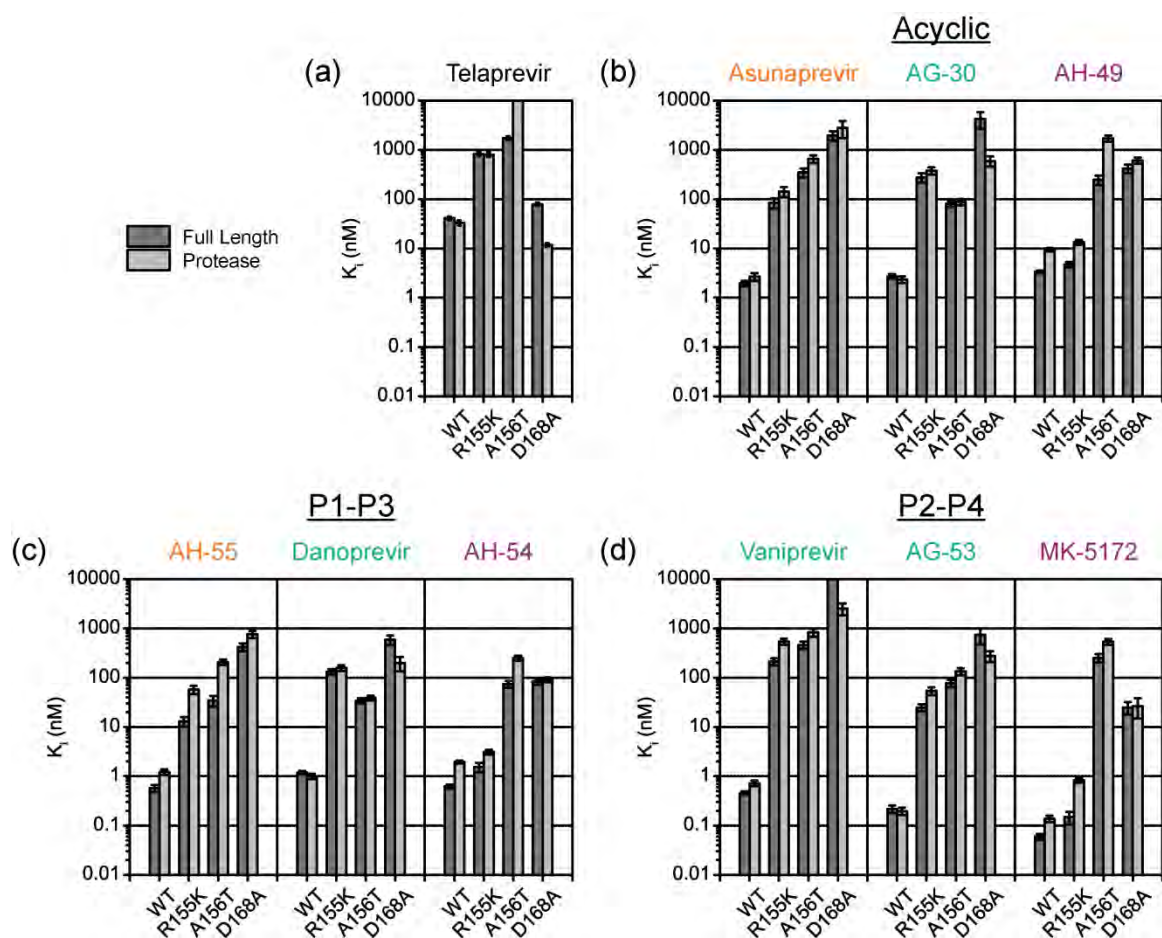
Macrocyclic compounds on the other hand are more potent, compared to acyclic compounds. P1-P3 macrocycles exhibit lower  $K_i$  values for both wildtype and drug resistant variants, with their potencies ranging from higher picomolar for the wildtype up to higher nanomolar for drug resistant mutants (Figure III.2c). P2-P4 macrocycles, except vaniprevir, were significantly more potent for the wildtype protease constructs, in the range of mid to high picomolar. Similar to P1-P3 macrocycles, potencies for the drug resistant variants were in the higher nanomolar range (Figure III.2d). Side by side comparison of vaniprevir against AG-53 exhibited significant loss of potency for vaniprevir; this signifies the importance of the terminal olefin in the P1 moiety where a potential pi-stacking interaction with the F154 is possible [42].

#### *Identity of the P2 Moiety is the Major Determinant in Drug Resistance*

The inhibitor potencies were normalized with respect to the corresponding wildtype construct to obtain drug resistance patterns for each inhibitor. The fold potency

**Figure III.2 – Potencies of protease inhibitors used in this study.** The inhibition constants ( $K_i$ 's) for wildtype and MDR variants for (a) telaprevir, (b) acyclic compounds, (c) P1-P3 macrocycles and (d) P2-P4 macrocycles. For the majority of MDR constructs, macrocyclic drugs are more active compared to acyclic drugs. For wildtype protein, the P2-P4 macrocyclic compounds are more potent compared to P1-P3, a trend which changes with MDRs. Vaniprevir is the exception due to the terminal methyl group in the P1 residue instead of the olefine. Error bars represented in this figure are standard errors of the mean ( $n \geq 4$ ).





losses for the compounds with respect to wildtype constructs for the full length versus protease domain are similar (Figure III.3), meaning that the helicase domain has minimal effect in drug resistance. However, when the data was grouped with respect to different P2 extensions, a clear trend in resistance profiles was observed. The fold loss potencies due to drug resistance mutations cluster well within drugs with different macrocyclization statuses either in the asunaprevir class (Figure III.3a), the danoprevir class (Figure III.3b) or the MK-5172 class (Figure III.3c). Hence, identity of P2 extension group primarily determines susceptibility to drug resistance mutations for protease inhibitors.

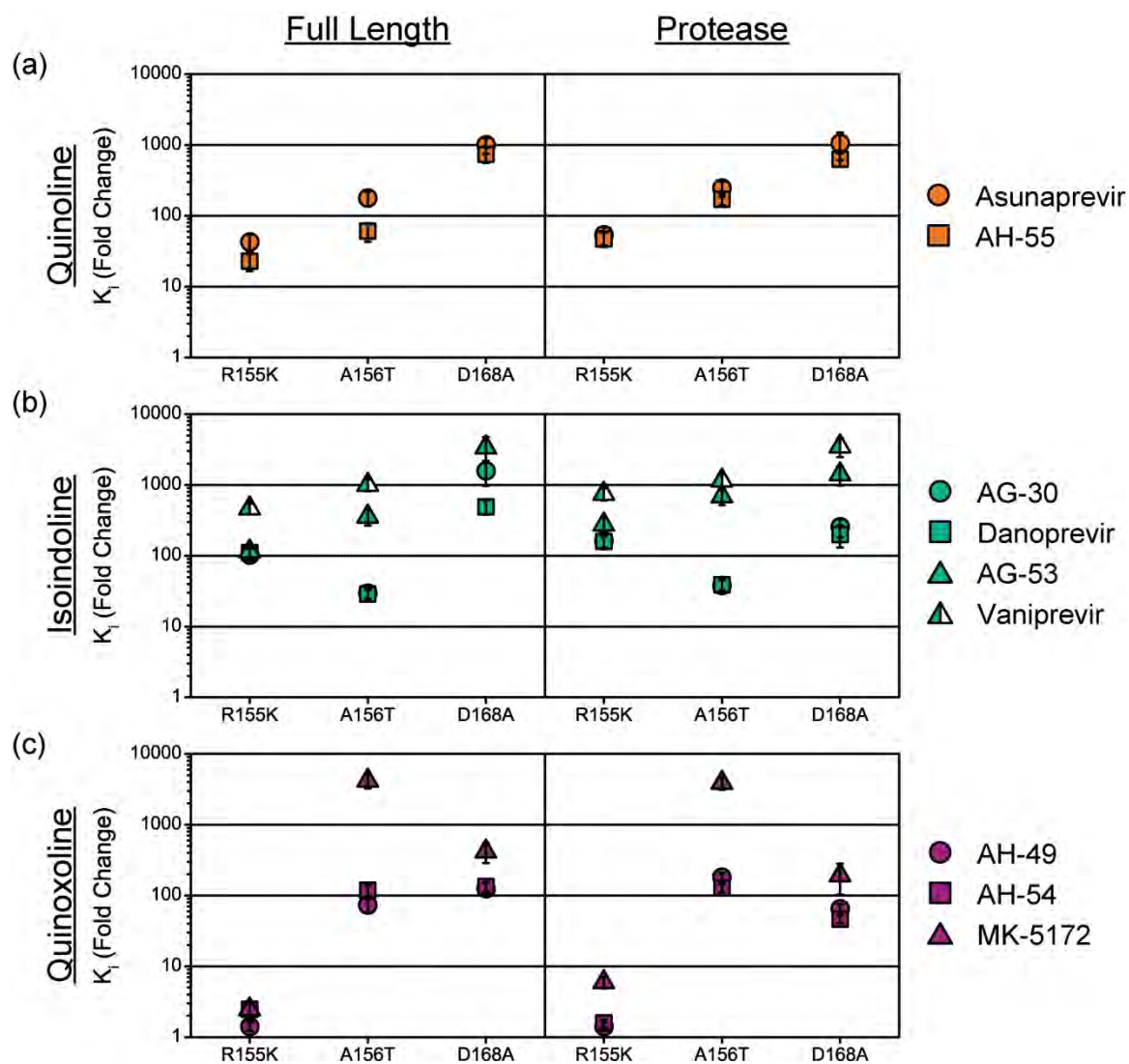
Two major exceptions have been observed. First, the P2-P4 macrocycles all exhibited increased susceptibility against the mutants A156T (more than an order of magnitude more susceptible) and D168A (less susceptibility compared to A156T). P2-P4 macrocycles restrict the movement of P2 extension, which may potentially affect the adaptability of P2 conformation [40, 43, 44]. Second, vaniprevir was more susceptible to drug resistance against each of the mutants analyzed, but the overall trend observed exhibited similar characteristics as AG-53 (Figure III.3b). In summary, the major factor affecting drug resistance was the P2 extension group.

*P2 Moiety is also the Major Determinant in Modulation of Inhibitor Potency by the Helicase Domain*

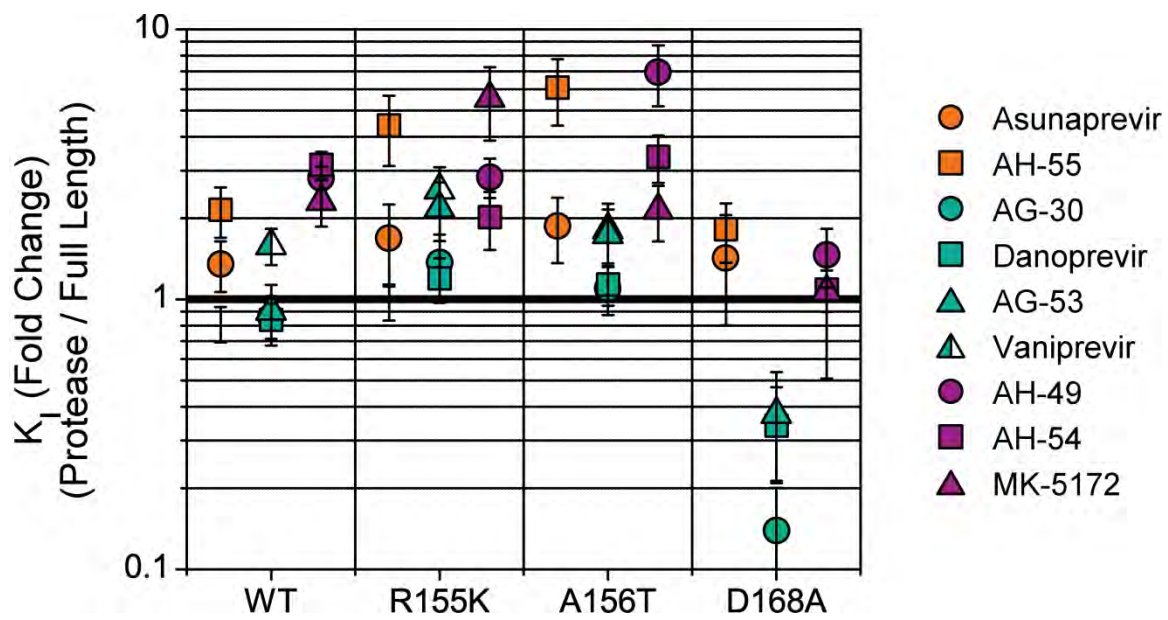
Inhibitor potencies for wildtype and drug resistant variants for the isolated protease domain were normalized with respect to their full length counterparts in order to determine the extent of modulation in inhibitor potencies when the helicase domain is present (Figure III.4). Up to an order of magnitude modulation was observed. For the

wildtype constructs, the modulation was minimal (up to 3 fold preference for the full length protein). For the mutants R155K and A156T, preference for the full length protein was more pronounced for asunaprevir and MK-5172 analogs (up to 8 fold), whereas limited modulation was observed for danoprevir analogs. The situation is reversed for D168A, where the full length preferences for the asunaprevir and MK-5172 were not present and, conversely, danoprevir analogs exhibited significant preference for the isolated protease domain (up to an order of magnitude). In addition, the potencies against the helicase mutants were normalized against full length wildtype construct. Similarly, an order of magnitude modulation with respect to different helicase mutations were observed, but a dominant trend was not present (Figure III.5). To conclude, helicase domain influences protease inhibitor potencies, albeit slightly.

**Figure III.3 – Drug resistance profiles of the inhibitors for multi drug resistant variants.** Resistance profiles of full length protein and isolated protease MDR constructs against (a) asunaprevir, (b) danoprevir and (c) MK-5172 analogs.  $K_i$  values were normalized to the respective wildtype constructs. The fold potency loss trends for the full length protein and the protease domain are similar. Except the P2-P4 macrocyclic drugs (triangles), the resistance trends are clustered with respect to the P2 group. For the P2-P4 macrocycles, the inhibitor potency is significantly lower for A156T and slightly lower for D168A. Error bars represented in this figure are the propagated errors from error-weighted divisions.

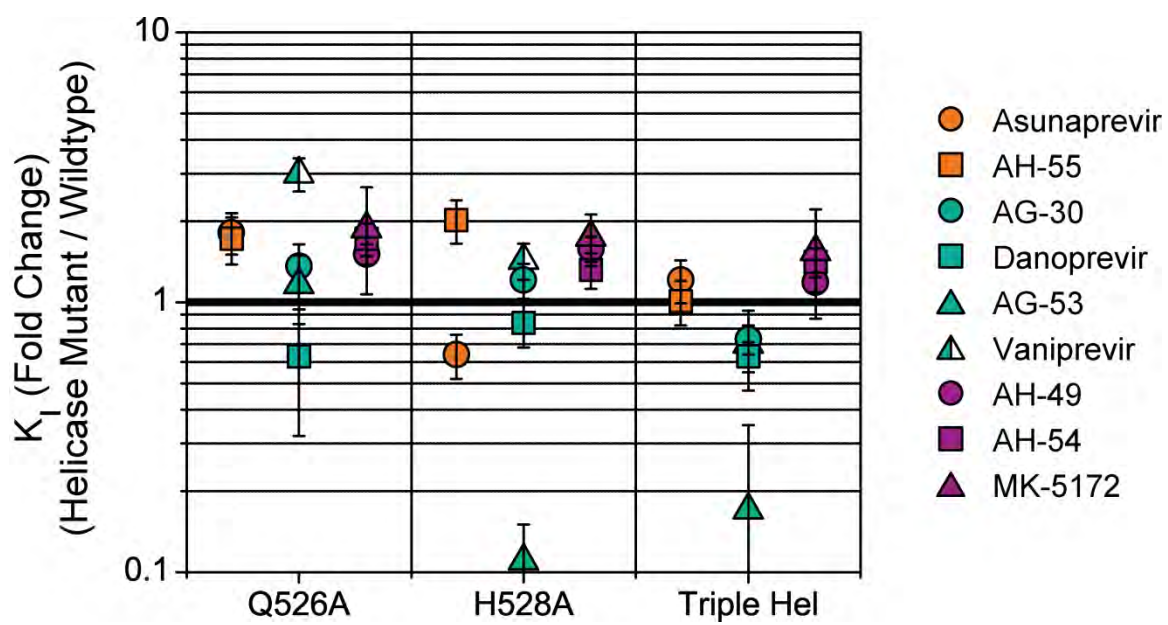


**Figure III.4 – The modulation of drug activities by the helicase domain in multi drug resistance mutants.** The fold change potencies are with respect to the full length constructs. Values higher than one indicate increased preference for the full length protein, and vice versa. The modulation of activity is more apparent for MDRs compared to the wildtype and the values cluster well according to the identity of the P2 moiety (different colors). The danoprevir class (teal) behaves differently; these drugs exhibit the least enhancement for WT, R155K and A156T whereas loses affinity for D168A when the helicase domain is present. Error bars represented in this figure are the propagated errors from error-weighted divisions.



**Figure III.5 – The effect of helicase mutations on inhibitor potency.** The fold changes are with respect to the full length wildtype construct. The helicase mutants affect inhibitor potency (to a smaller degree) but the trend is not consistent. Error bars represented in this figure are the propagated errors from error-weighted divisions.





## Discussion

In this study, differences in the structural features of a series of HCV NS3/4A protease inhibitors were evaluated for inhibitor potency, susceptibility to drug resistance, and modulation of inhibitor activity by the presence of the helicase. We have identified the macrocyclization status as the major determinant of inhibitor potency, and P2 moiety of drug resistance profile and modulation by helicase domain.

The macrocyclic drugs are more potent compared to their acyclic peers likely due to the restriction of dynamics of the molecule. P1-P3 macrocycles restrict the P1 terminus and the rotation around the core of the proline-like P2 moiety. P2-P4 macrocycles on the other hand restrict the movement of the P2 extension and the backbone of the drug extending from P2 proline to P4 capping. Restricting the movement around the P2 proline possibly aids in stabilizing the restrained conformation of the proline, which was observed as a conserved feature in substrate bound protease structures [42]. Further gain in potency observed for the P2-P4 macrocycles is probably caused by the restriction of the P2 extension, in addition to the restriction of the P2 backbone conformation. The bulky P2 extensions for the HCV NS3/4A protease inhibitors form specific contacts with the protease domain, especially in the vicinity of R155, A156 and D168 [40]. Hence, restricting the rotation of P2 extension would favor intermolecular interactions by decreasing the entropic penalty of binding. Restriction of the P1 cysteine-like moiety may be another reason for the increased potency in the case of P1-P3 macrocycles, since this moiety is buried deep within the core of the active site. Thus, by keeping the inhibitor

geometry restricted, the interactions between the inhibitor and the protease are stabilized more in the presence of macrocycles.

Drug resistance profile primarily depends on the identity of the P2 extension. The structural basis for resistance to HCV NS3/4A protease inhibitors for the multi drug resistance (MDR) mutations R155K, A156T and D168A has been explained in detail [40]. Although there are other mutations that confer drug resistance [45-47], these three mutations confer drug resistance to a large variety of protease inhibitors because the large P2 extension, which exists nearly in all of the HCV NS3/4A protease inhibitors, forms specific contacts with these residues [40]. However, the modulation of potencies varies between different drug classes, since the signature interactions between the drug and the protease primarily depend on the identity of the P2 group.

On the other hand, potency losses due to drug resistance mutations did not considerably depend on the macrocyclization status or the presence of the helicase domain. Compounds with P2-P4 macrocycles are more susceptible to resistance possibly due to the movement restriction enforced by macrocyclization. By restricting the geometry of the P2 extension, stronger interactions with the protein are formed – however, the ability for the inhibitor to adapt to a different environment is lowered in exchange. This inability leads to pronounced steric clashes (A156T), or disruption of the van der Waals network (D168A), which cannot be recuperated from [40]. The P1-P3 macrocyclization, on the other hand, does not have a pronounced role in drug resistance profiles for analyzed MDR mutations, possibly due to being away from the mutation sites, which are clustered together [40].

The effect of the helicase domain in potency loss due to drug resistance mutations is less pronounced, but correlates with the identity of the P2 extension. In a full length NS3/4A crystal structure with an inhibitor similar to danoprevir, P2 extension occupies the groove formed between the helicase and the protease active site [48]. Although no specific interactions between the inhibitor and the helicase were observed, solvent exclusion from the otherwise exposed hydrophobic P2 extension may be the reason for the observed potency increase in the asunaprevir and MK-5172 classes. The opposite trend observed in danoprevir is surprising and cannot be explained yet. The modulation of inhibition by mutations in the helicase on the other hand did not produce a consistent trend with respect to any inhibitor feature, consistent with a previous report [31]. In either case, modulation of protease inhibition by the helicase domain is observed, even though no specific recognition by the helicase was designed into any of these protease inhibitors.

In conclusion, this study elucidates the specific roles of different chemical structural features of protease inhibitors in association with HCV NS3/4A. P2 extension plays a pivotal role, and optimizing this moiety is the key to overcoming drug resistance. Another interesting feature is the extent of potency modulation in the full length protein via possibly non-specific interactions with the helicase domain. Accordingly, the helicase domain should be investigated as a possible recognition site for future inhibitor development. By considering inhibitor recognition by the helicase as well as the protease active site, better inhibitors that recognize both domains but stay within the substrate envelope can be generated. A dual-action inhibitor that recognizes both the helicase and the protease inhibits both domains, because the helicase domain must be liberated from

the protease domain to be fully active in RNA unwinding [49]. This, in turn, would lead to more effective treatments including monotherapy, where the cost and the length of combinatorial therapy may be reduced and side effects of the current treatment regime may be circumvented.

**Acknowledgements**

We would like to thank Keith Romano and Djade Soumana for valuable discussions; and Nese Kurt Yilmaz for proof-reading and editing the paper.

This work was supported by NIH grant R01 AI085051.

## Supplementary Methods

### *Inhibitor Design and Synthesis*

Danoprevir, asunaprevir, vaniprevir, and MK-5172 have similar core structures, but contain different heterocyclic moieties at P2 and varying macrocyclization status and position. These drugs exhibit differential susceptibilities to major drug resistant protease variants R155K, A156T, and D168A in both enzymatic and antiviral assays. To understand the effect of different P2 moieties and macrocyclization status on inhibitor activity against drug resistant NS3/4A protease variants, we designed and synthesized acyclic and macrocyclic analogues of these drugs. Asunaprevir, a linear compound with an ether-linked isoquinoline moiety at P2 proline, was prepared along with its P1–P3 macrocyclic analogue (AH-55). Danoprevir and vaniprevir are both macrocyclic compounds containing a carbamate-linked isoindoline moiety at P2 proline, but the direction of the macrocycle is different: danoprevir contains a P1–P3 macrocycle while vaniprevir contains a P2–P4 macrocycle, and a saturated P1 side chain [21]. For direct comparison with danoprevir, a vaniprevir P1 olefin analogue (AG-53) was prepared along with an acyclic analogue (AG-30) of these compounds. Finally, for MK-5172, which is a P2–P4 macrocyclic compound containing an ether-linked quinoxaline moiety, both the acyclic (AH-49) and the P1–P3 macrocyclic analogues (AH-53) were prepared.

The asunaprevir P1–P3 macrocyclic analogue (AH-55) was synthesized as the following. Briefly, the P2 fragment, (2*S*,4*R*)-1-(tert-butoxycarbonyl)-4-((7-chloro-4-methoxyisoquinolin-1-yl)oxy)pyrrolidine-2-carboxylic acid, obtained by the reaction of Boc-*trans*-4-hydroxy-L-proline and 1,7-dichloro-4-methoxyisoquinoline, was coupled

with the preassembled P1–P1' fragment, (1*R*,2*S*)-1-amino-*N*-(cyclopropylsulfonyl)-2-vinylcyclopropanecarboxamide [50], using *O*-(7-azabenzotriazol-1-yl)-1,1,3,3-tetramethyluroniumhexafluorophosphate (HATU)/ diisopropylethylamine (DIEA) to provide the P2–P1' intermediate. Boc deprotection followed by the coupling of the free amine with the Boc-protected amino acid (*S*)-2-(*tert*-butoxycarbonylamino)non-8-enoic acid using HATU/DIEA provided the *bis*-olefin precursor for ring-closing metathesis. Cyclization of the *bis*-olefin intermediate was accomplished using a highly efficient ring-closing metathesis catalyst Zhan 1B, and provided the asunaprevir P1–P3 macrocyclic analogue (AH-55).

The danoprevir acyclic analogue (AG-30) was synthesized as the following. Briefly, the P2 and P1–P1' fragments were preassembled and the final compound was generated by a four step reaction sequence, including Boc deprotection, P2–P3 amide coupling, ester hydrolysis, and finally coupling with the P1–P1' fragment. The P2–P3 fragment was assembled by coupling the Boc-*L-tert*-leucine with the preassembled P2 fragment, (3*R*,5*S*)-5-(methoxycarbonyl) pyrrolidin-3-yl 4-fluoroisindoline-2-carboxylate [40], using HATU/DIEA. Hydrolysis of the P2–P3 methyl ester with LiOH·H<sub>2</sub>O in a THF-MeOH-H<sub>2</sub>O mixture followed by coupling of the resulting acid with the P1–P1' fragment using HATU/DIEA provided the analogue AG-30.

Vaniprevir and its P1 olefin analogue (AG-53) were synthesized following the synthetic methods reported by McCauley *et al.* [24]. The MK-5172 acyclic and P1–P3 macrocyclic analogues were synthesized as the following. The intermediate (2*S*,4*R*)-1-*tert*-butyl 2-methyl 4-((3-chloro-7-methoxyquinoxalin-2-yl)oxy)pyrrolidine-1,2-



dicarboxylate was prepared from the quinoxaline derivative 3-chloro-7-methoxyquinoxalin-2-ol and the bosylated Boc-*cis*-4-hydroxy-L-proline, (2*S*,4*R*)-1-*tert*-butyl 2-methyl 4-(((4-bromophenyl)sulfonyl)oxy)pyrrolidine-1,2-dicarboxylate, using Cs<sub>2</sub>CO<sub>3</sub>-mediated S<sub>N</sub>2 displacement reaction [25]. Vinylation with potassium trifluoro(vinyl)borate via palladium catalysis followed by hydrogenation of the double bond provided the common P2 intermediate (2*S*,4*R*)-1-*tert*-butyl 2-methyl 4-((3-ethyl-7-methoxyquinoxalin-2-yl)oxy)pyrrolidine-1,2-dicarboxylate. The acyclic analogues (AH-49) was assembled from the P2 intermediate by a four step reaction sequence, including Boc deprotection, P2–P3 amide coupling, ester hydrolysis, and coupling with the P1–P1' fragment. the bis-olefin precursor for the P1–P3 macrocyclic analogue was synthesized from the P2 fragment in an analogous reaction sequence. Finally, cyclization of the *bis*-olefin intermediate using ring-closing metathesis catalyst Zhan 1B provided the MK5172 P1–P3 macrocyclic analogue (AH-54).

## References

1. Lavanchy, D., *The global burden of hepatitis C*. Liver Int, 2009. **29 Suppl 1**: p. 74-81.
2. Shepard, C.W., L. Finelli, and M.J. Alter, *Global epidemiology of hepatitis C virus infection*. Lancet Infect Dis, 2005. **5**(9): p. 558-67.
3. Kuiken, C. and P. Simmonds, *Nomenclature and numbering of the hepatitis C virus*. Methods Mol Biol, 2009. **510**: p. 33-53.
4. Simmonds, P., et al., *Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes*. Hepatology, 2005. **42**(4): p. 962-73.
5. Liang, T.J. and T. Heller, *Pathogenesis of hepatitis C-associated hepatocellular carcinoma*. Gastroenterology, 2004. **127**(5 Suppl 1): p. S62-71.
6. Ghany, M.G., et al., *An update on treatment of genotype 1 chronic hepatitis C virus infection: 2011 practice guideline by the American Association for the Study of Liver Diseases*. Hepatology, 2011. **54**(4): p. 1433-44.
7. Poordad, F., et al., *Boceprevir for untreated chronic HCV genotype 1 infection*. N Engl J Med, 2011. **364**(13): p. 1195-206.
8. Forestier, N. and S. Zeuzem, *Triple therapy with telaprevir: results in hepatitis C virus-genotype 1 infected relapsers and non-responders*. Liver Int, 2012. **32 Suppl 1**: p. 44-50.
9. Jacobson, I.M., et al., *Telaprevir for previously untreated chronic hepatitis C virus infection*. N Engl J Med, 2011. **364**(25): p. 2405-16.

10. Zeuzem, S., et al., *Telaprevir for retreatment of HCV infection*. N Engl J Med, 2011. **364**(25): p. 2417-28.
11. Bacon, B.R., et al., *Boceprevir for previously treated chronic HCV genotype 1 infection*. N Engl J Med, 2011. **364**(13): p. 1207-17.
12. Ge, D., et al., *Genetic variation in IL28B predicts hepatitis C treatment-induced viral clearance*. Nature, 2009. **461**(7262): p. 399-401.
13. Fried, M.W., *Side effects of therapy of hepatitis C and their management*. Hepatology, 2002. **36**(5 Suppl 1): p. S237-44.
14. Morikawa, K., et al., *Nonstructural protein 3-4A: the Swiss army knife of hepatitis C virus*. J Viral Hepat, 2011. **18**(5): p. 305-15.
15. Li, K., et al., *Immune evasion by hepatitis C virus NS3/4A protease-mediated cleavage of the Toll-like receptor 3 adaptor protein TRIF*. Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2992-7.
16. Foy, E., et al., *Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-1 signaling*. Proc Natl Acad Sci U S A, 2005. **102**(8): p. 2986-91.
17. Foy, E., et al., *Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease*. Science, 2003. **300**(5622): p. 1145-8.
18. Kwong, A.D., et al., *Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus*. Nat Biotechnol, 2011. **29**(11): p. 993-1003.

19. Venkatraman, S., et al., *Discovery and structure-activity relationship of P1-P3 ketoamide derived macrocyclic inhibitors of hepatitis C virus NS3 protease*. J Med Chem, 2009. **52**(2): p. 336-46.
20. McPhee, F., *Identification and preclinical profile of the novel HCV NS3 protease inhibitor BMS-650032*. J Hepatol, 2010. **52**: p. S296.
21. Llinas-Brunet, M., et al., *Discovery of a potent and selective noncovalent linear inhibitor of the hepatitis C virus NS3 protease (BI 201335)*. J Med Chem, 2010. **53**(17): p. 6466-76.
22. Seiwert, S.D., et al., *Discovery and Development of the HCV NS3/4A Protease Inhibitor Danoprevir (ITMN-191/RG7227)*, in *Antiviral Drugs*. 2011, John Wiley & Sons, Inc. p. 257-271.
23. Raboisson, P., et al., *Discovery and Development of the HCV Protease Inhibitor TMC435*, in *Antiviral Drugs*. 2011, John Wiley & Sons, Inc. p. 273-286.
24. McCauley, J.A., et al., *Discovery of vaniprevir (MK-7009), a macrocyclic hepatitis C virus NS3/4a protease inhibitor*. J Med Chem, 2010. **53**(6): p. 2443-63.
25. Harper, S., et al., *Discovery of MK-5172, a Macrocyclic Hepatitis C Virus NS3/4a Protease Inhibitor*. ACS Medicinal Chemistry Letters, 2012: p. 120302153334001.
26. Lam, A.M. and D.N. Frick, *Hepatitis C virus subgenomic replicon requires an active NS3 RNA helicase*. J Virol, 2006. **80**(1): p. 404-11.
27. Frick, D.N., *The hepatitis C virus NS3 protein: a model RNA helicase and potential drug target*. Curr Issues Mol Biol, 2007. **9**(1): p. 1-20.

28. Beran, R.K., V. Serebrov, and A.M. Pyle, *The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate*. J Biol Chem, 2007. **282**(48): p. 34913-20.
29. Beran, R.K. and A.M. Pyle, *Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase*. J Biol Chem, 2008. **283**(44): p. 29929-37.
30. Frick, D.N., et al., *The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex RNA efficiently*. J Biol Chem, 2004. **279**(2): p. 1269-80.
31. Thibeault, D., et al., *Use of the fused NS4A peptide-NS3 protease domain to study the importance of the helicase domain for protease inhibitor binding to hepatitis C virus NS3-NS4A*. Biochemistry, 2009. **48**(4): p. 744-53.
32. Dahl, G., et al., *Effects on protease inhibition by modifying of helicase residues in hepatitis C virus nonstructural protein 3*. FEBS J, 2007. **274**(22): p. 5979-86.
33. Dahl, G., O.G. Arenas, and U.H. Danielson, *Hepatitis C virus NS3 protease is activated by low concentrations of protease inhibitors*. Biochemistry, 2009. **48**(48): p. 11592-602.
34. Herrmann, E., et al., *Hepatitis C virus kinetics*. Antivir Ther, 2000. **5**(2): p. 85-90.
35. Rong, L., et al., *Rapid emergence of protease inhibitor resistance in hepatitis C virus*. Sci Transl Med, 2010. **2**(30): p. 30ra32.
36. Sarrazin, C. and S. Zeuzem, *Resistance to direct antiviral agents in patients with hepatitis C virus infection*. Gastroenterology, 2010. **138**(2): p. 447-62.

37. Kieffer, T.L., A.D. Kwong, and G.R. Picchio, *Viral resistance to specifically targeted antiviral therapies for hepatitis C (STAT-Cs)*. J Antimicrob Chemother, 2010. **65**(2): p. 202-12.
38. Hofmann, W.P. and S. Zeuzem, *Hepatitis C in 2011: A new standard of care and the race towards IFN-free therapy*. Nat Rev Gastroenterol Hepatol, 2012. **9**(2): p. 67-8.
39. Wittekind, M., et al., *Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes*, in *United States Patent Applications Publication*2002: United States of America. p. 26.
40. Romano, K.P., et al., *Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding*. Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20986-91.
41. Butt, A.A. and F. Kanwal, *Boceprevir and telaprevir in the management of hepatitis C virus-infected patients*. Clin Infect Dis, 2012. **54**(1): p. 96-104.
42. Romano, K.P., et al., *Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease*. J Virol, 2011. **85**(13): p. 6106-16.
43. King, N.M., et al., *Combating susceptibility to drug resistance: lessons from HIV-1 protease*. Chem Biol, 2004. **11**(10): p. 1333-8.
44. Prabu-Jeyabalan, M., E. Nalivaika, and C.A. Schiffer, *Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes*. Structure, 2002. **10**(3): p. 369-81.

45. Lim, S.R., et al., *Virologic escape during danoprevir (ITMN-191/RG7227) monotherapy is hepatitis C virus subtype dependent and associated with R155K substitution*. *Antimicrob Agents Chemother*, 2012. **56**(1): p. 271-9.
46. Lagace, L., et al., *In vitro resistance profile of the hepatitis C virus NS3 protease inhibitor BI 201335*. *Antimicrob Agents Chemother*, 2012. **56**(1): p. 569-72.
47. Lenz, O., et al., *In vitro resistance profile of the hepatitis C virus NS3/4A protease inhibitor TMC435*. *Antimicrob Agents Chemother*, 2010. **54**(5): p. 1878-87.
48. Schiering, N., et al., *A macrocyclic HCV NS3/4A protease inhibitor interacts with protease and helicase residues in the complex with its full-length target*. *Proc Natl Acad Sci U S A*, 2011. **108**(52): p. 21052-6.
49. Ding, S.C., A.S. Kohlway, and A.M. Pyle, *Unmasking the active helicase conformation of nonstructural protein 3 from hepatitis C virus*. *J Virol*, 2011. **85**(9): p. 4343-53.
50. Wang, X.A., et al., *Hepatitis C virus inhibitors*, 2006.

**CHAPTER IV**

**DISCUSSION**



## Discussion

### *Effect of the Domain-Domain Interface in Interdomain Dependence and Function*

To date, several crystal structures of the single chain NS3/4A protease – helicase complex (scNS3/4A) have been published [1-3]. In all of these structures, helicase domain covers the otherwise shallow, solvent exposed active site of the protease, and a 900 Å<sup>2</sup> interface is formed in this conformation. Most functional interfaces bury a surface area of at least 1500 Å<sup>2</sup> and are primarily formed between hydrophobic residues [4, 5]. The domain-domain interface in NS3/4A does not share these characteristics, and relevance of this interface has not been investigated before. Along with the semi-flexible linker region which covalently attaches the two domains together, this interface is the location through which domain-domain communication could happen. In Chapter II, a two-part approach was taken to validate this hypothesis. Molecular dynamics studies on the scNS3/4A and deletion constructs demonstrated significant interdomain correlations between helicase and protease. However, the observed interdomain correlations were not as strong as *intra*-domain correlations. In addition, these correlations were affected by deletions in the linker and/or C-terminus regions. These results suggest that the domain-domain interactions through the crystallographic interface is transient, and disruptions in the interface enhance protein function rather than disrupting. Biochemical assays probing multiple activities of full length scNS3/4A, interface mutants and isolated domains demonstrate that destabilizing the interface affect activities of both domains *in vitro*. However, the effect observed is possibly not through the domain-domain interface. Both in protease cleavage and DNA unwinding/binding experiments, disruption of the

interface caused increased activities compared to scNS3/4A protease – helicase complex. MD simulations and biochemical studies both indicate that the interface is transient and disruptions in the interface cause the two domains to become more independent.

The effect of the helicase domain on protease activity has been debated before, with opposing views. The first view is that helicase domain does not affect protease cleavage and the protease domain is more active in isolated form [6-8]. In our experiments, which utilized a similar construct – the single chain NS3/4A, similar results were observed. However, Pyle and coworkers demonstrated up to two orders of magnitude enhancement in protease function in the full length NS3/4A [9]. When these results are compared, kinetic efficiencies for both the single chain and full length constructs were comparable. However, the kinetic efficiency for the isolated protease construct from the Pyle lab was significantly lower. This construct is formed by the isolated protease domain and the unmodified NS4A, which contains the C-terminal product of the NS4A-NS4B cleavage junction. The C-terminus of the NS4A is in favorable position to occupy the active site of the protease, hence act as a competitive inhibitor of protease activity – which may explain the lower reaction rates observed with this construct. Overall, the helicase represses the protease function and disrupting the interface also disrupts the transient association between the two domains, leading to the liberation of the protease domain and the increase in cleavage activity.

Helicase DNA unwinding activity was modulated by the interface mutations in a similar manner as observed for protease activity, which indicates that the core functionalities of both domains are restricted when the domains are bound together.

However, RNA unwinding activity was highly enhanced when protease domain was present, and the modulation via the interface was minimal. In this case, presence of the protease domain is the main cause of this enhancement, rather than the interface.

Additionally, similar results were obtained from studies of the *Flaviviridae* NS3/2B complex, where the two domains form an alternate crystallographic interface [10-12].

However, the effect of the protease domain in RNA unwinding is unclear.

In conclusion, the interface between protease and helicase domains of HCV NS3/4A is likely more regulatory than functional. The domain-domain interface has to be disrupted in order to achieve optimum activity. This observation implies that when the two domains are liberated, they are free to rotate around the semi-flexible linker, making individual domains independent, yet in favorable distance to switch between a closed, inactive conformation and an open, active conformation. Disruptions in the interface increase the rate of cycling between conformations.

The crystallographic and solution structures of the *Flavivirus* single chain NS3/2B exhibits a different domain-domain arrangement compared to the HCV scNS3/4A. scNS3/2B constructs are similar to HCV scNS3/4A except they lack the C-terminal helix and the C-terminus product, which interacts with the HCV NS3/4A protease active site. Without the C-terminus, the *Flavivirus* scNS3/2B forms an extended conformation, where the protease domain is rotated away from the helicase and the N-terminal half of the protease interacts with the helicase subdomains 1 and 2 [10-12]. Therefore, will the absence of the C-terminus affect the activity of the HCV NS3/4A? Pyle and coworkers elegantly demonstrated that when the C-terminus was deleted, the incubation time for

maximal helicase activity was lowered [13]. This observation supports my hypothesis in two ways. First, the interface formed between the two domains is transient and mostly dependent on strong associations between the C-terminus and the protease. Second, with the C-terminus deletion, the conversion from the closed conformation to open is more rapid. In this context, the important question is whether the two domains associate and form an alternate interface in the open conformation or they rotate freely without any interaction. Presence of an alternate functional interface (compared to the HCV NS3/4A) was demonstrated by Lescar and coworkers for Dengue Virus NS3/2B [12]. Mutations in residues at the protease, which form contacts to the ATPase loop between helicase subdomains 1 and 2, affect ATP hydrolysis and RNA unwinding *in vitro*. Hence, a secondary interaction location between the two domains is also a possibility in the HCV NS3/4A. Mapping these mutations in the HCV NS3/4A and probing for modulations in the helicase activities may yield similar results, which would further complement the alternate conformational arrangement hypothesis.

Future studies should attempt to elucidate the different domain-domain arrangements with respect to various stages of cleavage and RNA unwinding. Crystal structures of HCV scNS3/4A to date demonstrate the apo form of the protein and various stages of RNA translocation with respect to ATP hydrolysis. However, both domains go through multiple steps to perform their respective activities. Reaction intermediates should be probed structurally, using solution-based biophysical methods, such as small angle x-ray scattering (SAXS), dynamic light scattering (DLS) and static light scattering (SLS), to obtain the size, shape and molecular weight of reaction intermediates.

Information obtained from these studies will provide the domain-domain orientation and oligomerization status of these species. An alternate method that can be utilized is deuterium exchange, which will provide solvent accessibility information. Alternate interface patterns can be obtained from deuterium exchange rates for reaction states. Lower deuteration rates mean lower solvent accessibility, hence a possible interaction between the domains. Furthermore, the rates of transition between different states can be obtained from single molecule experiments. If the two domains are labeled differentially with a FRET pair, the rate of transition between alternate states can be followed in real time using single molecule FRET methods. In both cases, the results obtained should elucidate how the two domains behave with respect to each other while performing their function.

#### *Effect of the Protease Domain in Helicase RNA Preference*

For both the *Flavivirus* NS3/2B and HCV NS3/4A, presence of the protease domain confers RNA preference over DNA and significantly enhances the RNA unwinding of the helicase domain [14-16]. In Chapter II, similar results were obtained for the scNS3/4A construct in comparison with the isolated helicase. This enhancement was mainly due to the presence of the protease domain, role of the domain-domain interface was limited and positive – disruptions in the interface caused further minor enhancements in unwinding activity. Comparable enhancement was observed for both the full length NS3/4A and scNS3/4A. Hence, the protease domain alone is the main reason behind the observed enhancement in RNA unwinding.

A direct interaction between NS3/4A protease and viral RNA can be the reason for the observed enhancement. Most DEAD-box RNA binding/processing proteins contain cofactors that enhance ATP hydrolysis, RNA binding and remodeling [17-19]. A third role, other than polyprotein processing and immune suppression, can be attributed to the protease domain – being a helicase cofactor. The protease may fulfill this role by potentially working as a RNA clamp by interacting with viral RNA. Previously, RNA binding activity for Rhinovirus serine protease 3C was demonstrated [20, 21]. Additionally, engineered RNA oligomers were shown to inhibit HCV NS3/4A protease specifically [22, 23]. In essence, possible RNA binding activity of the protease increases the efficiency and/or the processivity of the helicase.

A recent study by Ray and Das demonstrate that specific interactions between viral RNA and protease domain are possible [24]. In this study, the effect of NS3/4A protease on transcription of viral RNA and translation was thoroughly investigated. Specific binding was observed between HCV internal ribosome entry site (IRES) and the protease domain which can be antagonized by the host La protein – which also binds to the HCV IRES and function as a translation switch. Translation of HCV polyprotein was repressed when NS3/4A protease was overexpressed, possibly by competing with the La protein, and increases in the amount of HCV RNA transcripts were observed. This interaction is mapped to the active site of the protease due to the fact that proteolysis was inhibited when HCV IRES is bound to the protease. In addition, NS4A was eliminated as a possible cause of this interaction, only interactions with the C-terminal half of the

protease was observed. In essence, specific interactions between the HCV IRES and NS3/4A protease were shown to be important in HCV life cycle.

However, the above observation is not enough to determine the observed enhancement in RNA unwinding *in vitro*. Substrates for NS3 helicase unwinding reactions usually consist of a longer bottom strand and a shorter strand, which leaves a single stranded overhang stretch of rU (or dT in DNA substrates). Helicase domain associates with the substrate on this single stranded region, and the length of this single stranded stretch determines how many NS3/4A molecules associate with the RNA and perform the unwinding reaction (~6 nucleotides per molecule) [25]. Accordingly, if there is a specific interaction between helicase substrates and protease domain, dsRNA should be investigated for interactions with the protease domain, because if the helicase associates with single stranded region, protease domain, in turn, could contact the double stranded region. Enzymatic binding assays such as fluorescence anisotropy, gel-shift assays or surface plasmon resonance may be performed between single and double stranded HCV-derived RNA and protease domain, including non specific RNA from other organisms (e.g. Poliovirus RNA, which has no specificity towards the protease [24]) to probe for binding specificity and the mode of binding. Preferential binding for the HCV-derived RNA should be observed from these binding assays. Further experiments such as mutagenesis and crystallography can additionally be performed to pinpoint the location of specific interactions between the protease and RNA.

#### *Domain-Domain Arrangement in Vivo*

Biochemical and biophysical characterization of the HCV NS3/4A was primarily performed using individual domains or full length constructs which are expressed in a high yield expression system, purified and assayed in buffer systems. However, *in vivo*, NS3/4A protein integrates into the ER membrane, possibly via the N-terminal transmembrane helix of NS4A and helix  $\alpha 0$  of the protease domain [26]. Membrane integration is critical for viral propagation and disrupting membrane insertion results in the impairment of NS3/4A function [27, 28]. However, when the crystallographic structures of the scNS3/4A were juxtaposed onto the membrane-attached structural model of NS3/4A, the helicase domain clashes with the membrane [26]. Thus, the transient interface and cycling between open and close conformers observed *in vitro* may not be valid *in vivo*.

Helix  $\alpha 0$  possibly works as the initial targeting mechanism that directs the protein to the membrane either co-translationally or immediately after translation. Liberation of the NS4A, via the *cis* cleavage reaction in the NS3-NS4A junction, happens afterwards; hence integration of NS4A to the membrane is controlled by the initial interaction between protease helix  $\alpha 0$  and the membrane [26]. One major question is, whether this interaction is transient or permanent. If this interaction is transient, protease domain becomes relatively free to sample more conformations, possibly the closed conformation. One way to test this hypothesis is to employ an *in vitro* system involving liposomes and full length NS3/4A. This way, the activities of NS3/4A can be tested relatively easily in a membrane-like environment and by comparing results obtained from liposome studies to earlier unbound solution studies, effect of membrane binding can be evaluated.



Sub-genomic replicons of HCV can also be utilized as an alternate method to probe membrane binding. Sub-genomic replicons include the critical minimal machinery of the HCV (NS3/4A, NS4B, NS5A and NS5B) plus the HCV IRES that provide the essential functions of the virus, in a cell-based system. These replicons contain reporter systems to analyze protease (e.g. MAVS cleavage), replication (e.g. quantitative Real Time Polymerase Chain Reaction – qRT-PCR) and translation (e.g. luciferase activity). Interface mutations that were characterized biochemically can be cloned into these replicons to evaluate the domain-domain interface, to validate *in vitro* observations and eventually to ascertain closed and open conformations.

#### *Effect of Helicase Domain in Protease Inhibition*

HCV NS3/4A protease is the major target in combating HCV infection. The directly acting antivirals (DAAs) approved for treatment by Food and Drug Administration (FDA) – telaprevir and boceprevir – are antagonists of the protease domain and majority of the inhibitors in clinical trials, which exhibit up to low picomolar potencies, are also being developed against HCV NS3/4A protease [29, 30]. However, the role of the helicase domain in inhibitor recognition is largely undetermined. Conflicting results were reported that either favor the modulation of inhibitor potency by the helicase domain or argue against any effect of the helicase domain [7, 31, 32]. However, these studies utilize a limited range of drugs and protease constructs. In Chapter III, enzymatic potencies of ten inhibitors were determined for wildtype scNS3/4A and protease, multi drug resistance (MDR) mutants and helicase mutants. Either between the protease and the scNS3/4A constructs for wildtype and MDR variants

or between the wildtype scNS3/4A and helicase mutants, up to an order of magnitude modulation in drug potencies was observed. These results signify that, even without any designed recognition between protease inhibitors and helicase domain, helicase has a minor but discernible effect in inhibitor potency.

Modulation of inhibitor potency by the helicase domain primarily depends on the identity of the P2 extension of inhibitors. P2 extensions for the inhibitors are mostly bulky hydrophobic groups. In a recent crystal structure of scNS3/4A with a protease inhibitor, P2 extension occupies the groove formed between helicase domain and protease active site [3]. Specific interactions between helicase domain and the P2 extension were not observed, however non-specific interactions mostly hydrophobic in nature is a possibility. Observed modulations were more apparent in MDR mutants, which additionally signify that the loss of interactions between protease and the inhibitor can partially be recovered with helicase interactions. D168A mutant constructs however exhibit contradictory behavior, where danoprevir analogs significantly disfavor the helicase domain. Isoindoline moiety in danoprevir analogs is significantly dislocated in crystal structures of MDR variants [33] and change in conformation of the isoindoline extension may clash with the helicase domain, which may account for protease preference. In conclusion, helicase domain must be further explored for protease inhibition.

Using the helicase domain as a constraint in protease inhibitor development will open further avenues in drug discovery. Drugs that interact with both protease and helicase domains specifically will be advantageous. By strongly interacting with

domains, the interaction surface of the drug will be increased, resulting in stronger drug-protein interactions. In theory, these drugs should also inhibit the helicase domain and inhibit the competition between the inhibitor and protease substrates. By binding both domains simultaneously, the protein will be kept in the closed state. In the closed state, the helicase covers the protease active site, which should essentially render protease active site inaccessible. In addition, locking the two domains should also inhibit helicase function, since unwinding function of the helicase is most likely performed in the open conformation. Thus, dual-action protease inhibitors that recognize both domains strongly and specifically will result in repression of cleavage and unwinding activities of HCV NS3/4A, two essential functions in HCV life cycle.

Existing protease inhibitors can be altered to recognize both domains. MK-5172 acyclic analog (AH-49) is a good starting point for creating a proof-of-concept dual action NS3/4A inhibitor. MK-5172 and analogs are very potent against the protease domain and, except MK-5172 itself, not as susceptible to drug resistance compared to other inhibitors. Quinoxaline moiety of MK-5172 analogs favorably interact with H57 and D81, members of the catalytic triad which are strongly conserved, hence significant portion of the potency is retained because mutations in the catalytic triad essentially make the protein incompetent. This quinoxaline moiety can be further extended to the helicase by adding hydrophobic side chains or better occupation of this hydrophobic cleft. In addition, different P4-capping moieties can be explored, similar to the one used by Schiering *et al.* [3], to form more specific contacts with the helicase, especially residues

Q526 and H528. Potential gains in potency from these modifications can be tested both in enzymatic or replicon-based assays.

### **Conclusion**

HCV NS3/4A contains two important machineries essential for viral propagation. For this reason, both domains are still being investigated vigorously, for both scientific and pharmaceutical purposes. In addition, interdomain dependence between the domains poses an interesting scientific question. Why have these distinct domains evolved to work together? By better understanding the interdomain dependence, we can provide a better answer to this challenging question. Furthermore, we can find ways to exploit this phenomenon to produce better DAAs that may lead to the cure of Hepatitis C, using a single, highly specific, dual-action next generation drug that is effective regardless of the genotypical differences between the different HCV genotypes.

The lessons learned from HCV NS3/4A can also be applied to *Flaviviruses*. *Flaviviruses* are mostly disease causing viruses in humans and animals alike and, collectively, the annual death toll caused by *Flavivirus*-related diseases are significant, especially in the underdeveloped parts of the world. *Flavivirus* and HCV NS3 proteins are similar in composition and function. More interestingly, interdomain dependence observed in HCV NS3/4A also applies to *Flavivirus* NS3/2B. Essentially, similarities between these two proteins can be utilized to learn about the evolution of the *Flaviviridae* protease – helicase complexes and eventually produce effective DAAs against *Flaviviruses* to cure associated viral diseases.

## References

1. Yao, N., et al., *Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase*. *Structure*, 1999. **7**(11): p. 1353-63.
2. Appleby, T.C., et al., *Visualizing ATP-dependent RNA translocation by the NS3 helicase from HCV*. *J Mol Biol*, 2011. **405**(5): p. 1139-53.
3. Schiering, N., et al., *A macrocyclic HCV NS3/4A protease inhibitor interacts with protease and helicase residues in the complex with its full-length target*. *Proc Natl Acad Sci U S A*, 2011. **108**(52): p. 21052-6.
4. Nooren, I.M. and J.M. Thornton, *Structural characterisation and functional significance of transient protein-protein interactions*. *J Mol Biol*, 2003. **325**(5): p. 991-1018.
5. Wodak, S.J. and J. Janin, *Structural basis of macromolecular recognition*. *Adv Protein Chem*, 2002. **61**: p. 9-73.
6. Gallinari, P., et al., *Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus*. *J Virol*, 1998. **72**(8): p. 6758-69.
7. Thibeault, D., et al., *Use of the fused NS4A peptide-NS3 protease domain to study the importance of the helicase domain for protease inhibitor binding to hepatitis C virus NS3-NS4A*. *Biochemistry*, 2009. **48**(4): p. 744-53.
8. Taremi, S.S., et al., *Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease*. *Protein Sci*, 1998. **7**(10): p. 2143-9.

9. Beran, R.K. and A.M. Pyle, *Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase*. J Biol Chem, 2008. **283**(44): p. 29929-37.
10. Mastrangelo, E., et al., *Crystal structure and activity of Kunjin virus NS3 helicase; protease and helicase domain assembly in the full length NS3 protein*. J Mol Biol, 2007. **372**(2): p. 444-55.
11. Luo, D., et al., *Crystal structure of the NS3 protease-helicase from dengue virus*. J Virol, 2008. **82**(1): p. 173-83.
12. Luo, D., et al., *Flexibility between the protease and helicase domains of the dengue virus NS3 protein conferred by the linker region and its functional implications*. J Biol Chem, 2010. **285**(24): p. 18817-27.
13. Ding, S.C., A.S. Kohlway, and A.M. Pyle, *Unmasking the active helicase conformation of nonstructural protein 3 from hepatitis C virus*. J Virol, 2011. **85**(9): p. 4343-53.
14. Frick, D.N., et al., *The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex RNA efficiently*. J Biol Chem, 2004. **279**(2): p. 1269-80.
15. Chernov, A.V., et al., *The two-component NS2B-NS3 proteinase represses DNA unwinding activity of the West Nile virus NS3 helicase*. J Biol Chem, 2008. **283**(25): p. 17270-8.
16. Beran, R.K., V. Serebrov, and A.M. Pyle, *The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate*. J Biol Chem, 2007. **282**(48): p. 34913-20.

17. Fairman-Williams, M.E., U.P. Guenther, and E. Jankowsky, *SF1 and SF2 helicases: family matters*. *Curr Opin Struct Biol*, 2010. **20**(3): p. 313-24.
18. Jankowsky, E., *RNA helicases at work: binding and rearranging*. *Trends Biochem Sci*, 2011. **36**(1): p. 19-29.
19. Pyle, A.M., *RNA helicases and remodeling proteins*. *Curr Opin Chem Biol*, 2011. **15**(5): p. 636-42.
20. Matthews, D.A., et al., *Structure of human rhinovirus 3C protease reveals a trypsin-like polypeptide fold, RNA-binding site, and means for cleaving precursor polyprotein*. *Cell*, 1994. **77**(5): p. 761-71.
21. Leong, L.E., P.A. Walker, and A.G. Porter, *Human rhinovirus-14 protease 3C (3Cpro) binds specifically to the 5'-noncoding region of the viral RNA. Evidence that 3Cpro has different domains for the RNA binding and proteolytic activities*. *J Biol Chem*, 1993. **268**(34): p. 25735-9.
22. Fukuda, K., et al., *Isolation and characterization of RNA aptamers specific for the hepatitis C virus nonstructural protein 3 protease*. *Eur J Biochem*, 2000. **267**(12): p. 3685-94.
23. Umehara, T., et al., *Designing and analysis of a potent bi-functional aptamers that inhibit protease and helicase activities of HCV NS3*. *Nucleic Acids Symp Ser (Oxf)*, 2004(48): p. 195-6.
24. Ray, U. and S. Das, *Interplay between NS3 protease and human La protein regulates translation-replication switch of Hepatitis C virus*. *Sci Rep*, 2011. **1**: p. 1.

25. Mackintosh, S.G., et al., *Structural and biological identification of residues on the surface of NS3 helicase required for optimal replication of the hepatitis C virus*. J Biol Chem, 2006. **281**(6): p. 3528-35.
26. Brass, V., et al., *Structural determinants for membrane association and dynamic organization of the hepatitis C virus NS3-4A complex*. Proc Natl Acad Sci U S A, 2008. **105**(38): p. 14545-50.
27. Horner, S.M., H.S. Park, and M. Gale, Jr., *Control of innate immune signaling and membrane targeting by the Hepatitis C virus NS3/4A protease are governed by the NS3 helix alpha0*. J Virol, 2012. **86**(6): p. 3112-20.
28. He, Y., et al., *The N-terminal helix alpha(0) of hepatitis C virus NS3 protein dictates the subcellular localization and stability of NS3/NS4A complex*. Virology, 2012. **422**(2): p. 214-23.
29. Hofmann, W.P. and S. Zeuzem, *Hepatitis C in 2011: A new standard of care and the race towards IFN-free therapy*. Nat Rev Gastroenterol Hepatol, 2012. **9**(2): p. 67-8.
30. Vermehren, J. and C. Sarrazin, *New HCV therapies on the horizon*. Clin Microbiol Infect, 2011. **17**(2): p. 122-34.
31. Dahl, G., O.G. Arenas, and U.H. Danielson, *Hepatitis C virus NS3 protease is activated by low concentrations of protease inhibitors*. Biochemistry, 2009. **48**(48): p. 11592-602.
32. Dahl, G., et al., *Effects on protease inhibition by modifying of helicase residues in hepatitis C virus nonstructural protein 3*. FEBS J, 2007. **274**(22): p. 5979-86.



33. Romano, K.P., et al., *Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding*. Proc Natl Acad Sci U S A, 2010. **107**(49): p. 20986-91.