2011-05-31

Mechanisms of Substrate Recognition by HCV NS3/4A Protease Provide Insights Into Drug Resistance: A Dissertation

Keith P. Romano
University of Massachusetts Medical School

MECHANISMS OF SUBSTRATE RECOGNITION BY HCV NS3/4A PROTEASE PROVIDE INSIGHTS INTO DRUG RESISTANCE

A Dissertation Presented

By

Keith Patrick Romano

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

May 31, 2011

Biochemistry and Molecular Pharmacology

MECHANISMS OF SUBSTRATE RECOGNITION BY HCV NS3/4A PROTEASE PROVIDE INSIGHTS INTO DRUG RESISTANCE

A Dissertation Presented

By

Keith Patrick Romano

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

Phillip D. Zamore, Ph.D., Member of Committee

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

David N. Frick, 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.

Sean P. Ryder, 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

May 31, 2011

To the many teachers throughout my life…

“Critical comments by students should be taken in a friendly spirit – accumulation of material should not stifle the student's independence.”

— Albert Einstein

ACKNOWLEDGEMENTS

I must thank many people for their support over the years. First and foremost, I recognize the Department of Biochemistry and Molecular Pharmacology for fostering such a rich learning environment. The leadership starts with our Department Chair, Dr. Robert Matthews, whose encouragement of congeniality and collaboration permeates to every member of our department. Pursuing a doctorate – an extremely long-term commitment – becomes less about deriving knowledge and more about seeking it from others. The office doors of our long, arching hallways were always open, sometimes until very late at night. At one time or another, I approached nearly every faculty, post-doc, student and secretary for assistance and was received in every instance with enthusiasm and respect.

I must specifically recognize those who contributed directly to my research. My thesis committee, Drs. Sean Ryder, Phillip Zamore and David Lambright, were persistently supportive and critical over the years. When choosing my committee, I sought the smartest scientists available – a decision that paid off. My committee members challenged me to be meticulous in my work, critical of other studies and thoughtful in my justifications (even when trivial). Dr. Akbar Ali – our synthetic chemist whose talents are surpassed only by his pleasantness – provided me with nearly every drug that I crystallized. Dr. Herbert Klei of Bristol-Myers Squibb shared key conditions for protein expression, purification and crystallization. My external reviewer, Dr. David Frick, provided invaluable assistance throughout my thesis, initially driving here from New York to share his knowledge, reagents and protocols.

I also must thank my mentors and colleagues for making our laboratory such a wonderful work environment. In truth, my successes are but reflections of the great talents of those around me. First and foremost, the leadership and compassion of my mentor, Dr. Celia Schiffer, created the benevolent lab culture ripe with progress and collaboration. Dr. William Royer also devoted countless hours to discuss every intricacy of my project. Ellen Nalivaika, Jared Auclair, Moses Prabu and Madhavi Nalam were among the first to teach me the basics – cloning, protein expression and crystallization. Shiven Shandilya – thoughtful and capable in science – also spent hours managing our computers, and did so with a smile. Ayşegül Özen and Yufeng Cai enthusiastically (and patiently) assisted in calculations of the substrate envelope. And lastly, my bay-mates – Madhavi Kolli, Sagar Kathuria, Seema Mittal, Rajintha Bandaranayake and Djade Soumana – become my pseudo-siblings and great sources of support. We walked the line between work and play with balance, having fun while accomplishing so much.

I am also indebted to my family for their unconditional support. My parents and sister – to whom I owe my life's trajectory – supported my every step with love, encouragement and guidance. My uncle and aunt, Gary and Anne Romano, attended every major life event, and of course, my defense was no exception. My friends provided helpful entertainment and support, but also grounded me in my humble roots. My in-laws, the Saleemuddin family, welcomed me as a son and nurtured me with kindness and spices through the busiest times. And lastly, I remain thankful and committed to my beautiful wife Aasia – my best friend, soul mate and only true possession in the changing fortunes of time.

ABSTRACT

HCV afflicts many millions of people globally, and antiviral therapies are often ineffective and intolerable. The Food and Drug Administration approved the HCV protease inhibitors telaprevir and boceprevir in May 2011, marking an important milestone in anti-HCV research over the past two decades. Nevertheless, severe drug side effects of combination therapy – flu-like symptoms, depression and anemia – limit patient adherence to treatment regimens. The acquisition of resistance challenges the long-term efficacy of antiviral therapies, including protease inhibitors, as suboptimal dosing allows for the selection of drug resistant viral variants. A better understanding of the molecular basis of drug resistance is therefore central to developing future generation protease inhibitors that retain potency against a broader spectrum of HCV strains.

To this end, my research characterizes the molecular basis of drug resistance against HCV protease inhibitors. Chapter II defines the mode of substrate recognition by the common volume shared by NS3/4A substrate products – the substrate envelope. Chapter III then correlates patterns of drug resistance to regions where drugs protrude from the substrate envelope. Lastly, Chapter IV elucidates the molecular underpinnings of resistance against four leading protease inhibitors – telaprevir, danoprevir, vaniprevir and MK-5172 – and provides practical approaches to designing novel drugs that are less susceptible to resistance. I ultimately hope my work appeals to the broader biomedical community of virologists, medicinal chemists and clinicians, who struggle to understand HCV and other human pathogens in the face of rapid disease evolution.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABBREVIATIONS

β-ME –	β-mercaptoethanol
Cha –	β-cyclohexyl-L-alanine
Dif –	L-(3,3)-diphenlyalanine
DMF –	N,N′-dimethylformamide
DTT –	Dithiothreitol
HCV	Hepatitis C Virus
HEPES –	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HIV-1 –	Human Immunodeficiency Virus 1
HPLC –	High performance liquid chromatography
IPTG –	Isopropyl β-D-1-thiogalactopyranoside
MES –	2-(N-morpholino)ethanesulfonic acid
MWCO –	Molecular weight cutoff
PEG –	Polyethylene glycol
RMSD –	Root-mean-square deviation
vdW –	Van der Waals

PREFACE

Most will not find the patience or time to read my dissertation cover-to-cover. Thus each research chapter reads as an independent manuscript, comprising comprehensive background, discussion and references. The unintended consequences of such formatting, however, are periodic redundancies in content or theme. Nevertheless, I hope that my dissertation – in its current form – more practically conveys the implications my research in its broader medical context.

Chapter II has been published as:

Romano KP, Laine JM, Deveau LM, Cao H, Massi F, Schiffer CA. *Molecular mechanisms of viral and host-cell substrate recognition by HCV NS3/4A protease.* J. Virology. (2011): 85(13): 6106-16.

Chapter III has been published as:

Romano KP, Ali A, Royer WE, Schiffer CA. *Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding.* Proc Natl Acad Sci USA. (2010): 107(49): 20986-91.

Chapter IV has been published as:

Romano KP, Ali A, Soumana D, Özen A, Aydin C, Cao H, Deveau LM, Newton A, Petropoulos CJ, Huang W, Schiffer CA. *The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors.* [submitted].

CHAPTER I

INTRODUCTION

HEPATITIS C VIRUS

Since the discovery of Hepatitis C Virus (HCV) in 1989 (*1*), the World Health Organization estimates that the virus infects 3% of the global population and that 3–4 million new infections occur annually (*2*). About 80% of all infections evade immune detection and persist chronically in patients (*3*). HCV most commonly spreads through direct blood contact (*4*), though most infections go unrecognized due to mild disease symptoms (*3, 5*). Clinicians most often establish HCV diagnoses by screening patients for exposure risks rather than overt disease symptoms (*6*). In fact, approximately three-quarters of infected individuals are not formally diagnosed with the disease (*7*). An estimated 15–30% of chronically infected patients develop liver cirrhosis over a period of three decades (*8*). Most patients remain unaware of HCV infection until advanced-stage disease – a leading cause of cirrhosis, hepatocellular carcinoma and liver transplantation in the United States (*9, 10*).

HCV replicates rapidly in untreated patients, producing about one trillion virus particles daily (*11, 12*). Inaccuracy of viral replication accounts for the large genetic diversity among the six major genotypes globally, each subcategorized into several subtypes and many strains. Genotype 1 viruses cause the majority of all infections, predominating in North America, South America and Europe, while genotype 2 viruses are less common throughout the world. Genotype 3 viruses mainly afflict Southeast Asia and certain other countries, such as Russia. Genotype 4 viruses predominate in the Middle East, Egypt and central Africa. Genotype 5 viruses account for most infections in South Africa, while genotype 6 viruses are endemic to Asia (*13*).

Until recently, anti-HCV therapies comprised yearlong treatment with ribavirin and pegylated interferon-α, requiring strict patient adherence with weekly clinical appointments (*6*). In May 2011, the Food and Drug Administration approved the protease inhibitors telaprevir and boceprevir for use in combination with ribavirin and pegylated interferon-α. Nevertheless, severe drug side effects – flu-like symptoms, depression and anemia – limit patient adherence to treatment regimens. The high cost and adverse effects of antiviral drugs prohibit many clinicians from offering therapy to patients lacking the social, financial and emotional support systems necessary for strict adherence (*14*). Moreover, combination therapy with ribavirin and pegylated interferon-α cures only half of patients, depending on both host cell and viral genetic factors (*15-19*). Thus current anti-HCV treatment options are insufficient, and better antiviral therapies, diagnostic tools and vaccines must be developed and implemented throughout the world.

THE HCV LIFECYCLE

A *hepacivirus* of the *Flaviviridae* family, HCV contains an RNA genome with a single open reading frame flanked by 5′ and 3′ untranslated regions (UTRs). The positive-sense genome is packaged in an icosahedral scaffold of Core proteins, which are enveloped by phospholipid membranes derived from the host cell endoplasmic reticulum. The viral envelope glycoproteins, E1 and E2, insert within the outer envelope and facilitate virus fusion to human hepatocytes **(Figure 1.1)**. Specifically, E1 and E2 mediate fusion to the host cell membrane receptors Tetraspanin (CD81) and Scavenger receptor class B type I (SR-B1) (*20, 21*). The co-expression of CD81 and SR-B1 alone are insufficient for efficient HCV fusion, however, and additional factors further mediate the process, such as Claudin 1, Occludin, Low-density lipoprotein receptors and glycosaminoglycans *(22-24)*. After stable fusion to the host cell surface, the virus enters the cytosol by clathrin-mediated endocytosis (*25*). The acidic endoplasmic environment facilitates virus fusion to the endosomal membrane, leading to the release of the RNA genome into the cytoplasm (*26*).

Host cell machinery then translates the HCV genome in a cap-independent manner as a 3011-residue polyprotein, which is subsequently processed by host cell and viral proteases into structural (Core, E1, E2, p7) and non-structural (NS2, NS3, NS4A, NS4B, NS5A, NS5B) components (*27*). The structural proteins are inserted into the endoplasmic reticulum lumen and cleaved into functional units by host signal peptidases. The non-structural proteins – most exhibiting enzymatic functions – are translated along the cytosolic side of the endoplasmic reticulum and processed into functional units by

Figure 1.1: The HCV lifecycle. HCV virions fuse to human hepatocytes through interactions with CD81 and SR-B1 and enter by clathrin-mediated endocytosis; virus particles uncoat in the cytosol and host cell machinery translates the genomic RNA along the endoplasmic reticulum membrane; viral replication proceeds from a single duplex RNA template; nascent virions then assemble in the endoplasmic reticulum and are trafficked through the Golgi secretary pathway for extracellular release.

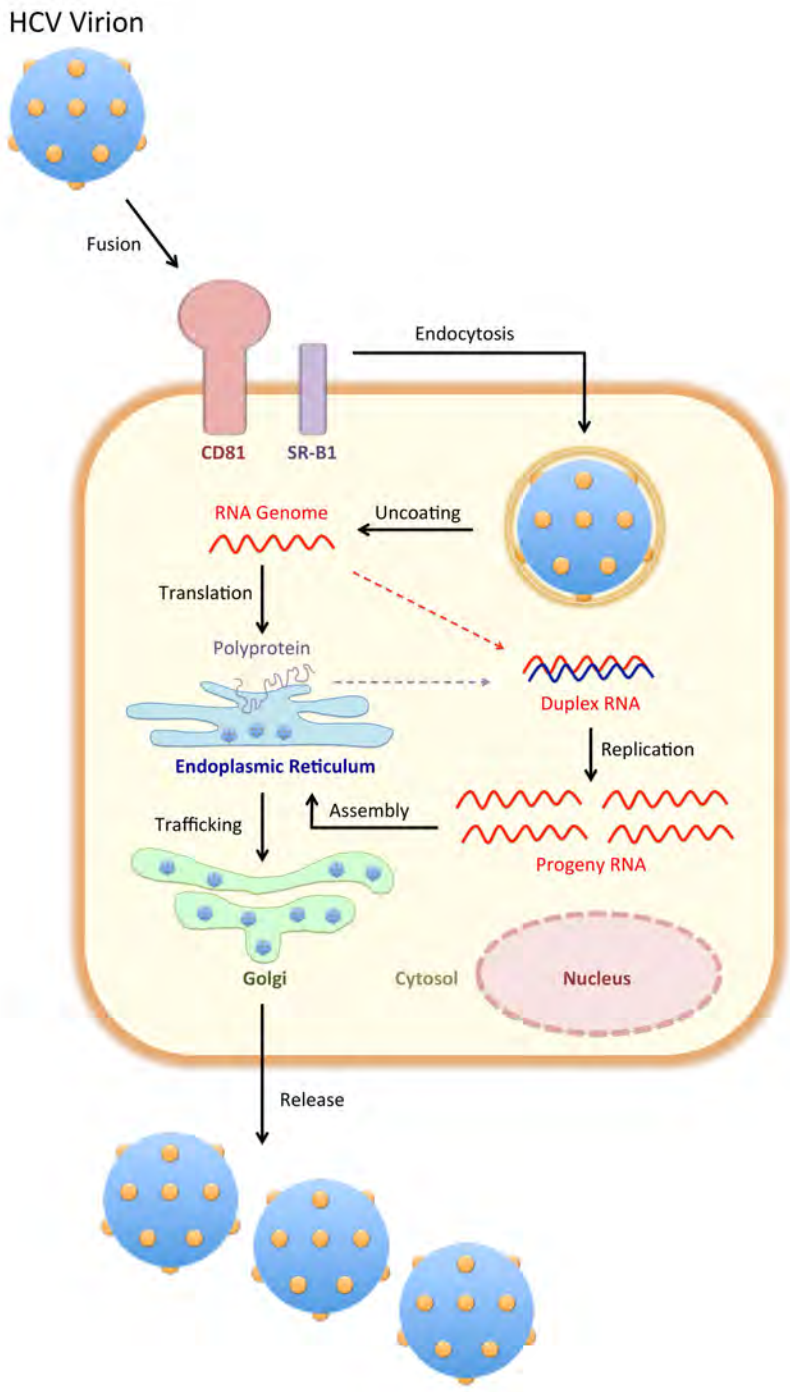

HCV Virion
Fusion
CD81
SR-B1
Endocytosis
RNA Genome
Uncoating
Translation
Polyprotein
Duplex RNA
Endoplasmic Reticulum
Replication
Trafficking
Assembly
Progeny RNA
Golgi
Cytosol
Nucleus
Release

Figure 1.1

two viral proteases, NS2/3 and NS3/4A (*28*). After translation of the necessary proteins, the viral replication complex forms along the cytoplasmic membrane, characterized as membranous webs by electron microscopy (*29*). The viral RNA-dependent RNA polymerase, NS5B, initiates HCV replication by synthesizing a single, negative-sense strand from the positive-sense RNA genome. NS5B copies the resulting duplex RNA molecule to generate many positive-sense progeny genomes (*30-32*). The HCV genomes then enter the endoplasmic reticulum and are packaged into virus particles, which are subsequently trafficked through the Golgi secretory pathway for extracellular release.

THE BIFUNCTIONAL NS3 PROTEIN

The NS3 protein comprises 631 amino acids organized into two functional domains **(Figure 1.2)**, which are each essential for viral replication (*33, 34*). The N-terminal protease domain, a chymotrypsin-like serine protease, heterodimerizes with the viral NS4A protein for efficient proteolytic activity (*35*). The C-terminal helicase domain, a super family II DEAD-box helicase (*36*), uses ATP hydrolysis to drive duplex RNA unwinding. Although the protease and helicase domains remain active in isolation (*37-41*), each domain depends on the other for optimal catalytic efficiency. The presence of the protease domain, for example, enhances the efficiency of duplex RNA unwinding (*38, 41*). Active duplex RNA unwinding also inhibits the concurrent hydrolysis of viral substrate peptides (*42*), suggesting that the protease domain may support RNA unwinding in a proteolytically-inhibited conformation. The helicase domain was also shown to enhance NS3/4A proteolytic efficiency (*40*), although others groups report comparable catalytic efficiencies between the full-length NS3 protein and protease domain alone (*39, 43*). Thus potential domain interdependency may influence NS3 protease and helicase activities, as well as HCV replication, evolution and virulence.

The NS3-mediated processing of viral and host cell targets is central to HCV replication and its evasion of the human innate immune response. NS3/4A protease cleaves four known sites in the viral polyprotein at junctions 3-4A, 4A4B, 4B5A and 5A5B. NS3 first cleaves the 3-4A junction *in cis* and subsequently forms a heterodimer with cofactor NS4A. The NS3/4A complex then cleaves the remaining three sites *in trans* (*35*). Schechter and Berger first formulated the nomenclature for peptide cleavage

Figure 1.2: Crystal structure of the bifunctional HCV NS3 protein. The NS3 helicase domain (pink) packs against the protease domain (cyan) with the essential cofactor NS4A shown in blue. The C-terminus of the helicase domain (magenta) – the post-cleavage product substrate 3-4A – binds to the protease active site with the catalytic triad residues shown in yellow.

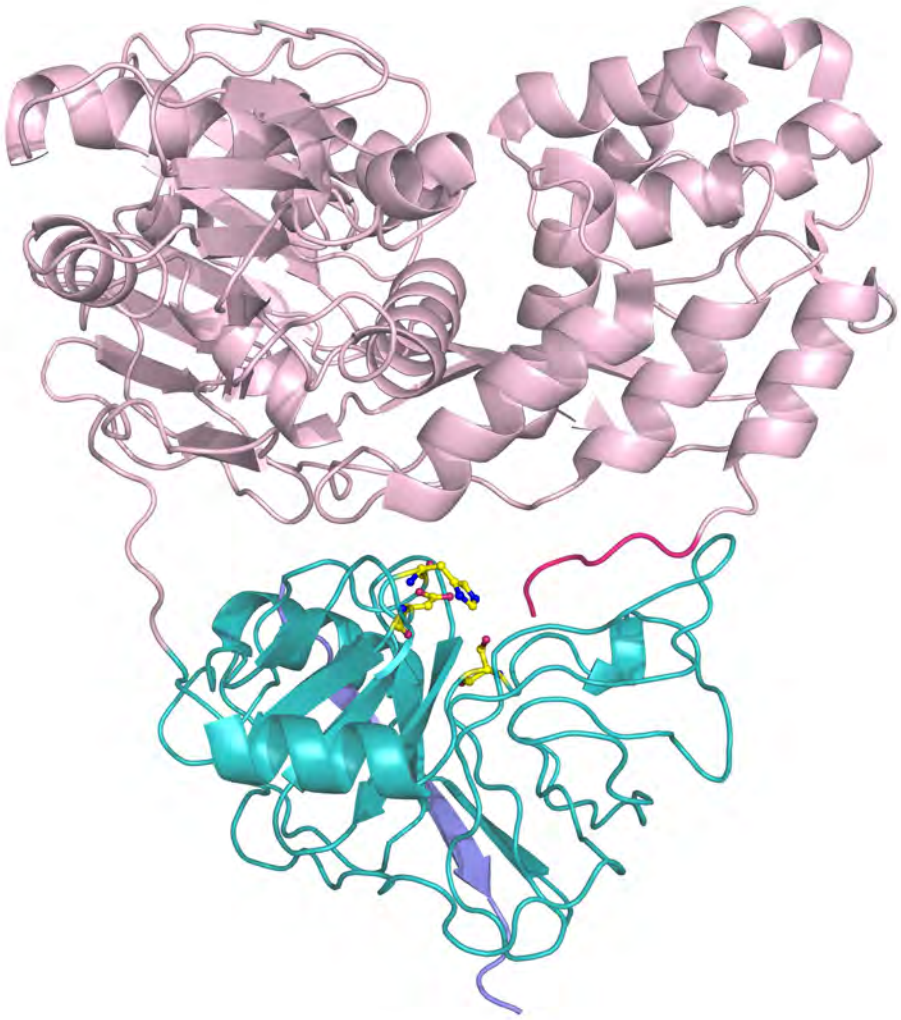


Figure 1.2

positions in 1967: they defined the cleavage site between residues P1–P1' (represented N-terminal to C-terminal) and designated upstream and downstream amino acids by incremental numbering (*44*). Notably, the four NS3/4A cleavage sites share little sequence homology spanning P6–P4', except of an acid at P6, cysteine or threonine at P1 and serine or alanine at P1' **(Table 1.1)**.

The NS3/4A protease also disrupts the innate immune response to viral infection by inactivating the transcription factors Interferon regulatory factor 3 (IRF-3) and Nuclear factor κB (NF-κB) (*45, 46*). The detection of viral RNA upregulates these host cell transcription factors through two distinct pathways, involving either Toll-like receptor 3 (TLR3) or Retinoic acid-inducible gene I (RIG-I) (*47, 48*). NS3/4A protease disrupts the TLR3 and RIG-I cascades by cleaving the essential adaptor proteins Toll-interleukin-1 receptor domain-containing adaptor-inducing interferon-β (TRIF) and Mitochondrial antiviral signaling protein (MAVS), respectively (*46, 49*). The TRIF and MAVS cleavage sites share little sequence homology with other viral substrates: TRIF contains cysteine at P1 and serine at P1', while MAVS contains glutamate at P6 and cysteine at P1 (**Table 1.1**). Notably, TRIF contains a tract of eight proline residues spanning P13–P6, which may play a unique role in the recognition process (*50*). Thus NS3/4A protease recognizes these viral and host cell sequences specifically as substrates despite their large sequence divergence, obscuring the molecular determinants of substrate recognition.

Table 1.1. Genotype 1a primary cleavage sequences of NS3/4A substrates.

Substrate	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'
TRIF	$P_{(8)}$	S	S	T	P	C	S	A	H	L
MAVS	E	R	E	V	P	C	H	R	P	S
3-4A	D	L	E	V	V	T	S	T	W	V
4A4B	D	E	M	E	E	C	S	Q	H	L
4B5A	E	C	T	T	P	C	S	G	S	W
5A5B	E	D	V	V	C	C	S	M	S	Y

EMERGING ANTIVIRAL THERAPIES

Since the protease domain was first crystallized in 1997 (*51*), pharmaceutical companies have invested considerable resources in developing NS3/4A protease inhibitors. Early drug design efforts were hampered by the relatively shallow, featureless architecture of the protease active site. The eventual observation of N-terminal product inhibition served as a stepping-stone for the discovery of more potent hexapeptide inhibitors (*52, 53*). Over the past decade, pharmaceutical companies have further developed these lead compounds by evaluating the effect of functional moieties at positions P4–P1′ on antiviral potency (*54-60*). The resulting protease inhibitor class contains both linear and macrocyclic drug candidates. Macrocyclic drugs are constrained to adopt the secondary structure of bound hexapeptides, forcing *trans* geometry of the P2–P3 amide bond (*61, 62*). Linear drugs achieve similar geometry with bulky P3 substituents, which disfavor *cis* P2–P3 bonds (*63*). Most protease inhibitors include either P1–P1′ ketoamide moieties, which form reversible covalent bonds with the protease, or sulfonamide groups.

Boehringer Ingelheim first demonstrated protease inhibitor efficacy in the 2005 clinical trial with its leading drug, ciluprevir (*64, 65*). Though later discontinued due to cardiotoxicity (*66*), ciluprevir provided proof-of-principle that protease inhibitors could achieve potent antiviral effects. Several other companies, including Vertex, Merck, Bristol-Myers Squibb, Roche, Tibotec and Medavir, have since developed promising clinical candidates **(Figure 1.3)**. Most notably, boceprevir (Merck) and telaprevir (Vertex) – linear ketoamide compounds – were approved in May 2011 by the Food and

Figure 1.3: HCV NS3/4A protease inhibitors in development. The protease inhibitor class contains both linear and macrocyclic compounds. Ciluprevir first demonstrated efficacy in clinical trials, but was later discontinued due to cardiotoxicity. Telaprevir and boceprevir were approved by the Food and Drug Administration in May 2011 for use in combination with ribavirin and pegylated interferon-α. Many other drug candidates are in various stages of clinical development.

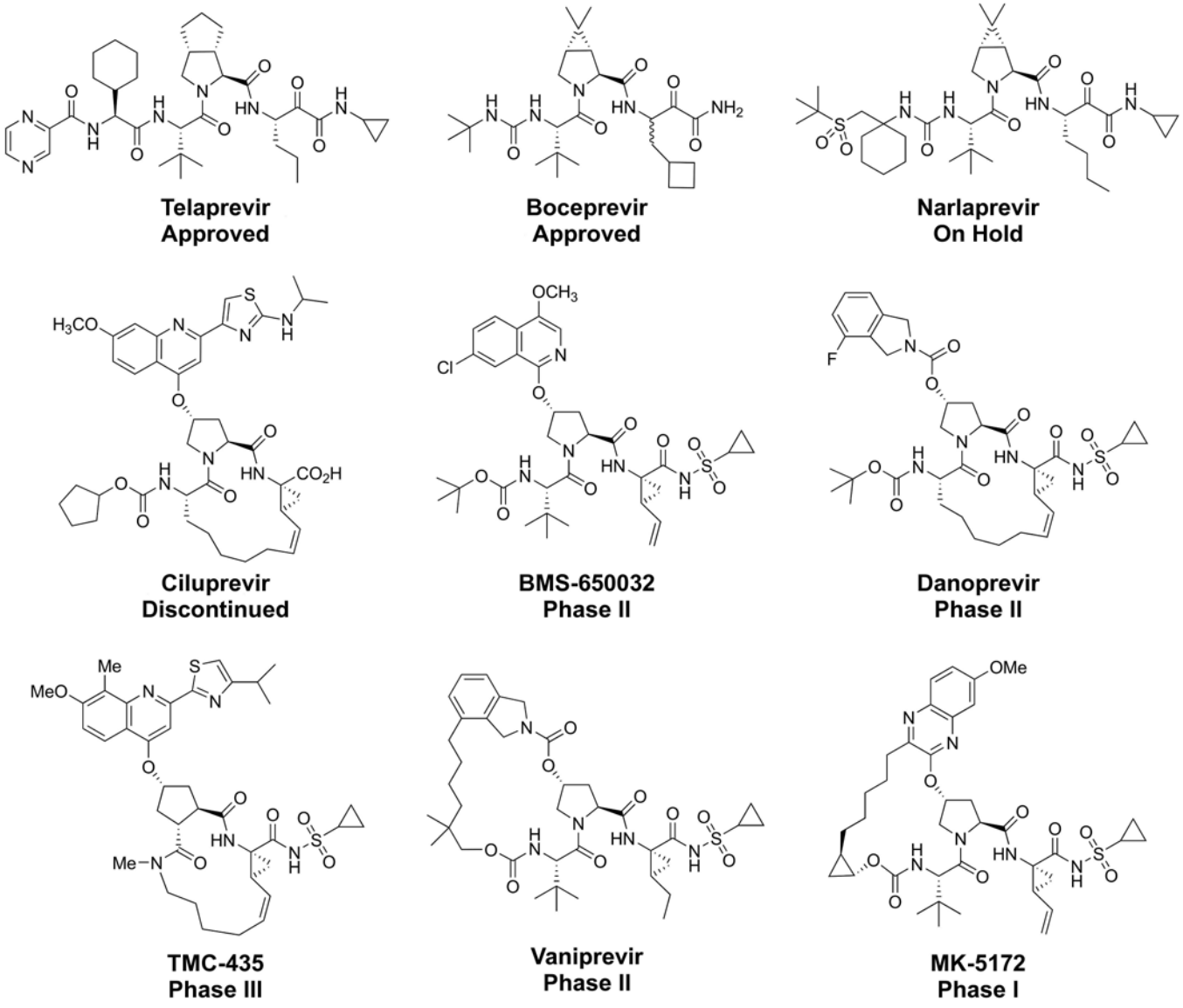

Telaprevir
Approved
Boceprevir
Approved
Narlaprevir
On Hold
Ciluprevir
Discontinued
BMS-650032
Phase II
Danoprevir
Phase II
TMC-435
Phase III
Vaniprevir
Phase II
MK-5172
Phase I

Figure 1.3

Drug Administration, marking an important milestone in anti-HCV research and drug development over the past two decades. Other drugs are also advancing through clinical trials: *(i)* the macrocyclic sulfonamide TMC-435 (Tibotec/Medavir) in phase 3, *(ii)* the macrocyclic sulfonamide drugs danoprevir (Roche) and vaniprevir (Merck) in phase 2, *(iii)* the linear sulfonamide BMS-650032 (Bristol-Myers Squibb) and linear carboxylate BI201335 (Boehringer Ingelheim) in phase 2, and *(iv)* the macrocyclic sulfonamide MK-5172 (Merck) in phase 1 clinical trials. Thus more anti-HCV treatment options will become available to patients in the coming decade, led by many clinical candidates in the protease inhibitor class.

Pharmaceutical companies have also developed nucleoside and non-nucleoside inhibitors of the NS5B polymerase. Nucleoside inhibitors mimic the natural viral substrates of the RNA-dependent RNA polymerase and act as terminators in nascent RNA chains. Though these drugs display marked antiviral potencies, severe adverse effects have halted the development of several, such as R1626 (Roche) (*67, 68*). Nevertheless, some nucleoside inhibitors, such as RG7128 (Roche/Pharmasset) and PSI-7977 (Pharmasset), have advanced into phase 2 clinical trials. Non-nucleoside inhibitors bind to allosteric sites away from the NS5B active site and induce inhibited enzyme conformations. Though resistance mutations often limit the efficacy of these compounds (*69*), many non-nucleoside inhibitors are in early phases of human clinical trials, such as filibuvir (Pfizer), BI207127 (Boehringer Ingelheim), GS-9190 (Gilead) and ABT-333 (Abbott) (*70*).

Difficulties in propagating HCV in cell culture have hampered the identification

of other drug targets and delayed the discovery of novel antiviral drugs. Nearly two decades after the discovery of HCV, researchers in 2005 established the first cell culture systems that robustly supported viral replication in human-derived cells (*71, 72*). The improved understanding of HCV biology has led to the identification of many other antiviral targets, including the viral proteins NS5A, NS4B and p7 (*73-75*). The viral NS5A protein plays an essential role in viral replication, assembly and release. The highly potent NS5A inhibitor BMS-790052 (Bristol-Myers Squibb) is showing promise in phase 2 clinical trials (*69*). In addition, pharmaceutical companies are targeting host cell proteins central to HCV replication, such as cyclophilins, which regulate HCV replication by chaperoning protein folding. The cyclophilin inhibitor alisporivir (Novartis) demonstrated efficacy against several HCV genotypes during monotherapy and in combination with pegylated interferon-α (*76*). Thus diverse antiviral therapies are emerging in the collective effort to combat HCV in the face of its global disease burden.

PROTEASE INHIBITOR RESISTANCE

The absence of proofreading activity makes the HCV NS5B RNA-dependent RNA polymerase highly inaccurate. Misincorporation of bases during replication causes high mutation rates in all viral proteins (*77*). Some of these mutations are neutral, while others alter the viability of the virus. Patients infected with HCV for a period of time will develop heterogeneous virus populations, known as quasispecies (*78-80*). As patients begin antiviral treatment, selective drug pressures favor resistant viral strains, particularly in the setting of monotherapy (*81*). To reduce the likelihood of resistance, treatment regimens utilize multiple antiviral agents simultaneously that limit viral evolution by distinct mechanisms.

Nevertheless, combination therapies with ribavirin and pegylated interferon-α are only effective for a subset of patients, depending on both human and viral genetic factors (*15-19*). Moreover, sustained treatment adherence remains a critical predictor for treatment outcome, as suboptimal dosing allows for the selection of drug resistant viral variants (*82*). Adverse drug effects, including flu-like symptoms, depression and anemia, are commonly reported and often hinder the successful completion of full-dose treatment regimens (*82, 83*). Though high treatment completion rates are reported in clinical trials, these patients are routinely screened prior to enrollment and actively supported throughout therapy to maximize adherence (*84*). In the actual clinical setting, however, several small cohort studies suggest completion rates as low as 50%, highlighting the influence of disease co-morbidities, socioeconomic status, emotional support and related factors on treatment outcome (*85, 86*). Furthermore, dominant polymerase and protease

inhibitor resistance mutations have been detected in a minority of treatment naïve individuals, suggesting that drug resistant strains pre-exist and may be directly transmissible (*87*). Thus despite the benefits of combination therapy in limiting viral evolution, the emergence of drug resistance still challenges the long-term efficacy of antiviral therapies.

Indeed, the rapid rise of resistance has reduced the efficacy of many potent protease inhibitors **(Table 1.2)**. Most notably, mutations at residues R155, A156 and D168 – located in the protease P2 subsite – confer resistance against many leading drugs. In particular, the mutations A156T/V and R155K confer high-level resistance against nearly all drugs reported in the literature (*88-93*). Resistance mutations at residues V36, T54 and V170 are more distant from the protease site and implicate indirect effects, such as protein dynamics, in drug binding. In addition, the S489L mutation – within the NS3 helicase domain – confers resistance against danoprevir, suggesting that the helicase domain may influence drug binding (*94*). Drug resistance occurs when mutations in the target disrupt inhibitor binding while preserving target function (*95*). Thus drug resistant protease variants retain hydrolytic efficiency, highlighting the need to understand the molecular underpinnings of resistance in the context of substrate recognition. Ultimately, structural studies may help elucidate the mechanistic pathways of resistance, facilitating additional approaches for developing protease inhibitors that are equally potent but less susceptible to resistance.

Table 1.2. Reported sites of drug resistance mutations

Res.	Mutations	Drugs
V36	A, M, L, G	Boceprevir, telaprevir
Q41	R	Boceprevir, danoprevir
F43	S, C, V, I	Boceprevir, telaprevir, danoprevir, TMC435
V55	A	Boceprevir
T54	A, S	Boceprevir, telaprevir
Q80	K, R, H, G, L	TMC435
S138	T	Danoprevir, TMC435†
R155	K, T, I, M, G, L, S, Q	Boceprevir, telaprevir, danoprevir, ciluprevir, TMC435
A156	V, T, S, I, G	Boceprevir, telaprevir, danoprevir, ciluprevir, TMC435
V158	I	Boceprevir
D168	A, V, E, G, N, T, Y, H, I	Danoprevir, ciluprevir, TMC435
V170	A	Boceprevir, telaprevir
M175	L	Boceprevir

†-TMC435 displays reduced activity against S138T, but the mutation was not observed in selection experiments.

INSIGHTS FROM HIV-1 PROTEASE

The acquisition of HIV-1 protease inhibitor resistance provides important insights into the HCV drug resistance. The nine HIV-1 protease inhibitors approved by the Food and Drug Administration contain similar functional groups and form similar interactions with the protease (*95-97*). The natural viral substrates – though highly diverse in primary sequence – bind to the active site in a common three-dimensional volume, termed the substrate envelope (*96*). The superposition of drug and substrate complexes reveals that most resistance mutations occur where the inhibitors protrude from the substrate envelope **(Figure 1.4)** (*95*). Resistance mutations in regions outside the substrate envelope – space not occupied by viral substrates – selectively prevent drug binding with minimal effect on substrate processing.

In theory, inhibitors designed to fit within the substrate envelope will be less susceptible to resistance, as mutations preventing drug binding will concurrently disrupt the binding of viral substrates. To test this theory, novel protease inhibitors were designed to fit within the substrate envelope based on the scaffold of the highly potent drug darunavir (*98-101*). The effects of chemical substituents on antiviral potency were evaluated at several positions on the darunavir scaffold. After several rounds of optimization, ten compounds exhibited greater antiviral activity than darunavir in cell culture infected with highly resistant HIV-1 strains **(Figure 1.5)**. Thus the substrate envelope-based design strategy led to highly potent, broadly active HIV-1 protease inhibitors, and may also prove useful in developing of robust therapies against other rapidly evolving pathogens, such as HCV.

Figure 1.4: HIV-1 protease substrate and inhibitor envelopes. (A) The HIV-1 protease substrate envelope, depicted in blue, is defined as the consensus vdW volume shared by any four of the six viral substrates. (B) The consensus vdW shared by the protease inhibitors defines the inhibitor envelope, shown in red. (C) Superposition of the inhibitor envelope reveals regions where the drugs collectively protrude from the substrate envelope, which coincide with sites of drug resistance mutations.

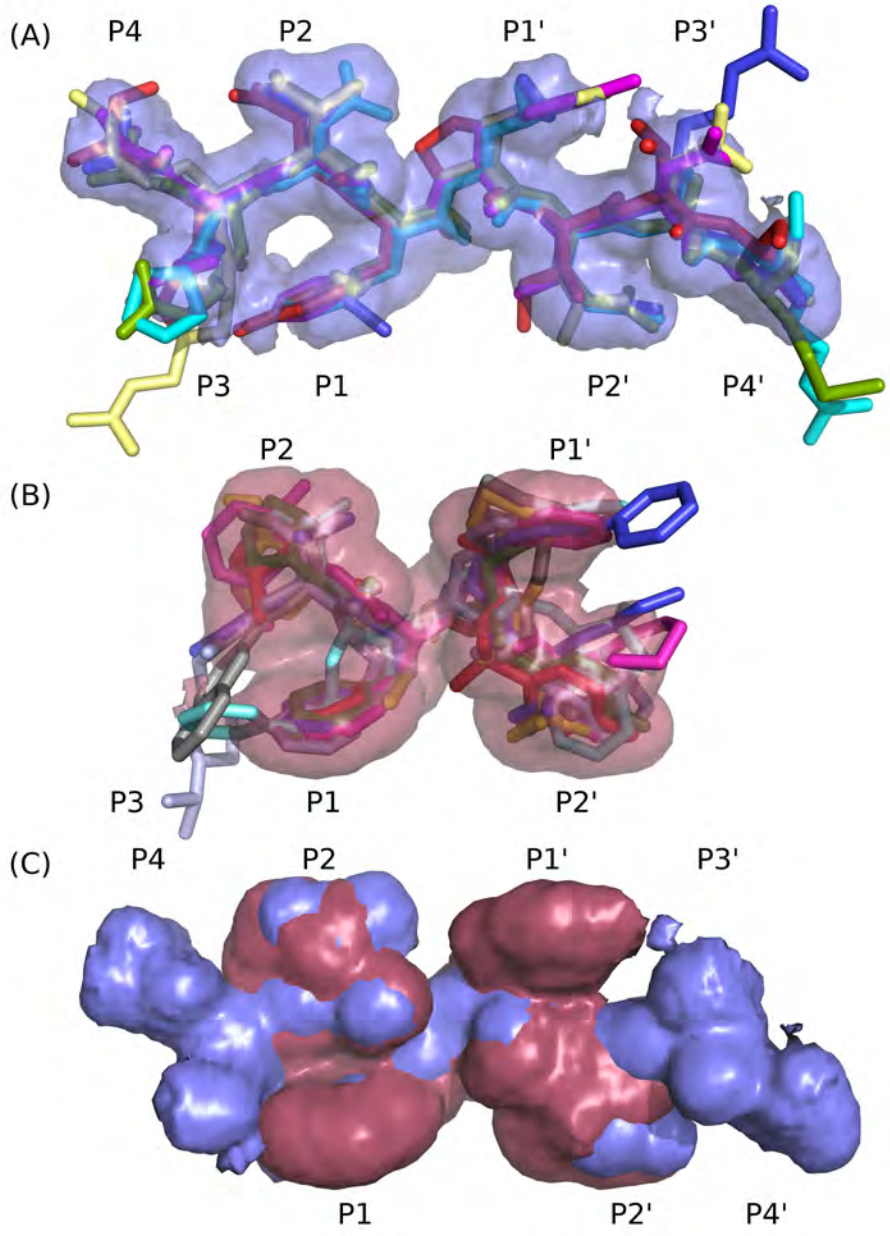


Figure 1.4

Figure 1.5: HIV-1 substrate envelope-based inhibitor design. Ten drug candidates – designed to fit within the substrate envelope – retain high potency against a panel of drug resistant HIV strains relative to darunavir (black circles), the most potent drug currently available to patients.

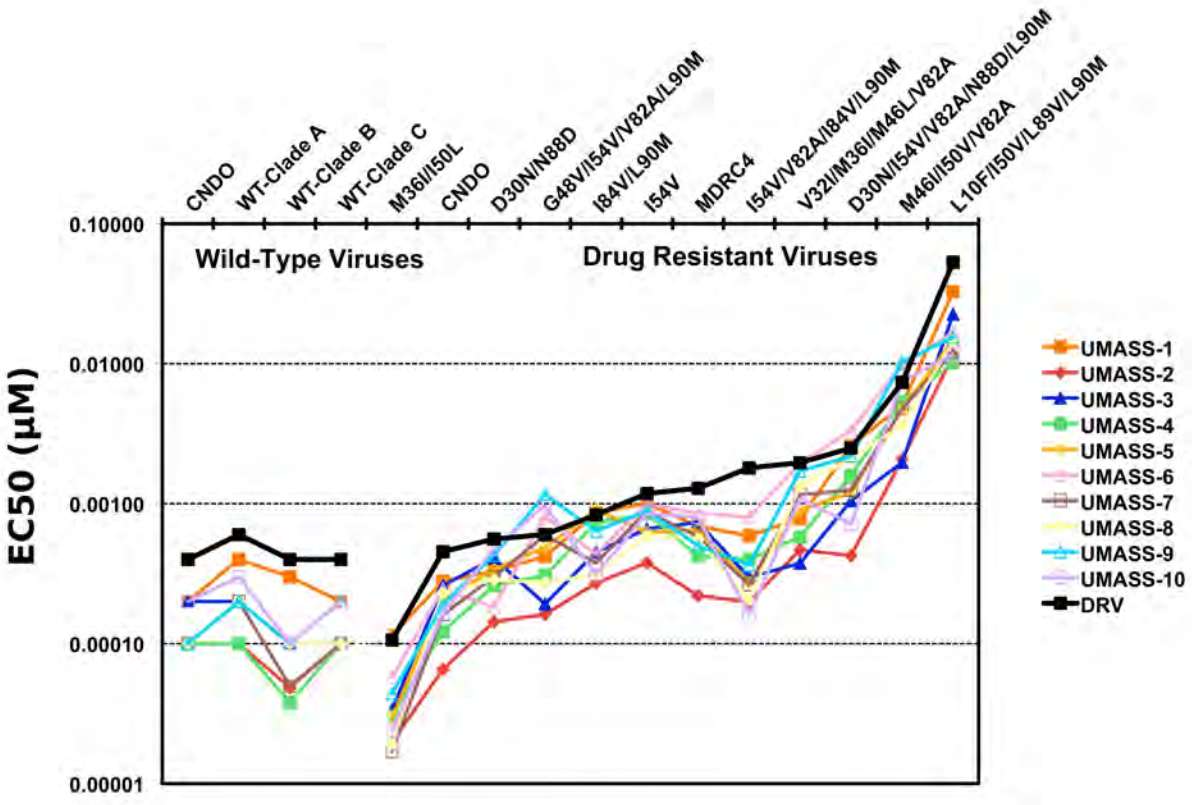

CNDO
WT-Clade A
WT-Clade B
WT-Clade C
M36I/I50L
CNDO
D30N/N88D
G48V/I54V/V82A/L90M
I84V/L90M
I54V
MDRC4
I54V/V82A/I84V/L90M
V32I/M36I/M46L/V82A
D30N/I54V/V82A/N88D/L90M
M46I/I50V/V82A
L10F/I50V/L89V/L90M
0.10000
0.01000
0.00100
0.00010
0.00001
EC50 (µM)
Wild-Type Viruses
Drug Resistant Viruses
UMASS-1
UMASS-2
UMASS-3
UMASS-4
UMASS-5
UMASS-6
UMASS-7
UMASS-8
UMASS-9
UMASS-10
DRV

Figure 1.5

SCOPE OF THESIS

I must first address gaps in the current body of HCV literature before pursuing substrate enveloped-based drug design approaches. At the conception of this research, no crystal structures were available of the NS3/4A protease in complex with authentic viral substrates or post-cleavage products. In addition, despite several drugs in clinical trials, only the crystal structure of boceprevir with genotype 1b protease was available. My research addresses these gaps by determining crystal structures of NS3/4A protease bound to authentic substrate products and leading protease inhibitors. Together these studies elucidate the molecular underpinnings of drug resistance, and may someday provide a more direct path for developing novel drugs that retain activity against a broader spectrum of HCV strains.

REFERENCES

1. Choo, Q. L., Kuo, G., Weiner, A. J., Overby, L. R., Bradley, D. W., and Houghton, M. (1989) Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome, *Science 244*, 359-362.
2. WHO. (2010) Vaccine Research: Hepatitis C, In *Hepatitis C Virus: Disease Burden* (Barnes, E., Ed.).
3. Hoofnagle, J. H. (2002) Course and outcome of hepatitis C, *Hepatology 36*, S21-29.
4. Sharma, S. D. (2010) Hepatitis C virus: molecular biology & current therapeutic options, *Indian J Med Res 131*, 17-34.
5. Seeff, L. B. (2002) Natural history of chronic hepatitis C, *Hepatology 36*, S35-46.
6. Strader, D. B., Wright, T., Thomas, D. L., and Seeff, L. B. (2004) Diagnosis, management, and treatment of hepatitis C, *Hepatology 39*, 1147-1171.
7. Hagan, H., Campbell, J., Thiede, H., Strathdee, S., Ouellet, L., Kapadia, F., Hudson, S., and Garfein, R. S. (2006) Self-reported hepatitis C virus antibody status and risk behavior in young injectors, *Public Health Rep 121*, 710-719.
8. Thein, H. H., Yi, Q., Dore, G. J., and Krahn, M. D. (2008) Estimation of stage-specific fibrosis progression rates in chronic hepatitis C virus infection: a meta-analysis and meta-regression, *Hepatology 48*, 418-431.
9. Kim, W. R., Brown, R. S., Jr., Terrault, N. A., and El-Serag, H. (2002) Burden of liver disease in the United States: summary of a workshop, *Hepatology 36*, 227-242.
10. Davis, G. L., Albright, J. E., Cook, S. F., and Rosenberg, D. M. (2003) Projecting future complications of chronic hepatitis C in the United States, *Liver Transpl 9*, 331-338.
11. Herrmann, E., Neumann, A. U., Schmidt, J. M., and Zeuzem, S. (2000) Hepatitis C virus kinetics, *Antivir Ther 5*, 85-90.
12. Neumann, A. U., Lam, N. P., Dahari, H., Gretch, D. R., Wiley, T. E., Layden, T. J., and Perelson, A. S. (1998) Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy, *Science 282*, 103-107.
13. Mondelli, M. U., and Silini, E. (1999) Clinical significance of hepatitis C virus genotypes, *J Hepatol 31 Suppl 1*, 65-70.
14. Butt, A. A., Wagener, M., Shakil, A. O., and Ahmad, J. (2005) Reasons for non-treatment of hepatitis C in veterans in care, *J Viral Hepat 12*, 81-85.
15. Manns, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M., and Albrecht, J. K. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial, *Lancet 358*, 958-965.
16. Fried, M. W., Shiffman, M. L., Reddy, K. R., Smith, C., Marinos, G., Goncales, F. L., Jr., Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J., and Yu, J. (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection, *N Engl J Med 347*, 975-982.
17. Zeuzem, S., Buti, M., Ferenci, P., Sperl, J., Horsmans, Y., Cianciara, J., Ibranyi, E., Weiland, O., Noviello, S., Brass, C., and Albrecht, J. (2006) Efficacy of 24 weeks treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C infected with genotype 1 and low pretreatment viremia, *J Hepatol 44*, 97-103.
18. Scotto, G., Fazio, V., Palumbo, E., Cibelli, D. C., Saracino, A., and Angarano, G. (2005) Treatment of genotype 1b HCV-related chronic hepatitis: efficacy and toxicity of three different interferon alfa-2b/ribavirin combined regimens in naive patients, *New Microbiol 28*, 23-29.

19. Hadziyannis, S. J., Sette, H., Jr., Morgan, T. R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H., Jr., Bernstein, D., Rizzetto, M., Zeuzem, S., Pockros, P. J., Lin, A., and Ackrill, A. M. (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose, *Ann Intern Med 140*, 346-355.
20. Scarselli, E., Ansuini, H., Cerino, R., Roccasecca, R. M., Acali, S., Filocamo, G., Traboni, C., Nicosia, A., Cortese, R., and Vitelli, A. (2002) The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus, *EMBO J 21*, 5017-5025.
21. Pileri, P., Uematsu, Y., Campagnoli, S., Galli, G., Falugi, F., Petracca, R., Weiner, A. J., Houghton, M., Rosa, D., Grandi, G., and Abrignani, S. (1998) Binding of hepatitis C virus to CD81, *Science 282*, 938-941.
22. Owen, D. M., Huang, H., Ye, J., and Gale, M., Jr. (2009) Apolipoprotein E on hepatitis C virion facilitates infection through interaction with low-density lipoprotein receptor, *Virology 394*, 99-108.
23. Barth, H., Schafer, C., Adah, M. I., Zhang, F., Linhardt, R. J., Toyoda, H., Kinoshita-Toyoda, A., Toida, T., Van Kuppevelt, T. H., Depla, E., Von Weizsacker, F., Blum, H. E., and Baumert, T. F. (2003) Cellular binding of hepatitis C virus envelope glycoprotein E2 requires cell surface heparan sulfate, *J Biol Chem 278*, 41003-41012.
24. Bartosch, B., Vitelli, A., Granier, C., Goujon, C., Dubuisson, J., Pascale, S., Scarselli, E., Cortese, R., Nicosia, A., and Cosset, F. L. (2003) 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 278*, 41624-41630.
25. Blanchard, E., Belouzard, S., Goueslain, L., Wakita, T., Dubuisson, J., Wychowski, C., and Rouille, Y. (2006) Hepatitis C virus entry depends on clathrin-mediated endocytosis, *J Virol 80*, 6964-6972.
26. Tscherne, D. M., Jones, C. T., Evans, M. J., Lindenbach, B. D., McKeating, J. A., and Rice, C. M. (2006) Time- and temperature-dependent activation of hepatitis C virus for low-pH-triggered entry, *J Virol 80*, 1734-1741.
27. Major, M. E., and Feinstone, S. M. (1997) The molecular virology of hepatitis C, *Hepatology 25*, 1527-1538.
28. Moradpour, D., Penin, F., and Rice, C. M. (2007) Replication of hepatitis C virus, *Nat Rev Microbiol 5*, 453-463.
29. Egger, D., Wolk, B., Gosert, R., Bianchi, L., Blum, H. E., Moradpour, D., and Bienz, K. (2002) Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex, *J Virol 76*, 5974-5984.
30. Chang, M., Williams, O., Mittler, J., Quintanilla, A., Carithers, R. L., Jr., Perkins, J., Corey, L., and Gretch, D. R. (2003) Dynamics of hepatitis C virus replication in human liver, *Am J Pathol 163*, 433-444.
31. Quinkert, D., Bartenschlager, R., and Lohmann, V. (2005) Quantitative analysis of the hepatitis C virus replication complex, *J Virol 79*, 13594-13605.
32. Targett-Adams, P., Boulant, S., and McLauchlan, J. (2008) Visualization of double-stranded RNA in cells supporting hepatitis C virus RNA replication, *J Virol 82*, 2182-2195.
33. Kolykhalov, A. A., Mihalik, K., Feinstone, S. M., and Rice, C. M. (2000) Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo, *J Virol 74*, 2046-2051.

34. Lam, A. M., and Frick, D. N. (2006) Hepatitis C virus subgenomic replicon requires an active NS3 RNA helicase, *J Virol 80*, 404-411.
35. Reed, K. E., and Rice, C. M. (2000) Overview of hepatitis C virus genome structure, polyprotein processing, and protein properties, *Curr Top Microbiol Immunol 242*, 55-84.
36. Gorbalenya, A. E., Donchenko, A. P., Koonin, E. V., and Blinov, V. M. (1989) N-terminal domains of putative helicases of flavi- and pestiviruses may be serine proteases, *Nucleic Acids Res 17*, 3889-3897.
37. Lam, A. M., Keeney, D., Eckert, P. Q., and Frick, D. N. (2003) Hepatitis C virus NS3 ATPases/helicases from different genotypes exhibit variations in enzymatic properties, *J Virol 77*, 3950-3961.
38. Frick, D. N., Rypma, R. S., Lam, A. M., and Gu, B. (2004) The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex RNA efficiently, *J Biol Chem 279*, 1269-1280.
39. Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkuhler, C., and De Francesco, R. (1998) Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus, *J Virol 72*, 6758-6769.
40. Beran, R. K., and Pyle, A. M. (2008) Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase, *J Biol Chem 283*, 29929-29937.
41. Beran, R. K., Serebrov, V., and Pyle, A. M. (2007) The serine protease domain of hepatitis C viral NS3 activates RNA helicase activity by promoting the binding of RNA substrate, *J Biol Chem 282*, 34913-34920.
42. Frick, D. N., Frenz, C. M., and Lam, A. M. I. (2006) RNA unwinding by the hepatitis C virus NS3 inhibits its ability to cleave the viral polyprotein, *FASEB J. 20*, A43-c-.
43. Taremi, S. S., Beyer, B., Maher, M., Yao, N., Prosise, W., Weber, P. C., and Malcolm, B. A. (1998) Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease, *Protein Sci 7*, 2143-2149.
44. Schechter, I., and Berger, A. (1967) On the size of the active site in proteases. I. Papain, *Biochem Biophys Res Commun 27*, 157-162.
45. Foy, E., Li, K., Wang, C., Sumpter, R., Jr., Ikeda, M., Lemon, S. M., and Gale, M., Jr. (2003) Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease, *Science 300*, 1145-1148.
46. Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr., and Lemon, S. M. (2005) 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 102*, 2992-2997.
47. Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y. M., Gale, M., Jr., Akira, S., Yonehara, S., Kato, A., and Fujita, T. (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity, *J Immunol 175*, 2851-2858.
48. Alexopoulou, L., Holt, A. C., Medzhitov, R., and Flavell, R. A. (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3, *Nature 413*, 732-738.
49. Foy, E., Li, K., Sumpter, R., Jr., Loo, Y. M., Johnson, C. L., Wang, C., Fish, P. M., Yoneyama, M., Fujita, T., Lemon, S. M., and Gale, M., Jr. (2005) Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling, *Proc Natl Acad Sci U S A 102*, 2986-2991.

50. Ferreon, J. C., Ferreon, A. C., Li, K., and Lemon, S. M. (2005) Molecular determinants of TRIF proteolysis mediated by the hepatitis C virus NS3/4A protease, *J Biol Chem 280*, 20483-20492.
51. Love, R. A., Parge, H. E., Wickersham, J. A., Hostomsky, Z., Habuka, N., Moomaw, E. W., Adachi, T., and Hostomska, Z. (1996) The crystal structure of hepatitis C virus NS3 proteinase reveals a trypsin-like fold and a structural zinc binding site, *Cell 87*, 331-342.
52. Llinas-Brunet, M., Bailey, M., Fazal, G., Goulet, S., Halmos, T., Laplante, S., Maurice, R., Poirier, M., Poupart, M. A., Thibeault, D., Wernic, D., and Lamarre, D. (1998) Peptide-based inhibitors of the hepatitis C virus serine protease, *Bioorg Med Chem Lett 8*, 1713-1718.
53. Steinkuhler, C., Biasiol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A., and De Francesco, R. (1998) Product inhibition of the hepatitis C virus NS3 protease, *Biochemistry 37*, 8899-8905.
54. Arasappan, A., Njoroge, F. G., Chen, K. X., Venkatraman, S., Parekh, T. N., Gu, H., Pichardo, J., Butkiewicz, N., Prongay, A., Madison, V., and Girijavallabhan, V. (2006) P2-P4 macrocyclic inhibitors of hepatitis C virus NS3-4A serine protease, *Bioorg Med Chem Lett 16*, 3960-3965.
55. Bogen, S., Arasappan, A., Pan, W., Ruan, S., Padilla, A., Saksena, A. K., Girijavallabhan, V., and Njoroge, F. G. (2008) Hepatitis C virus NS3-4A serine protease inhibitors: SAR of new P1 derivatives of SCH 503034, *Bioorg Med Chem Lett 18*, 4219-4223.
56. Malancona, S., Colarusso, S., Ontoria, J. M., Marchetti, A., Poma, M., Stansfield, I., Laufer, R., Di Marco, A., Taliani, M., Verdirame, M., Gonzalez-Paz, O., Matassa, V. G., and Narjes, F. (2004) SAR and pharmacokinetic studies on phenethylamide inhibitors of the hepatitis C virus NS3/NS4A serine protease, *Bioorg Med Chem Lett 14*, 4575-4579.
57. Nilsson, M., Belfrage, A. K., Lindstrom, S., Wahling, H., Lindquist, C., Ayesa, S., Kahnberg, P., Pelcman, M., Benkestock, K., Agback, T., Vrang, L., Terelius, Y., Wikstrom, K., Hamelink, E., Rydergard, C., Edlund, M., Eneroth, A., Raboisson, P., Lin, T. I., de Kock, H., Wigerinck, P., Simmen, K., Samuelsson, B., and Rosenquist, S. (2010) Synthesis and SAR of potent inhibitors of the Hepatitis C virus NS3/4A protease: exploration of P2 quinazoline substituents, *Bioorg Med Chem Lett 20*, 4004-4011.
58. Venkatraman, S., Velazquez, F., Wu, W., Blackman, M., Chen, K. X., Bogen, S., Nair, L., Tong, X., Chase, R., Hart, A., Agrawal, S., Pichardo, J., Prongay, A., Cheng, K. C., Girijavallabhan, V., Piwinski, J., Shih, N. Y., and Njoroge, F. G. (2009) Discovery and structure-activity relationship of P1-P3 ketoamide derived macrocyclic inhibitors of hepatitis C virus NS3 protease, *J Med Chem 52*, 336-346.
59. Raboisson, P., de Kock, H., Rosenquist, A., Nilsson, M., Salvador-Oden, L., Lin, T. I., Roue, N., Ivanov, V., Wahling, H., Wickstrom, K., Hamelink, E., Edlund, M., Vrang, L., Vendeville, S., Van de Vreken, W., McGowan, D., Tahri, A., Hu, L., Boutton, C., Lenz, O., Delouvroy, F., Pille, G., Surleraux, D., Wigerinck, P., Samuelsson, B., and Simmen, K. (2008) 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 18*, 4853-4858.
60. Perni, R. B., Pitlik, J., Britt, S. D., Court, J. J., Courtney, L. F., Deininger, D. D., Farmer, L. J., Gates, C. A., Harbeson, S. L., Levin, R. B., Lin, C., Lin, K., Moon, Y. C., Luong, Y. P., O'Malley, E. T., Rao, B. G., Thomson, J. A., Tung, R. D., Van Drie, J. H., and Wei, Y. (2004) Inhibitors of hepatitis C virus NS3.4A protease 2. Warhead SAR and optimization, *Bioorg Med Chem Lett 14*, 1441-1446.

61. Tsantrizos, Y. S., Bolger, G., Bonneau, P., Cameron, D. R., Goudreau, N., Kukolj, G., LaPlante, S. R., Llinas-Brunet, M., Nar, H., and Lamarre, D. (2003) Macrocyclic inhibitors of the NS3 protease as potential therapeutic agents of hepatitis C virus infection, *Angew Chem Int Ed Engl 42*, 1356-1360.
62. LaPlante, S. R., Cameron, D. R., Aubry, N., Lefebvre, S., Kukolj, G., Maurice, R., Thibeault, D., Lamarre, D., and Llinas-Brunet, M. (1999) Solution structure of substrate-based ligands when bound to hepatitis C virus NS3 protease domain, *J Biol Chem 274*, 18618-18624.
63. Tsantrizos, Y. S. (2008) Peptidomimetic therapeutic agents targeting the protease enzyme of the human immunodeficiency virus and hepatitis C virus, *Acc Chem Res 41*, 1252-1263.
64. Lamarre, D., Anderson, P. C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bos, M., Cameron, D. R., Cartier, M., Cordingley, M. G., Faucher, A. M., Goudreau, N., Kawai, S. H., Kukolj, G., Lagace, L., LaPlante, S. R., Narjes, H., Poupart, M. A., Rancourt, J., Sentjens, R. E., St George, R., Simoneau, B., Steinmann, G., Thibeault, D., Tsantrizos, Y. S., Weldon, S. M., Yong, C. L., and Llinas-Brunet, M. (2003) An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus, *Nature 426*, 186-189.
65. Hinrichsen, H., Benhamou, Y., Wedemeyer, H., Reiser, M., Sentjens, R. E., Calleja, J. L., Forns, X., Erhardt, A., Cronlein, J., Chaves, R. L., Yong, C. L., Nehmiz, G., and Steinmann, G. G. (2004) Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients, *Gastroenterology 127*, 1347-1355.
66. Vanwolleghem, T., Meuleman, P., Libbrecht, L., Roskams, T., De Vos, R., and Leroux-Roels, G. (2007) Ultra-rapid cardiotoxicity of the hepatitis C virus protease inhibitor BILN 2061 in the urokinase-type plasminogen activator mouse, *Gastroenterology 133*, 1144-1155.
67. Pockros, P. J., Nelson, D., Godofsky, E., Rodriguez-Torres, M., Everson, G. T., Fried, M. W., Ghalib, R., Harrison, S., Nyberg, L., Shiffman, M. L., Najera, I., Chan, A., and Hill, G. (2008) R1626 plus peginterferon Alfa-2a provides potent suppression of hepatitis C virus RNA and significant antiviral synergy in combination with ribavirin, *Hepatology 48*, 385-397.
68. Roberts, S. K., Cooksley, G., Dore, G. J., Robson, R., Shaw, D., Berns, H., Hill, G., Klumpp, K., Najera, I., and Washington, C. (2008) Robust antiviral activity of R1626, a novel nucleoside analog: a randomized, placebo-controlled study in patients with chronic hepatitis C, *Hepatology 48*, 398-406.
69. Vermehren, J., and Sarrazin, C. (2011) New HCV therapies on the horizon, *Clin Microbiol Infect 17*, 122-134.
70. Lange, C. M., Sarrazin, C., and Zeuzem, S. (2010) Review article: specifically targeted anti-viral therapy for hepatitis C - a new era in therapy, *Alimentary pharmacology & therapeutics 32*, 14-28.
71. Lindenbach, B. D., Evans, M. J., Syder, A. J., Wolk, B., Tellinghuisen, T. L., Liu, C. C., Maruyama, T., Hynes, R. O., Burton, D. R., McKeating, J. A., and Rice, C. M. (2005) Complete replication of hepatitis C virus in cell culture, *Science 309*, 623-626.
72. Zhong, J., Gastaminza, P., Cheng, G., Kapadia, S., Kato, T., Burton, D. R., Wieland, S. F., Uprichard, S. L., Wakita, T., and Chisari, F. V. (2005) Robust hepatitis C virus infection in vitro, *Proc Natl Acad Sci U S A 102*, 9294-9299.
73. Rai, R., and Deval, J. (2011) New opportunities in anti-hepatitis C virus drug discovery: Targeting NS4B, *Antiviral Res 90*, 93-101.

74. Khaliq, S., Jahan, S., and Hassan, S. (2011) Hepatitis C virus p7: molecular function and importance in hepatitis C virus life cycle and potential antiviral target, *Liver Int 31*, 606-617.
75. Foster, T. L., Verow, M., Wozniak, A. L., Bentham, M. J., Thompson, J., Atkins, E., Weinman, S. A., Fishwick, C., Foster, R., Harris, M., and Griffin, S. (2011) Resistance mutations define specific antiviral effects for inhibitors of the hepatitis C virus (HCV) p7 ion channel, *Hepatology*.
76. Flisiak, R., Feinman, S. V., Jablkowski, M., Horban, A., Kryczka, W., Pawlowska, M., Heathcote, J. E., Mazzella, G., Vandelli, C., Nicolas-Metral, V., Grosgurin, P., Liz, J. S., Scalfaro, P., Porchet, H., and Crabbe, R. (2009) The cyclophilin inhibitor Debio 025 combined with PEG IFNalpha2a significantly reduces viral load in treatment-naive hepatitis C patients, *Hepatology 49*, 1460-1468.
77. Qureshi, S. A. (2007) Hepatitis C virus--biology, host evasion strategies, and promising new therapies on the horizon, *Med Res Rev 27*, 353-373.
78. Cabot, B., Martell, M., Esteban, J. I., Piron, M., Otero, T., Esteban, R., Guardia, J., and Gomez, J. (2001) Longitudinal evaluation of the structure of replicating and circulating hepatitis C virus quasispecies in nonprogressive chronic hepatitis C patients, *J Virol 75*, 12005-12013.
79. Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J., and Gomez, J. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution, *J Virol 66*, 3225-3229.
80. Farci, P., Strazzera, R., Alter, H. J., Farci, S., Degioannis, D., Coiana, A., Peddis, G., Usai, F., Serra, G., Chessa, L., Diaz, G., Balestrieri, A., and Purcell, R. H. (2002) Early changes in hepatitis C viral quasispecies during interferon therapy predict the therapeutic outcome, *Proc Natl Acad Sci U S A 99*, 3081-3086.
81. Paolucci, S., Baldanti, F., Campanini, G., Zavattoni, M., Cattaneo, E., Dossena, L., and Gerna, G. (2001) Analysis of HIV drug-resistant quasispecies in plasma, peripheral blood mononuclear cells and viral isolates from treatment-naive and HAART patients, *J Med Virol 65*, 207-217.
82. Fried, M. W. (2002) Side effects of therapy of hepatitis C and their management, *Hepatology 36*, S237-244.
83. Liu, S. S., Schneekloth, T. D., Talwalkar, J. A., Kim, W. R., Poterucha, J. J., Charlton, M. R., Wiesner, R. H., and Gross, J. B. (2010) Impact of depressive symptoms and their treatment on completing antiviral treatment in patients with chronic hepatitis C, *J Clin Gastroenterol 44*, e178-185.
84. Butt, A. A., McGinnis, K. A., Skanderson, M., and Justice, A. C. (2009) Hepatitis C treatment completion rates in routine clinical care, *Liver Int 30*, 240-250.
85. Lee, S. S., Bain, V. G., Peltekian, K., Krajden, M., Yoshida, E. M., Deschenes, M., Heathcote, J., Bailey, R. J., Simonyi, S., and Sherman, M. (2006) Treating chronic hepatitis C with pegylated interferon alfa-2a (40 KD) and ribavirin in clinical practice, *Aliment Pharmacol Ther 23*, 397-408.
86. Dollarhide, A. W., Loh, C., Leckband, S. G., Endow-Eyer, R., Robinson, S., and Meyer, J. M. (2007) Psychiatric comorbidity does not predict interferon treatment completion rates in hepatitis C seropositive veterans, *J Clin Gastroenterol 41*, 322-328.
87. Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A. M., Young, S. K., Lee, B., Heckerman, D., Carlson, J., Reyor, L. L., Kleyman, M., McMahon, C. M., Birch, C., Schulze Zur Wiesch, J., Ledlie, T., Koehrsen, M., Kodira, C., Roberts, A. D., Lauer, G.

M., Rosen, H. R., Bihl, F., Cerny, A., Spengler, U., Liu, Z., Kim, A. Y., Xing, Y., Schneidewind, A., Madey, M. A., Fleckenstein, J. F., Park, V. M., Galagan, J. E., Nusbaum, C., Walker, B. D., Lake-Bakaar, G. V., Daar, E. S., Jacobson, I. M., Gomperts, E. D., Edlin, B. R., Donfield, S. M., Chung, R. T., Talal, A. H., Marion, T., Birren, B. W., Henn, M. R., and Allen, T. M. (2008) Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients, *Hepatology 48*, 1769-1778.

88. Tong, X., Bogen, S., Chase, R., Girijavallabhan, V., Guo, Z., Njoroge, F. G., Prongay, A., Saksena, A., Skelton, A., Xia, E., and Ralston, R. (2008) Characterization of resistance mutations against HCV ketoamide protease inhibitors, *Antiviral Res 77*, 177-185.
89. He, Y., King, M. S., Kempf, D. J., Lu, L., Lim, H. B., Krishnan, P., Kati, W., Middleton, T., and Molla, A. (2008) Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system, *Antimicrob Agents Chemother 52*, 1101-1110.
90. Sarrazin, C., Rouzier, R., Wagner, F., Forestier, N., Larrey, D., Gupta, S. K., Hussain, M., Shah, A., Cutler, D., Zhang, J., and Zeuzem, S. (2007) SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders, *Gastroenterology 132*, 1270-1278.
91. Kieffer, T. L., Sarrazin, C., Miller, J. S., Welker, M. W., Forestier, N., Reesink, H. W., Kwong, A. D., and Zeuzem, S. (2007) Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients, *Hepatology 46*, 631-639.
92. Yi, M., Ma, Y., Yates, J., and Lemon, S. M. (2009) Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation, *PLoS Pathog 5*, e1000403.
93. Lenz, O., Verbinnen, T., Lin, T. I., Vijgen, L., Cummings, M. D., Lindberg, J., Berke, J. M., Dehertogh, P., Fransen, E., Scholliers, A., Vermeiren, K., Ivens, T., Raboisson, P., Edlund, M., Storm, S., Vrang, L., de Kock, H., Fanning, G. C., and Simmen, K. A. (2010) In vitro resistance profile of the HCV NS3/4A protease inhibitor TMC435, *Antimicrob Agents Chemother*.
94. Seiwert. (2006) Sequence variation of NS3 and NS4A in hepatitis C virus (HCV) replicons following exposure to ITMN-191 concentrations likely to encompass those achieved in human liver following clinical dosing., In *First International Workshop in Hepatitis C Resistance and New Compounds*, Boston, MA.
95. King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease, *Chem Biol 11*, 1333-1338.
96. Prabu-Jeyabalan, M., Nalivaika, E., and Schiffer, C. A. (2002) Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes, *Structure 10*, 369-381.
97. Prabu-Jeyabalan, M., Nalivaika, E. A., King, N. M., and Schiffer, C. A. (2003) Viability of a drug-resistant human immunodeficiency virus type 1 protease variant: structural insights for better antiviral therapy, *J Virol 77*, 1306-1315.
98. Ali, A., Reddy, G. S., Cao, H., Anjum, S. G., Nalam, M. N., Schiffer, C. A., and Rana, T. M. (2006) Discovery of HIV-1 protease inhibitors with picomolar affinities incorporating N-aryl-oxazolidinone-5-carboxamides as novel P2 ligands, *J Med Chem 49*, 7342-7356.
99. Chellappan, S., Kiran Kumar Reddy, G. S., Ali, A., Nalam, M. N., Anjum, S. G., Cao, H., Kairys, V., Fernandes, M. X., Altman, M. D., Tidor, B., Rana, T. M., Schiffer, C. A., and

Gilson, M. K. (2007) Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis, *Chem Biol Drug Des 69*, 298-313.

100. Reddy, G. S., Ali, A., Nalam, M. N., Anjum, S. G., Cao, H., Nathans, R. S., Schiffer, C. A., and Rana, T. M. (2007) Design and synthesis of HIV-1 protease inhibitors incorporating oxazolidinones as P2/P2' ligands in pseudosymmetric dipeptide isosteres, *J Med Chem 50*, 4316-4328.
101. Altman, M. D., Ali, A., Reddy, G. S., Nalam, M. N., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., Gilson, M. K., Schiffer, C. A., Rana, T. M., and Tidor, B. (2008) HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants, *J Am Chem Soc 130*, 6099-6113.

CHAPTER II

MOLECULAR MECHANISMS OF VIRAL AND HOST CELL SUBSTRATE RECOGNITION BY HCV NS3/4A PROTEASE

Author Contributions

This work was performed in collaboration – I performed the crystallization and structural analyses; Jennifer Laine and Francesca Massi performed the NMR experiments and processed the data; Laura Deveau designed and cloned all protease mutants; Hong Cao and I performed the fluorescence polarization experiments; and Celia Schiffer and I wrote the paper.

ABSTRACT

Hepatitis C NS3/4A protease is a prime therapeutic target responsible for cleaving the viral polyprotein at junctions 3-4A, 4A4B, 4B5A and 5A5B, and two host cell adapter proteins of the innate immune response, TRIF and MAVS. In this study, NS3/4A crystal structures of both host cell cleavage sites are determined and compared to the crystal structures of viral substrates. Two distinct protease conformations are observed and correlate with substrate specificity: *(i)* 3-4A, 4A4B, 5A5B and MAVS, which are processed more efficiently by the protease, form extensive electrostatic networks when in complex with the protease, and *(ii)* TRIF and 4B5A – containing polyproline motifs in their full-length sequences – do not form electrostatic networks in their crystal complexes. These findings provide mechanistic insights into NS3/4A substrate recognition, which may assist in a more rational approach to inhibitor design in the face of the rapid acquisition of resistance.

INTRODUCTION

Hepatitis C virus (HCV) is a genetically diverse *hepacivirus* of the *Flaviviridae* family infecting over 180 million people worldwide (*1*). The positive-sense, single-stranded RNA genome encodes a single polyprotein, which is translated along the endoplasmic reticulum. Host cell and viral proteases cleave the viral polyprotein into structural (C, E1, E2) and non-structural components (p7, NS2, NS3, NS4A, NS4B, NS5A, NS5B) (*2*). The NS3/4A protein, a bifunctional protease/helicase enzyme formed by the non-covalent association of NS3 and NS4A, hydrolyzes four known sites along the viral polyprotein, thereby liberating non-structural proteins essential for viral replication. Previous kinetic data suggest that the first cleavage at the 3-4A junction occurs *in cis,* while processing of the remaining junctions occurs *in trans* (*3, 4*). Interestingly, heterodimerization with NS4A is essential for the cleavage of junction 4B5A. These viral substrates share little sequence similarity, except for an acid at P6, cysteine or threonine at P1 and serine or alanine at P1' **(Table 2.1)**. We have shown that the diverse set of NS3/4A substrate sequences are recognized in a conserved three-dimensional shape, defining a consensus van der Waals (vdW) volume, or substrate envelope (*5*). This conserved mode of substrate recognition regulates polyprotein processing and thus the biology of HCV replication.

In addition to its essential role in processing the viral polyprotein, NS3/4A protease also confounds the innate immune response to viral infection by disrupting activation of the transcription factors interferon regulatory factor 3 (IRF-3) and nuclear factor κB (NF-κB) (*6, 7*). The detection of viral RNA induces these transcription factors

Table 2.1. Genotype 1a primary cleavage sequences of NS3/4A substrates.

Substrate	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'
TRIF	$P_{(8)}$	S	S	T	P	C	S	A	H	L
MAVS	E	R	E	V	P	C	H	R	P	S
3-4A	D	L	E	V	V	T	S	T	W	V
4A4B	D	E	M	E	E	C	S	Q	H	L
4B5A	E	C	T	T	P	C	S	G	S	W
5A5B	E	D	V	V	C	C	S	M	S	Y

through two distinct pathways, involving signaling through Toll-like receptor 3 (TLR3) or retinoic acid-inducible gene I (RIG-I) (*8, 9*). NS3/4A protease disrupts the TLR3 and RIG-I cascades by cleaving the essential adaptor proteins Toll-interleukin-1 receptor domain-containing adaptor-inducing interferon-β (TRIF) and mitochondrial antiviral signaling protein (MAVS), respectively (*7, 10*). The TRIF and MAVS cleavage sites share little sequence homology with other viral substrates: TRIF contains cysteine at P1 and serine at P1', while MAVS contains glutamate at P6 and cysteine at P1 **(Table 2.1)**. Notably, in place of an acid at position P6, TRIF consists of a tract of eight proline residues spanning P13–P6, which has been previously implicated as part of the substrate recognition motif for NS3/4A (*11*). The NS3-mediated processing of both viral and host cell targets is central to the interplay between HCV replication and the innate immune response, thus highlighting the importance of better elucidating the mechanisms of substrate recognition.

Despite great efforts devoted to the development of NS3/4A protease inhibitors, the rapid rise of drug resistance in human clinical trials has limited the efficacy of the most promising drug candidates. Drug resistance mutations in the protease emerge as molecular changes that prevent the binding of drugs, but still permitting the recognition and cleavage of substrates. A more detailed understanding of the molecular details underlying substrate recognition is therefore critical for explaining patterns of drug resistance and for designing novel drugs that are less susceptible to resistance. Here we analyze crystal structures of NS3/4A protease in complex with N-terminal products of viral substrates 3-4A, 4A4B, 4B5A and 5A5B. TRIF and MAVS crystal complexes

further reveal that these host cell products bind to the protease active site in a conserved three-dimensional manner similar to that of the viral products. Notably, extensive electrostatic networks involving protease residues D81, R155, D168 and R123 form in product complexes 3-4A, 4A4B, 5A5B and MAVS while these networks are absent in the 4B5A and TRIF product complexes. Short peptides corresponding to the immediate cleavage sequences of TRIF and 4B5A have significantly weaker affinities for NS3/4, which correlate with their inability to form of such electrostatic networks with the protease. Taken together, our findings support previous biochemical studies implicating the role of polyproline II helices in TRIF cleavage by NS3/4A (*11*), and provide a structural basis for future studies aimed at better elucidating the detailed mechanism NS3/4A substrate recognition and cleavage.

MATERIALS AND METHODS

Mutagenesis and gene information

The HCV genotype 1a NS3/4A protease gene described in a Bristol-Meyers Squibb patent (*12*) was synthesized by GenScript and cloned into the pET28a expression vector (Novagen). The gene encodes a highly soluble form of the NS3/4A protease domain as a single-chain, with 11 core amino acids of NS4A located at the N-terminus. The inactive S139A protease variant was subsequently constructed using the QuikChange site-directed mutagenesis kit from Stratagene and sequenced by Davis Sequencing for confirmation.

Viral substrate and peptide product purchase and storage

Thirty milligrams of each substrate peptide and the corresponding N-terminal cleavage product (TRIF and MAVS) were purchased from 21st Century Biochemicals (Marlboro, MA). The TRIF and MAVS peptides were synthesized as a 13mer (P13–P1) and 7mer (P7–P1), respectively. The C-termini of all substrate peptides were blocked with amide groups. All peptides were stored as solids at −20ºC and dissolved in DMF to a final concentration of 50–100mM for crystallization trials.

Expression and purification of NS3/4A protease constructs

NS3/4A expression and purification were carried out as described previously (*12, 13*). Briefly, transformed BL21(DE3) *E. coli* cells were grown at 37ºC and induced at an optical density of 0.6 by the addition of 1mM IPTG. Cells were harvested after 5 hours of expression, pelleted and frozen at −80ºC for storage. Cell pellets were thawed, resuspended in 5mL/g of resuspension buffer (50mM phosphate buffer at pH 7.5, 500mM

NaCl, 10% glycerol, 2mM β-ME) and lysed with a cell disruptor. The soluble fraction was retained, applied to a nickel column (Qiagen), washed with resuspension buffer, and eluted with resuspension buffer supplemented with 200mM imidazole. The eluant was dialyzed overnight (MWCO 10kD) to remove the imidazole and the his-tag was simultaneously removed with thrombin treatment. The nickel-purified protein was then flash-frozen and stored at −80ºC for up to six months.

Crystallization of product complexes and apo enzyme

For crystallization, the protein solution was thawed, concentrated to ~3mg/mL and loaded on a HiLoad Superdex75 16/60 column equilibrated with gel filtration buffer (25mM MES at pH 6.5, 500mM NaCl, 10% glycerol, 30μM zinc chloride, and 2mM DTT). The protease fractions were pooled and concentrated to 20–25mg/mL with an Amicon Ultra-15 10kD device (Millipore). The concentrated samples were either used for crystallization of apo structure or incubated for one hour with 2–20 molar excess of substrate product TRIF or MAVS. Diffraction-quality crystals were obtained overnight by mixing equal volume of concentrated protein solution with precipitant solution (20–26% PEG-3350, 0.1M sodium MES buffer at pH 6.5, and 4% ammonium sulfate) in 24-well VDX hanging drop trays.

Data collection and structure solution

Crystals large enough for data collection were flash-frozen in liquid nitrogen for storage. The TRIF, MAVS and apo crystals were mounted under constant cryostream and X-ray diffraction data were collected at Advanced Photon Source LS-CAT 21-ID-F, BioCARS 14-BMC and our in-house RAXIS IV X-ray system, respectively. Diffraction

intensities of product complexes were indexed, integrated and scaled using the program HKL2000 (*14*). All structure solutions were generated using simple isomorphous molecular replacement with PHASER (*15*). The B chain model of viral substrate product 4A4B (3M5M) (*5*) was used as the starting model for all structure solutions. Initial refinement was carried out in the absence of modeled ligand, which was subsequently built in during later stages of refinement. Upon obtaining the correct molecular replacement solutions, the phases were improved by building solvent molecules using ARP/wARP (*16*). Subsequent crystallographic refinement was carried out within the CCP4 program suite with iterative rounds of TLS and restrained refinement until convergence was achieved (*17*). The final structures were evaluated with MolProbity (*18*) prior to deposition in the protein data bank. 5% of the data was reserved for the free R-value calculation to limit the possibility of model bias throughout the refinement process (*19*). Interactive model building and electron density viewing was carried out using the program COOT (*20*).

Double-difference plots and global analysis

Double-difference plots were computed as described previously (*21*). Briefly, the atomic distances were calculated between each Cα of a given protease molecule and every other Cα in the same molecule. The differences of these Cα-Cα distances were then calculated between each pair of protease molecules and plotted as a contour graph for visualization. These analyses allowed for effective structural comparisons without the biases associated with superimpositions and space group differences. Double-difference plots were used to determine the structurally invariant regions of the protease, consisting

of residues 32–36, 42–47, 50–54, 84–86 and 140–143. Structural superimpositions were carried out in PyMOL (*22*) using the Cα atoms of these residues for all protease molecules. The apo structure was used as the reference structure for the alignments of the TRIF and MAVS product complexes. The Cα RMSD for each residue was subsequently calculated to access the degree of structural variation throughout the protein. The B-factor column of a representative structure was replaced with these values and used to generate the rainbow color spectrum to visualize these variations.

The viral substrate envelope

All active site alignments were performed with PyMOL using the Cα atoms of protease residues 137–139 and 154–160. For each alignment, the B chain of complex 4A4B was used as the reference structure. The viral substrate envelope – representing the consensus vdW volume shared by any three of the four viral substrate products – was computed as described previously using the full-length NS3/4A structure (1CU1) (*23*) and product complexes 4A4B (3M5M), 4B5A (3M5N) and 5A5B (3M5O) (*5*).

NMR spectroscopy and data processing

NMR data were recorded using 650μL of 390μM [U-^{15}N] NS3/4A protease (95% H_2O/5% D_2O, 25mM sodium phosphate at pH 7.2, 150mM KCl, 5μM zinc chloride, and 1mM TCEP). Backbone ^{1}H and ^{15}N resonance assignments of NS3/4A protease were kindly provided by Herbert Klei of Bristol-Myers Squibb, and confirmed using nuclear Overhauser enhancement spectroscopy (NOESY) experiments. Backbone ^{1}H and ^{15}N resonance assignments of NS3/4A protease bound to the N-terminal cleavage product of substrate 4A4B were obtained from the assignment of the free protein by following the

chemical shift changes upon titration of the ligand. All experiments were performed at 298K using a Varian Inova spectrometer operating at 600 MHz (14.1 T). Spectra were processed using nmrPipe (*24*) and Sparky (*25*).

Binding of the unlabeled peptide corresponding to the N-terminal cleavage product of substrate 4A4B to [U-^{15}N] NS3/4A protease was monitored using a series of two-dimensional ^{15}N-^{1}H HSQC spectra collected at increasing concentrations of the peptide to a final concentration of 2.5mM. The change of the cross-peak positions for NS3/4A residues was recorded as a function of titrated peptide concentration. The normalized change in the chemical shift was calculated for each protease residue using the following equation:

$$\Delta = \sqrt{\Delta\delta_H^2 + \left(\Delta\delta_N \gamma_N / \gamma_H\right)^2}$$

where $\Delta\delta$ is the chemical shift change observed between the free and bound states, γ is the gyromagnetic ratio, and the subscript indicates either ^{1}H or ^{15}N nuclei. The b-factors for each residue of the apo enzyme crystal structure were then replaced by the maximal chemical shift changes from the titration data. The crystal structure was colored in PyMOL (*22*) according to the chemical shift magnitudes to graphically depict the locations of the shifting residues.

VdW contact energy

VdW contact energies between protease residues and peptide products were computed using a simplified Lennard-Jones potential as described previously (*26*). Briefly, the Lennard-Jones potential (V_r) was calculated for each protease-product atom

pair where r, ε and σ represent the interatomic distance, vdW well depth and atomic diameter, respectively:

$$V_r = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

V_r was computed for all possible protease-product atom pairs within 5Å, and potentials for non-bonded pairs separated by less than the distance at the minimum potential were equated to −ε. Using this simplified potential value for each non-bonded protease-product atom pair, the total vdW contact energy (ΣV_r) was computed for each peptide residue. For graphical convenience, VdW energy indexes were then calculated by multiplying the raw values by a factor of –10.

Fluorescence polarization

For fluorescence polarization experiments, the NS3/4A protease domain was purified as described previously in purification buffer (25mM HEPES, pH 7.5, 150mM NaCl, 20% glycerol, 4mM DTT) and subsequently concentrated to 200–400μM. The concentrated stocks were then two-thirds serially diluted in 384-well plates (Corning) in reaction buffer using the Genesis Workstation (Tecan). An equal volume of substrate buffer (25mM HEPES, pH 7.5, 0mM NaCl, 20% glycerol, 4mM DTT) containing 10nM of fluorescein-tagged substrate product (4A4B, 4B5A or 5A5B) was added to each well to make a final well volume of 60μL. The final condition constituted 5nM fluorescein-tagged cleavage products in 25mM HEPES (pH 7.5), 75mM NaCl, 20% glycerol and 4mM DTT. The plates were incubated at room temperature for two hours and five fluorescence polarization measurements were taken for each well using the Victor-3 plate

reader (Perkins Elmer). Five sets of binding data were collected for each substrate product and each trial was processed independently. The average and standard deviations were then calculated from the results of these five trials. Fluorescence polarization data (in milli-polarization units, *mP*) were fit to the Langmuir Equation, where E_T is the total NS3/4A concentration, K_d is the equilibrium binding constant, *b* is the baseline fluorescence polarization and *m* is the fluorescence polarization maximum:

$$mP = b + (m - b)\left[\frac{E_T}{E_T + K_d}\right]$$

RESULTS

Structure determination of apo NS3/4A and product complexes

The apo NS3/4A protease domain and host cell product complexes TRIF and MAVS all crystallized in the space group $P2_12_12_1$ with one molecule in the asymmetric unit **(Table 2.2)**. For all structures, we utilized the highly soluble NS3/4A protease domain described previously, containing the essential residues of cofactor NS4A covalently linked at the N-terminus (*12*). This NS3/4A construct also contains the inactivating mutation S139A, designed to further enhance protein stability by minimizing autoproteolysis during the crystallization process. The partially inactivated variant still exhibits residual proteolytic activity, as observed for other serine proteases (*27, 28*), likely facilitated by the nucleophilic attack of water. Thus the complete characterization of full-length substrate peptides was not possible and short peptides corresponding to the N-terminal cleavage products of authentic substrates were used for all crystallization trials. Peptide products 4A4B, 4B5A, 5A5B and MAVS spanned P7–P1, while the TRIF sequence spanned P13–P1 with a tract of eight proline residues at the N-terminus **(Table 2.1)**. The entire peptide sequence spanning P7–P1 could be modeled in each structure except for the TRIF complex, which revealed electron density for the residues spanning P6–P1 but not the polyproline tract.

Tertiary structure analysis

Structural analyses of the NS3/4A apo enzyme were carried out in conjunction with: *(i)* the host cell product complexes TRIF and MAVS, *(ii)* the viral product complexes 4A4B, 4B5A and 5A5B (*5*), and *(iii)* the full-length NS3/4A structure (*23*), in

Table 2.2. X-ray data collection and crystallographic refinement statistics.

Dataset	**TRIF**	**MAVS**	**APO**
PDB ID	3RC4	3RC5	3RC6
Modeled Ligand	PSSTPC	QEREVPC	*(unliganded)*
Resolution (Å)	1.50 (1.50–1.55)	1.60 (1.60–1.66)	1.30 (1.30–1.35)
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Molecules in AU[†]	1	1	1
Cell dimensions:			
a (Å)=	53.9	54.1	54.8
b (Å)=	58.1	58.2	58.6
c (Å)=	61.3	61.3	60.9
Completeness (%)	87.4 (93.1)	99.3 (99.1)	93.1 (97.1)
Total reflections	136929	97504	190364
Unique reflection	27770	26041	45595
Average I/σ	13.5 (4.4)	18.8 (5.4)	14.8 (4.1)
Redundancy	4.9 (4.7)	3.7 (3.7)	4.2 (4.0)
R_{sym} (%)[‡]	4.8 (30.1)	3.3 (19.3)	4.1 (29.3)
RMSD[∂] **in:**			
Bonds (Å)	0.009	0.009	0.009
Angles (°)	1.28	1.26	1.29
R_{factor} (%)[§]	18.7	17.3	16.2
R_{free} (%)[§]	21.5	19.3	19.1

[†] AU, asymmetric unit.

[‡] $R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I = observed intensity, <I> = average intensity over symmetry equivalent.

[∂] RMSD, root mean square deviation.

[§] $R_{work} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$. R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

( ) Denotes statistics for the highest resolution shell.

which the C-terminus represents the post-cleavage product 3-4A. The seven structures constitute a total of twelve NS3/4A protease monomers, which all adopt the same chymotrypsin-like tertiary fold defined by the labeling-scheme for trypsin (*29*). The N-terminal distorted β-barrel subdomain contains two α-helices (α0 and α1) and seven β-strands (A0–F1), while the C-terminal β-barrel subdomain comprises of two α-helices (α2 and α3) and six β-strands (A2–F2). The cofactor NS4A contributes a single β-strand to the N-terminal distorted β-barrel, which is essential for efficient catalytic function (*4*). The catalytic triad is located in the cleft between these subdomains, with the N-terminal β-barrel contributing residues H57 and D81 and the C-terminal β-barrel contributing the nucleophilic S139. The active site residues in all crystal structures share similar architecture, defined by the catalytic triad residues (S139A, H57, D81) and backbone nitrogens of the oxyanion hole (G137, S138, S139A).

Certain global differences are observed between these structures when superpositions are performed. Double-difference plots were therefore generated between each co-complex and the apo enzyme to determine the most invariant regions **(Figure 2.1)**. Product 3-4A varied most extensively from the apo state, which likely reflects differences in genotype, protein size and crystal packing of the full-length construct. The remaining product complexes derive from the same NS3/4A protease domain construct, and in general, vary less extensively from the apo enzyme. Notably, the host cell product complexes are most similar to the apo enzyme, while viral product complexes 4A4B, 4B5A and 5A5B vary more extensively. Taken together, these findings suggest that

Figure 2.1: Double-difference plots of product complexes. Double-difference plots were computed for viral and host-cell product complexes relative to the apo enzyme structure. NS3 protease residues are numbered according to the conventional system, whereas negative numbers indicate residues of the cofactor NS4A. Red and blue contour lines represent positive differences of 1.0Å and 0.5Å, respectively, while black and green lines represent negative differences of the same magnitudes.

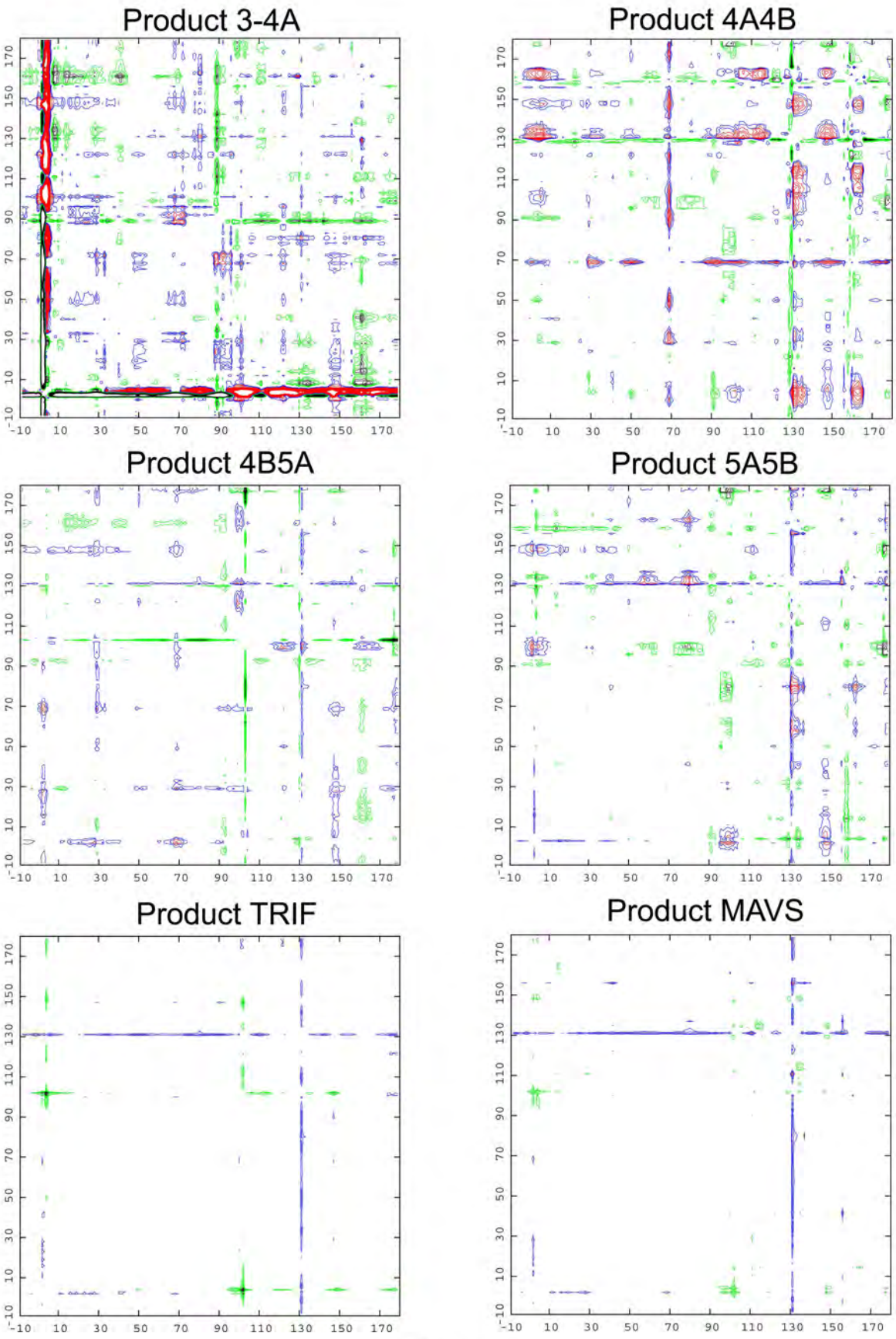

Product 3-4A
Product 4A4B
Product 4B5A
Product 5A5B
Product TRIF
Product MAVS
-10 10 30 50 70 90 110 130 150 170

Figure 2.1

structural differences likely reflect the inherent flexibility in certain regions of the protease.

The set of double difference plots were further analyzed and the most invariant regions of the protease were determined to contain residues 32–36, 42–47, 50–54, 84–86, and 140–143 **(Figure 2.1)**. Product complexes 4A4B, 4B5A, 5A5B, TRIF and MAVS consist of ten protease molecules, which were subsequently superposed onto the apo enzyme using the Cα atoms of the structurally invariant residues. The Cα average root mean square (RMS) deviation was calculated for each residue and the seven most variable regions of the protease were determined to be **(Figure 2.2A)**: *(i)* the linker connecting cofactor 4A at the N-terminus, *(ii)* the loop containing residues 65–70, *(iii)* the zinc-binding site containing residues 95–105, *(iv)* the 3_{10} helix region spanning residues 128–136, *(v)* the zinc-binding site containing residues 145–148, *(vi)* the active site anti-parallel β-sheet containing residues 156–168, and *(vii)* the C-terminal α3 helix. These regions are solvent-exposed and likely influenced by both crystal packing effects and inherent flexibility.

Extensive structural differences are observed near the active site as indicated by large RMS deviations for residues 156–168. These differences are most pronounced for the β-strands E2 and F2, which form the anti-parallel β-sheet constituting the majority of the active site. This region is least variable near the catalytic triad, while the average RMS deviations increase significantly toward the loop connecting these β-strands **(Figure 2.2B)**. Though the architecture of the protease catalytic triad is conserved, these observations suggest a potential dynamic interaction between the anti-parallel β-sheet of

Figure 2.2: Average RMS deviations of product complexes. All protease molecules from product complexes 4A4B, 4B5A, 5A5B, TRIF and MAVS were superposed onto the apo enzyme structure using the most invariant core residues 32–36, 42–47, 50–54, 84–86 and 140–143. (A) The average Cα RMS deviations were plotted versus residue number and (B) mapped onto a representative protease molecule with the most variable regions depicted in red and the most invariant regions depicted in blue. The seven most variable regions of the protease are labeled in both panels.

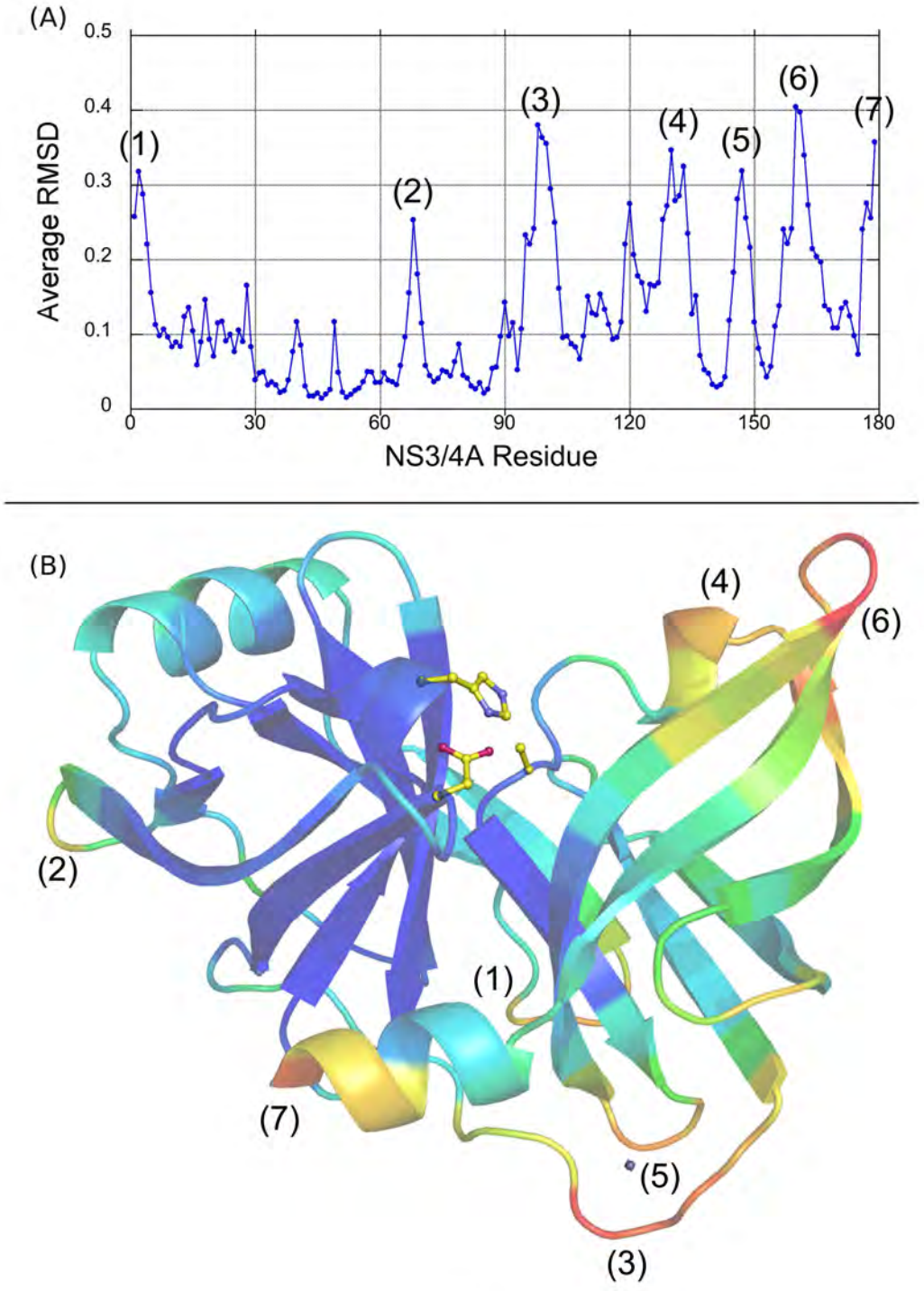

(A)
Average RMSD
0.5
0.4
0.3
0.2
0.1
0
0
30
60
90
120
150
180
NS3/4A Residue
(1)
(2)
(3)
(4)
(5)
(6)
(7)
(B)
(1)
(2)
(3)
(4)
(5)
(6)
(7)

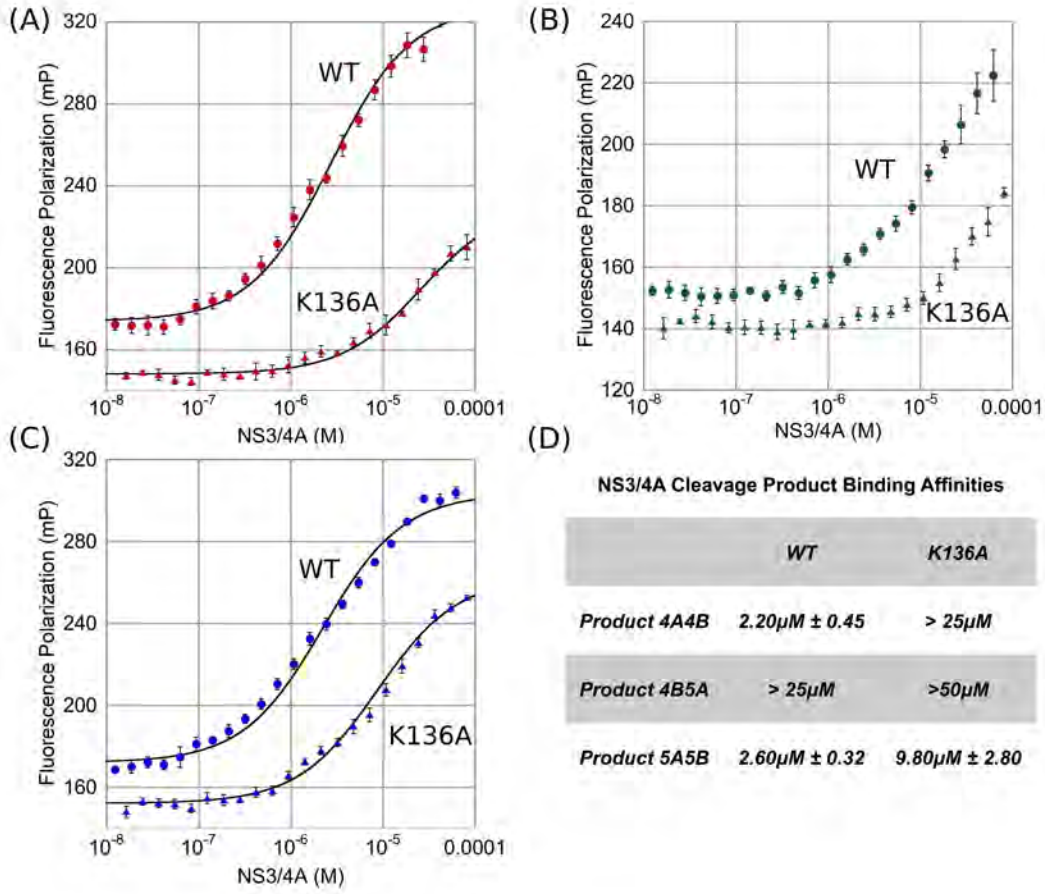
Figure 2.2

the protease active site and substrate products. Further studies are necessary to probe the nature and extent of such dynamic interplay. Nevertheless, though the Cα atoms of active site residues shift relative to the protease core, these residues superpose well onto the APO enzyme with a RMS deviation range of 0.3–0.5Å. Moreover, the residue side chains adopt similar rotamer conformations and interact with the same surrounding residues, suggesting that potential flexibility in the protease active site would not disrupt its tertiary structure.

Analysis of viral product binding

Fluorescence polarization experiments reveal that the viral products bind with different affinities to the protease, with the K_d for 4B5A over 10-fold weaker than those values for 4A4B and 5A5B **(Figure 2.3)**. Crystal structures reveal that these viral products bind in a conserved manner, forming an anti-parallel β-sheet with protease residues 154–160 and burying 500–600$Å^2$ of solvent accessible surface area (*30*). The peptide product backbone torsion angles are very similar, with positions P1–P4 being the most similar and residues P5–P7 deviating progressively toward the C-terminus. A constrained P2 ϕ torsion angle of about –60° is observed in product complexes 3-4A, 4A4B and 5A5B. Interestingly, these P2 residues could sterically tolerate the substitution of proline, which is found at the P2 position in substrates 4B5A, TRIF and MAVS. The ability of the P2 residue to adopt this constrained backbone torsion is a likely determinant in recognition process, allowing for the proper positioning of the P1 cysteine for catalysis in the active site.

Figure 2.3: Fluorescence polarization of viral product binding. Fluorescence polarization binding experiments were conducted with inactive S139A protease (circles) and the S139A/K136A variant (triangles) with fluorescein-tagged peptide products (A) 4A4B, (B) 4B5A and (C) 5A5B. For all conditions, trace amounts of peptide were incubated for two hours with increasing concentrations of NS3/4A protease in 25mM HEPES (pH 7.2), 75mM potassium chloride, 20% glycerol, 0.5μM zinc chloride and 4mM DTT. The panels depict a single representative trial, though each experiment was repeated five times. (D) Data for each trial were processed independently by least-squares regression fitting with the Langmuir Equation and the average and standard deviation of equilibrium dissociation constants (K_d) were calculated from the five repeats.



NS3/4A Cleavage Product Binding Affinities

	WT	*K136A*
Product 4A4B	***2.20μM ± 0.45***	***> 25μM***
Product 4B5A	***> 25μM***	***>50μM***
Product 5A5B	***2.60μM ± 0.32***	***9.80μM ± 2.80***

Figure 2.3

There are many conserved features in viral product binding, involving both backbone and side chain interactions **(Figure 2.4)**. For example, eight hydrogen bonds between backbone amide and carbonyl groups are completely conserved in the product complexes, involving protease residues G137, S138, S139A, R155, A157 and S159. S159 (C159 in product complex 3-4A) and A157 each contribute two hydrogen bonds with bound products at positions P5 and P3, respectively. The P1 residue, cysteine in all substrates but 3-4A, interacts favorably with the π-system of electrons of F154. All P1 terminal carboxyl groups sit in the oxyanion hole, hydrogen bonding with the Nε nitrogen of H57 and the amide nitrogens of residues 137–139. Though the coordinates of the P1 terminal oxygen atom are not included in the full-length NS3/4A structure, geometric restraints would position it similarly to the other peptide products. Thus the same set of protease residues contacts all peptide sequences, although the precise nature of these interactions varies depending on the particular residue involved in each contact.

Despite these similarities, there are also unique features that likely underlie the particular specificity of NS3/4A for each substrate **(Figure 2.4)**. For example, the four acidic residues in product 4A4B lead to a highly charged peptide in solution. In the bound state, however, the atomic geometry suggests that the P5 glutamate is protonated and hydrogen bonding with the carboxyl group of the P3 glutamate, which itself forms an ionic interaction with the terminal nitrogen of K136. In fact, K136 interacts differently with all four viral products, forming: *(i)* a hydrogen bond with the P2 carbonyl oxygen of 3-4A, *(ii)* a salt bridge with the P3 glutamate of 4A4B, and *(iii)* an extended conformation that does not interact considerably in product complexes 4B5A and 5A5B.

Figure 2.4: Stereo view of viral product binding to NS3/4A protease. N-terminal cleavage products (A) 3-4A, (B) 4A4B, (C) 4B5A, and (D) 5A5B are shown bound to the protease active site with the electrostatic interactions of backbone and sidechain atoms depicted in black and red, respectively.

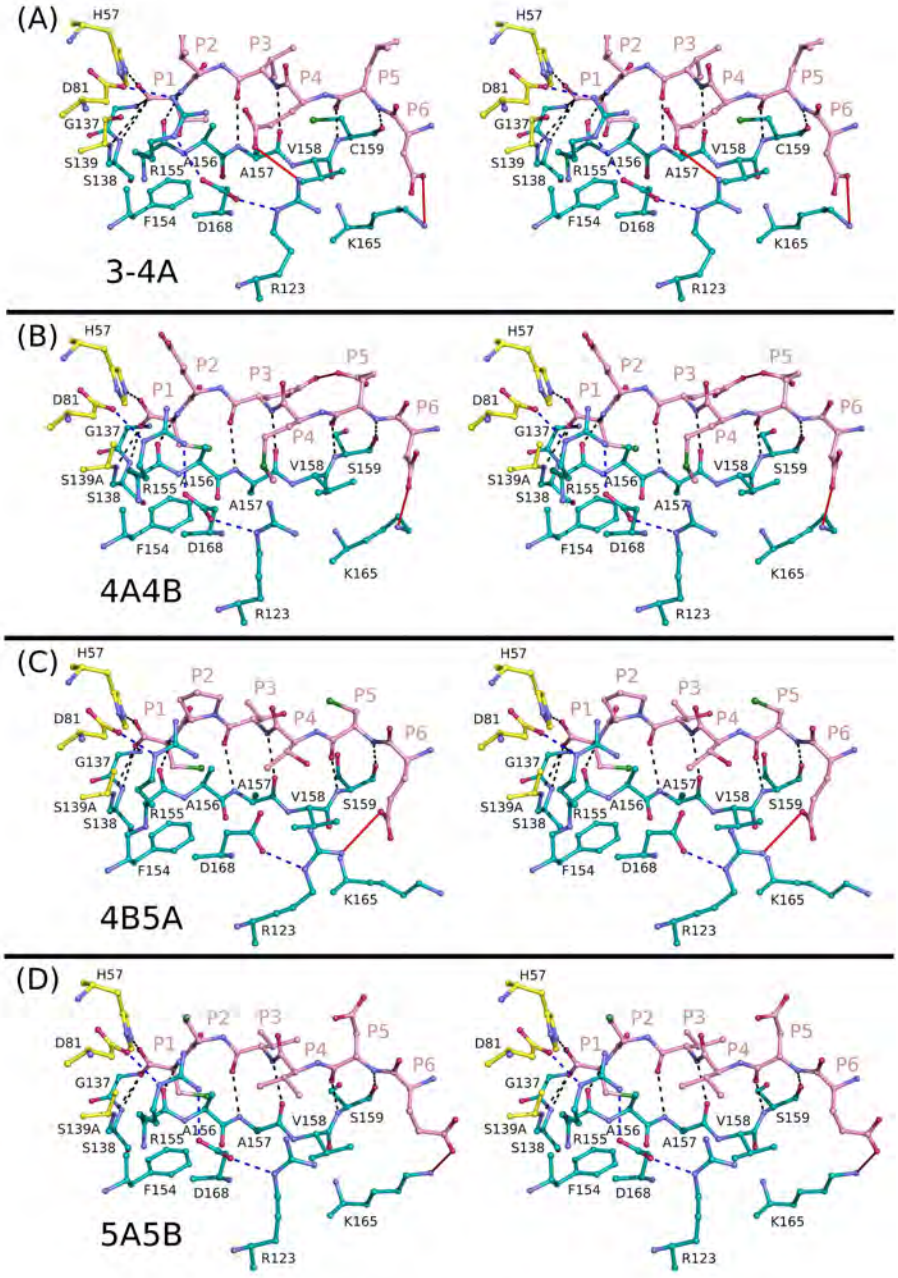

(A)
3-4A
(B)
4A4B
(C)
4B5A
(D)
5A5B
H57
D81
G137
S139
S138
S139A
R155
A156
A157
V158
C159
S159
F154
D168
K165
R123
P1
P2
P3
P4
P5
P6

Figure 2.4

Fluorescence polarization data confirms a more significant loss in binding affinity of product 4A4B for the K136A protease variant compared to products 4B5A and 5A5B **(Figure 2.3)**. Thus the affinity of a substrate product likely arises from the side chain interactions unique to that particular product.

NMR solution studies of product 4A4B binding

The NS3/4A protease active site is located on the surface of the protein and thus highly solvent-exposed. The analysis of viral and host cell product binding is therefore complicated by the proximity of symmetry-related molecules within the crystal lattice. To investigate the possibility of crystal packing effects confounding structural observations, we carried out NMR HSQC titration experiments using peptide product 4A4B spanning P7–P1 **(Figure 2.5A)**. HSQC chemical shift perturbations upon product titration are consistent with the molecular interactions observed in the crystal complex. The normalized chemical shift perturbations for each protease residue were compared to the buried surface area calculated directly from product complex 4A4B **(Figure 2.5B)**. These data indicate that the same set of protease residues with large chemical shift perturbations are also observed to interact extensively with product 4A4B in the crystal structure. The NMR solution studies recapitulate our structural observations and suggest that our crystal structure analyses are indeed representative of the interactions occurring in solution.

Analysis of host cell product complexes TRIF and MAVS

Active site superpositions (residues 137–139; 154–160) reveal that TRIF and MAVS peptide products both bind to the protease active site in a conserved three-

Figure 2.5: NS3/4A H1, N15 NMR HSQC titration data. (A) NMR spectra reveal N15 labeled NS3/4A chemical shift perturbations upon titration of increasing concentrations of substrate 4A4B product peptide. Cross-peaks are colored according to the concentration of peptide product 4A4B, ranging from 0mM (blue) to 2.5mM (red). For clarity, the titration data for the highest peptide concentration (2.5mM) are only depicted in the insets. (B) The apo NS3/4A crystal structure was colored according to the chemical shift changes between the apo enzyme state and 2.5mM product 4A4B. Protease residues undergoing chemical shifts are colored by a spectrum ranging from pink (small shifts) to red (large shifts). All unassigned residues are shown in white.

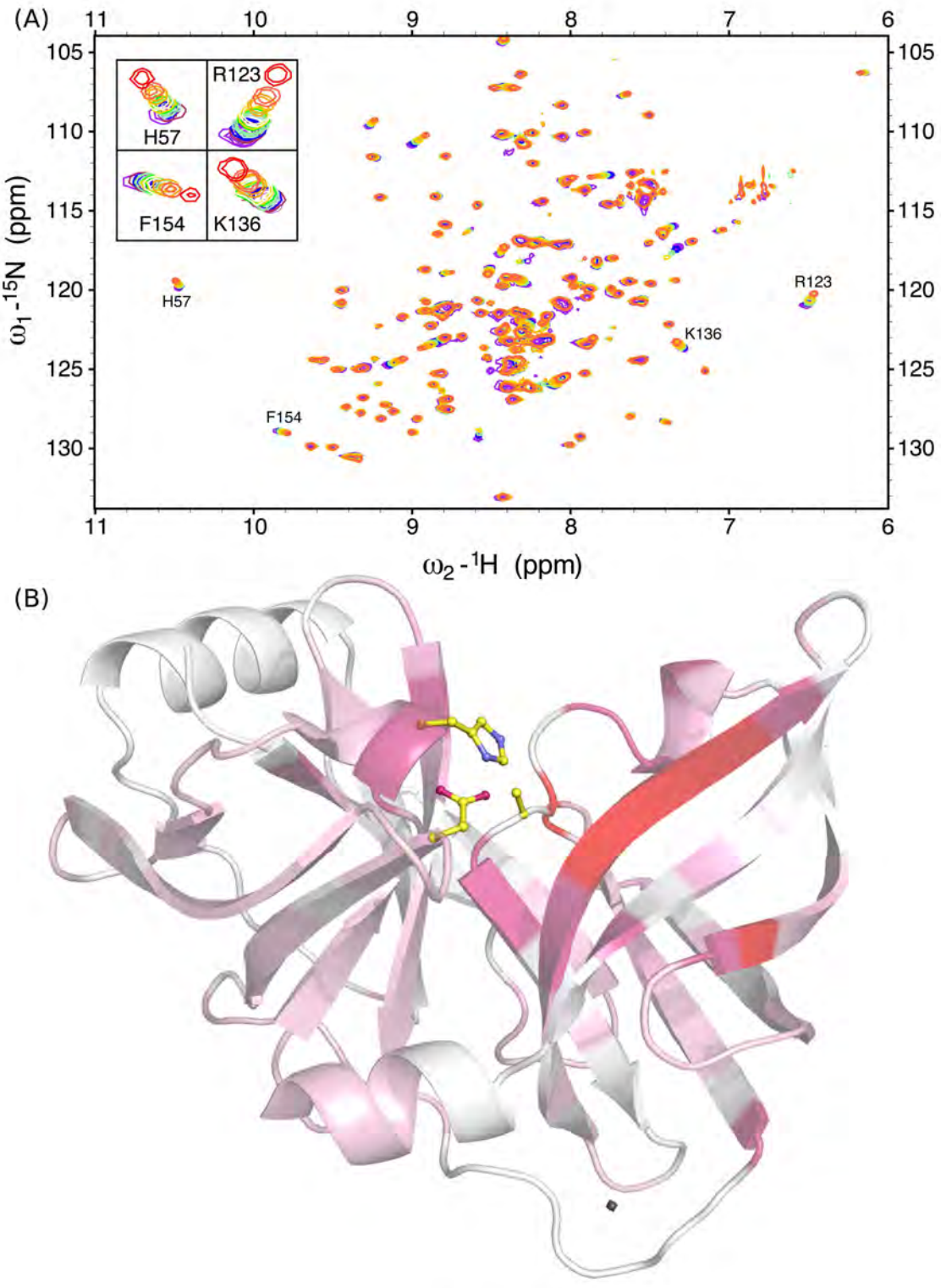

(A)
11
10
9
8
7
6
105
110
115
120
125
130
ω_1 - ^{15}N (ppm)
ω_2 - ^{1}H (ppm)
H57
R123
F154
K136
(B)

Figure 2.5

dimensional shape **(Figure 2.6)**. Both substrate products form anti-parallel β-sheets with protease residues 154–160. There is no clear electron density in the TRIF complex for the proline residues spanning P13–P7. Nevertheless, the residues from P6–P1 overlap closely with the corresponding residues in the MAVS complex. The P1 cysteine residues of both substrate products interact with the aromatic ring of F154. Eight hydrogen bonds are observed in both structures, involving the amide nitrogens or carboxyl oxygens of residues G137, S138, S139A, R155, A157 and S159. A157 and S159 each contribute two hydrogen bonds with the P3 and P5 residues, respectively. In both structures, the carbonyl groups at position P1 interact with the protease oxyanion hole, defined by the backbone amide nitrogens of residues 137–139. Thus, the post-cleavage products of the cellular substrates TRIF and MAVS bind to the protease active site in a conserved manner despite their large variations in primary sequence.

There are also many differences in the binding of TRIF and MAVS involving mainly sidechain interactions with the protease **(Figure 2.6)**. For example, MAVS interacts closely with the protease electrostatic network formed by residues D81, R155, D168 and R123. The P4 glutamate of MAVS interacts with R155 and R123 in this network, while the P6 glutamate forms a salt bridge with R123. The TRIF peptide product, however, lacks such extended residues on this surface of the molecule, and the electrostatic network is notably absent with the participating residues adopting conformations observed in the apo structure. Previous studies demonstrate that a large fraction of full-length TRIF exists as polyproline II helices, and that the interaction of NS3/4A with a polyproline II helix facilitates TRIF cleavage (*11*). The absence of clear

Figure 2.6: Stereo view of host cell product binding to NS3/4A protease. (A) The apo enzyme structure and N-terminal cleavage products (B) TRIF and (C) MAVS are shown bound to the protease active site with the electrostatic interactions of backbone and sidechain atoms depicted in black and red, respectively.

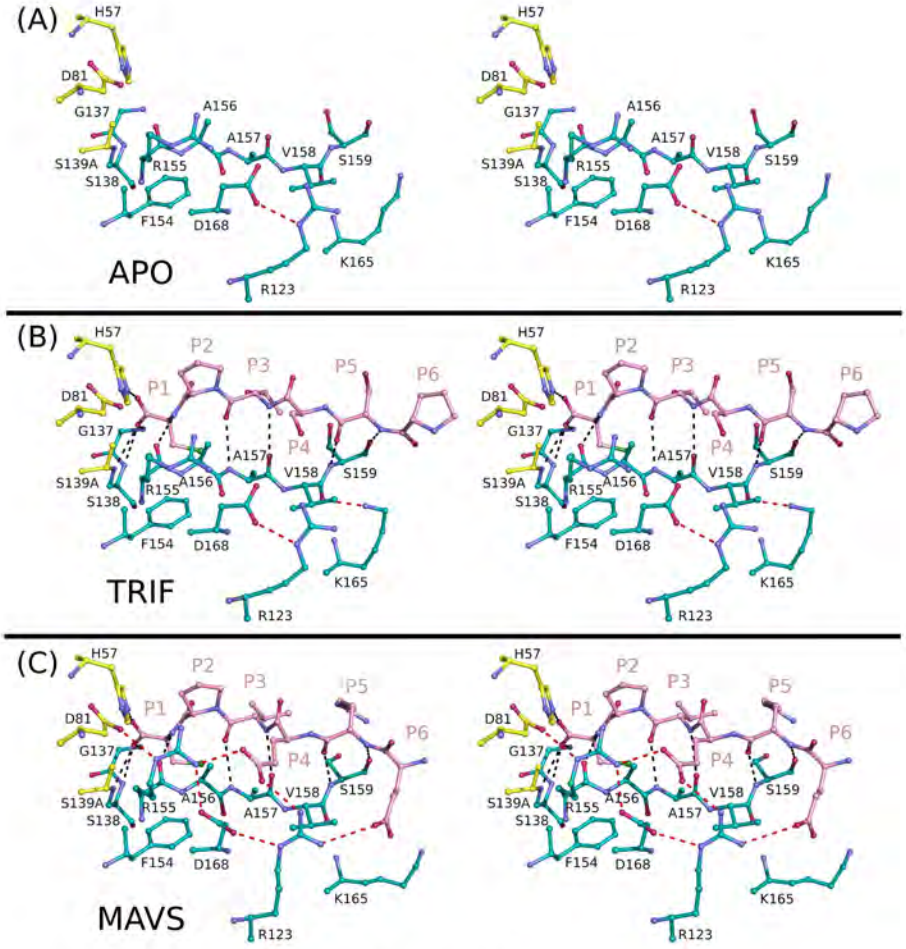

(A)
APO
(B)
TRIF
(C)
MAVS
H57
D81
G137
S139A
S138
R155
A156
A157
V158
S159
F154
D168
K165
R123
P1
P2
P3
P4
P5
P6

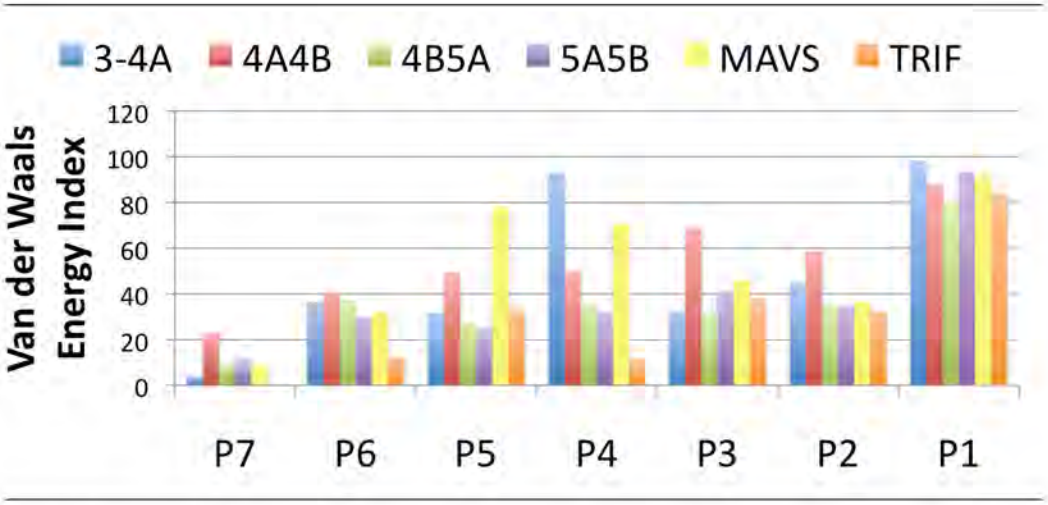
Figure 2.6

electron density for the proline residues suggests that full-length TRIF may be necessary to stabilize the polyproline tract in a conformation capable of specific interaction with NS3/4A.

Comparison of host cell and viral product binding

Host cell product binding was analyzed on a broader basis by comparison with the binding of viral substrates, previously reported by our group (*5*). Both viral (3-4A, 4A4B, 4B5A, 5A5B) and host cell (TRIF, MAVS) substrate products bind to the protease active site in a conserved three-dimensional shape. The peptide backbone torsional angles are very similar, being most conserved at position P1 and deviating slightly toward position P6. All peptide products adopt constrained P2 Ψ torsion angles, even those containing non-proline residues at this site. However, vdW analyses of substrate products indicate large variations in sidechain interactions with the protease. All of the NS3/4A substrates contain either cysteine or threonine at position P1, while five of the six contain an acid at position P6. The P1 and P6 substrate residues each contribute the same amount of vdW energies in all product complexes **(Figure 2.7)**. The amino acid makeup of viral cleavage sequences is much more diverse at positions P5–P2, and in general, the vdW energies at each position correlate with amino acid size. For example, the P4 glutamates in substrates 3-4A and MAVS, and to a lesser extent the P4 methionine in substrate 4A4B, are associated with larger vdW energies relative to the substrates with smaller P4 residues. Likewise, the larger glutamate residues at both P3 and P2 of substrate 4A4B also correlate with greater contact energies compared to the other substrates, which contain smaller amino acids at these positions.

Figure 2.7: vdW energies of substrate product binding. The vdW binding energies by residue for each NS3/4A cleavage products (3-4A, 4A4B, 4B5A, 5A5B, TRIF and MAVS) are graphically depicted with their primary sequences tabulated.



3-4A	**A**	**D**	**L**	**E**	**V**	**V**	**T**
4A4B	**F**	**D**	**E**	**M**	**E**	**E**	**C**
4B5A	**S**	**E**	**C**	**T**	**T**	**P**	**C**
5A5B	**T**	**E**	**D**	**V**	**V**	**C**	**C**
MAVS	**Q**	**E**	**R**	**E**	**V**	**P**	**C**
TRIF	**P**	**P**	**S**	**S**	**T**	**P**	**C**

Figure 2.7

Though similar in shape, protease substrate binding can be further categorized into two groups: *(i)* product complexes 3-4A, 4A4B, 5A5B and MAVS bind with an intact electrostatic network involving residues D81, R155, D168 and R123, while *(ii)* product complexes 4B5A and TRIF bind without this network such that R155, D168 and R123 maintain conformations observed in the apo enzyme. Notably, the 4B5A and TRIF cleavage sites contain fewer charged residues compared to the other substrates, which may underlie their inability to form the electrostatic network. Binding studies demonstrate that both 4B5A and TRIF have relatively weaker affinities for NS3/4A compared to the other substrates (*11*, *31*). However, most biochemical studies have been conducted on small peptides corresponding to the immediate cleavage sequences of TRIF and 4B5A. Indeed, kinetic studies revealed that full-length TRIF is processed more efficiently than peptides corresponding to the cleavage sequence (*11*). Additional molecular interactions by the full-length proteins or by adaptor proteins in the authentic cellular environment may better facilitate substrate binding. Nevertheless, the current structural analyses suggest that NS3/4A substrates vary in specificity by their ability to form and stabilize the protease electrostatic network; the sequence specificity is particularly influenced by amino acid variation at positions P6, P4 and P2.

Analysis using the viral substrate envelope

The binding of NS3/4A cellular substrates was analyzed in terms of the viral substrate envelope **(Figure 2.8)**, which was previously defined as the vdW volume shared by any three of four viral products (*5*). This shape could not be predicted by the primary sequences alone and highlights the conserved mode of viral substrate recognition despite

Figure 2.8: Host-cell product binding and the NS3/4A substrate envelope. (A) The NS3/4A substrate envelope is calculated from the consensus van der Waals volume shared by any three of the four viral products. (B) The TRIF cleavage product is confined within the substrate envelope, (C) while MAVS is mostly located within the substrate envelope except for the sidechain atoms spanning P6–P4.

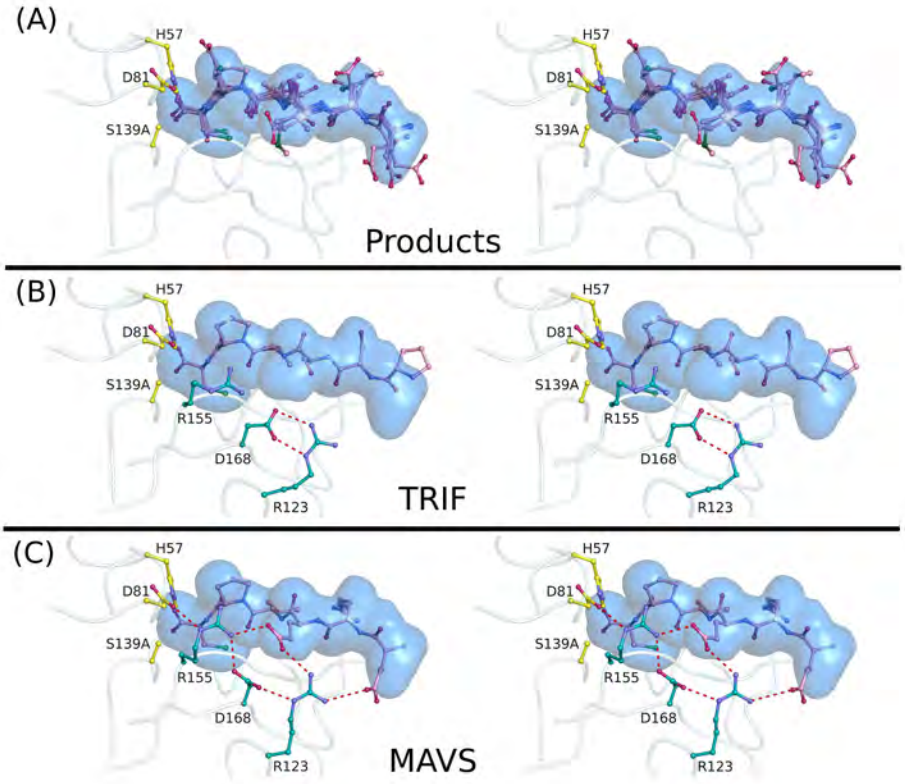

(A)
H57
D81
S139A
Products
(B)
H57
D81
S139A
R155
D168
R123
TRIF
(C)
H57
D81
S139A
R155
D168
R123
MAVS

Figure 2.8

their high sequence diversity. The backbone chains of both TRIF and MAVS fit entirely within the substrate envelope, as well as the sidechains of TRIF spanning P5–P1. The sidechains of MAVS are also mostly confined within the substrate envelope, except for the longer side chains of the P6 glutamate, P5 arginine and P4 glutamate. The carboxylic acids of these glutamate residues interact extensively with the protease electrostatic network, while the P5 arginine packs against loop residues 159–162. As these interactions occur outside the viral substrate envelope, we speculate that mutations that disrupt the electrostatic network, such as R155K and D168A, would preferentially reduce the proteolytic processing of MAVS compared to TRIF.

DISCUSSION

The recognition and proteolysis of the viral polyprotein and host cell adaptor proteins by NS3/4A protease play an integral role in the ability of HCV to replicate and evade the innate immune response to viral infection (*7, 10*). In this study, crystal structures of the NS3/4A protease domain reveal that viral and host cell products bind to the protease active site in a similar three-dimensional shape, defined by the viral substrate envelope reported previously. The MAVS product complex reveals the formation of an extensive electrostatic network involving protease residues D81, R155, D168 and R123, which also form in viral product complexes 3-4A, 4A4B and 5A5B. No such networks form in the TRIF and 4B5A complexes, and residues in this region of the protease adopt the same conformations observed in the apo state. The absence or presence of electrostatic networks also correlate with the affinities of product binding, with the K_d of 4B5A being 10-fold weaker compared to products 4A4B and 5A5B. The greater catalytic efficacies of NS3/4A for substrates 4A4B and 5A5B relative to 4B5A (*31*) may also derive from the formation of electrostatic networks. However, short peptide may only partially mimic how the viral cleavage sequences are processed along the viral polyprotein in the natural cellular environment. Additional molecular features may further modulate the binding of TRIF and 4B5A, perhaps facilitated by the proline-rich regions contained in both proteins (*11*). Thus, the specificity of substrate processing by NS3/4A protease seems to arise from at least two distinct molecular interaction patterns, which likely influence the order and kinetics of polyprotein processing during the HCV lifecycle.

In fact, these structural observations can be further linked to the known biology of NS3/4A processing during viral replication. Previous NS3-mediated cleavage assays of HCV polyprotein substrates revealed that NS4A is essential for the *trans* cleavage of junction 4B5A, but not required for the processing of junctions 4A4B and 5A5B (*3, 4*). Our structural analyses provide further insight into the molecular interactions underlying these previous findings. The NS4A cofactor likely stabilizes the tertiary protease fold required for the binding of NS3/4A substrates, and the binding of substrate 4B5A may absolutely depend on this particular protease conformation. Substrates 4A4B and 5A5B, however, may be able to induce these conformational changes through charge interactions, even in the absence of cofactor NS4A. Thus our findings support the previous published data for NS3 HCV polyprotein processing, and future research is warranted to better ascertain the dynamic mechanisms of substrate recognition.

The ability for HCV to establish chronic human infections is highly dependent on the viruses ability to effectively replicate while simultaneously evading the host cell immune response. The virally encoded NS3/4A protein plays an integral role in this process by mediating the cleavage of essential viral proteins and antiviral host cell adaptors. NS3/4A protease is thus a prime therapeutic target and great efforts have been devoted in the development of protease inhibitors, which have demonstrated efficacy in late phases of human clinical trials. Nevertheless, the high rate and error-prone nature of HCV replication has led to the emergence of drug resistance against the most promising protease inhibitors to date, such as boceprevir, telaprevir and danoprevir (*32-37*). Inhibitor potency often derives from molecular interactions that are not essential for

substrate recognition and cleavage. Mutations in these regions of the protease can selectively prevent drug binding while still allowing for the recognition and cleavage of viral and host cell substrates. Thus identification of the protease residues that are important for substrate binding is crucial and will ultimately facilitate the design of drugs that target these particular residues. A more detailed understanding of the mechanisms underlying viral and host cell substrate recognition is therefore essential in facilitating a more rationale approach to the design of more robust NS3/4A protease inhibitors.

ACKNOWLEDGEMENTS

NMR chemical assignments were kindly provided by Herbert Klei of Bristol-Myers Squibb. We thank David Smith of the LS-CAT beamline at Argonne National Laboratory for data collection of the apo enzyme; we also thank Shivender Shandilya and Vukica Šrajer for data collection of the TRIF complex at BioCARS; Madhavi Nalam and Rajintha Bandaranayake for assistance with structural refinement; and Aysegul Ozen, Seema Mittal and Madhavi Kolli and for their computational support. Use of the Advanced Photon Source was supported by the U. S. Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was supported by the National Institutes of Health, National Center for Research Resources, under grant number RR007707. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817). The NIH grants R01-GM65347 and R01-AI085051 supported this work.

REFERENCES

1. WHO. (2010) Vaccine Research: Hepatitis C, In *Hepatitis C Virus: Disease Burden* (Barnes, E., Ed.).
2. Major, M. E., and Feinstone, S. M. (1997) The molecular virology of hepatitis C, *Hepatology 25*, 1527-1538.
3. Bartenschlager, R., Ahlborn-Laake, L., Mous, J., and Jacobsen, H. (1994) Kinetic and structural analyses of hepatitis C virus polyprotein processing, *J Virol 68*, 5045-5055.
4. Lin, C., Pragai, B. M., Grakoui, A., Xu, J., and Rice, C. M. (1994) Hepatitis C virus NS3 serine proteinase: trans-cleavage requirements and processing kinetics, *J Virol 68*, 8147-8157.
5. Romano, K. P., Ali, A., Royer, W. E., and Schiffer, C. A. (2010) 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*.
6. Foy, E., Li, K., Wang, C., Sumpter, R., Jr., Ikeda, M., Lemon, S. M., and Gale, M., Jr. (2003) Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease, *Science 300*, 1145-1148.
7. Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr., and Lemon, S. M. (2005) 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 102*, 2992-2997.
8. Yoneyama, M., Kikuchi, M., Matsumoto, K., Imaizumi, T., Miyagishi, M., Taira, K., Foy, E., Loo, Y. M., Gale, M., Jr., Akira, S., Yonehara, S., Kato, A., and Fujita, T. (2005) Shared and unique functions of the DExD/H-box helicases RIG-I, MDA5, and LGP2 in antiviral innate immunity, *J Immunol 175*, 2851-2858.
9. Alexopoulou, L., Holt, A. C., Medzhitov, R., and Flavell, R. A. (2001) Recognition of double-stranded RNA and activation of NF-kappaB by Toll-like receptor 3, *Nature 413*, 732-738.
10. Foy, E., Li, K., Sumpter, R., Jr., Loo, Y. M., Johnson, C. L., Wang, C., Fish, P. M., Yoneyama, M., Fujita, T., Lemon, S. M., and Gale, M., Jr. (2005) Control of antiviral defenses through hepatitis C virus disruption of retinoic acid-inducible gene-I signaling, *Proc Natl Acad Sci U S A 102*, 2986-2991.
11. Ferreon, J. C., Ferreon, A. C., Li, K., and Lemon, S. M. (2005) Molecular determinants of TRIF proteolysis mediated by the hepatitis C virus NS3/4A protease, *J Biol Chem 280*, 20483-20492.
12. Wittekind, M., Weinheirner, S., Zhang, Y., and Goldfarb, V. (2002) Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes, In *United States Patent Applications Publication*, p 26, United States of America.
13. Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkuhler, C., and De Francesco, R. (1998) Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus, *J Virol 72*, 6758-6769.
14. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, (C.W. Carter, J. R. M. S., Eds., Ed.), pp Volume 276: Macromolecular Crystallography, part A, p.307-326, Methods in Enzymology.
15. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.

16. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2002) ARP/wARP's model-building algorithms. I. The main chain, *Acta Crystallogr D Biol Crystallogr 58*, 968-975.
17. COLLABORATIVE COMPUTATIONAL PROJECT, N. (1994) The CCP4 Suite: Programs for Protein Crystallography., *Acta Crystallographica 50*, 760-763.
18. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res 35*, W375-383.
19. Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.
20. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.
21. Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006) Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate, *J Virol 80*, 3607-3616.
22. DeLano, W. L. (2008) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA.
23. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. C. (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase, *Structure 7*, 1353-1363.
24. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J Biomol NMR 6*, 277-293.
25. Goddard, T. D., and Kneller, D. G. SPARKY 3, University of California, San Francisco.
26. Nalam, M. N., Ali, A., Altman, M. D., Reddy, G. S., Chellappan, S., Kairys, V., Ozen, A., Cao, H., Gilson, M. K., Tidor, B., Rana, T. M., and Schiffer, C. A. (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance, *J Virol 84*, 5368-5378.
27. Carter, P., and Wells, J. A. (1988) Dissecting the catalytic triad of a serine protease, *Nature 332*, 564-568.
28. Krishnan, R., Sadler, J. E., and Tulinsky, A. (2000) Structure of the Ser195Ala mutant of human alpha--thrombin complexed with fibrinopeptide A(7--16): evidence for residual catalytic activity, *Acta Crystallogr D Biol Crystallogr 56*, 406-410.
29. Bode, W., and Huber, R. (1978) Crystal structure analysis and refinement of two variants of trigonal trypsinogen: trigonal trypsin and PEG (polyethylene glycol) trypsinogen and their comparison with orthorhombic trypsin and trigonal trypsinogen, *FEBS Lett 90*, 265-269.
30. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol 372*, 774-797.
31. Steinkuhler, C., Biasiol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A., and De Francesco, R. (1998) Product inhibition of the hepatitis C virus NS3 protease, *Biochemistry 37*, 8899-8905.
32. Tong, X., Bogen, S., Chase, R., Girijavallabhan, V., Guo, Z., Njoroge, F. G., Prongay, A., Saksena, A., Skelton, A., Xia, E., and Ralston, R. (2008) Characterization of resistance mutations against HCV ketoamide protease inhibitors, *Antiviral Res 77*, 177-185.
33. He, Y., King, M. S., Kempf, D. J., Lu, L., Lim, H. B., Krishnan, P., Kati, W., Middleton, T., and Molla, A. (2008) Relative replication capacity and selective advantage profiles of

protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system, *Antimicrob Agents Chemother 52*, 1101-1110.

34. Sarrazin, C., Rouzier, R., Wagner, F., Forestier, N., Larrey, D., Gupta, S. K., Hussain, M., Shah, A., Cutler, D., Zhang, J., and Zeuzem, S. (2007) SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders, *Gastroenterology 132*, 1270-1278.
35. Kieffer, T. L., Sarrazin, C., Miller, J. S., Welker, M. W., Forestier, N., Reesink, H. W., Kwong, A. D., and Zeuzem, S. (2007) Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients, *Hepatology 46*, 631-639.
36. Yi, M., Ma, Y., Yates, J., and Lemon, S. M. (2009) Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation, *PLoS Pathog 5*, e1000403.
37. Lenz, O., Verbinnen, T., Lin, T. I., Vijgen, L., Cummings, M. D., Lindberg, J., Berke, J. M., Dehertogh, P., Fransen, E., Scholliers, A., Vermeiren, K., Ivens, T., Raboisson, P., Edlund, M., Storm, S., Vrang, L., de Kock, H., Fanning, G. C., and Simmen, K. A. (2010) In vitro resistance profile of the HCV NS3/4A protease inhibitor TMC435, *Antimicrob Agents Chemother*.

CHAPTER III

DRUG RESISTANCE AGAINST HCV NS3/4A PROTEASE INHIBITORS:
THE BALANCE OF SUBSTRATE RECOGNITION AND INHIBITOR BINDING

Author Contributions

This work was performed in collaboration – Akbar Ali and I share first authorship. I performed the crystallization and structural analyses; Akbar Ali synthesized danoprevir; William Royer assisted in structure refinement; and Celia Schiffer, Akbar Ali and I wrote the paper.

ABSTRACT

Hepatitis C virus infects an estimated 180 million people worldwide, prompting enormous efforts to develop inhibitors targeting the essential NS3/4A protease. Resistance against the most promising protease inhibitors, telaprevir, boceprevir and danoprevir, has emerged in clinical trials. In this study, crystal structures of the NS3/4A protease domain reveal that viral substrates bind to the protease active site in a conserved manner defining a consensus volume, or substrate envelope. Mutations that confer the greatest resistance occur where inhibitors protrude from the substrate envelope, as these changes selectively weaken inhibitor binding without compromising the binding of substrates. These findings suggest a general model for predicting the susceptibility of protease inhibitors to resistance: drugs designed to fit within the substrate envelope will be less susceptible to resistance, as mutations affecting inhibitor binding would simultaneously interfere with the recognition of viral substrates.

INTRODUCTION

Drug resistance impedes the treatment of quickly evolving diseases. Hepatitis C virus (HCV), a genetically diverse *Flavivirus*, infects an estimated 180 million people worldwide (*1*). The viral RNA genome encodes a single polyprotein that is subsequently cleaved by host cell and viral proteases into structural (C, E1, E2, p7) and non-structural (NS2, NS3, NS4A, NS4B, NS5A, NS5B) proteins (*2*). The viral RNA-dependent RNA polymerase, NS5B, is inherently inaccurate and misincorporation of bases accounts for a very high mutation rate (*3*). While some mutations are neutral, others will alter the viability of the virus and propagate with varying efficiencies in each patient. Thus HCV infected individuals will develop a heterogeneous population of virus variants known as quasispecies (*4*). As patients begin treatment, the selective pressures of antiviral drugs will favor drug resistant variants (*5*). Therefore, an inhibitor must not only recognize one protein variant, but an ensemble of related enzymes. A detailed understanding of the atomic mechanisms of resistance is therefore essential to effectively combat drug resistance against HCV antivirals.

The essential HCV NS3/4A protease is an attractive therapeutic target responsible for cleaving at least four sites along the viral polyprotein. These sites share little sequence homology except for an acid at position P6, cysteine or threonine at P1, and serine or alanine at P1' **(Table 3.1)**. The first cleavage occurs at the 3-4A junction *in cis*, while the remaining substrates are processed *in trans*. The NS3/4A protease also cleaves the human cellular targets TRIF and MAVS, which has been shown to confound the innate immune response to viral infection (*6-8*). Early drug design efforts were hampered

Table 3.1. Genotype 1a NS3/4A substrate cleavage sequences

Substrate	P6	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'
3-4A	D	L	E	V	V	T	S	T	W	V
4A4B	D	E	M	E	E	C	S	Q	H	L
4B5A	E	C	T	T	P	C	S	G	S	W
5A5B	E	D	V	V	C	C	S	M	S	Y

by the relatively shallow, featureless architecture of the protease active site. The eventual observation of N-terminal product inhibition served as a stepping-stone for the discovery of more potent peptidomimetic inhibitors (*9, 10*). Over the past decade, pharmaceutical companies have further developed these lead compounds. Many structure-activity-relationship (SAR) studies have been performed to evaluate the effect of different functional moieties on protease inhibition at positions P4–P1' (*11-17*). Crystal structures have been determined of NS3/4A protease bound to a variety of inhibitors as well as of several drug resistant protease variants, such as R155K and V36M (*18, 19*). These data elucidate the molecular interactions of NS3/4A with inhibitors and the effect of specific drug resistance mutations on binding. These efforts, conducted in parallel by several pharmaceutical companies, led to the discovery of many protease inhibitors. Proof-of-concept for the successful clinical activity of this drug class was first demonstrated by the macrocyclic inhibitor ciluprevir (Boehringer Ingelheim) (*20, 21*), which was later dropped from clinical trials due to cardiotoxicity (*22*). Many other NS3/4A protease inhibitors continue to advance in clinical trials, and telaprevir (Vertex) and boceprevir (Merck) are currently approved for clinical use in humans **(Figure 3.1A)**.

Despite these successes, the rapid acquisition of drug resistance has limited the efficacy of the most potent NS3/4A protease inhibitors in both replicon studies and human clinical trials **(Figure 3.1B; Table 3.2)**. In this study, we show that mutations conferring the most severe resistance occur where the protease extensively contacts the inhibitors but not the natural viral substrates. Four crystal structures of the NS3/4A protease domain in complex with the N-terminal products of viral substrates reveal a

Figure 3.1: NS3/4A protease inhibitors and sites of drug resistance. (A) The leading protease inhibitors in development mimic the N-terminal side of the viral substrates. (B) The majority of reported drug resistance mutations cluster around the protease active site with the catalytic triad depicted in yellow.

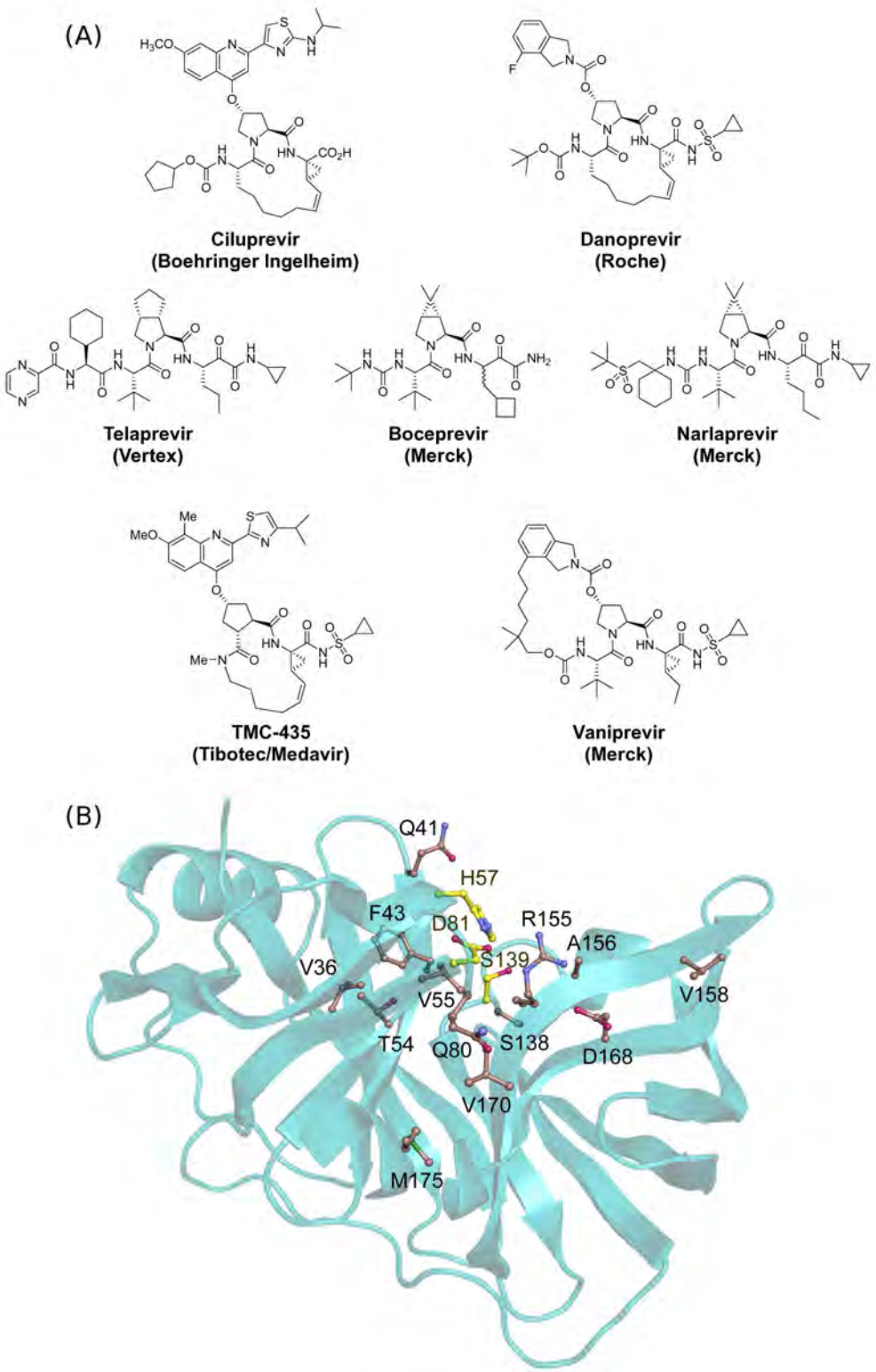

(A)
Ciluprevir
(Boehringer Ingelheim)
Danoprevir
(Roche)
Telaprevir
(Vertex)
Boceprevir
(Merck)
Narlaprevir
(Merck)
TMC-435
(Tibotec/Medavir)
Vaniprevir
(Merck)
(B)
Q41
H57
F43
D81
R155
A156
S139
V36
V55
V158
T54
Q80
S138
D168
V170
M175

Figure 3.1

Table 3.2. Reported sites of drug resistance mutations

Res.	Mutation	Drug
V36	A, M, L, G	Boceprevir, telaprevir
Q41	R	Boceprevir, danoprevir
F43	S, C, V, I	Boceprevir, telaprevir, danoprevir, TMC435
V55	A	Boceprevir
T54	A, S	Boceprevir, telaprevir
Q80	K, R, H, G, L	TMC435
S138	T	Danoprevir, TMC435†
R155	K, T, I, M, G, L, S, Q	Boceprevir, telaprevir, danoprevir, ciluprevir, TMC435
A156	V, T, S, I, G	Boceprevir, telaprevir, danoprevir, ciluprevir, TMC435
V158	I	Boceprevir
D168	A, V, E, G, N, T, Y, H, I	Danoprevir, ciluprevir, TMC435
V170	A	Boceprevir, telaprevir
M175	L	Boceprevir

†-TMC435 displays reduced activity against S138T, but the mutation was not observed in selection experiments.

conserved mode of substrate binding, with the consensus volume defining the substrate envelope. The protease inhibitors danoprevir (3M5L), TMC435 (3KEE) (*23*) and boceprevir (2OC8) (*24*) protrude extensively from the substrate envelope in regions that correlate with known sites of resistance mutations. Most notably, the P2 moieties of all three drugs protrude to contact A156 and R155, which mutate to confer high-level resistance against nearly all drugs reported in the literature (*25-30*). These findings suggest that drug resistance results from a change in molecular recognition and imply that drugs designed to fit within the substrate envelope will be less susceptible to resistance, as mutations altering inhibitor binding will simultaneously interfere with the binding of substrates.

MATERIALS AND METHODS

Mutagenesis and gene information

The HCV genotype 1a NS3/4A protease gene described in a Bristol-Meyers Squibb patent (*31*) was synthesized by GenScript and cloned into the pET28a expression vector (Novagen). The gene encodes a highly soluble form of the NS3/4A protease domain as a single-chain, with 11 core amino acids of NS4A located at the N-terminus. The inactive S139A protease variant was subsequently constructed using the QuikChange Site-Directed Mutagenesis Kit from Stratagene and sequenced by Davis Sequencing for confirmation.

Viral peptide product purchase and storage

Thirty milligrams of each substrate peptide and the corresponding N-terminal cleavage product (4A4B, 4B5A, 5A5B) were purchased from 21st Century Biochemicals (Marlboro, MA) at 95% purity. The purity was determined by HPLC and mass spectrometry profiles as provided by the company. The substrates were synthesized as 12-mers, extending the substrate sequence by an additional amino acid on each side of the minimal cleavage sequence of the HCV polyprotein. Peptide products were synthesized at 7-mers spanning P7–P1. The N-termini of all peptides were acetylated, while the C-termini of substrate peptides were blocked with amide groups. All peptides were stored as solids at −20ºC and dissolved in DMF to a final concentration of 50–100mM for crystallization trials.

Protein expression and purification

NS3/4A expression and purification were carried out as described previously (*31, 32*). Briefly, transformed BL21(DE3) *E. coli* cells were grown at 37ºC to an OD of 0.8 and subsequently induced with the addition of 1mM IPTG. Cells were harvested after 5 hours of expression, pelleted and frozen at −80ºC for storage. Cell pellets were thawed, resuspended in 5mL/g of Resuspension Buffer (50mM phosphate buffer at pH 7.5, 500mM NaCl, 10% glycerol, 2mM β-ME) and lysed with a cell disruptor. The soluble fraction was retained, applied to a nickel column (Qiagen), washed with Resuspension Buffer, and eluted with Resuspension Buffer supplemented with 200mM imidazole. The eluant was dialyzed overnight (MWCO 10kD) to remove the imidazole and the his-tag was simultaneously removed with thrombin treatment. The nickel-purified protein was then flash-frozen and stored at −80ºC for up to six months.

Protein crystallization

The NS3/4A protease construct described previously (*31*) was expressed and purified as reported previously (*31, 32*), detailed in the supporting information. Purified protein was concentrated to ~3mg/mL and loaded on a HiLoad Superdex75 16/60 column equilibrated with 25mM MES at pH 6.5, 500mM NaCl, 10% glycerol, 30μM zinc chloride, and 2mM DTT. The protease fractions were pooled and concentrated to 20–25mg/mL with an Amicon Ultra-15 device (MWCO 10kD; Millipore). The concentrated samples were then incubated at 4°C for one hour with 2–20 molar excess of viral substrate 4A4B, peptide products 4B5A or 5A5B, or danoprevir. Information about the synthesis of viral peptides and danoprevir is provided in the supporting information.

Diffraction-quality crystals were obtained overnight for all ligands by mixing equal volume of concentrated protein solution with precipitant solution (20–26% PEG-3350, 0.1M sodium MES buffer at pH 6.5, and 4% ammonium sulfate) in 24-well VDX hanging drop trays.

Data collection and structure solution

Crystals were flash-frozen in liquid nitrogen and mounted under constant cryostream. X-ray diffraction data were collected at Advanced Photon Source BioCARS 14-IDB, 14-BMC, and LS-CAT 21-ID-F. Diffraction intensities of the complexes were indexed, integrated and scaled using the programs HKL2000 (*33*) and XDS (*34*). 5% of the data was used to calculate R-free (*35*). All structure solutions were generated using isomorphous molecular replacement with PHASER (*36*) or AMORE. The NS3/4A protease domain (PDB code 2A4G) (*37*) was used for molecular replacement in solving the product 4A4B structure, and this structure was subsequently used for solving the other complexes. In all cases, initial refinement was carried out in the absence of modeled ligand, which was subsequently built in. Phases were improved using ARP/wARP (*38*). Iterative rounds of TLS and restrained refinement with CCP4 (*39*) and graphical model building with COOT (*40*) until R-value convergence. All structures were evaluated with MolProbity (*41*) before deposition in the protein data bank.

Structural analysis

Double-difference plots (*42*) were used to determine the structurally invariant regions of the protease, consisting of residues 32–36, 42–47, 50–54, 84–86 and 140–143. Structures were superimposed with PyMOL (*43*) using the Cα atoms of these residues for

all protease molecules from the solved structures (nine total). The B chain of product complex 4A4B was used as the reference structure in all alignments. Fit of individual inhibitors into the substrate envelope was quantified by mapping the substrate envelope and the van der Waals (vdW) volume of each inhibitor on a three-dimensional grid with spacing of 0.5Å. V_{out} for each drug moiety was computed by counting the grid cells, which were occupied by any inhibitor atom of that site but not the substrate envelope, and multiplied by the grid cell size, 0.125$Å^3$ (*44*, *45*).

Substrate envelope and inhibitor analyses

NS3/4A substrate envelope was computed using product complexes 4A4B (B chain), 4B5A (D chain) and 5A5B (A chain). In structures with multiple protease molecules in the asymmetric unit, the one containing the most ordered peptide product was used for the alignment. The protease domain of the full-length NS3/4A structure (A chain; PDB code 1CU1) (*46*), including the C-terminal six amino acids, was included as a product complex 3-4A. All active site alignments were performed in PyMOL using Cα atoms of protease residues 137–139 and 154–160. After superposition, gaussian object maps were generated in PyMOL for each cleavage product. Four consensus gaussian maps were then calculated, representing the intersecting volume of a group of three object maps. Finally, the summation of these four consensus maps was generated to construct the substrate envelope, depicting the vdW volume shared by any three of the four products. The previously determined boceprevir complex (PDB code 2OC8) (*24*) and TMC435 complex (PDB code 3KEE) (*23*) were used in this study (*47*).

RESULTS

Synthesis of danoprevir

We synthesized the macrocyclic inhibitor danoprevir using the convergent reaction sequence described in the Supplemental Experimental Procedures. Briefly, the P2 and P1–P1' fragments were preassembled and the macrocyclic drug compound was generated by a four-step reaction sequence, including P2–P3 amide coupling, ester hydrolysis, coupling with the P1–P1' fragment, and ring-closing metathesis. The P2–P3 fragment was assembled by coupling the commercially available Boc-protected amino acid (*S*)-2-(*tert*-butoxycarbonylamino)non-8-enoic acid (Acme Biosciences, Inc., CA) with the preassembled P2 fragment, (3*R*,5*S*)-5-(methoxycarbonyl)pyrrolidin-3-yl 4-fluoroisoindoline-2-carboxylate (*48*), using HATU/DIPEA conditions. Hydrolysis of the P2–P3 methyl ester with LiOH.H_2O in a mixture of THF-MeOH-H_2O followed by coupling of the resulting acid under HATU/DIPEA conditions with the preassembled P1–P1' fragment, (1*R*,2*S*)-1-amino-N-(cyclopropylsulfonyl)-2-vinylcyclopropane-carboxamide (*49*), 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 protease inhibitor danoprevir.

Structure determination of inhibitor and substrate complexes

Although NS3/4A cleaves the 3011-residue viral polyprotein at four specific sites *in vivo*, we focused on the local interactions of the protease domain with short peptide sequences corresponding to the immediate cleavage sites. All structural studies were carried out with the highly soluble, single-chain construct of the NS3/4A protease domain

described previously (*31*), which contains a fragment of the essential cofactor NS4A covalently linked at the N-terminus by a flexible linker. A similar protease construct was shown to retain comparable catalytic activity to the authentic protein complex (*50*). Crystallization trials were initially carried out using the inactive (S139A) protease variant in complex with substrate peptides spanning P7–P5'. The 4A4B substrate complex revealed cleavage of the scissile bond and no ordered regions for the C-terminal fragment of the substrate. Similar observations were previously described for two other serine proteases where catalytic activity was observed, presumably facilitated by water, despite alanine substitutions of the catalytic serine (*51, 52*). Thus all subsequent crystallization trials with the NS3/4A protease were performed using N-terminal cleavage products of the viral substrates spanning P7–P1.

NS3/4A crystal structures in complex with danoprevir and peptide products 4A4B, 4B5A and 5A5B were determined and refined at 1.25Å, 1.70Å, 1.90Å and 1.60Å resolution, respectively **(Table 3.3)**. The complexes crystallized in the space groups $P2_12_12_1$ and $P2_1$ with between one and four molecules in the asymmetric unit. The average B factors range from 16.8–29.7Å^2, and there are no outliers in the Ramachandran plots. These structures represent the highest resolution crystal structures of NS3/4A protease reported to date.

Overall structure analysis

The NS3/4A protease domain adopts a tertiary fold characteristic of serine proteases of the chymotrypsin family (*53, 54*). A total of nine protease molecules were modeled in the four crystal structures solved in this study with an overall root-mean-

Table 3.3. X-ray data collection and crystallographic refinement statistics.

Dataset	Product 4A4B	Product 4B5A	Product 5A5B	Danoprevir
PDB ID	3M5M	3M5N	3M5O	3M5L
Modeled Ligand	FDEMEEC	SECTTPC	TEDVVCC	Danoprevir
Resolution (Å)	1.70	1.90	1.60	1.25
Space group	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Molecules in AU[§]	2	4	2	1
Cell dimensions:				
a (Å)=	58.2	67.3	47.3	55.2
b (Å)=	60.1	58.9	58.9	58.7
c (Å)=	95.7	98.3	67.1	61.1
β (°)=	90.0	101.0	99.2	90.0
Completeness (%)	99.7	99.4	94.7	99.7
Total reflections	278535	213010	150463	317137
Unique reflection	37615	59622	45685	52645
Average I/σ	10.2	12.4	17.4	20.7
Redundancy	7.4	3.6	3.3	6.0
R_{sym} (%)[¶]	6.2	5.3	6.7	4.6
RMSD[†] **in:**				
Bonds (Å)	0.009	0.009	0.009	0.009
Angles (°)	1.25	1.53	1.26	1.48
R_{factor} (%)[*]	17.7	18.8	18.0	15.0
R_{free} (%)[*]	20.6	24.0	22.6	16.8

§- AU, asymmetric unit.

¶- $R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I = observed intensity, $\langle I \rangle$ = average intensity over symmetry equivalent.

†- RMSD, root mean square deviation.

*- $R_{work} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$. R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

square deviation (RMSD) of 0.28Å. The RMSDs reveal the five most variable regions of the protease to be **(Figure 3.2)**: *(i)* the linker connecting cofactor 4A at the N-terminus, *(ii)* the loop containing residues 65–70, *(iii)* the zinc-binding site containing residues 95–105, *(iv)* the 3_{10} helix region spanning residues 128–136, and *(v)* the active site anti-parallel β-sheet containing residues 156–168. These structural differences likely indicate inherent flexibility in the protease and do not appear to correlate with ligand type or active site occupancy.

Analysis of product complexes

Product complexes 4A4B, 4B5A and 5A5B were further analyzed with the C-terminus of the full-length NS3/4A structure (1CU1), which contains the N-terminal cleavage product of viral substrate 3-4A (*46*). All four products bind to the protease active site in a conserved manner **(Figure 3.3)**, forming an anti-parallel β-sheet with residues 154–160 and burying 500–600$Å^2$ of solvent accessible surface area as calculated by PISA (*55*). The peptide backbone torsions are very similar, being most conserved at position P1 and deviating slightly toward position P4. Eight hydrogen bonds between backbone amide and carbonyl groups are completely conserved, involving protease residues S159 (C159 in product 3-4A), A157, R155, S139A, S138, and G137. S159 (C159 in product 3-4A) and A157 each contribute two hydrogen bonds with the P5 and P3 peptide residues, respectively. All P1 terminal carboxyl groups sit in the oxyanion hole, hydrogen bonding with the Nε atom of H57 and the amide nitrogens of residues 137–139. Although only product 4B5A contains a proline at P2, the other substrate

Figure 3.2: Structural variations between protease molecules of solved crystal complexes. (A) The Cα RMSDs of all nine protease molecules from the four crystal structures determined in this study are plotted as a function of residue number after superimposition using the core residues 33–37, 51–54, 81–86 and 140–144. (B) The RMSD values were then used to color the NS3/4A structure with a rainbow gradient. Blue indicates structurally conserved regions of the protein, while red regions are the most variable from structure to structure.

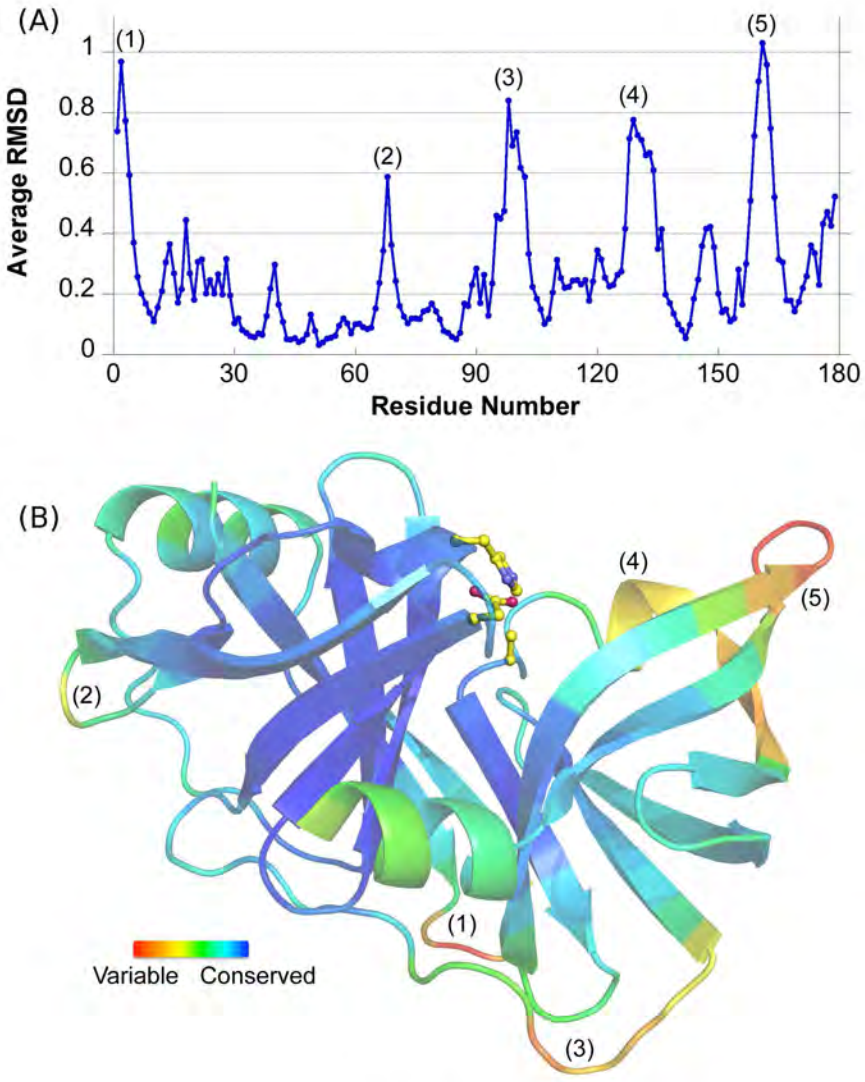


Figure 3.2

Figure 3.3: Stereo view of cleavage product binding to NS3/4A protease. N-terminal protease cleavage products (A) 3-4A, (B) 4A4B, (C) 4B5A and (D) 5A5B are depicted as they bind to the protease active site. All conserved interactions are indicated by black dashes, while red lines depict interactions that are not present in all structures. The electrostatic networks involving residues R123, D168, R155 and D81 are indicated by blue dashes.

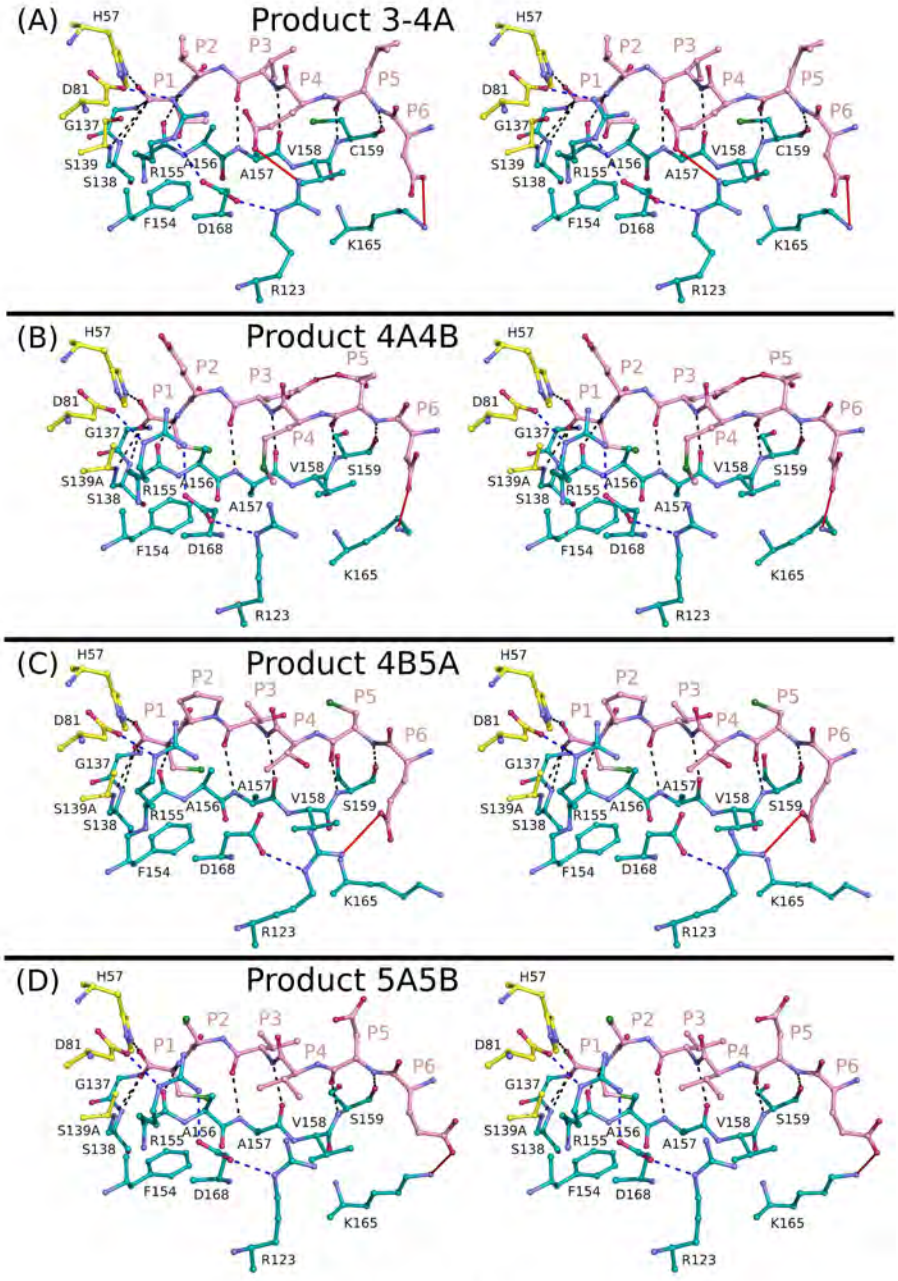

(A)
Product 3-4A
(B)
Product 4A4B
(C)
Product 4B5A
(D)
Product 5A5B
H57
D81
G137
S139
S138
S139A
R155
A156
A157
V158
C159
S159
F154
D168
K165
R123
P1
P2
P3
P4
P5
P6

Figure 3.3

sequences still adopt constrained P2 ϕ torsion angles. Thus products bind similarly despite their high sequence diversity.

The P1 and P6 residues are most conserved among the substrate sequences, as are most of their interactions with the protease. The P1 side chains interact with the aromatic ring of F154. In all structures but product complex 4B5A, K165 forms salt-bridges with the P6 acids, while residues R123, D168, R155, and the catalytic D81 form an ionic network along one surface of the bound products **(Figure 3.3)**. In complex 4B5A, R123 interacts directly with the P6 acid, while D168 reorients and no longer contacts R155. Other molecular interactions in the product complexes are more diverse. Notably, K136 interacts differently with the cleavage products, forming: *(i)* a hydrogen bond with the P2 backbone carbonyl oxygen of 3-4A, *(ii)* a salt-bridge with the P3 glutamate of product 4A4B, and *(iii)* nonspecific vdW interactions with the P2 and P3 side chains of products 4B5A and 5A5B. Also, in product complex 4A4B, an intramolecular hydrogen bond forms between the P3 and P5 glutamate residues, while the unique P4 acid of product 3-4A forms salt-bridges with the guanidinium groups of R123 and R155. Thus distinct patterns of side chain interactions underlie the set of conserved features involved in NS3/4A cleavage product binding.

The substrate envelope

To further analyze the structural similarities of the four NS3/4A product complexes, the active sites were superposed on the Cα atoms of residues 137–139 and 154–160, revealing that both the active site residues and substrate products spanning P6–P1 align closely with an average Cα RMSD of 0.24Å and 0.35Å, respectively. The

consensus vdW volume shared by any three of the four cleavage products was then calculated to generate the NS3/4A substrate envelope **(Figure 3.4A)**. This shape could not be predicted by the primary sequences alone and highlights the conserved mode of viral substrate recognition despite their high sequence diversity.

Analysis of inhibitor complexes

Danoprevir, TMC435 and boceprevir are all peptidomimetic NS3/4A protease inhibitors. Active site superpositions of these drug complexes reveal that the inhibitors interact with many of the same protease residues as the cleavage products. Despite the P3–P1 cyclization of danoprevir and TMC435, the functional groups are positioned similarly in all three inhibitor complexes. The P1 cysteine surrogates interact with the aromatic ring of F154, while the P2 and P3 moieties overlap closely. Although TMC435 does not contain a P4 substituent, the P4 *tert*-butyl groups of danoprevir and boceprevir also align closely. In addition, the P1 and P3 backbone atoms of all inhibitors hydrogen bond with the carbonyl oxygens of R155 and A157, respectively. These observations verify the peptidomimetic nature of these drugs and support their observed mechanism as competitive active site inhibitors.

The largest variation between these three protease inhibitors occurs at P2 where the aromatic rings of danoprevir and TMC435 stack against the guanidinium group of R155 **(Figure 3.4)**. This molecular interaction alters the electrostatic network involving R123, D168, R155, and D81. R155 rotates nearly 180° around Cδ relative to its conformation observed in product complexes, losing its hydrogen bond with D81 but maintaining interaction with D168. Mutations at R155 or D168 would disrupt the

Figure 3.4: Stereo view of protease inhibitors in the substrate envelope. (A) After active site superpositions, the overlapping vdW volume shared by any three of the four cleavage products defines the substrate envelope, depicted in blue. Protease residues that mutate to confer drug resistance are shown in brown. (B) Danoprevir, (C) boceprevir and (D) TMC435 protrude from the substrate envelope at several locations, which correlate with known sites of drug resistance mutations to each inhibitor, shown in red.

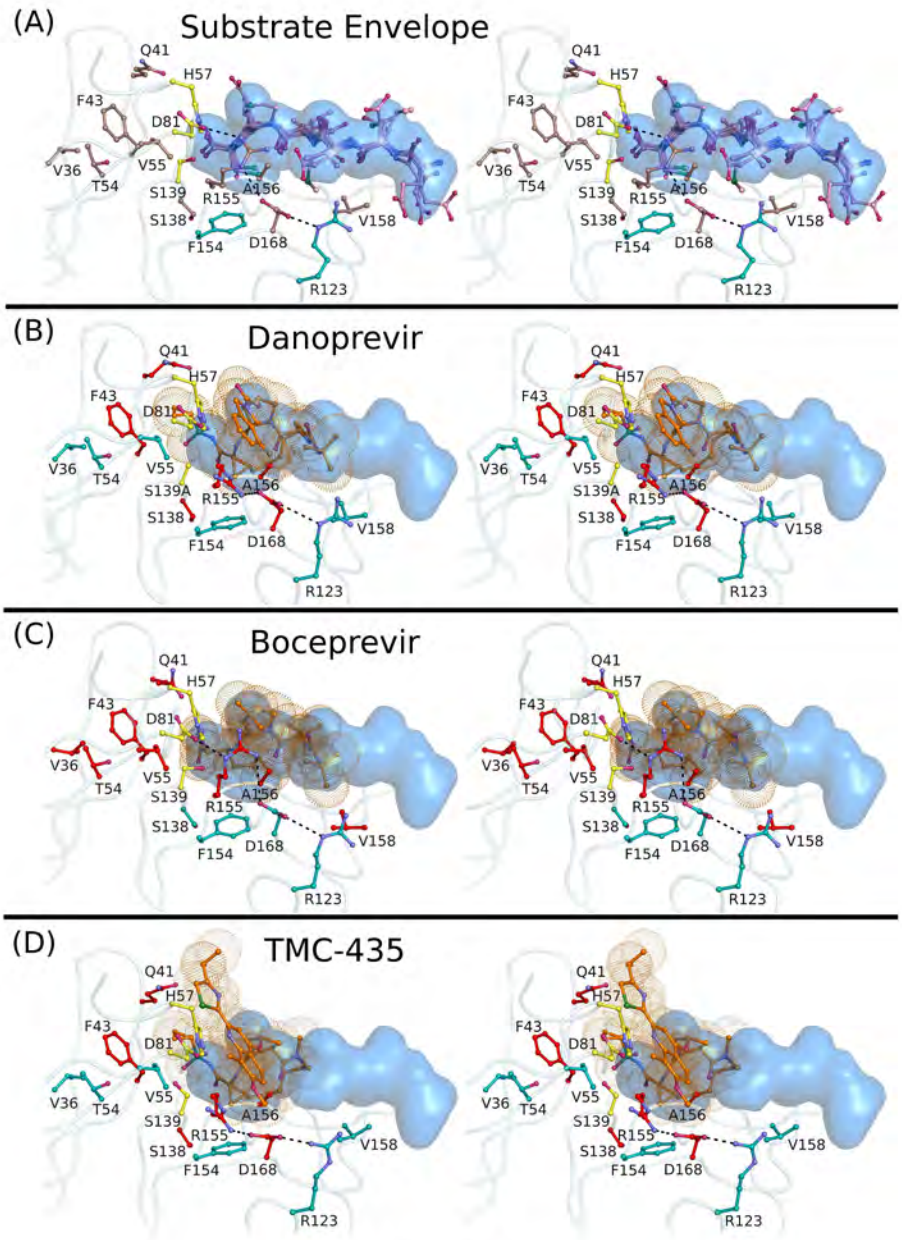

(A)
Substrate Envelope
(B)
Danoprevir
(C)
Boceprevir
(D)
TMC-435
Q41
H57
F43
D81
V36
T54
V55
S139
S139A
S138
R155
A156
F154
D168
V158
R123

Figure 3.4

electrostatic network and destabilize this packing thereby lowering the affinity for these macrocyclic drugs. This provides a structural rationale for the drug resistance mutations R155K, as previously proposed (*19*), and D168A/V, which both confer a selective advantage *in vitro* in the presence of danoprevir or TMC435 (*26, 30*). In addition, the TMC435 complex reveals that R155 is stabilized by a hydrogen bond with Q80, which also mutates to confer resistance to TMC435 (*30*). Thus many of the primary drug resistance mutations can be explained by the disruption of the interactions involving the functional groups at P2 of the drugs.

Insights into drug resistance

To determine the locations where the inhibitors protrude from the substrate envelope, the inhibitor and product complexes were also superposed using residues 137–139 and 154–160. The vdW volumes of inhibitor protrusion from the substrate envelope (V_{out}) (*44, 45*) were calculated for each drug and compared with published EC_{50} fold-change data for drug resistance variants (*30*). The P2 moieties of boceprevir, danoprevir and TMC435 protrude most extensively from the substrate envelope with V_{out} values of 105, 294 and 496Å^3, respectively **(Table 3.4)**. The magnitudes of the EC_{50} fold-change data determined for each NS3/4A mutant generally trend with the V_{out} values for the three drugs.

However, the precise level of drug resistance observed is also determined by the particular change in molecular interaction occurring for a given mutation. A156 and R155 pack with the P2 moieties of these three inhibitors where they protrude beyond the substrate envelope. Mutations of A156 to bulkier side chains would result in a steric

Table 3.4. Drug activities† and volume of protrusion from the substrate envelope (V_{out}).

		Boceprevir		Danoprevir		TMC435	
Subsite	**Resistance Mutation**	**V_{out} ($Å^3$)**	**EC_{50} FC**	**V_{out} ($Å^3$)**	**EC_{50} FC**	**V_{out} ($Å^3$)**	**EC_{50} FC**
Total		292		500		649	
P1		76		67		64	
P2		105		294		496	
	Q80R		0.5		3.5		6.9
	Q80K		0.8		2.3		7.7
	R155K		4.7		**447**		30
	A156V		**75**		63		177
	A156T		**65**		41		44
	D168A		0.7		153		**594**
	D168E		0.8		75		40
P3		34		70		67	
P4		76		69		0	
	V158I		3.3‡		ND		ND

†- Antiviral activity was reported previously by Lenz et al., 2010.
‡- Fold-change in EC_{50} reported in replicon assay by Qiu et al., 2009.

clash with the P2 drug moieties. For example, the rigid dimethylcyclopropane group of boceprevir protrudes from the substrate envelope at the P2 subsite, and A156V or A156T confer 65 and 75 fold-changes in EC_{50}, respectively **(Table 3.4)**. Similarly, molecular changes at R155 and D168 would result in a substantial loss of interactions with P2. The most extensive protrusions of danoprevir and TMC435 at P2 trend with their greatest fold-change in potency of nearly 450 and 600, respectively, from mutations in this subsite. Thus the extent by which an inhibitor protrudes from the substrate envelope in a given subsite is indicative of the inhibitor's vulnerability to drug resistance.

Further structural analyses with the substrate envelope provide insights into other NS3/4A drug resistance mutations. The P1' sulfonamide groups of danoprevir and TMC435, as well as the P1' ketoamide of boceprevir, protrude from the substrate envelope near residues Q41 and F43, which both mutate to confer low-level resistance to these drugs (*25, 30, 56*). The keto group of boceprevir also projects outside the substrate envelope near T54 and V55. T54A/S confers low-level resistance to boceprevir, while V55A was recently identified in patient isolates after treatment with boceprevir (*57*). The analogous carbonyl groups of danoprevir and TMC435, however, are orientated in the opposite direction and protrude toward S138. In fact, *in vitro* studies reveal reduced activity for danoprevir and TMC435 against S138T variants, while boceprevir remains fully active (*30, 56*). The bulky P4 *tert*-butyl group of boceprevir extends outside the substrate envelope contacting V158; V158I variant has lower affinity, likely due to a steric clash (*58*). This variant may also impact the affinity of danoprevir, as its P4 *tert*-butyl also protrudes at the same location. These findings demonstrate that in regions

outside the P2 subsite, positions where danoprevir, TMC435 and boceprevir protrude from the substrate envelope also correlate with many other known sites for drug resistance mutations.

DISCUSSION

The emergence of drug resistance is a major obstacle in modern medicine that limits the long-term usefulness of the most promising therapeutics. By considering how HCV NS3/4A protease inhibitors bind relative to natural viral substrates, we discovered that primary sites of resistance occur in regions of the protease where drugs protrude from the substrate envelope. In particular, R155 and A156, which mutate to confer severe resistance against danoprevir, TMC435 and boceprevir, interact closely with the P2 drug moieties where they protrude most extensively from the substrate envelope. Molecular changes at these residues confer resistance by selectively weakening inhibitor binding without compromising the binding of viral substrates. We further speculate that these mutations will not considerably affect the binding of the host cellular substrates TRIF and MAVS, which likely fit well within the substrate envelope as they share many features with the viral substrates. However, TRIF contains a tract of eight proline residues instead of an acidic residue at position P6, which may modulate its binding. Further structural studies are warranted to better ascertain the molecular details of how these cellular substrates are recognized by the NS3/4A protease.

Although this study focuses on danoprevir, TMC435 and boceprevir, other NS3/4A protease inhibitors in clinical trials, including telaprevir, narlaprevir and vaniprevir **(Figure 3.1A)**, contain similar functional groups that likely protrude from the substrate envelope. Most notably, all these drug candidates contain bulky P2 moieties and are therefore susceptible to cross-resistance against mutations at R155 and A156. R155 and A156 mutations have been shown to confer telaprevir resistance in treated

patients (*59*). Cross-resistance studies have also shown that narlaprevir displays similar fold losses in activity against most of the known drug resistance mutations for telaprevir and boceprevir (*60*). Ultimately, to slow the emergence of multi-drug resistant viral strains, inhibitors should be confined within the substrate envelope, particularly at the P2 position. To compensate for the loss of binding affinity that will likely accompany these changes, additional interactions could potentially be optimized spanning the S4–S6 subsites of the protease and the catalytic triad.

Our findings further suggest a general model for using the substrate envelope to predict patterns of drug resistance in other quickly evolving diseases. For drug resistance to occur, mutations must selectively weaken target affinity for an inhibitor without significantly altering its natural biological function. Mutations occurring outside the substrate envelope are better able to achieve this effect, as these molecular changes can selectively impair inhibitor binding without compromising substrate recognition. When the interaction of a drug target with its biological substrates can be structurally characterized, we predict that drugs contained within the substrate envelope will be less susceptible to resistance. Structure-based design strategies can utilize this model as an added constraint to develop inhibitors that fit within the substrate envelope. In fact, previous work in our laboratory provides proof-of-concept for the successful incorporation of the substrate envelope in the design of novel HIV protease inhibitors, which maintain potent activities against a series of highly resistant HIV strains (*45, 61-66*). As a general paradigm, substrate envelope-based design efforts can facilitate the rationale development of drug candidates that are less susceptible to resistance.

ACKNOWLEDGEMENTS

We thank Herbert Klei for helpful discussions. We also thank Zdzislaw Wawrzak, Mike Bolbat and Keith Brister of the LS-CAT beamline at Argonne National Laboratory for data collection of the danoprevir complex; Madhavi Nalam and Rajintha Bandaranayake for assistance with structural refinement; Aysegul Ozen for providing V_{out} calculations; and Shivender Shandilya and Yufeng Cai for computational support. The NIH grants R01-GM65347 and R01-AI085051 supported this work. Use of APS was supported by the U.S. DOE, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Use of the BioCARS Sector 14 was supported by NIH-NCRR RR007707. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor under grant number 085P1000817.

REFERENCES

1. WHO. (2010) Vaccine Research: Hepatitis C, In *Hepatitis C Virus: Disease Burden* (Barnes, E., Ed.).
2. Major, M. E., and Feinstone, S. M. (1997) The molecular virology of hepatitis C, *Hepatology 25*, 1527-1538.
3. Qureshi, S. A. (2007) Hepatitis C virus--biology, host evasion strategies, and promising new therapies on the horizon, *Med Res Rev 27*, 353-373.
4. Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J., and Gomez, J. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution, *J Virol 66*, 3225-3229.
5. Paolucci, S., Baldanti, F., Campanini, G., Zavattoni, M., Cattaneo, E., Dossena, L., and Gerna, G. (2001) Analysis of HIV drug-resistant quasispecies in plasma, peripheral blood mononuclear cells and viral isolates from treatment-naive and HAART patients, *J Med Virol 65*, 207-217.
6. Chen, Z., Benureau, Y., Rijnbrand, R., Yi, J., Wang, T., Warter, L., Lanford, R. E., Weinman, S. A., Lemon, S. M., Martin, A., and Li, K. (2007) GB virus B disrupts RIG-I signaling by NS3/4A-mediated cleavage of the adaptor protein MAVS, *J Virol 81*, 964-976.
7. Li, X. D., Sun, L., Seth, R. B., Pineda, G., and Chen, Z. J. (2005) Hepatitis C virus protease NS3/4A cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity, *Proc Natl Acad Sci U S A 102*, 17717-17722.
8. Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr., and Lemon, S. M. (2005) 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 102*, 2992-2997.
9. Llinas-Brunet, M., Bailey, M., Fazal, G., Goulet, S., Halmos, T., Laplante, S., Maurice, R., Poirier, M., Poupart, M. A., Thibeault, D., Wernic, D., and Lamarre, D. (1998) Peptide-based inhibitors of the hepatitis C virus serine protease, *Bioorg Med Chem Lett 8*, 1713-1718.
10. Steinkuhler, C., Biasiol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A., and De Francesco, R. (1998) Product inhibition of the hepatitis C virus NS3 protease, *Biochemistry 37*, 8899-8905.
11. Arasappan, A., Njoroge, F. G., Chen, K. X., Venkatraman, S., Parekh, T. N., Gu, H., Pichardo, J., Butkiewicz, N., Prongay, A., Madison, V., and Girijavallabhan, V. (2006) P2-P4 macrocyclic inhibitors of hepatitis C virus NS3-4A serine protease, *Bioorg Med Chem Lett 16*, 3960-3965.
12. Bogen, S., Arasappan, A., Pan, W., Ruan, S., Padilla, A., Saksena, A. K., Girijavallabhan, V., and Njoroge, F. G. (2008) Hepatitis C virus NS3-4A serine protease inhibitors: SAR of new P1 derivatives of SCH 503034, *Bioorg Med Chem Lett 18*, 4219-4223.
13. Malancona, S., Colarusso, S., Ontoria, J. M., Marchetti, A., Poma, M., Stansfield, I., Laufer, R., Di Marco, A., Taliani, M., Verdirame, M., Gonzalez-Paz, O., Matassa, V. G., and Narjes, F. (2004) SAR and pharmacokinetic studies on phenethylamide inhibitors of the hepatitis C virus NS3/NS4A serine protease, *Bioorg Med Chem Lett 14*, 4575-4579.
14. Nilsson, M., Belfrage, A. K., Lindstrom, S., Wahling, H., Lindquist, C., Ayesa, S., Kahnberg, P., Pelcman, M., Benkestock, K., Agback, T., Vrang, L., Terelius, Y., Wikstrom, K., Hamelink, E., Rydergard, C., Edlund, M., Eneroth, A., Raboisson, P., Lin,

T. I., de Kock, H., Wigerinck, P., Simmen, K., Samuelsson, B., and Rosenquist, S. (2010) Synthesis and SAR of potent inhibitors of the Hepatitis C virus NS3/4A protease: exploration of P2 quinazoline substituents, *Bioorg Med Chem Lett 20*, 4004-4011.

15. Venkatraman, S., Velazquez, F., Wu, W., Blackman, M., Chen, K. X., Bogen, S., Nair, L., Tong, X., Chase, R., Hart, A., Agrawal, S., Pichardo, J., Prongay, A., Cheng, K. C., Girijavallabhan, V., Piwinski, J., Shih, N. Y., and Njoroge, F. G. (2009) Discovery and structure-activity relationship of P1-P3 ketoamide derived macrocyclic inhibitors of hepatitis C virus NS3 protease, *J Med Chem 52*, 336-346.
16. Raboisson, P., de Kock, H., Rosenquist, A., Nilsson, M., Salvador-Oden, L., Lin, T. I., Roue, N., Ivanov, V., Wahling, H., Wickstrom, K., Hamelink, E., Edlund, M., Vrang, L., Vendeville, S., Van de Vreken, W., McGowan, D., Tahri, A., Hu, L., Boutton, C., Lenz, O., Delouvroy, F., Pille, G., Surleraux, D., Wigerinck, P., Samuelsson, B., and Simmen, K. (2008) 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 18*, 4853-4858.
17. Perni, R. B., Pitlik, J., Britt, S. D., Court, J. J., Courtney, L. F., Deininger, D. D., Farmer, L. J., Gates, C. A., Harbeson, S. L., Levin, R. B., Lin, C., Lin, K., Moon, Y. C., Luong, Y. P., O'Malley, E. T., Rao, B. G., Thomson, J. A., Tung, R. D., Van Drie, J. H., and Wei, Y. (2004) Inhibitors of hepatitis C virus NS3.4A protease 2. Warhead SAR and optimization, *Bioorg Med Chem Lett 14*, 1441-1446.
18. Zhou, Y., Bartels, D. J., Hanzelka, B. L., Muh, U., Wei, Y., Chu, H. M., Tigges, A. M., Brennan, D. L., Rao, B. G., Swenson, L., Kwong, A. D., and Lin, C. (2008) Phenotypic characterization of resistant Val36 variants of hepatitis C virus NS3-4A serine protease, *Antimicrob Agents Chemother 52*, 110-120.
19. Zhou, Y., Muh, U., Hanzelka, B. L., Bartels, D. J., Wei, Y., Rao, B. G., Brennan, D. L., Tigges, A. M., Swenson, L., Kwong, A. D., and Lin, C. (2007) Phenotypic and structural analyses of hepatitis C virus NS3 protease Arg155 variants: sensitivity to telaprevir (VX-950) and interferon alpha, *J Biol Chem 282*, 22619-22628.
20. Lamarre, D., Anderson, P. C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bos, M., Cameron, D. R., Cartier, M., Cordingley, M. G., Faucher, A. M., Goudreau, N., Kawai, S. H., Kukolj, G., Lagace, L., LaPlante, S. R., Narjes, H., Poupart, M. A., Rancourt, J., Sentjens, R. E., St George, R., Simoneau, B., Steinmann, G., Thibeault, D., Tsantrizos, Y. S., Weldon, S. M., Yong, C. L., and Llinas-Brunet, M. (2003) An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus, *Nature 426*, 186-189.
21. Hinrichsen, H., Benhamou, Y., Wedemeyer, H., Reiser, M., Sentjens, R. E., Calleja, J. L., Forns, X., Erhardt, A., Cronlein, J., Chaves, R. L., Yong, C. L., Nehmiz, G., and Steinmann, G. G. (2004) Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients, *Gastroenterology 127*, 1347-1355.
22. Vanwolleghem, T., Meuleman, P., Libbrecht, L., Roskams, T., De Vos, R., and Leroux-Roels, G. (2007) Ultra-rapid cardiotoxicity of the hepatitis C virus protease inhibitor BILN 2061 in the urokinase-type plasminogen activator mouse, *Gastroenterology 133*, 1144-1155.
23. Cummings, M. D., Lindberg, J., Lin, T. I., de Kock, H., Lenz, O., Lilja, E., Fellander, S., Baraznenok, V., Nystrom, S., Nilsson, M., Vrang, L., Edlund, M., Rosenquist, A., Samuelsson, B., Raboisson, P., and Simmen, K. (2010) Induced-fit binding of the macrocyclic noncovalent inhibitor TMC435 to its HCV NS3/NS4A protease target, *Angew Chem Int Ed Engl 49*, 1652-1655.

24. Prongay, A. J., Guo, Z., Yao, N., Pichardo, J., Fischmann, T., Strickland, C., Myers, J., Jr., Weber, P. C., Beyer, B. M., Ingram, R., Hong, Z., Prosise, W. W., Ramanathan, L., Taremi, S. S., Yarosh-Tomaine, T., Zhang, R., Senior, M., Yang, R. S., Malcolm, B., Arasappan, A., Bennett, F., Bogen, S. L., Chen, K., Jao, E., Liu, Y. T., Lovey, R. G., Saksena, A. K., Venkatraman, S., Girijavallabhan, V., Njoroge, F. G., and Madison, V. (2007) Discovery of the HCV NS3/4A protease inhibitor (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3- [2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl] - 6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (Sch 503034) II. Key steps in structure-based optimization, *J Med Chem 50*, 2310-2318.
25. Tong, X., Bogen, S., Chase, R., Girijavallabhan, V., Guo, Z., Njoroge, F. G., Prongay, A., Saksena, A., Skelton, A., Xia, E., and Ralston, R. (2008) Characterization of resistance mutations against HCV ketoamide protease inhibitors, *Antiviral Res 77*, 177-185.
26. He, Y., King, M. S., Kempf, D. J., Lu, L., Lim, H. B., Krishnan, P., Kati, W., Middleton, T., and Molla, A. (2008) Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system, *Antimicrob Agents Chemother 52*, 1101-1110.
27. Sarrazin, C., Rouzier, R., Wagner, F., Forestier, N., Larrey, D., Gupta, S. K., Hussain, M., Shah, A., Cutler, D., Zhang, J., and Zeuzem, S. (2007) SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders, *Gastroenterology 132*, 1270-1278.
28. Kieffer, T. L., Sarrazin, C., Miller, J. S., Welker, M. W., Forestier, N., Reesink, H. W., Kwong, A. D., and Zeuzem, S. (2007) Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients, *Hepatology 46*, 631-639.
29. Yi, M., Ma, Y., Yates, J., and Lemon, S. M. (2009) Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation, *PLoS Pathog 5*, e1000403.
30. Lenz, O., Verbinnen, T., Lin, T. I., Vijgen, L., Cummings, M. D., Lindberg, J., Berke, J. M., Dehertogh, P., Fransen, E., Scholliers, A., Vermeiren, K., Ivens, T., Raboisson, P., Edlund, M., Storm, S., Vrang, L., de Kock, H., Fanning, G. C., and Simmen, K. A. (2010) In vitro resistance profile of the HCV NS3/4A protease inhibitor TMC435, *Antimicrob Agents Chemother*.
31. Wittekind, M., Weinheirner, S., Zhang, Y., and Goldfarb, V. (2002) Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes, In *United States Patent Applications Publication*, p 26, United States of America.
32. Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkuhler, C., and De Francesco, R. (1998) Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus, *J Virol 72*, 6758-6769.
33. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, (C.W. Carter, J. R. M. S., Eds., Ed.), pp Volume 276: Macromolecular Crystallography, part A, p.307-326, Methods in Enzymology.
34. Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants, *Journal of Applied Crystallography 26*, 795-800.
35. Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.

36. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.
37. Arasappan, A., Njoroge, F. G., Chan, T. Y., Bennett, F., Bogen, S. L., Chen, K., Gu, H., Hong, L., Jao, E., Liu, Y. T., Lovey, R. G., Parekh, T., Pike, R. E., Pinto, P., Santhanam, B., Venkatraman, S., Vaccaro, H., Wang, H., Yang, X., Zhu, Z., McKittrick, B., Saksena, A. K., Girijavallabhan, V., Pichardo, J., Butkiewicz, N., Ingram, R., Malcolm, B., Prongay, A., Yao, N., Marten, B., Madison, V., Kemp, S., Levy, O., Lim-Wilby, M., Tamura, S., and Ganguly, A. K. (2005) Hepatitis C virus NS3-4A serine protease inhibitors: SAR of P'2 moiety with improved potency, *Bioorg Med Chem Lett 15*, 4180-4184.
38. Morris, R. J., Perrakis, A., and Lamzin, V. S. (2002) ARP/wARP's model-building algorithms. I. The main chain, *Acta Crystallogr D Biol Crystallogr 58*, 968-975.
39. COLLABORATIVE COMPUTATIONAL PROJECT, N. (1994) The CCP4 Suite: Programs for Protein Crystallography., *Acta Crystallographica 50*, 760-763.
40. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.
41. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res 35*, W375-383.
42. Prabu-Jeyabalan, M., Nalivaika, E. A., Romano, K., and Schiffer, C. A. (2006) Mechanism of substrate recognition by drug-resistant human immunodeficiency virus type 1 protease variants revealed by a novel structural intermediate, *J Virol 80*, 3607-3616.
43. DeLano, W. L. (2008) The PyMOL Molecular Graphics System, DeLano Scientific LLC, San Carlos, CA.
44. Nalam, M. N., Ali, A., Altman, M. D., Reddy, G. S., Chellappan, S., Kairys, V., Ozen, A., Cao, H., Gilson, M. K., Tidor, B., Rana, T. M., and Schiffer, C. A. (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance, *J Virol 84*, 5368-5378.
45. Chellappan, S., Kairys, V., Fernandes, M. X., Schiffer, C., and Gilson, M. K. (2007) Evaluation of the substrate envelope hypothesis for inhibitors of HIV-1 protease, *Proteins 68*, 561-567.
46. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. C. (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase, *Structure 7*, 1353-1363.
47. Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N., and Bourne, P. E. (2000) The Protein Data Bank, *Nucleic Acids Res 28*, 235-242.
48. Arasappan, A., Njoroge, F. G., and Girijavallabhan, V., M. (2005) Substituted prolines as inhibitors of Hepatitis C virus NS3 serine protease.
49. Wang, X. A., Sun, L.-Q., Sit, S.-Y., Sin, N., Scola, P. M., Hewawasam, P., Good, A., C., Chen, Y., and Campbell, A. (2006) Hepatitis C virus inhibitors.
50. Taremi, S. S., Beyer, B., Maher, M., Yao, N., Prosise, W., Weber, P. C., and Malcolm, B. A. (1998) Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease, *Protein Sci 7*, 2143-2149.
51. Carter, P., and Wells, J. A. (1988) Dissecting the catalytic triad of a serine protease, *Nature 332*, 564-568.

52. Krishnan, R., Sadler, J. E., and Tulinsky, A. (2000) Structure of the Ser195Ala mutant of human alpha--thrombin complexed with fibrinopeptide A(7--16): evidence for residual catalytic activity, *Acta Crystallogr D Biol Crystallogr 56*, 406-410.
53. Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A. (1996) Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide, *Cell 87*, 343-355.
54. Bode, W., and Huber, R. (1978) Crystal structure analysis and refinement of two variants of trigonal trypsinogen: trigonal trypsin and PEG (polyethylene glycol) trypsinogen and their comparison with orthorhombic trypsin and trigonal trypsinogen, *FEBS Lett 90*, 265-269.
55. Krissinel, E., and Henrick, K. (2007) Inference of macromolecular assemblies from crystalline state, *J Mol Biol 372*, 774-797.
56. Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A. M., Young, S. K., Lee, B., Heckerman, D., Carlson, J., Reyor, L. L., Kleyman, M., McMahon, C. M., Birch, C., Schulze Zur Wiesch, J., Ledlie, T., Koehrsen, M., Kodira, C., Roberts, A. D., Lauer, G. M., Rosen, H. R., Bihl, F., Cerny, A., Spengler, U., Liu, Z., Kim, A. Y., Xing, Y., Schneidewind, A., Madey, M. A., Fleckenstein, J. F., Park, V. M., Galagan, J. E., Nusbaum, C., Walker, B. D., Lake-Bakaar, G. V., Daar, E. S., Jacobson, I. M., Gomperts, E. D., Edlin, B. R., Donfield, S. M., Chung, R. T., Talal, A. H., Marion, T., Birren, B. W., Henn, M. R., and Allen, T. M. (2008) Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients, *Hepatology 48*, 1769-1778.
57. Susser, S., Welsch, C., Wang, Y., Zettler, M., Domingues, F. S., Karey, U., Hughes, E., Ralston, R., Tong, X., Herrmann, E., Zeuzem, S., and Sarrazin, C. (2009) Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients, *Hepatology*.
58. Qiu, P., Sanfiorenzo, V., Curry, S., Guo, Z., Liu, S., Skelton, A., Xia, E., Cullen, C., Ralston, R., Greene, J., and Tong, X. (2009) Identification of HCV protease inhibitor resistance mutations by selection pressure-based method, *Nucleic Acids Res 37*, e74.
59. Sarrazin, C., Kieffer, T. L., Bartels, D., Hanzelka, B., Muh, U., Welker, M., Wincheringer, D., Zhou, Y., Chu, H. M., Lin, C., Weegink, C., Reesink, H., Zeuzem, S., and Kwong, A. D. (2007) Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir, *Gastroenterology 132*, 1767-1777.
60. Tong, X., Arasappan, A., Bennett, F., Chase, R., Feld, B., Guo, Z., Hart, A., Madison, V., Malcolm, B., Pichardo, J., Prongay, A., Ralston, R., Skelton, A., Xia, E., Zhang, R., and Njoroge, F. G. (2010) Preclinical characterization of the antiviral activity of SCH 900518 (Narlaprevir), a novel mechanism-based inhibitor of hepatitis C virus NS3 protease, *Antimicrob Agents Chemother 54*, 2365-2370.
61. Chellappan, S., Kiran Kumar Reddy, G. S., Ali, A., Nalam, M. N., Anjum, S. G., Cao, H., Kairys, V., Fernandes, M. X., Altman, M. D., Tidor, B., Rana, T. M., Schiffer, C. A., and Gilson, M. K. (2007) Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis, *Chem Biol Drug Des 69*, 298-313.
62. Ali, A., Reddy, G. S. K. K., Nalam, M. N. L., Anjum, S. G., Cao, H., Altman, M. D., Tidor, B., Rana, T. M., and Schiffer, C. A. (2009) Substrate envelope based design of new HIV-1 protease inhibitors active against drug-resistant HIV-1, In *238th ACS*

National Meeting, Washington, DC, United States, pp MEDI-102, American Chemical Society.

63. King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease, *Chem Biol 11*, 1333-1338.
64. Prabu-Jeyabalan, M., King, N. M., Nalivaika, E. A., Heilek-Snyder, G., Cammack, N., and Schiffer, C. A. (2006) Substrate envelope and drug resistance: crystal structure of RO1 in complex with wild-type human immunodeficiency virus type 1 protease, *Antimicrob Agents Chemother 50*, 1518-1521.
65. Altman, M. D., Nalivaika, E. A., Prabu-Jeyabalan, M., Schiffer, C. A., and Tidor, B. (2008) Computational design and experimental study of tighter binding peptides to an inactivated mutant of HIV-1 protease, *Proteins 70*, 678-694.
66. Altman, M. D., Ali, A., Reddy, G. S., Nalam, M. N., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., Gilson, M. K., Schiffer, C. A., Rana, T. M., and Tidor, B. (2008) HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants, *J Am Chem Soc 130*, 6099-6113.

CHAPTER IV

THE MOLECULAR BASIS OF DRUG RESISTANCE AGAINST HCV NS3/4A PROTEASE INHIBITORS

AUTHOR CONTRIBUTIONS

This work was performed in collaboration – I performed the crystallization and structural analyses; Akbar Ali synthesized danoprevir, vaniprevir and MK-5172; Djade Soumana assisted in protein crystallization and structure solution; Ayşegül Özen assisted in structure solution and van der Waals energy calculations; Cihan Aydin performed the *in vitro* enzymatic inhibition experiments; Laura Deveau designed and cloned all protease mutants; Wei Huang, Alicia Newton and Christos Petropoulos performed replicon-based drug susceptibility testing; and Celia Schiffer and I wrote the paper.

ABSTRACT

Hepatitis C virus infects 180 million people worldwide but antiviral therapies are only effective for a subset of patients. Despite extensive efforts to develop inhibitors targeting the essential viral NS3/4A protease, the rapid acquisition of resistance challenges the long-term efficacy of this drug class. We determined high-resolution crystal structures of four chemically diverse protease inhibitors – telaprevir, danoprevir, vaniprevir and MK-5172 – in complex with the wildtype protease and three major drug resistant variants. The drugs exhibit differential susceptibilities to these mutations; telaprevir, danoprevir and vaniprevir interact directly with sites that mutate to confer resistance, while MK-5172 interacts in a unique conformation with the catalytic triad. This novel mode of MK-5172 binding explains its retained potency against two multi-drug resistant variants, R155K and D168A. Together our findings define the molecular basis of HCV N3/4A protease inhibitor resistance, providing a direct strategy for designing robust therapies against this rapidly evolving virus.

INTRODUCTION

Hepatitis C virus (HCV) infects an estimated 180 million people worldwide and is the leading cause for liver transplantation in the United States (*1, 2*). The HCV genome encodes a single polyprotein that is cleaved into structural (C, E1, E2, p7) and non-structural (NS2, NS3, NS4A, NS4B, NS5A, NS5B) proteins by host cell and viral proteases (*3*). The viral NS3/4A protease, a chymotrypsin-like serine protease, is a prime therapeutic target that cleaves four known sites along the polyprotein (*4, 5*). The NS3/4A protease also hydrolyzes two human proteins of the innate immune system, TRIF and MAVS, thereby confounding the immune response to viral infection (*6, 7*). Pharmaceutical companies have invested considerable resources in the development of NS3/4A protease inhibitors. Proof-of-concept for the efficacy of this drug class was first demonstrated in 2004 by the inhibitor ciluprevir (*8, 9*), though later discontinued due to cardiotoxicity (*10*). Many other protease inhibitors have since advanced into later stages of human clinical trials **(Figure 4.1)**. The Food and Drug Administration recently approved the protease inhibitors telaprevir and boceprevir for clinical use in combination with ribavirin and pegylated interferon-α, marking an important milestone in antiviral research and drug development over the past two decades.

HCV replicates rapidly in untreated individuals, producing an estimated one trillion virus particles daily (*11, 12*). Viral replication is inherently error-prone and misincorporation of bases leads to genetically diverse virus populations over time, known as quasispecies (*13*). As patients begin antiviral treatment, selective drug pressures favor resistant viral strains, particularly in the setting of monotherapy (*14*). To reduce the

Figure 4.1: The chemical structures of nine NS3/4A protease inhibitors. The canonical nomenclature for drug moiety positioning is indicated using telaprevir. Telaprevir, danoprevir, vaniprevir and MK- 5172 are representative of many other protease inhibitors in development. Telaprevir, recently approved for clinical use, is an acyclic ketoamide inhibitor that forms a reversible, covalent bond with the protease. Danoprevir, currently in phase II clinical trials, is a sulfonamide inhibitor with a P1–P3 macrocycle. Vaniprevir and MK-5172 are also sulfonamide inhibitors, but contain P2–P4 macrocycles. Vaniprevir and MK-5172 differ in the construction of their $P2^*$ moieties: vaniprevir contains a carbamate linkage between the P2 proline and $P2^*$ isoindoline moiety, whereas MK-5172 contains a shorter ether linkage between its P2 proline and $P2^*$ quinoxaline moiety.

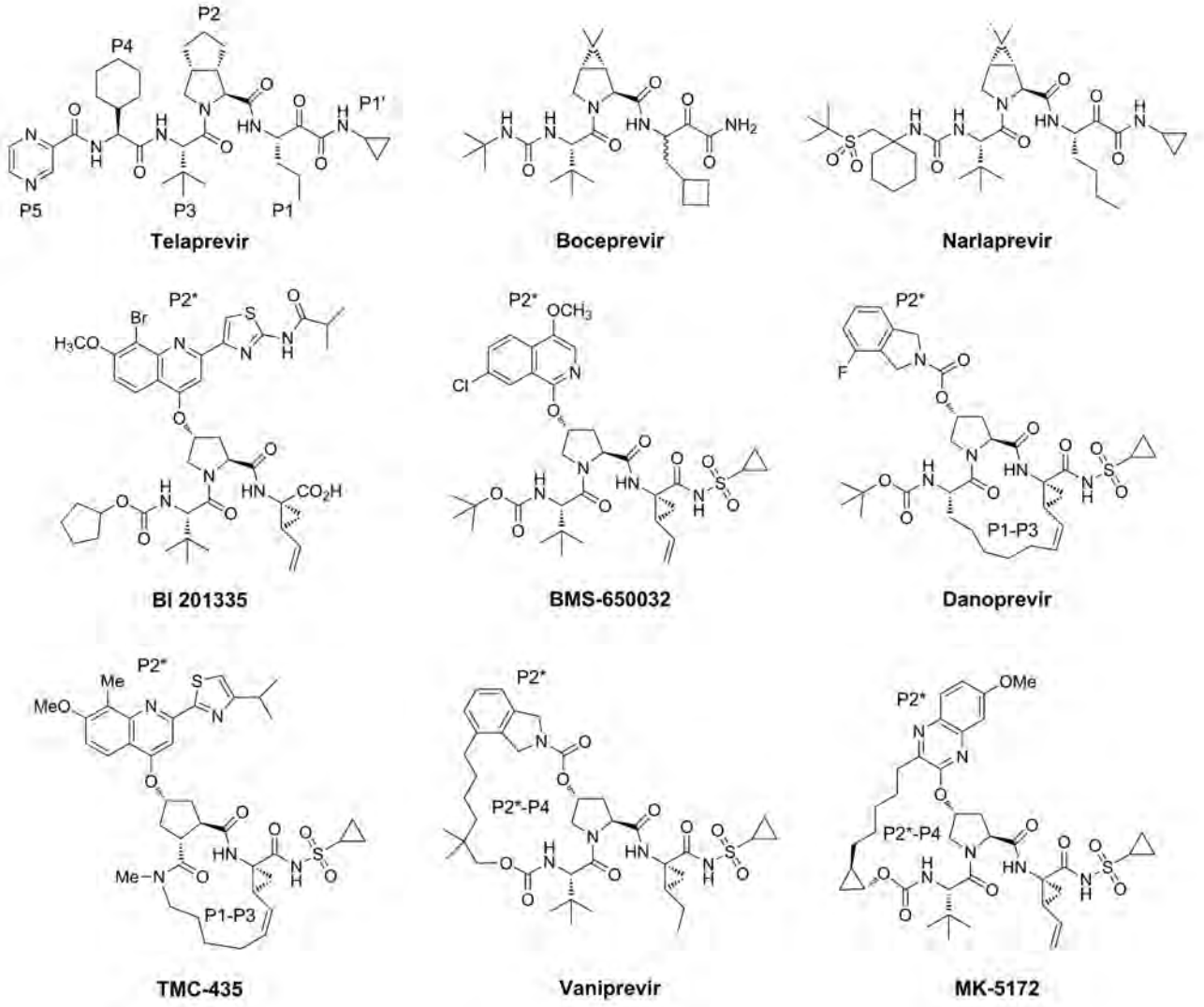

P2
P4
P1'
P5
P3
P1
Telaprevir
Boceprevir
Narlaprevir
P2*
Br
H3CO
BI 201335
P2*
OCH3
Cl
BMS-650032
P2*
F
P1-P3
Danoprevir
P2*
Me
MeO
Me–N
P1-P3
TMC-435
P2*
P2*-P4
Vaniprevir
P2*
OMe
P2*-P4
MK-5172

Figure 4.1

likelihood of resistance, treatment regimens utilize combination therapies that limit viral evolution through several distinct mechanisms simultaneously. However, combination therapy with ribavirin and pegylated interferon-α is only effective for a subset of patients, depending on both human and viral genetic factors (*15-19*). Sustained treatment adherence also remains a critical predictor for treatment outcome, as suboptimal dosing allows for the selection of drug resistant viral variants (*20*). Adverse drug effects, including flu-like symptoms, depression and anemia, are commonly reported and often hinder the successful completion of full-dose treatment regimens (*20, 21*). Though high treatment completion rates are reported in clinical trials, these patients are routinely screened prior to enrollment and actively supported throughout therapy to maximize adherence (*22*). However, several small cohort studies report completion rates as low as 50% in the clinical setting, highlighting the impact of disease co-morbidities, socioeconomic status, emotional support and related factors on treatment outcome (*23, 24*). Although new antiviral agents are becoming available, dominant polymerase and protease inhibitor resistance mutations have been detected in treatment naïve individuals, indicating that drug resistant strains may be directly transmissible(*25*). Thus despite the benefits of combination therapy in limiting viral evolution, the emergence of drug resistance still challenges the long-term efficacy of antiviral therapies in the clinic.

The rapid rise of resistance has reduced the efficacy of the most potent NS3/4A protease inhibitors in human clinical trials. Most notably, single site mutations at residues R155, A156 and D168 confer multi-drug resistance (*26-31*). In this study, four chemically diverse protease inhibitors – telaprevir, danoprevir, vaniprevir and MK-5172

– exhibit distinct susceptibilities to the protease variants R155K, D168A and A156T **(Table 4.1)**. Sixteen high-resolution crystal structures of the wildtype protease and these drug resistant variants reveal the molecular basis underlying the resistance profiles of these four drugs **(Table 4.2)**. The A156T mutation clashes with all four drugs, although flexibility in danoprevir accommodates this clash. The R155K mutation confers telaprevir resistance by reducing drug packing with this protease variant. The R155K and D168A mutations also confer resistance to danoprevir and vaniprevir by disrupting favorable stacking interactions. In contrast, MK-5172 packs against the catalytic triad in a novel conformation, explaining its retained potency against R155K and D168A. Taken together, the differential drug susceptibilities to these major mutations arise from unique drug interactions, highlighting direct strategies for designing novel protease inhibitors that are highly potent but less susceptible to resistance.

Table 4.1. NS3/4A protease inhibitor activities from enzymatic inhibition experiments and drug susceptibility assays against wildtype protease and variants R155K, D168A and A156T.

	K_i (nM)	Drug Susceptibilities - IC_{50}† (nM)			
Drug	**WT**	**WT**	**R155K**	**D168A**	**A156T**
Telaprevir	66 ± 3	1030	5300 **(5.1)**	420 **(0.4)**	>50,000 **(>49)**
Danoprevir	0.46 ± 0.03	0.24	>100 **(>416)**	48 **(200)**	5.7 **(24)**
Vaniprevir	0.13 ± 0.02	0.34	>400 **(>1176)**	>400 **(>1176)**	176 **(518)**
MK-5172	0.0024 ± 0.005	0.11	0.55 **(5)**	13 **(118)**	108 **(982)**

† – Numbers in parentheses reflect fold-change relative to wildtype.

Table 4.2. X-ray data collection and crystallographic refinement statistics.

Drug	**Telaprevir**				**Danoprevir**				**Vaniprevir**				**MK-5172**			
Protease variant	WT	R155K	D168A	A156T	WT^{i}	R155K^{i}	D168A^{i}	A156T^{i}	WT^{i}	R155K^{i}	D168A^{i}	A156T^{i}	WT^{t}	R155K^{t}	D168A^{t}	A156T^{i}
PDB ID	3SV6	3SV7	3SV8	3SV9	3M5L	3SU0	3SU1	3SU2	3SU3	3SU4	3SU5	3SU6	3SUD	3SUE	3SUF	3SUG
Resolution (Å)	1.40	1.55	2.50	1.60	1.25	1.16	1.40	1.50	1.30	2.25	1.55	1.10	1.96	2.20	2.30	1.80
Space group	$P2_12_12_1$	$P2_12_12_1$	$P4_12_12$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P6_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_1$	$P2_1$	$P2_1$	$P2_12_12_1$
Twin Law	-	-	-	-	-	-	-	-	-	-	-	-	-h, -k, h+l	-h, -k, h+l	-h, -k, h+l	-
Twin Fraction	-	-	-	-	-	-	-	-	-	-	-	-	0.42	0.42	0.43	-
Molecules in AU^{a}	1	1	1	1	1	1	1	1	1	2	1	1	4	4	4	1
Cell dimensions:																
a (Å)=	55.1	55.3	69.5	54.8	55.2	55.3	55.0	54.9	55.1	85.8	55.1	55.0	56.1	56.3	56.0	53.9
b (Å)=	58.8	58.8	69.5	58.7	58.7	58.5	58.5	58.5	58.5	85.8	58.8	58.5	102.7	103.3	103.6	58.2
c (Å)=	60.3	60.4	79.1	60.7	61.1	60.6	60.0	60.0	60.3	97.4	60.0	59.8	73.3	73.5	73.5	62.0
β (°)=	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	90.0	120.0	90.0	90.0	112.5	112.6	112.0	90.0
Completeness (%)	99.9	99.6	99.8	99.8	99.7	97.9	100.0	100.0	95.7	99.9	95.8	96.6	91.6	91.6	95.9	90.1
Total reflections	39340	183407	91636	147126	317137	394396	233176	228689	196059	152278	190711	400120	101245	77906	107542	62460
Unique reflection	218617	29054	7137	26564	52645	67429	38857	31822	46251	19213	27508	76300	50198	36194	37671	16890
Average I/σ	13.8	11.1	7.5	12.0	20.7	11.2	9.8	10.0	12.0	7.2	8.8	12.1	10.0	9.5	9.4	19.0
Redundancy	5.6	6.3	12.8	5.5	6.0	5.8	6.0	7.2	4.2	7.9	6.9	5.2	2.0	2.2	2.9	3.7
R_{sym} (%)^{b}	3.5	6.1	11.0	4.5	4.6	4.1	4.3	4.3	4.1	7.7	6.9	4.4	6.0	6.7	6.3	2.8
RMSD^{c} in:																
Bonds (Å) =	0.009	0.009	0.012	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009	0.009
Angles (°) =	1.32	1.30	1.47	1.31	1.48	1.32	1.37	1.40	1.44	1.26	1.39	1.36	1.33	1.31	1.36	1.40
R_{factor} (%)^{d} =	15.8	16.5	20.5	17.2	15.0	15.3	15.7	15.4	16.4	16.9	16.5	15.0	18.3	18.5	19.7	19.4
R_{free} (%)^{d} =	17.1	19.9	28.3	19.7	16.8	17.2	17.8	17.7	18.2	22.3	18.3	16.5	23.8	22.7	25.6	23.0

i, S139A protease mutant used for crystallization.
t, pseudo-merohedral twin.
a, AU, asymmetric unit.
b, $R_{sym} = \Sigma|I - \langle I \rangle|/\Sigma I$, where I = observed intensity, <I> = average intensity over symmetry equivalent.
c, RMSD, root mean square deviation.
d, $R_{work} = \Sigma||F_o| - |F_c||/\Sigma|F_o|$. R_{free} was calculated from 5% of reflections, chosen randomly, which were omitted from the refinement process.

MATERIALS AND METHODS

Reagent synthesis, purchase and storage

Danoprevir was synthesized using the convergent reaction scheme described previously (*32*). Vaniprevir and MK-5172 were prepared following the previously published procedures (*33, 34*). One hundred milligrams of telaprevir were purchased from A ChemTek, Inc. (Worcester, MA). All drugs were reconstituted in DMF to a final concentration of 30mM and placed at −20ºC for storage. Thirty milligrams of the previously reported peptide inhibitor (DE-Dif-I-Cha-C) (*35*) were purchased from 21st Century Biochemicals (Marlboro, MA) at 95% purity with 5-carboxyfluorescein covalently bonded to the N-terminus. The purity was confirmed by HPLC and mass spectrometry profiles provided by the company. The fluorescent peptide probe was stored as a solid at −20ºC and dissolved in reaction buffer to a final concentration of 5–15nM for competition experiments.

Mutagenesis and gene information

The HCV genotype 1a NS3/4A protease gene described in a Bristol-Meyers Squibb patent (*36*) was synthesized by GenScript and cloned into the pET28a expression vector (Novagen). The gene encodes a highly soluble form of the NS3/4A protease domain as a single-chain, with 11 core amino acids of NS4A located at the N-terminus. The protease variants S139A, R155K, D168A and A156T were subsequently constructed using the QuikChange site-directed mutagenesis kit from Stratagene and sequenced by Davis Sequencing for confirmation.

Expression and purification of NS3/4A protease constructs

Protein expression and purification were carried out as described previously (*36, 37*). Briefly, transformed BL21(DE3) *E. coli* cells were grown at 37ºC and induced at an optical density of 0.6 by the addition of 1mM IPTG. Cells were harvested after 5 hours of expression, pelleted and frozen at −80ºC for storage. Cell pellets were thawed, resuspended in 5mL/g of resuspension buffer (50mM phosphate buffer at pH 7.5, 500mM NaCl, 10% glycerol, 2mM β-ME) and lysed with a cell disruptor. The soluble fraction was retained, applied to a nickel column (Qiagen), washed with resuspension buffer, and eluted with resuspension buffer supplemented with 200mM imidazole. The eluant was dialyzed overnight (MWCO 10kD) to remove the imidazole and the his-tag was simultaneously removed with thrombin treatment. The nickel-purified protein was then flash-frozen and stored at −80ºC for up to six months.

Crystallization of inhibitor complexes

For crystallization, the protein solution was thawed, concentrated to ~3mg/mL and loaded on a HiLoad Superdex75 16/60 column equilibrated with gel filtration buffer (25mM MES at pH 6.5, 500mM NaCl, 10% glycerol, 30μM zinc chloride, and 2mM DTT). The protease fractions were pooled and concentrated to 20–25mg/mL with an Amicon Ultra-15 10kD device (Millipore). The concentrated samples were incubated for one hour with 1–3 molar excess of inhibitor. Diffraction-quality crystals were obtained overnight by mixing equal volume of concentrated protein solution with precipitant solution (20–26% PEG-3350, 0.1M sodium MES buffer at pH 6.5, and 4% ammonium sulfate) in 24-well VDX hanging drop trays.

Data collection and structure solution

Crystals large enough for data collection were flash-frozen in liquid nitrogen for storage. Crystals were mounted under constant cryostream and X-ray diffraction data were collected at Advanced Photon Source LS-CAT 21-ID-F, GM/CA-CAT 23-ID-D or our in-house RAXIS IV X-ray system. Diffraction intensities were indexed, integrated and scaled using the program HKL2000 (*38*). All structure solutions were generated using simple isomorphous molecular replacement with PHASER (*39*). The B chain model of viral substrate product 4A4B (3M5M) (*32*) was used as the starting model for all structure solutions. Initial refinement was carried out in the absence of modeled ligand, which was subsequently built in during later stages of refinement. Upon obtaining the correct molecular replacement solutions, subsequent crystallographic refinement was carried out within the CCP4 program suite with iterative rounds of TLS and restrained refinement until convergence was achieved (*40*). Amplitude-based twinned refinement was carried out for all pseudo-merohedral twins (crystal complexes with MK-5172 with the exception of A156T). The final structures were evaluated with MolProbity (*41*) prior to deposition in the protein data bank. 5% of the data was reserved for the free R-value calculation to limit the possibility of model bias throughout the refinement process (*42*). Interactive model building and electron density viewing was carried out using the program COOT (*43*).

Inhibitor complex analysis and the substrate envelope

All active site alignments were performed in PyMOL using the Cα atoms of protease residues 137–139 and 154–160. For each alignment, the wildtype-danoprevir

complex was used as the reference structure. The viral substrate envelope – representing the consensus vdW volume shared by any three of the four viral substrate products – was computed as described previously using the full-length NS3/4A structure (1CU1) (*44*) and product complexes 4A4B (3M5M), 4B5A (3M5N) and 5A5B (3M5O) (*32*).

Drug susceptibility assays

Single mutations (R155K, D168A, or A156T) were introduced into the NS3 region of an HCV Con1 luciferase reporter replicon using the mega-primer method of mutagenesis(*45*). Replicon RNA of each protease variant was introduced into Huh7 cells by electroporation. Replication was then assessed in the presence of increasing concentration of protease inhibitors (telaprevir, danoprevir, vaniprevir or MK-5172) by measuring luciferase activity (relative light units) 96 hours after electroporation. The drug concentrations required to inhibit replicon replication by 50% (IC_{50}) were calculated directly from the drug inhibition curves.

Enzymatic inhibition assays

For enzymatic inhibition experiments, 5nM of the NS3/4A protease domain or full-length protein were incubated with increasing amounts of drug for 15 minutes (90 minutes for telaprevir) in 25mM Tris (pH 7.5), 2.5% glycerol, 5mM TCEP, 0.1% OβG and 2% DMSO. Proteolysis reactions were initiated by the addition of 200nM of fluorogenic substrate RET S1(*46*) (AnaSpec) and monitored using the EnVision plate reader (Perkin Elmer) at excitation and emission wavelengths of 340nm and 490nm, respectively. The initial cleavage velocities were determined from sections of the progress curves corresponding to less than 15% substrate cleavage. Drug inhibition

constants (K_i) were obtained by nonlinear regression fitting to the Morrison equation of initial velocity versus inhibitor concentration using Prism (GraphPad Software). Data for each drug were generated in triplicate and processed independently to calculate the average inhibition constant and standard deviation.

VdW contact energy

VdW contact energies between protease residues and peptide products were computed using a simplified Lennard-Jones potential as described previously(*47*). Briefly, the Lennard-Jones potential (V_r) was calculated for each protease-drug atom pair where r, ε and σ represent the interatomic distance, vdW well depth and atomic diameter, respectively:

$$V_r = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

V_r was computed for all possible protease-drug atom pairs within 5Å, and potentials for non-bonded pairs separated by less than the distance at the minimum potential were equated to −ε. Using this simplified potential value for each non-bonded protease-product atom pair, the total vdW contact energy was computed for each peptide residue. For graphical convenience, VdW energy indexes were then calculated by multiplying the raw values by a factor of −10.

RESULTS

The Substrate Envelope

The wildtype drug complexes were analyzed with the substrate envelope to characterize drug interactions relative to natural substrate products **(Figures 4.2A–4.5A)**. The substrate envelope was constructed as reported previously and represented the volume shared by any three of the four N-terminal viral substrate products (*32*). Superposition of the drug complexes revealed that the four inhibitors protruded from the substrate envelope in a region termed the P2 subsite, packing against residues R155, A156 and D168. The relatively small P2 cyclopentylproline moiety of telaprevir interacted less extensively with these residues relative to the other drugs, which contain bulkier $P2^*$ moieties. The carbamate-linked $P2^*$ isoindoline moieties of danoprevir and vaniprevir formed cation-π stacking interactions with R155, and also interacted extensively with A156 and D168. The isoindoline moiety of danoprevir bound in two conformations to the wildtype protease, with its fluorine atom oriented either towards the catalytic triad or the P2 subsite. The $P2^*$ moiety of vaniprevir adopted just one conformation and formed greater interactions with D168 due the positioning of its C4 methylene linker in the P2–P4 macrocycle. The longer C5 methylene linker in the P2–P4 macrocycle of MK-5172 extended away and thus contributed less van der Waals (vdW) contact with D168. Notably, the ether-linked $P2^*$ quinoxaline moiety of MK-5172 protruded uniquely from the substrate envelope towards the catalytic triad, avoiding direct contact with R155 and D168. Thus these drugs make different interactions with the protease outside the substrate envelope, and molecular changes in these regions can

Figure 4.2: Stereo view of the telaprevir complexes. (A) Telaprevir is shown bound to the wildtype protease with the substrate envelope in blue. Telaprevir is also shown bound to the drug-resistant variants (B) R155K, (C) D168A and (D) A156T with the transparent coordinates representing the wildtype structure to better highlight the molecular changes of each mutation. In all cases, catalytic residues are depicted in yellow, the P2 subsite in pink, and the drug molecules in orange.

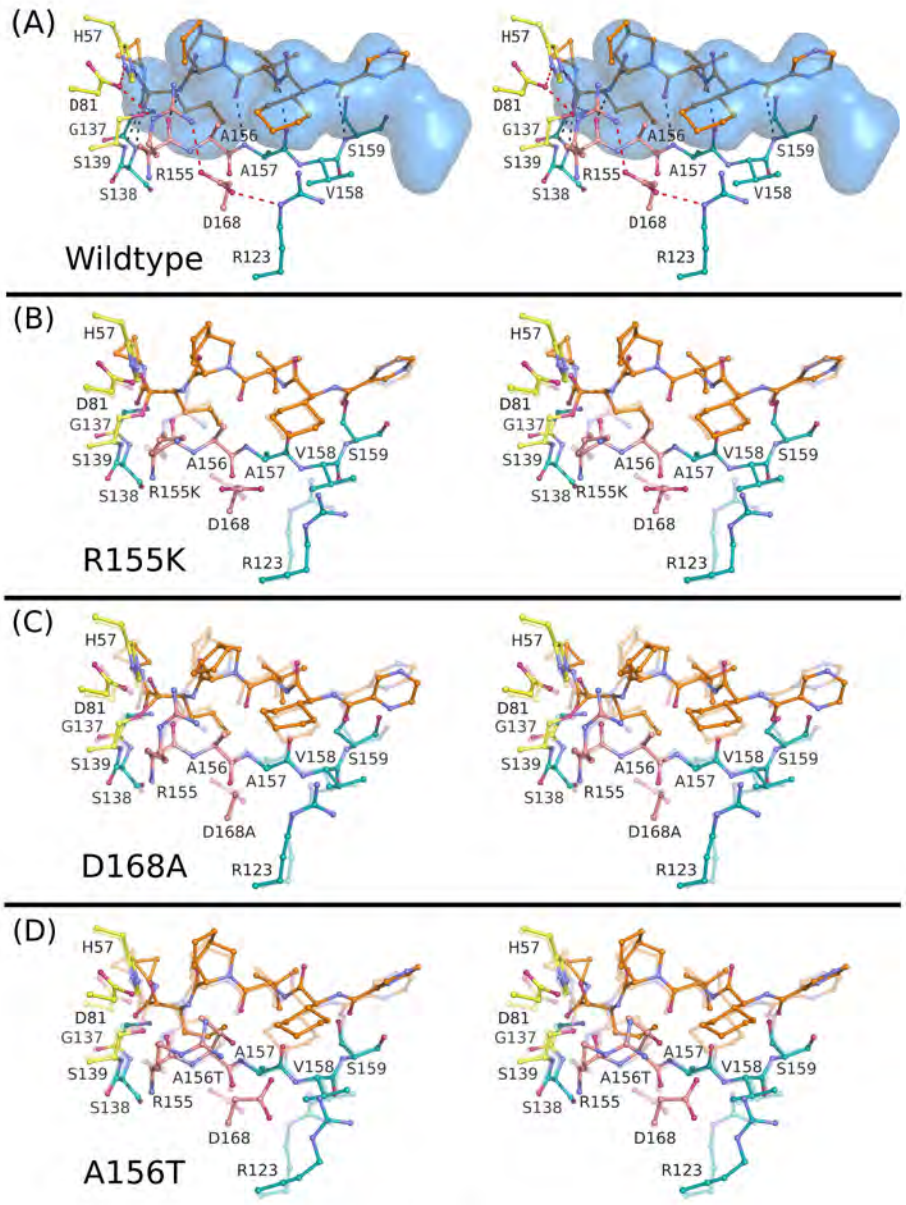

(A)
H57
D81
G137
S139
S138
R155
A156
A157
S159
V158
D168
R123
Wildtype
(B)
R155K
(C)
D168A
(D)
A156T

Figure 4.2

facilitate resistance by selectively disrupting drug binding without substantially affecting the recognition and processing of authentic viral substrates.

Wildtype Drug Binding

Danoprevir, vaniprevir and MK-5172 bound with similar affinities to the wildtype protease, while telaprevir binding was 15–20 fold weaker **(Table 4.1)**. Many structural features underlying these affinities were common for all drug complexes. For example, the four drugs made the same three hydrogen bonds with the protease backbone **(Table 4.3)**: *(i)* the P1 amide nitrogen with the carbonyl oxygen of R155, *(ii)* the P3 carbonyl oxygen with the amide nitrogen of A157, and *(iii)* the P3 amide nitrogen with the carbonyl oxygen of A157. The P5 amide nitrogen of telaprevir formed an additional hydrogen bond with the carbonyl oxygen of S159. In the telaprevir complex, the ketoamide oxygen sat in the oxyanion hole and interacted with the backbone amide nitrogens of protease residues 137–139, while the Nε nitrogen of H57 hydrogen bonded with the keto oxygen **(Figure 4.2A)**. The sulfonamide groups of danoprevir, vaniprevir and MK-5172 were also positioned in the oxyanion hole, hydrogen-bonding with the same set of backbone nitrogens as the macrocyclic drugs. Meanwhile the Nε nitrogen of H57 interacted with the sulfonamide nitrogen in these complexes, suggesting that the Nε atoms were deprotonated in these complexes **(Figure 4.3A–4.5A)**.

Mechanisms of Drug Resistance

Drug susceptibility data and high-resolution crystal structures were utilized to analyze the sixteen combinations of wildtype and multi-drug resistant protease variants with the four protease inhibitors: telaprevir, danoprevir, vaniprevir and MK-5172.

Table 4.3. Drug hydrogen bonds (Hbonds) and vdW contacts with wildtype protease.

Moiety	*Telaprevir*		*Danoprevir*		*Vaniprevir*		*MK-5172*	
	Hbonds	***vdW***	***Hbonds***	***vdW***	***Hbonds***	***vdW***	***Hbonds***	***vdW***
P1'	G137 S138 S139	Q41 T42 F43 K136	G137 S138 S139A	Q41 T42 F43 K136	G137 S138 S139A	Q41 T42 F43 K136	G137 S138 S139A	Q41 T42 F43 K136
P1	H57 R155	F154	H57 R155	F154	H57 R155	F154	H57 R155	F154
P2		R155 A156 D168		R155 A156 D168 D79 K80		R155 A156 D168 D79 K80		Y56 H57 D81 R155 A156
P3	A157†	I132 L135 A157	A157†	I132 L135 A157	A157	I132 L135 A157	A157	I132 L135 A157
P4		R123 A156 V158 D168		A156 D168		R123 D168		R123 V158
P5		S159	∂	∂	∂	∂	∂	∂

†- Contributes two hydrogen bonds
∂- Inhibitors do not contain P5 moiety

Figure 4.3: Stereo view of the danoprevir complexes. (A) Danoprevir is shown bound to the wildtype protease with the substrate envelope in blue. Danoprevir is also shown bound to the drug-resistant variants (B) R155K, (C) D168A and (D) A156T with the transparent coordinates representing the wildtype structure to better highlight the molecular changes of each mutation. In all cases, catalytic residues are depicted in yellow, the P2 subsite in pink, and the drug molecules in orange.

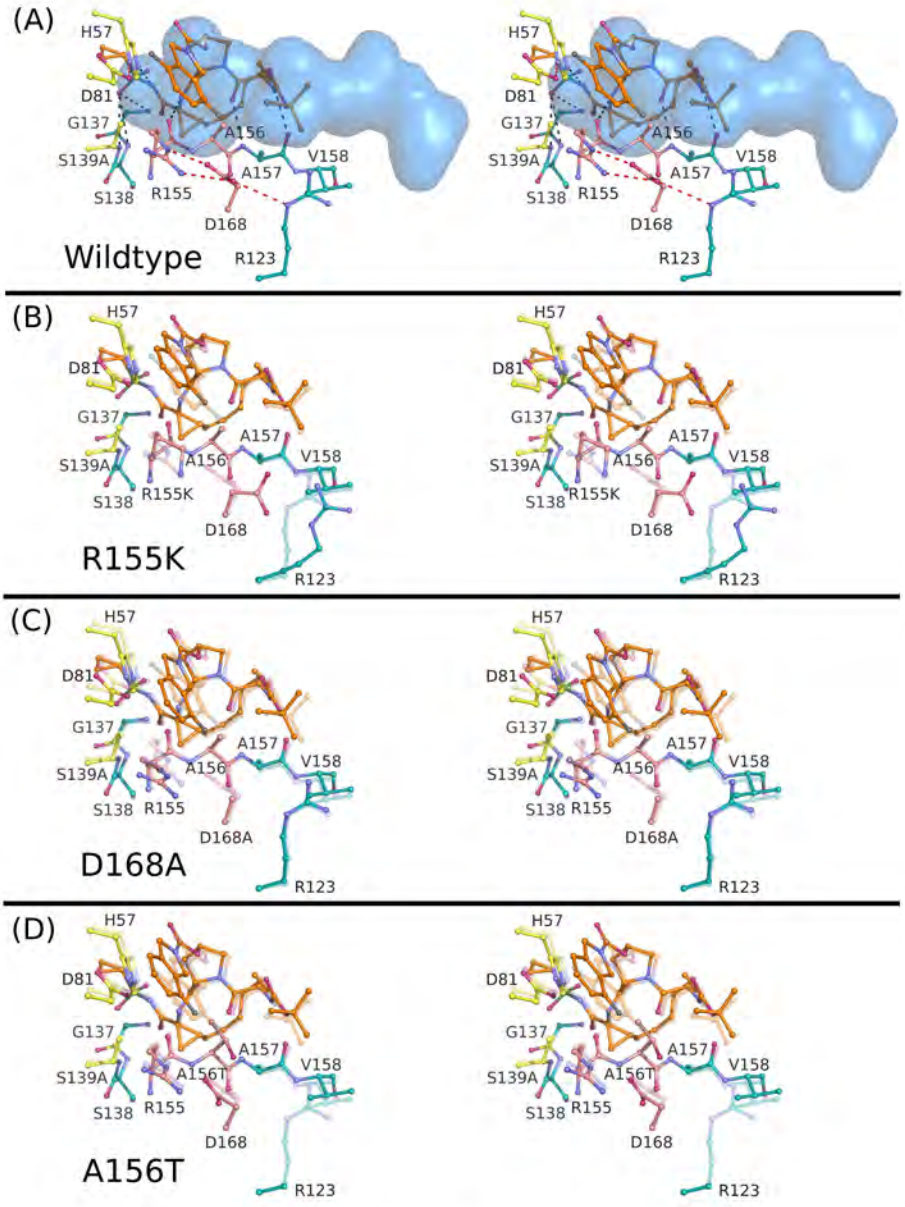

(A)
H57
D81
G137
S139A
S138
R155
A156
A157
V158
D168
R123
Wildtype
(B)
R155K
(C)
D168A
(D)
A156T

Figure 4.3

- *Telaprevir resistance*

Telaprevir bound to the R155K variant in a conformation similar to the wildtype complex **(Figure 4.2B)**. However, this mutation lost considerable vdW contact with telaprevir **(Figure 4.6A)** and also disrupted the electrostatic network spanning R123, D168, R155 and D81, forming along the drug binding surface in the wildtype complex. Specifically, R155K no longer interacted with D168, which rotated slightly to maintain contact with R123. These rearrangements disrupted the charge landscape along this surface of the protease, which in turn disrupted interactions with the P2 cyclopentylproline and P4 cyclohexylalanine moieties of telaprevir. Thus these structural changes disrupt many features observed in the wildtype telaprevir complex, explaining the 5-fold reduction in telaprevir potency against R155K in drug susceptibility assays **(Table 4.1)**.

The D168A mutant complex revealed a complete shift of telaprevir by approximately 0.5Å relative to its wildtype conformation **(Figure 4.2C)**. The P2 cyclopentylproline moiety of telaprevir bent considerably towards D168A, resulting in increased vdW interaction with both R155 and A156. These structural changes also reduced vdW contact with H57, but this catalytic base still maintained its hydrogen bond with the keto oxygen of the ketoamide group. The altered packing of telaprevir increased interactions in the P2 subsite, explaining its enhanced potency against the D168A in drug susceptibility assays **(Table 4.1)**.

The A156T-telaprevir complex revealed in a direct clash between A156T and the P2 cyclopentylproline, causing the position of telaprevir to shift relative to the wildtype

(Figure 4.2D). H57 also shifted to maintain its hydrogen bond with the keto oxygen of the ketoamide group. The P2 cyclopentylproline of telaprevir moved away from R155, losing considerable vdW contact with the drug. In addition, the nucleophilic oxygen of S139 was over 2Å from the keto carbon of the ketoamide group of telaprevir. Although this carbon was modeled in the tetrahedral state, indicating covalent bond formation with the protease, this increased inter-atomic distance suggested that the electron density was the average of the tetrahedral state and the planar state of the unmodified drug form. Thus the A156T mutation reduced the capacity of telaprevir to react covalently with the catalytic S139, contributing considerably to the loss in potency. These structural changes are consistent with the large fold-loss in potency of telaprevir against the A156T variant **(Table 4.1)**.

- *Danoprevir and vaniprevir resistance*

The $P2^*$ moieties of danoprevir and vaniprevir shifted in the R155K complexes relative to the wildtype coordinates, disrupting the favorable cation-π stacking interactions observed in the wildtype complex **(Figure 4.3B)**. Danoprevir bound in only one conformation to the R155K variant with the fluorine of its isoindoline moiety orientated towards the P2 subsite. In both drug complexes, R155K lost considerable vdW interactions with the $P2^*$ isoindoline moieties relative to the wildtype coordinates **(Figure 4.6B)**. Although these changes also caused greater vdW interactions with H57 and D81, these gains did not compensate fully for the disruption of the favorable cation-π stacking between R155 and the aromatic rings of the $P2^*$ moieties. Indeed, these structural observations are consistent with the 400-fold reductions in potency of

danoprevir and vaniprevir against the R155K variant **(Table 4.1)**.

In the D168A complexes, the P2* isoindoline moieties of danoprevir and vaniprevir also shifted towards D81 of the catalytic triad, thereby disrupting the cation-π stacking interactions observed in the wildtype complexes **(Figure 4.4C)**. The 4-fluoroisoindoline moiety of danoprevir also adopted just one conformation in the D168A complex. The D168A mutation also reduced the drug vdW interactions with due to the smaller volume of alanine compared to aspartate. However, these structural changes were minor relative to the large losses in potency caused by this mutation **(Table 4.1)**. Other molecular forces, such as electrostatics, may play crucial roles in the binding of danoprevir and vaniprevir. In particular, the electrostatic interactions between R155 and D168, observed in the wildtype complexes, may stabilize R155 for optimal cation-π stacking with the P2* isoindoline moieties. Thus the D168A mutation may disrupt these interactions without substantially altering the bound drug conformations **(Table 4.1)**.

In both A156T complexes, the P2* moieties of danoprevir and vaniprevir shifted away from the P2 subsite and toward the catalytic triad. The shifted positioning of the P2* moieties directly disrupted the cation-π stacking interactions with R155, observed in the wildtype complexes. The Oγ oxygens of A156T also formed hydrogen bonds with D168 in both drug complexes, which by competing for interaction with D168, may indirectly disrupt the cationic of R155 necessary for optimal stacking with the P2* isoindoline moieties. As in the other mutant complexes, the 4-fluoroisoindoline moiety of danoprevir adopted just one conformation in the A156T complex. However, the flexibility of the P2* moiety of danoprevir allowed for its large shift and retained packing

Figure 4.4: Stereo view of the vaniprevir complexes. (A) Vaniprevir is shown bound to the wildtype protease with the substrate envelope in blue. Vaniprevir is shown bound to the drug-resistant variants (B) R155K, (C) D168A and (D) A156T with the transparent coordinates representing the wildtype structure to better highlight the molecular changes of each mutation. In all cases, catalytic residues are depicted in yellow, the P2 subsite in pink, and the drug molecules in orange.

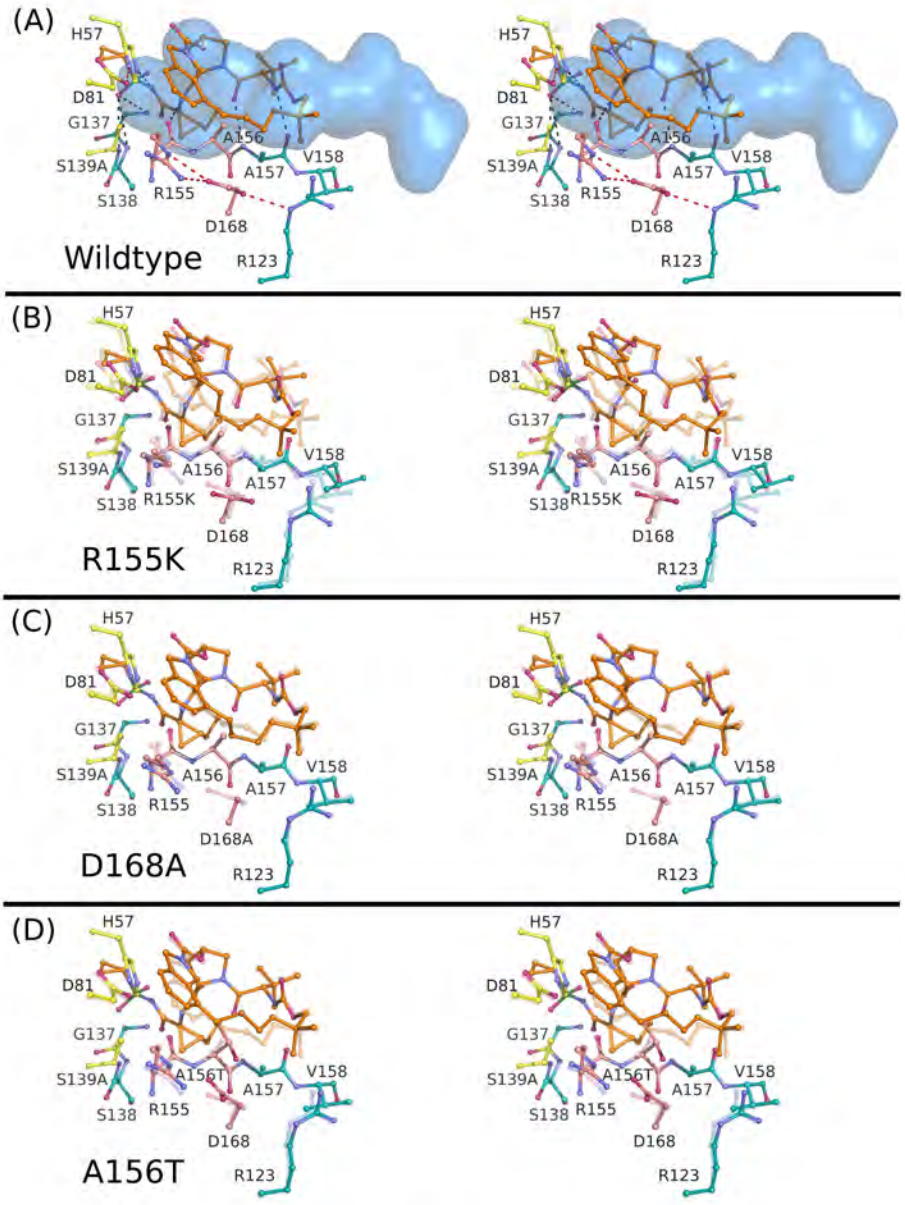

(A)
H57
D81
G137
S139A
S138
R155
A156
A157
V158
D168
R123
Wildtype
(B)
R155K
(C)
D168A
(D)
A156T

Figure 4.4

against A156T, partially compensating for the steric burden of this mutation **(Figure 4.3D)**. In contrast, the P2–P4 macrocycle of vaniprevir restrained its ability to accommodate this steric burden, consistent with the smaller shift of its $P2^*$ moiety relative to that of danoprevir. Thus the rigidity of the $P2^*$ moiety of vaniprevir likely minimizes its compensatory potential, explaining its 500-fold loss in potency against the A156T variant, while danoprevir remained highly active against this protease variant **(Table 4.1)**.

- *MK-5172 resistance*

R155 and D168 did not interact extensively with the $P2^*$ quinoxaline moiety of MK-5172 in the wildtype complex. Thus single mutations at these residues, R155K or D168A, did not substantially alter drug binding **(Figures 4.5B and 4.5C)**. In fact, R155K contributed more vdW contact with MK-5172 relative to R155 in the wildtype complex **(Figure 4.6D)**. The less bulky R155K also allowed for a slight shift of MK-5172 toward the catalytic triad, resulting in increased vdW contact of the $P2^*$ moiety with A156, D81 and H57. Similar changes were observed in the D168A complex, although this mutation also led to an extended conformation of R123 in this case, losing vdW contact with the P4 cyclopropyl moiety of MK-5172. The slight loss in potency of MK-5172 against D168A likely arose from this minor loss in contact energy. Nevertheless, these structural changes explain the high potency of MK-5172 against the R155K and D168A protease variants in drug susceptibility assays **(Table 4.1)**.

The A156T mutation resulted in a steric clash with the benzylic methylene of the P2–P4 macrocycle of MK-5172, causing the drug to shift considerably relative to the

Figure 4.5: Stereo view of the MK-5172 complexes. (A) MK-5172 is shown bound to the wildtype protease with the substrate envelope in blue. MK-5172 is shown bound to the drug-resistant variants (B) R155K, (C) D168A and (D) A156T with the transparent coordinates representing the wildtype structure to better highlight the molecular changes of each mutation. In all cases, catalytic residues are depicted in yellow, the P2 subsite in pink, and the drug molecules in orange.

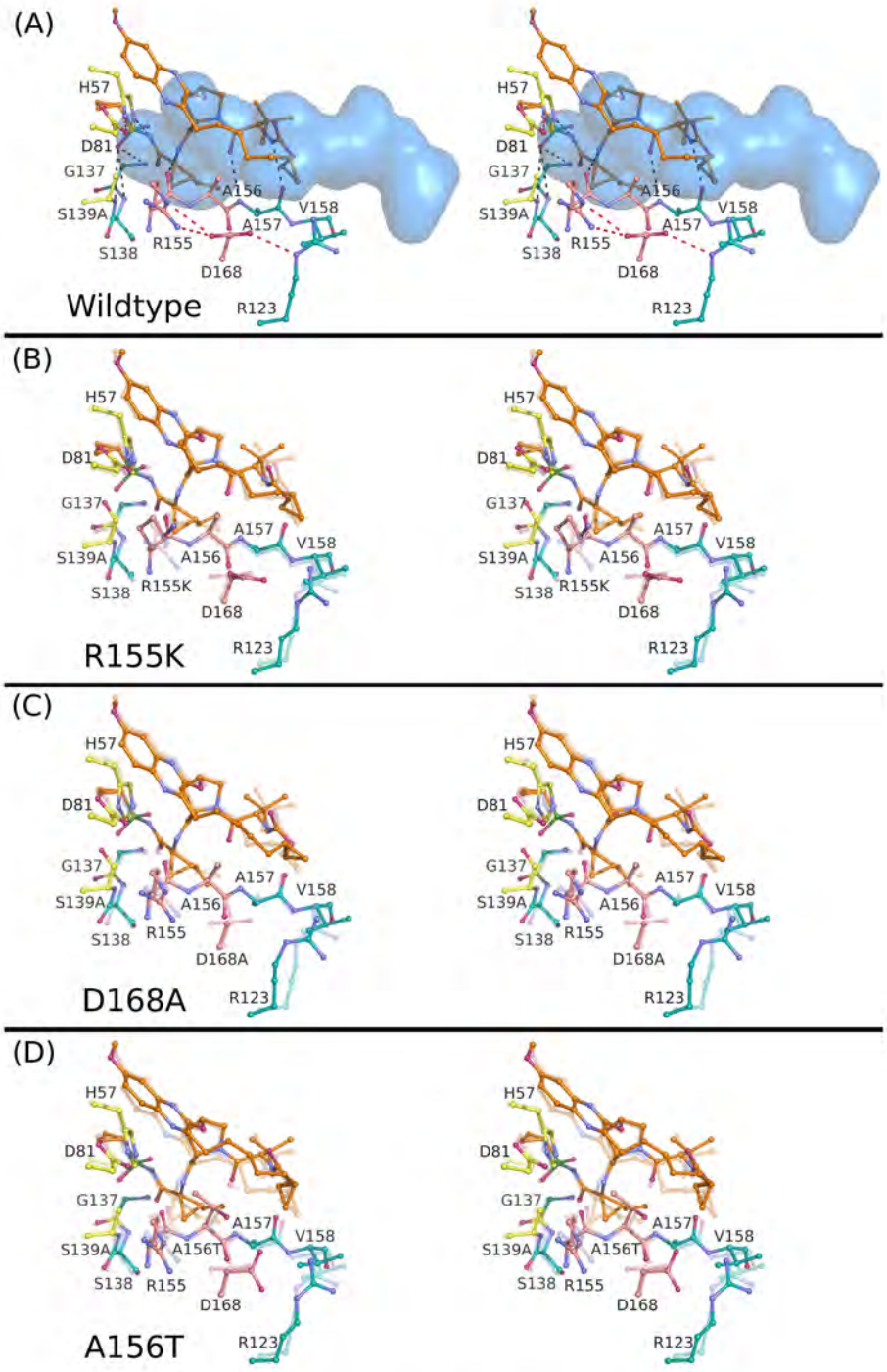

(A)
H57
D81
G137
S139A
S138
R155
A156
A157
V158
D168
R123
Wildtype
(B)
R155K
(C)
D168A
(D)
A156T

Figure 4.5

wildtype coordinates **(Figure 4.5D)**. H57 also shifted by the same magnitude to maintain its hydrogen bond with the amide nitrogen of the sulfonamide group. Despite increased vdW contact with A156T (due to steric volume), these changes also caused losses in vdW contact with D81 and R155. Together these structural changes explain the 50-fold loss in potency of MK-5172 against the A156T **(Table 4.1)**.

Figure 4.6: Drug vdW interactions in wildtype and mutant complexes. The vdW energy indexes for the wildtype protease and mutant variants R155K, D168A and A156T are shown by residue for (A) telaprevir, (B) danoprevir, (C) vaniprevir and (D) MK-5172.

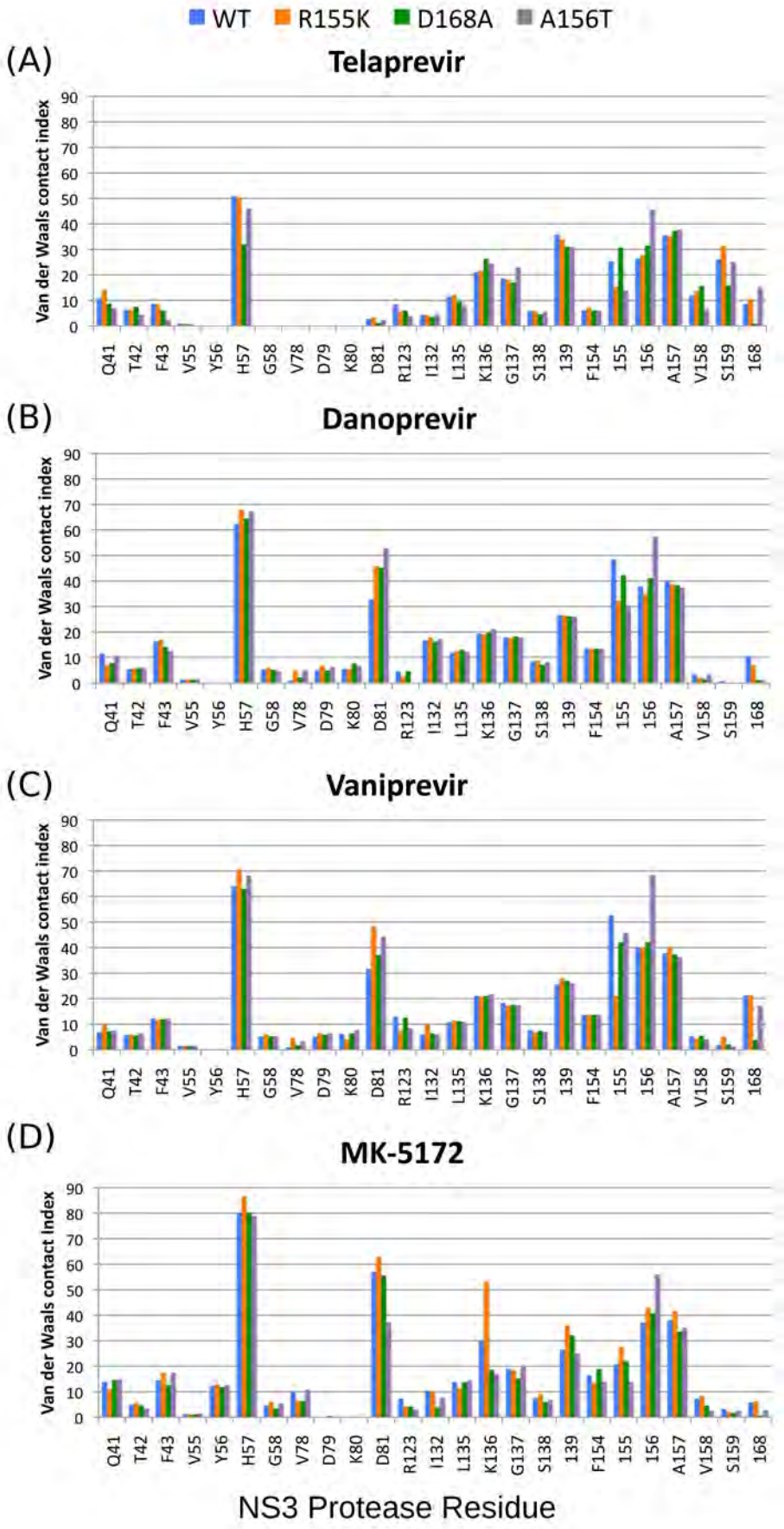

WT
R155K
D168A
A156T
(A)
Telaprevir
(B)
Danoprevir
(C)
Vaniprevir
(D)
MK-5172
Van der Waals contact index
Van der Waals contact Index
0
10
20
30
40
50
60
70
80
90
Q41
T42
F43
V55
Y56
H57
G58
V78
D79
K80
D81
R123
I132
L135
K136
G137
S138
139
F154
155
156
A157
V158
S159
168
NS3 Protease Residue

Figure 4.6

DISCUSSION

Drug resistance challenges the long-term effectiveness of anti-HCV therapies, highlighting the need for novel inhibitors that are less susceptible to resistance. Despite considerable advances in the development of NS3/4A protease inhibitors, the rapid rise of resistance has limited the efficacy of many potent drugs in this class (*26-31*). By determining sixteen high-resolution crystal structures, we have elucidated the molecular basis of drug resistance against four chemically diverse protease inhibitors: telaprevir, danoprevir, vaniprevir and MK-5172. These drugs exhibit unique patterns of resistance against the multi-drug resistant variants R155K, D168A and A156T. The A156T mutation substantially reduces the potencies of telaprevir, vaniprevir and MK-5172 through steric clashes. In contrast, the mobility of the $P2^*$ isoindoline moiety of danoprevir – not restrained by a P2–P4 macrocycle – allows its retained potency against this variant. The R155K and D168A mutations also confer resistance to danoprevir and vaniprevir by disrupting favorable cation-π stacking interactions of R155 with their $P2^*$ isoindoline moieties. Telaprevir displays enhanced potency against D168A through increased packing in the P2 subsite, but loses potency against R155K through decreased packing. In contrast to the other drugs, the $P2^*$ quinoxaline moiety of MK-5172 adopts in a unique conformation and interacts with H57 and D81 of the catalytic triad, rather than residues R155 and D168. This novel mode of MK-5172 binding explains its robust activity against the R155K and D168A variants, as MK-5172 potency derives from contacting residues of the catalytic triad that are essential for HCV NS3/4A protease function.

Our results also provide predictions of drug activities against other genotypes and strains. Genotype 3 viruses contain the NS3 protease residues Q168 and T123, instead of D168 and R123 found in other genotypes. We predict that the terminal amide group of Q168 will be unable to stabilize R155 for stacking against the $P2^*$ moieties of danoprevir and vaniprevir, but may interact with T123 instead. Thus we expect danoprevir and vaniprevir will exhibit reduced potencies against genotype 3 viruses, while MK-5172 will remain fully active. In addition, for genotype 1 strains, our results indicate: *(i)* MK-5172 is active against R155K variants, *(ii)* telaprevir and MK-5172 are active against D168A variants, and *(iii)* danoprevir remains highly potent against A156T variants. Thus as new inhibitors are developed and HCV resistance testing becomes more available, our findings can help guide anti-HCV treatment regimens for individual patients in the future.

For drug resistance to occur against HCV protease inhibitors, mutations must selectively reduce drug binding without disrupting the recognition and cleavage of the natural viral substrates. We show that most primary drug resistance mutations occur in regions of the protease outside the space normally occupied by natural viral substrates, defined as the substrate envelope. Mutations in these regions of the protease selectively weaken drug binding without substantially compromising the recognition and cleavage of viral substrates. Moreover, we demonstrate that inhibitors protruding from the substrate envelope must contact sites that are critical for catalytic activity. The high potency of MK-5172, for example, derives from interactions with the NS3/4A catalytic triad, which cannot mutate without disrupting protease activity, and thus viral replication. Future drug candidates that incorporate ether-linked $P2^*$ quinoxaline moieties, or related functional

groups that pack on the catalytic triad, will likely retain potency against R155K and D168A variants. In addition, drugs that incorporate adaptable $P2^*$ moieties, as in danoprevir, will better retain potency against the A156T mutation or other sterically bulky mutations. Thus utilizing the substrate envelope, in combination with targeting the catalytic triad and incorporating flexibility in the $P2^*$ moieties, provides a focused approach for designing robust NS3/4A inhibitors that retain potency against a wide spectrum of drug resistant viral strains.

Modern drug design strategies often utilize high-throughput screening and along with iterative analyses of the relationship of chemical structure to activity. As these techniques focus primarily on disrupting target activity, drug potencies often derive from contacting residues that are not essential for target activity, which can mutate to confer resistance without altering biological function. Indeed, the emergence of drug resistance is particularly concerning when treating HCV in the clinical setting, as drug side effects, disease co-morbidities, and many other factors limit patient adherence to antiviral therapy. These high-resolution data, in conjunction with the substrate envelope, provide a powerful tool for the more rapid and rational evaluation of drug candidates in the face of drug resistance.

ACKNOWLEDGEMENTS

We thank Nukri Sanishvili of the GM/CA-CAT beamline, Markus Bohn and Andrei Korostelev for data collection of the D168A-MK-5172 complex; David Smith of the LS-CAT beamline for data collection of all other drug complexes; Herbert Klei, Madhavi Kolli and William Royer for helpful discussions; and Shivender Shandilya, Seema Mittal and Madhavi Nalam for their computational support. Use of the Advanced Photon Source was supported by the U.S. Department of Energy, Basic Energy Sciences, Office of Science, under Contract No. DE-AC02-06CH11357. Use of the LS-CAT Sector 21 was supported by the Michigan Economic Development Corporation and the Michigan Technology Tri-Corridor for the support of this research program (Grant 085P1000817). GM/CA-CAT has been funded in whole or in part with Federal funds from the National Cancer Institute (Y1-CO-1020) and the National Institute of General Medical Sciences (Y1-GM-1104). This work was supported by the National Institute of Health grants R01-GM65347 and R01-AI085051.

REFERENCES

1. WHO. (2010) Vaccine Research: Hepatitis C, In *Hepatitis C Virus: Disease Burden* (Barnes, E., Ed.).
2. Brown, R. S. (2005) Hepatitis C and liver transplantation, *Nature 436*, 973-978.
3. Major, M. E., and Feinstone, S. M. (1997) The molecular virology of hepatitis C, *Hepatology 25*, 1527-1538.
4. Lam, A. M., and Frick, D. N. (2006) Hepatitis C virus subgenomic replicon requires an active NS3 RNA helicase, *J Virol 80*, 404-411.
5. Kolykhalov, A. A., Mihalik, K., Feinstone, S. M., and Rice, C. M. (2000) Hepatitis C virus-encoded enzymatic activities and conserved RNA elements in the 3' nontranslated region are essential for virus replication in vivo, *J Virol 74*, 2046-2051.
6. Li, K., Foy, E., Ferreon, J. C., Nakamura, M., Ferreon, A. C., Ikeda, M., Ray, S. C., Gale, M., Jr., and Lemon, S. M. (2005) 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 102*, 2992-2997.
7. Foy, E., Li, K., Wang, C., Sumpter, R., Jr., Ikeda, M., Lemon, S. M., and Gale, M., Jr. (2003) Regulation of interferon regulatory factor-3 by the hepatitis C virus serine protease, *Science 300*, 1145-1148.
8. Lamarre, D., Anderson, P. C., Bailey, M., Beaulieu, P., Bolger, G., Bonneau, P., Bos, M., Cameron, D. R., Cartier, M., Cordingley, M. G., Faucher, A. M., Goudreau, N., Kawai, S. H., Kukolj, G., Lagace, L., LaPlante, S. R., Narjes, H., Poupart, M. A., Rancourt, J., Sentjens, R. E., St George, R., Simoneau, B., Steinmann, G., Thibeault, D., Tsantrizos, Y. S., Weldon, S. M., Yong, C. L., and Llinas-Brunet, M. (2003) An NS3 protease inhibitor with antiviral effects in humans infected with hepatitis C virus, *Nature 426*, 186-189.
9. Hinrichsen, H., Benhamou, Y., Wedemeyer, H., Reiser, M., Sentjens, R. E., Calleja, J. L., Forns, X., Erhardt, A., Cronlein, J., Chaves, R. L., Yong, C. L., Nehmiz, G., and Steinmann, G. G. (2004) Short-term antiviral efficacy of BILN 2061, a hepatitis C virus serine protease inhibitor, in hepatitis C genotype 1 patients, *Gastroenterology 127*, 1347-1355.
10. Vanwolleghem, T., Meuleman, P., Libbrecht, L., Roskams, T., De Vos, R., and Leroux-Roels, G. (2007) Ultra-rapid cardiotoxicity of the hepatitis C virus protease inhibitor BILN 2061 in the urokinase-type plasminogen activator mouse, *Gastroenterology 133*, 1144-1155.
11. Herrmann, E., Neumann, A. U., Schmidt, J. M., and Zeuzem, S. (2000) Hepatitis C virus kinetics, *Antivir Ther 5*, 85-90.
12. Neumann, A. U., Lam, N. P., Dahari, H., Gretch, D. R., Wiley, T. E., Layden, T. J., and Perelson, A. S. (1998) Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy, *Science 282*, 103-107.
13. Martell, M., Esteban, J. I., Quer, J., Genesca, J., Weiner, A., Esteban, R., Guardia, J., and Gomez, J. (1992) Hepatitis C virus (HCV) circulates as a population of different but closely related genomes: quasispecies nature of HCV genome distribution, *J Virol 66*, 3225-3229.
14. Paolucci, S., Baldanti, F., Campanini, G., Zavattoni, M., Cattaneo, E., Dossena, L., and Gerna, G. (2001) Analysis of HIV drug-resistant quasispecies in plasma, peripheral blood mononuclear cells and viral isolates from treatment-naive and HAART patients, *J Med Virol 65*, 207-217.

15. Manns, M. P., McHutchison, J. G., Gordon, S. C., Rustgi, V. K., Shiffman, M., Reindollar, R., Goodman, Z. D., Koury, K., Ling, M., and Albrecht, J. K. (2001) Peginterferon alfa-2b plus ribavirin compared with interferon alfa-2b plus ribavirin for initial treatment of chronic hepatitis C: a randomised trial, *Lancet 358*, 958-965.
16. Fried, M. W., Shiffman, M. L., Reddy, K. R., Smith, C., Marinos, G., Goncales, F. L., Jr., Haussinger, D., Diago, M., Carosi, G., Dhumeaux, D., Craxi, A., Lin, A., Hoffman, J., and Yu, J. (2002) Peginterferon alfa-2a plus ribavirin for chronic hepatitis C virus infection, *N Engl J Med 347*, 975-982.
17. Zeuzem, S., Buti, M., Ferenci, P., Sperl, J., Horsmans, Y., Cianciara, J., Ibranyi, E., Weiland, O., Noviello, S., Brass, C., and Albrecht, J. (2006) Efficacy of 24 weeks treatment with peginterferon alfa-2b plus ribavirin in patients with chronic hepatitis C infected with genotype 1 and low pretreatment viremia, *J Hepatol 44*, 97-103.
18. Scotto, G., Fazio, V., Palumbo, E., Cibelli, D. C., Saracino, A., and Angarano, G. (2005) Treatment of genotype 1b HCV-related chronic hepatitis: efficacy and toxicity of three different interferon alfa-2b/ribavirin combined regimens in naive patients, *New Microbiol 28*, 23-29.
19. Hadziyannis, S. J., Sette, H., Jr., Morgan, T. R., Balan, V., Diago, M., Marcellin, P., Ramadori, G., Bodenheimer, H., Jr., Bernstein, D., Rizzetto, M., Zeuzem, S., Pockros, P. J., Lin, A., and Ackrill, A. M. (2004) Peginterferon-alpha2a and ribavirin combination therapy in chronic hepatitis C: a randomized study of treatment duration and ribavirin dose, *Ann Intern Med 140*, 346-355.
20. Fried, M. W. (2002) Side effects of therapy of hepatitis C and their management, *Hepatology 36*, S237-244.
21. Liu, S. S., Schneekloth, T. D., Talwalkar, J. A., Kim, W. R., Poterucha, J. J., Charlton, M. R., Wiesner, R. H., and Gross, J. B. (2010) Impact of depressive symptoms and their treatment on completing antiviral treatment in patients with chronic hepatitis C, *J Clin Gastroenterol 44*, e178-185.
22. Butt, A. A., McGinnis, K. A., Skanderson, M., and Justice, A. C. (2009) Hepatitis C treatment completion rates in routine clinical care, *Liver Int 30*, 240-250.
23. Lee, S. S., Bain, V. G., Peltekian, K., Krajden, M., Yoshida, E. M., Deschenes, M., Heathcote, J., Bailey, R. J., Simonyi, S., and Sherman, M. (2006) Treating chronic hepatitis C with pegylated interferon alfa-2a (40 KD) and ribavirin in clinical practice, *Aliment Pharmacol Ther 23*, 397-408.
24. Dollarhide, A. W., Loh, C., Leckband, S. G., Endow-Eyer, R., Robinson, S., and Meyer, J. M. (2007) Psychiatric comorbidity does not predict interferon treatment completion rates in hepatitis C seropositive veterans, *J Clin Gastroenterol 41*, 322-328.
25. Kuntzen, T., Timm, J., Berical, A., Lennon, N., Berlin, A. M., Young, S. K., Lee, B., Heckerman, D., Carlson, J., Reyor, L. L., Kleyman, M., McMahon, C. M., Birch, C., Schulze Zur Wiesch, J., Ledlie, T., Koehrsen, M., Kodira, C., Roberts, A. D., Lauer, G. M., Rosen, H. R., Bihl, F., Cerny, A., Spengler, U., Liu, Z., Kim, A. Y., Xing, Y., Schneidewind, A., Madey, M. A., Fleckenstein, J. F., Park, V. M., Galagan, J. E., Nusbaum, C., Walker, B. D., Lake-Bakaar, G. V., Daar, E. S., Jacobson, I. M., Gomperts, E. D., Edlin, B. R., Donfield, S. M., Chung, R. T., Talal, A. H., Marion, T., Birren, B. W., Henn, M. R., and Allen, T. M. (2008) Naturally occurring dominant resistance mutations to hepatitis C virus protease and polymerase inhibitors in treatment-naive patients, *Hepatology 48*, 1769-1778.

26. Tong, X., Bogen, S., Chase, R., Girijavallabhan, V., Guo, Z., Njoroge, F. G., Prongay, A., Saksena, A., Skelton, A., Xia, E., and Ralston, R. (2008) Characterization of resistance mutations against HCV ketoamide protease inhibitors, *Antiviral Res 77*, 177-185.
27. He, Y., King, M. S., Kempf, D. J., Lu, L., Lim, H. B., Krishnan, P., Kati, W., Middleton, T., and Molla, A. (2008) Relative replication capacity and selective advantage profiles of protease inhibitor-resistant hepatitis C virus (HCV) NS3 protease mutants in the HCV genotype 1b replicon system, *Antimicrob Agents Chemother 52*, 1101-1110.
28. Sarrazin, C., Rouzier, R., Wagner, F., Forestier, N., Larrey, D., Gupta, S. K., Hussain, M., Shah, A., Cutler, D., Zhang, J., and Zeuzem, S. (2007) SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders, *Gastroenterology 132*, 1270-1278.
29. Kieffer, T. L., Sarrazin, C., Miller, J. S., Welker, M. W., Forestier, N., Reesink, H. W., Kwong, A. D., and Zeuzem, S. (2007) Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients, *Hepatology 46*, 631-639.
30. Yi, M., Ma, Y., Yates, J., and Lemon, S. M. (2009) Trans-complementation of an NS2 defect in a late step in hepatitis C virus (HCV) particle assembly and maturation, *PLoS Pathog 5*, e1000403.
31. Lenz, O., Verbinnen, T., Lin, T. I., Vijgen, L., Cummings, M. D., Lindberg, J., Berke, J. M., Dehertogh, P., Fransen, E., Scholliers, A., Vermeiren, K., Ivens, T., Raboisson, P., Edlund, M., Storm, S., Vrang, L., de Kock, H., Fanning, G. C., and Simmen, K. A. (2010) In vitro resistance profile of the HCV NS3/4A protease inhibitor TMC435, *Antimicrob Agents Chemother*.
32. Romano, K. P., Ali, A., Royer, W. E., and Schiffer, C. A. (2010) 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*.
33. McCauley, J. A., McIntyre, C. J., Rudd, M. T., Nguyen, K. T., Romano, J. J., Butcher, J. W., Gilbert, K. F., Bush, K. J., Holloway, M. K., Swestock, J., Wan, B. L., Carroll, S. S., DiMuzio, J. M., Graham, D. J., Ludmerer, S. W., Mao, S. S., Stahlhut, M. W., Fandozzi, C. M., Trainor, N., Olsen, D. B., Vacca, J. P., and Liverton, N. J. (2010) Discovery of vaniprevir (MK-7009), a macrocyclic hepatitis C virus NS3/4a protease inhibitor, *J Med Chem 53*, 2443-2463.
34. Harper, S., Summa, V., Liverton, N. J., and McCauley, J. A. (2010) Macrocyclic Quinoxaline Compounds as HCV NS3 Protease Inhibitors.
35. Ingallinella, P., Altamura, S., Bianchi, E., Taliani, M., Ingenito, R., Cortese, R., De Francesco, R., Steinkuhler, C., and Pessi, A. (1998) Potent peptide inhibitors of human hepatitis C virus NS3 protease are obtained by optimizing the cleavage products, *Biochemistry 37*, 8906-8914.
36. Wittekind, M., Weinheirner, S., Zhang, Y., and Goldfarb, V. (2002) Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes, In *United States Patent Applications Publication*, p 26, United States of America.
37. Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkuhler, C., and De Francesco, R. (1998) Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus, *J Virol 72*, 6758-6769.
38. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray Diffraction Data Collected in Oscillation Mode, (C.W. Carter, J. R. M. S., Eds., Ed.), pp Volume 276: Macromolecular Crystallography, part A, p.307-326, Methods in Enzymology.

39. McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C., and Read, R. J. (2007) Phaser crystallographic software, *J Appl Crystallogr 40*, 658-674.
40. COLLABORATIVE COMPUTATIONAL PROJECT, N. (1994) The CCP4 Suite: Programs for Protein Crystallography., *Acta Crystallographica 50*, 760-763.
41. Davis, I. W., Leaver-Fay, A., Chen, V. B., Block, J. N., Kapral, G. J., Wang, X., Murray, L. W., Arendall, W. B., 3rd, Snoeyink, J., Richardson, J. S., and Richardson, D. C. (2007) MolProbity: all-atom contacts and structure validation for proteins and nucleic acids, *Nucleic Acids Res 35*, W375-383.
42. Brunger, A. T. (1992) Free R value: a novel statistical quantity for assessing the accuracy of crystal structures, *Nature 355*, 472-475.
43. Emsley, P., and Cowtan, K. (2004) Coot: model-building tools for molecular graphics, *Acta Crystallogr D Biol Crystallogr 60*, 2126-2132.
44. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. C. (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase, *Structure 7*, 1353-1363.
45. Sarkar, G., and Sommer, S. S. (1990) The "megaprimer" method of site-directed mutagenesis, *Biotechniques 8*, 404-407.
46. Matayoshi, E. D., Wang, G. T., Krafft, G. A., and Erickson, J. (1990) Novel fluorogenic substrates for assaying retroviral proteases by resonance energy transfer, *Science 247*, 954-958.
47. Nalam, M. N., Ali, A., Altman, M. D., Reddy, G. S., Chellappan, S., Kairys, V., Ozen, A., Cao, H., Gilson, M. K., Tidor, B., Rana, T. M., and Schiffer, C. A. (2010) Evaluating the substrate-envelope hypothesis: structural analysis of novel HIV-1 protease inhibitors designed to be robust against drug resistance, *J Virol 84*, 5368-5378.

CHAPTER V

DISCUSSION

THESIS IMPLICATIONS

Rationally Evaluating Drug Candidates

Modern drug design strategies often incorporate high-throughput screening in combination with structure-activity relationship (SAR) studies – analyses of the effect of chemical structure on activity. These techniques focus primarily on disrupting target activity, and the resulting drugs often contact unessential target residues that can mutate without altering biological activity. Thus alternative approaches – particularly in the face of target evolution – must exploit drug interactions with residues that are essential for target function. MK-5172 exemplifies this approach by stacking extensively against the catalytic triad residues, which are strictly required for efficient protease activity. When applicable, genetic analyses of target sequences can conveniently identify invariant residues across genotypes, subtypes or strains. However, genetic invariance may not constitute functional necessity in the absence of selective drug pressures. Thus mutational studies, such as alanine substitutions, are required to rigorously demonstrate the relative importance of specific residues to biological activity.

Such mutational experiments, however, require time-consuming cloning and efficient assay systems. Moreover, many target residues play intermediate roles and thus are not strictly required for activity. The substrate envelope provides a practical alternative for evaluating drug candidates in the face of target evolution. My research shows that resistance mutations arise in regions outside the substrate envelope, where NS3/4A protease extensively contacts drugs but not substrate products (*1*). Thus the binding of drugs relative to the substrate envelope can predict their susceptibilities to

resistance mutations. Such analyses – performed with inhibitor crystal structures or computational models – are rapid and circumvent the need for laborious resistance testing in the early stages of development. I hope that the substrate envelope becomes a useful, accessible tool for evaluating the clinical potential of novel HCV protease inhibitors in light of rapidly evolving resistance.

Designing Novel Protease Inhibitors

Structure-based design strategies can incorporate the substrate envelope as a constraint in the construction of novel drug libraries. In fact, our laboratory has extensive expertise in this area from the design project of HIV-1 protease inhibitors (*2-5*). In theory, drugs designed to fit within the HCV substrate envelope will be less susceptible to resistance, as mutations disrupting drug binding will simultaneously reduce viral fitness by preventing substrate binding. We can design novel compounds to fit within the substrate envelope by modifying scaffolds of the most potent drugs, such as danoprevir, vaniprevir and MK-5172. SAR studies – comprising inhibitor crystal structures and activity assays – can assess the effects of various chemical substituents on drug potency and binding relative to the substrate envelope. The best lead compounds can direct subsequent rounds of SAR studies, facilitating the discovery of more potent drug candidates that fit within the substrate envelope. We can then perform drug susceptibility testing – in viral culture or replicon-based assays – to evaluate these drugs against major resistant HCV strains.

The potencies of many NS3/4A protease inhibitors, however, derive from favorable interactions outside the substrate envelope, most notably in the P2 subsite (*1, 6,*

7). Thus constraining new compounds to fit within the substrate envelope will likely reduce their binding affinities. Efforts must explore many substituents in parallel, including large P2 moieties, which may enhance drug potency despite protruding from the substrate envelope. Specifically, the ether-linked P2 quinoxaline moiety of MK-5172 should be incorporated in novel compounds, as it extends away from the P2 subsite and interacts with the catalytic residues H57 and D81. Drug resistance mutations are unlikely to disrupt these interactions – despite their location outside the substrate envelope – because mutations at these catalytic sites would severely compromise viral fitness by disrupting proteolytic efficiency. In addition, we should carefully consider and meticulously explore the location of macrocycles throughout the design process. P2–P4 macrocycles, like in vaniprevir and MK-5172, have rigidifying effects that increase susceptibility to bulky mutations at A156. Drugs without P2 macrocycles, such as danoprevir, can better accommodate steric mutations due to the increased flexibility of its P2 substituent. Ultimately, I hope substrate envelope-based SAR studies lead to the discovery of future generation NS3/4A protease inhibitors that remain active against highly resistant HCV variants.

Prime Side Substrate Interactions

Inactive serine proteases – containing single, double or triple alanine mutations of the catalytic triad – still have residual catalytic activity (*8*). Moreover, C-terminal substrate cleavage fragments diffuse from the active site and cannot be modeled in crystal structures (*1*, *9*). As a result, crystal structures only contain electron density for N-terminal cleavage products, which bind to the protease active site with higher affinities

than their parent substrates (*10*). Thus the residual catalytic activity of inactive NS3/4A protease prevents the complete characterization of viral substrate binding.

My research assumes that N-terminal cleavage products make the same interactions with the protease as non-cleaved substrates. Though this assumption – in my opinion – is likely valid, it should be rigorously demonstrated by crystallizing the protease with non-hydrolyzable substrate mimics. Non-cleavable P1–P1' dipeptide fragments will prevent NS3-mediated scissile bond hydrolysis. Proline at P1' would most easily achieve this effect, although proline is not found naturally at this position and would kink in the peptide substrate. Thus we should also explore two additional approaches: *(i)* replacing the P1 amide bond with an azapeptide linkage, and *(ii)* incorporating P1–P1' aminomethylene or ketomethylene dipeptide isosteres. Characterizing the binding of these non-hydrolyzable substrate mimics would elucidate the interactions of C-terminal substrate residues, as well as N-terminal residues in the setting of full-length substrate binding.

Protease Dynamics and Substrate Recognition

The crystal structures of substrate products, presented in Chapter II, raise questions about the role of protein dynamics in the substrate recognition. The protease undergoes a conformational change upon binding certain substrate products, forming an extensive electrostatic network spanning residues H57, D81, R155, D168 and R123. These networks are absent in the crystal structures of products 4B5A and TRIF, however, which are less negatively charged relative to the other substrates. Thus charge interactions may facilitate the recognition process of some – but not all – viral substrates.

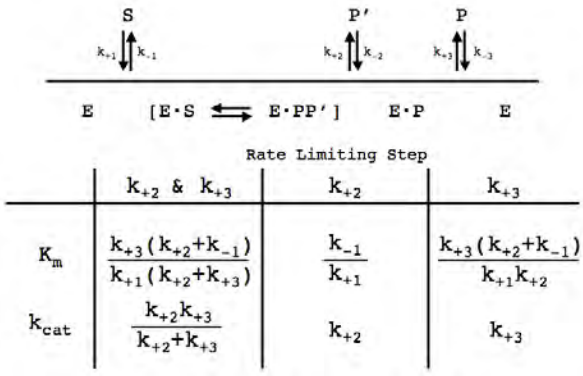
NMR relaxation experiments of the unliganded and substrate-bound protease variants can probe changes in protein flexibility upon binding. N^{15}-labeled protease is highly soluble and easily expressed (*11*), making these experiments very feasible in our own laboratory. Relaxation experiments of drug-bound protease variants may also reveal differences in drug binding relative to authentic substrates, perhaps highlighting drug resistance mechanisms of non-active site mutations, such as V36M, T54A and V170A. These experiments would complement our structural data, together providing a more comprehensive view of dynamic effects on protease function and drug resistance.

The Helicase Domain in Protease Function

The helicase domain was shown to enhance NS3/4A proteolytic efficiency (*12*), although others groups report comparable catalytic efficiencies between full-length NS3 and the protease domain alone (*13, 14*). Alanine substitutions at residues Q526 and H528 – positioned along the helicase interface with the protease active site – reduce the affinities of peptide inhibitors relative to the wildtype enzyme (*15*). In addition, the helicase mutation S489L confers resistance against danoprevir, suggesting that the helicase domain influences drug binding (*16*). Taken together, these findings implicate the helicase domain as a putative factor in substrate recognition and inhibitor binding.

Substrate-bound crystal structures of full-length NS3 would most completely elucidate potential helicase domain interactions with the protease. As such crystals are exceedingly difficult to grow, small-angle X-ray scattering and hydrogen-deuterium exchange experiments – on both bound and unbound states – may more reliably probe conformational NS3 ensembles. Well-designed kinetic experiments can also provide

detailed molecular views of domain interdependency. Full-length NS3 may undergo autoinhibition by the 3-4A post-cleavage product, which remains bound in the active site of the full-length crystal structure (*17*). Thus kinetic measurements may underestimate the proteolytic efficiency of full-length NS3, as autoinhibition would increase the apparent K_m values. Changes in apparent K_m of T631C and T631N variants – designed to increase and decrease 3-4A product binding, respectively – would depend on the relative autoinhibition of the wildtype protein. The presence of the helicase domain could alter the elementary rate constants, however, potentially confounding data interpretation compared to the protease domain alone. Serine proteases exhibit two-step reaction mechanisms with covalently bound intermediates, and thus k_{cat} and K_m can represent various expressions depending on the rate-limiting step:



Similar kinetic parameters of the protease domain and full-length protein may derive from different mathematical expressions of these elementary rate constants. Pre-steady state kinetic studies are therefore necessary to rigorously differentiate catalytic differences between full-length NS3 and protease domain variants.

Other HCV Genotypes

Though my research focuses on genotype 1a HCV, it also provides a blueprint for investigating drug efficacy against other HCV genotypes. In particular, the NS3/4A protease of genotype 3a viruses, endemic in Southeast Asia (*18*), diverges considerably in primary sequence from genotype 1a strains. Most notably, the protease contains threonine instead of arginine at residue 123, and glutamine instead of aspartate at residue 168. These residues participate in the electrostatic networks in many product complexes, and thus substantial mechanistic differences may exist in both substrate recognition and inhibitor binding. In fact, telaprevir, boceprevir, danoprevir, vaniprevir and TMC-435 exhibit markedly reduced activities against genotype 3 virus relative to other genotypes (*19*). Structural and biochemical studies of substrate and drug binding should therefore be repeated for genotype 3 protease variants. These genotypic differences will likely have important medical implications as protease inhibitors become integrated into anti-HCV treatment regimens worldwide.

Targeting Other Rapidly Evolving Diseases

Drug resistance arises from target mutations that selectively weaken inhibitor potency without significantly altering its natural biological function. We observe that HCV and HIV resistance mutations arise where drugs protrude from the substrate envelope, as changes in these regions disrupt drug binding but not substrate recognition and cleavage (*1, 20*). We can also define substrate envelopes in other disease systems whenever substrate interactions can be structurally characterized with drug targets. Potential systems include other *Flaviviridae*, such as Dengue, West Nile Virus and

Yellow Fever, which encode serine proteases homologous to HCV NS3/4A (*21-23*). The protozoan parasite causing malaria – *Plasmodium falciparum* – also encodes aspartyl proteases called plasmepsins (*24, 25*). We can further expand our reach beyond the realm of protease enzymes. The substrate envelope of *influenzavirus A* Neuraminidase – a glycoside hydrolase that cleaves neuraminic acids on infected cell surfaces – may facilitate the discovery of broadly acting neuraminidase inhibitors against diverse influenza serotypes. Perhaps substrate envelope-based design approaches can someday help combat the emergence of drug resistance in other disease systems in the face of rapid disease evolution.

CONCLUDING REMARKS

The relevance of my research rests on the claim that drug resistance is a major obstacle in modern medicine. Not all scientists share this view – opponents argue that combination therapies eliminate the likelihood of resistance altogether, as multiple drugs limit disease evolution through several distinct mechanisms simultaneously. Indeed, this reasoning is sound and well supported for patients with the social, economic and logistical support necessary for strict treatment adherence. Unfortunately, drug side effects, high cost and lacking social support often prevent patients from adhering to therapies. In such instances, patients may routinely miss doses or fail to complete treatment regimens altogether. Insufficient dosing causes pathogens to persist in patients, allowing for the evolution and selection of drug resistant strains. Thus in the actual clinical setting, the acquisition of resistance challenges the long-term efficacy of most drugs – whether given as monotherapy or in combination.

The emergence of drug resistance will continue until combination therapies comprise tolerable, accessible and affordable drugs that retain potency against broad spectrums of pathogen strains. These are difficult challenges – drug tolerability, accessibility and affordability are often uncontrollable variables for basic scientists, and become matters of public health and social justice. On the contrary, scientists have complete jurisdiction over optimizing drug potency in the face of evolving targets. Understanding the molecular underpinnings of resistance is therefore central to developing potent therapies against genetically diverse pathogen strains. To this end, my research comprehensively defines the molecular basis of drug resistance against HCV

NS3/4A protease inhibitors. I ultimately hope my work helps generate more efficacious and tolerable anti-HCV therapies, and eventually serves to combat rising resistance in other infectious diseases causing human suffering throughout the world.

REFERENCES

1. Romano, K. P., Ali, A., Royer, W. E., and Schiffer, C. A. (2010) 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*.
2. Chellappan, S., Kiran Kumar Reddy, G. S., Ali, A., Nalam, M. N., Anjum, S. G., Cao, H., Kairys, V., Fernandes, M. X., Altman, M. D., Tidor, B., Rana, T. M., Schiffer, C. A., and Gilson, M. K. (2007) Design of mutation-resistant HIV protease inhibitors with the substrate envelope hypothesis, *Chem Biol Drug Des 69*, 298-313.
3. Altman, M. D., Ali, A., Reddy, G. S., Nalam, M. N., Anjum, S. G., Cao, H., Chellappan, S., Kairys, V., Fernandes, M. X., Gilson, M. K., Schiffer, C. A., Rana, T. M., and Tidor, B. (2008) HIV-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants, *J Am Chem Soc 130*, 6099-6113.
4. Reddy, G. S., Ali, A., Nalam, M. N., Anjum, S. G., Cao, H., Nathans, R. S., Schiffer, C. A., and Rana, T. M. (2007) Design and synthesis of HIV-1 protease inhibitors incorporating oxazolidinones as P2/P2' ligands in pseudosymmetric dipeptide isosteres, *J Med Chem 50*, 4316-4328.
5. Ali, A., Reddy, G. S. K. K., Nalam, M. N. L., Anjum, S. G., Cao, H., Altman, M. D., Tidor, B., Rana, T. M., and Schiffer, C. A. (2009) Substrate envelope based design of new HIV-1 protease inhibitors active against drug-resistant HIV-1, In *238th ACS National Meeting, Washington, DC, United States*, pp MEDI-102, American Chemical Society.
6. Cummings, M. D., Lindberg, J., Lin, T. I., de Kock, H., Lenz, O., Lilja, E., Fellander, S., Baraznenok, V., Nystrom, S., Nilsson, M., Vrang, L., Edlund, M., Rosenquist, A., Samuelsson, B., Raboisson, P., and Simmen, K. (2010) Induced-fit binding of the macrocyclic noncovalent inhibitor TMC435 to its HCV NS3/NS4A protease target, *Angew Chem Int Ed Engl 49*, 1652-1655.
7. Prongay, A. J., Guo, Z., Yao, N., Pichardo, J., Fischmann, T., Strickland, C., Myers, J., Jr., Weber, P. C., Beyer, B. M., Ingram, R., Hong, Z., Prosise, W. W., Ramanathan, L., Taremi, S. S., Yarosh-Tomaine, T., Zhang, R., Senior, M., Yang, R. S., Malcolm, B., Arasappan, A., Bennett, F., Bogen, S. L., Chen, K., Jao, E., Liu, Y. T., Lovey, R. G., Saksena, A. K., Venkatraman, S., Girijavallabhan, V., Njoroge, F. G., and Madison, V. (2007) Discovery of the HCV NS3/4A protease inhibitor (1R,5S)-N-[3-amino-1-(cyclobutylmethyl)-2,3-dioxopropyl]-3- [2(S)-[[[(1,1-dimethylethyl)amino]carbonyl]amino]-3,3-dimethyl-1-oxobutyl] - 6,6-dimethyl-3-azabicyclo[3.1.0]hexan-2(S)-carboxamide (Sch 503034) II. Key steps in structure-based optimization, *J Med Chem 50*, 2310-2318.
8. Carter, P., and Wells, J. A. (1988) Dissecting the catalytic triad of a serine protease, *Nature 332*, 564-568.
9. Krishnan, R., Sadler, J. E., and Tulinsky, A. (2000) Structure of the Ser195Ala mutant of human alpha--thrombin complexed with fibrinopeptide A(7--16): evidence for residual catalytic activity, *Acta Crystallogr D Biol Crystallogr 56*, 406-410.
10. Steinkuhler, C., Biasiol, G., Brunetti, M., Urbani, A., Koch, U., Cortese, R., Pessi, A., and De Francesco, R. (1998) Product inhibition of the hepatitis C virus NS3 protease, *Biochemistry 37*, 8899-8905.
11. Romano, K. P., Laine, J. M., Deveau, L. M., Cao, H., Massi, F., and Schiffer, C. A. (2011) Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease, *Journal of virology 85*, 6106-6116.

12. Beran, R. K., and Pyle, A. M. (2008) Hepatitis C viral NS3-4A protease activity is enhanced by the NS3 helicase, *J Biol Chem 283*, 29929-29937.
13. Gallinari, P., Brennan, D., Nardi, C., Brunetti, M., Tomei, L., Steinkuhler, C., and De Francesco, R. (1998) Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus, *J Virol 72*, 6758-6769.
14. Taremi, S. S., Beyer, B., Maher, M., Yao, N., Prosise, W., Weber, P. C., and Malcolm, B. A. (1998) Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis C virus protease, *Protein Sci 7*, 2143-2149.
15. Dahl, G., Sandstrom, A., Akerblom, E., and Danielson, U. H. (2007) Effects on protease inhibition by modifying of helicase residues in hepatitis C virus nonstructural protein 3, *FEBS J 274*, 5979-5986.
16. Seiwert. (2006) Sequence variation of NS3 and NS4A in hepatitis C virus (HCV) replicons following exposure to ITMN-191 concentrations likely to encompass those achieved in human liver following clinical dosing., In *First International Workshop in Hepatitis C Resistance and New Compounds*, Boston, MA.
17. Yao, N., Reichert, P., Taremi, S. S., Prosise, W. W., and Weber, P. C. (1999) Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis C virus bifunctional protease-helicase, *Structure 7*, 1353-1363.
18. Mondelli, M. U., and Silini, E. (1999) Clinical significance of hepatitis C virus genotypes, *J Hepatol 31 Suppl 1*, 65-70.
19. Gottwein, J. M., Scheel, T. K., Jensen, T. B., Ghanem, L., and Bukh, J. (2011) Differential Efficacy of Protease Inhibitors Against HCV Genotypes 2a, 3a, 5a, and 6a NS3/4A Protease Recombinant Viruses, *Gastroenterology*.
20. King, N. M., Prabu-Jeyabalan, M., Nalivaika, E. A., and Schiffer, C. A. (2004) Combating susceptibility to drug resistance: lessons from HIV-1 protease, *Chem Biol 11*, 1333-1338.
21. Phong, W. Y., Moreland, N. J., Lim, S. P., Wen, D., Paradkar, P. N., and Vasudevan, S. G. (2011) Dengue protease activity: the structural integrity and interaction of NS2B with NS3 protease and its potential as a drug target, *Biosci Rep*.
22. Robin, G., Chappell, K., Stoermer, M. J., Hu, S. H., Young, P. R., Fairlie, D. P., and Martin, J. L. (2009) Structure of West Nile virus NS3 protease: ligand stabilization of the catalytic conformation, *Journal of molecular biology 385*, 1568-1577.
23. Kondo, M. Y., Oliveira, L. C., Okamoto, D. N., de Araujo, M. R., Duarte dos Santos, C. N., Juliano, M. A., Juliano, L., and Gouvea, I. E. (2011) Yellow fever virus NS2B/NS3 protease: hydrolytic properties and substrate specificity, *Biochemical and biophysical research communications 407*, 640-644.
24. Kim, Y. M., Lee, M. H., Piao, T. G., Lee, J. W., Kim, J. H., Lee, S., Choi, K. M., Jiang, J. H., Kim, T. U., and Park, H. (2006) Prodomain processing of recombinant plasmepsin II and IV, the aspartic proteases of Plasmodium falciparum, is auto- and trans-catalytic, *J Biochem 139*, 189-195.
25. Clemente, J. C., Govindasamy, L., Madabushi, A., Fisher, S. Z., Moose, R. E., Yowell, C. A., Hidaka, K., Kimura, T., Hayashi, Y., Kiso, Y., Agbandje-McKenna, M., Dame, J. B., Dunn, B. M., and McKenna, R. (2006) Structure of the aspartic protease plasmepsin 4 from the malarial parasite Plasmodium malariae bound to an allophenylnorstatine-based inhibitor, *Acta crystallographica. Section D, Biological crystallography 62*, 246-252.