VIRAL PROTEASES AS DRUG TARGETS AND THE MECHANISMS OF DRUG RESISTANCE

A Dissertation Presented

By

Kuan-Hung Lin

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

September 1st, 2016

Biochemistry and Molecular Pharmacology

VIRAL PROTEASES AS DRUG TARGETS AND THE MECHANISMS OF DRUG RESISTANCE

A Dissertation Presented

By

Kuan-Hung Lin

This work was undertaken in the Graduate School of Biomedical Sciences

Biochemistry and Molecular Pharmacology

The signature of the Thesis Advisor signifies validation of Dissertation content

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

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

William R. Kobertz, Ph.D., Member of Committee

Sharone Green, M.D., Member of Committee

Brian A. Kelch, Ph.D., Member of Committee

Peter Chien, Ph.D., External Member of Committee

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

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

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

Anthony Carruthers, Ph.D., Dean of the Graduate School of Biomedical Sciences

September 1st, 2016

Acknowledgements

The passion for science motivated me to join the graduate program at UMass Medical School, and the support from faculties, colleagues, friends and family made my graduate study enjoyable.

In December 2010, I walked into my thesis advisor Dr. Celia Schiffer's office without knowing anything about structural biology. The only thing I knew was people said she is a good mentor. After hearing about my naivety of science, she smiled and started to give me a lecture about structural biology and research of the lab. I did not pick up much during that one hour meeting, but I decided to join the lab because of her kindness and passion for science. Celia gave me the freedom to work on the projects I was interested in and provided me all the guidance I needed as a student. She gave me full support even after I could not get any data on the dengue project for three years, and after all, she still had faith in me. Celia is more than a good mentor; she not only provided me a wonderful research environment but also made sure I could blend in the American culture properly as a foreigner. I appreciate her friendship over the past years.

I would like to thank my TRAC committee members: Dr. William Royer, Dr. William Kobertz, Dr. Sharone Green, Dr. Brian Kelch and Dr. Manuel Navia. You were all generous with your time when I need guidance, and you always encouraged me and make me feel confident even when I thought I did not accomplish much on my research. Your insights in science greatly helped me with the progress of my research, and also motivated me to improve myself all the time.

I would also like to thank all the people in the Schiffer lab. I truly appreciate Dr. Nese Kurt-Yilmaz; she always spends time helping me structure my project, interpret the data, troubleshoot, and improve my writing. Madhavi Nalam and Seema Mittal taught me all the protein expression/purification techniques and crystallography when I was a rotation student. Ellen Nalivaika helped me with all the projects and always helped me out when I got stuck in experiments. Cihan Aydin helped me set up the enzyme inhibition assay of dengue protease. Dr. Akbar Ali gave me a huge support for my research from the chemistry prospect, and I was lucky to learn compound synthesis from him. Akbar made lots of protease inhibitors for the dengue project, and Dr. Linah Rusere and Jacqueto Zephyr also contributed a great amount of work to this research. Dr. Madhavi Kolli and Dr. Sagar Kathuria helped me with every aspect of my research and will always be good friends of mine.

I am so grateful to my family in the U.S., Djade Soumana, Kristina Prachanronarong and Tania Silvas. We went through lots of up and down together, and your support and sense of humor helped me enjoy the process.

To my family, Tsuo-Pin Lin, Mu-Tan Hsieh, Pei-Jun Lin, Keng-Jan Hsiao, Tzu-Chi Hsiao and Chiao-Yun Hsiao, for having my back all the time and making me feel warm. You made me feel good about myself even when nothing seemed to work.

Last but not least, to my wife Ping-I Chiang, for always believing in me but still challenging me, for helping me know myself and being humble. You are the reason I always try my best.

Abstract

Viral proteases have been shown to be effective targets of anti-viral therapies for human immunodeficiency virus (HIV) and hepatitis C virus (HCV). However, under the pressure of therapy including protease inhibitors, the virus evolves to select drug resistance mutations both in the protease and substrates. In my thesis study, I aimed to understand the mechanisms of how this protease−substrate co-evolution contributes to drug resistance. Currently, there are no approved drugs against dengue virus (DENV); I investigated substrate recognition by DENV protease and designed cyclic peptides as inhibitors targeting the prime site of dengue protease.

First, I used X-ray crystallography and subsequent structural analysis to investigate the molecular basis of HIV-1 protease and p1-p6 substrate coevolution. I found that co-evolved p1-p6 substrates rescue the HIV-1 I50V protease's binding activity by forming more van der Waals contacts and hydrogen bonds, and that co-evolution restores the dynamics at the active site for all three mutant substrates.

Next, I used aprotinin as a platform to investigate DENV protease–substrate recognizing pattern, which revealed that the prime side residues significantly modulate substrate affinity to protease and the optimal interactions at each residue position. Based on these results, I designed cyclic peptide inhibitors that target the prime site pocket of DENV protease. Through optimizing the length and sequence, the best inhibitor achieved a 2.9 micromolar K_i value against DENV3 protease. Since dengue protease does not share substrate sequence with human serine proteases, these cyclic peptides can be used as scaffolds for inhibitor design with higher specificity.

Table of Contents

Abstract ..... **v**

List of Tables ..... **ix**

List of Figures ..... **xi**

List of Third Party Copyrighted Materials ..... **xiv**

List of Abbreviations ..... **xv**

Chapter I **1**

Introduction **1**

1.1 Viral protease as drug target **2**

1.2 Human immunodeficiency virus **3**

1.2.1 HIV-1 viral proteins and viral lifecycle 4

1.2.2 HIV-1 protease and substrate recognition 8

1.2.3 HIV-1 protease inhibitors as antivirals and drug resistance 13

1.2.4 HIV-1 protease-substrate co-evolution 17

1.3 Dengue virus **18**

1.3.1 DENV lifecycle and viral proteins 19

1.3.2 Therapeutic strategies against dengue virus 22

1.3.3 Dengue NS3/2B protease and structural features 23

1.3.4 Dengue NS3/2B protease substrate specificity 25

1.3.5 Inhibitors targeting dengue NS3/2B protease 29

1.4 Thesis scope **32**

Chapter II **34**

Structural Basis and Distal Effects of Gag Substrate Co-evolution in Drug Resistance to HIV-1 Protease **34**

PREFACE **35**

2.1 Abstract **36**

2.2 Introduction **37**

2.3 Results **41**

2.3.1 The overall structure and substrate envelope is conserved in co-evolved complexes 41

2.3.2 Substrate mutations have distal effects and enhance packing at the active site 47

2.3.3 The RP4'S substrate has less packing interactions but an additional hydrogen bond 53

2.3.4 Active site dynamics is restored by coevolution 56

2.4 Discussion **58**

2.5 Conclusion **62**

2.6 Materials and Methods **63**

2.6.1 Nomenclature 63

2.6.2 Substrate peptides 63

2.6.3 Protease gene construction 63

2.6.4 Protein expression, purification and crystallization 64

2.6.5 Data collection and structure solution 64
2.6.6 Structural analysis 66
2.6.7 Molecular dynamics simulations 66

Chapter III 69

Dengue Protease Substrate Recognition: Binding of the Prime Side 69
PREFACE 70
3.1 Abstract 71
3.2 Introduction 72
3.3 Results 75
3.3.1 Design of aprotinin constructs mimicking dengue protease substrates 75
3.3.2 Substrate sequences have widely varying affinity to DENV3 protease 80
3.3.3 The binding interactions are solely entropy driven 82
3.3.4 Binding loop fluctuations correlate with affinity 86
3.3.5 Substrate sequences with higher affinity have better packing at the protease active site 89
3.3.6 Tightest binders have more hydrogen bonds 94
3.3.7 Interactions at individual positions reveal amino acids for optimal packing ... 97
3.4 Discussion 100
3.5 Conclusion 104
3.6 Materials and Methods 105
3.6.1 Protease gene construction, protein expression and purification 105
3.6.2 Aprotinin gene construction, protein expression and purification 105
3.6.3 Enzyme inhibition assay 105
3.6.4 Molecular dynamic simulations 106
3.6.5 MD results analysis 107
3.6.6 Isothermal titration calorimetry experiment 107

Chapter IV 109

Inhibitors of Dengue Protease: Exploiting the Primed Side 109
PREFACE 110
4.1 Abstract 111
4.2 Introduction 112
4.3 Results 115
4.3.1 Optimizing the linker and peptide length 119
4.3.2 Optimizing residues for P1 and P2’ positions of the peptides 122
4.3.3 Optimal C terminal residue depends on the peptide length 123
4.3.4 N terminal capping 124
4.3.5 Targeting Arg54 on the protease 125
4.3.6 Cyclic peptide is more flexible compared to binding loop in aprotinin 126
4.3.7 Binding loop residues lose interactions in cyclic peptide 128
4.4 Discussion 131
4.5 Conclusion 133
4.6 Material and Methods 134

4.6.1 Protein and peptides 134
4.6.2 Enzyme inhibition assay 134
4.6.3 Molecular dynamics simulations 135
4.6.4 Structural analysis 136

Chapter V 137

Discussion 137
5.1 Viral protease substrate recognition and drug design 138
5.2 Dynamics of protease substrate/inhibitor interactions 142
5.3 Optimization of dengue protease inhibitors 145
5.4 Design of dengue inhibitors with higher specificity 148
5.5 Implications for Zika virus 150

Appendix A: Distal Mutation V36M Allosterically Modulates the Active Site to Accentuate Drug Resistance in HCV NS3/4A Protease 151
A.1 Preface 151
A.1.1 Hepatitis C virus 153
A.1.2 HCV NS3/4A protease and substrate recognition 153
A.1.3 HCV NS3/4A protease inhibitors as antivirals 155
A.1.4 Drug resistance mutation selected by HCV protease's inhibitor usage 155
A.2 Methods 160
A.2.1 Protein Expression and purification 160
A.2.2 Crystallization 160
A.2.3 Data collection and structure solution 161
A.3 Results 163
A.3.1 V36M Further Decreases Susceptibility of R155K variants to PIs 163
A.3.2 Resistance Mutations Cause Changes in the Electrostatic Network 164
A.3.3 The Distal V36M Mutation Changes the Active Site via F43 164

Appendix B: Engineering DENV2 NS2B/3 protease for crystallization 173
B.1 Preface 173
B.2 Methods and Results 175

Appendix C: Virtual fragment screening against DENV3 NS3/2B protease 181
C.1 Preface 181
C.2 Methods 182
C.3 Summary of findings 183

Appendix D: Design of P site cyclic peptide inhibitors against dengue NS3/2B protease 186
D.1 Preface 186
D.2 Methods and Results 186

Appendix E: Design of linear P site inhibitors against dengue NS3/2B protease .. 191
E.1 Preface 191
E.2 Methods and results 191

References 196

List of Tables

Table 1.1. HIV-1 protease cleavage site sequences based on the HXB2 sequence. 11

Table 1.2. Selected peptidiomimetic dengue NS3/2B protease inhibitors..... 31

Table 2.1. Crystallographic statistics for HIV-1 protease with p1-p6 substrate co-crystal structures. 43

Table 2.2. Hydrogen bonds between the p1-p6 substrate and HIV-1 protease in cocrystal structures. 55

Table 3.1. The sequence alignments of dengue and Zika proteases' polyprotein cleavage sites..... 81

Table 3.2. The values derived from ITC experiments of WT-AP and aprotinin constructs against DENV3 protease. 83

Table 3.3. The values of binding loop van der Waals contacts calculated from MD simulations of WT-AP and aprotinin constructs against DENV3 protease. .. 92

Table 3.4. The hydrogen bonds between substrate residues at the binding loop and DENV3 protease for the various aprotinin constructs. 96

Table 4.1. The K_i values of different length of cyclic peptides against DENV3 NS3/2B protease. 120

Table 4.2. The K_i values of cyclic peptides with sequence optimization against DENV3 NS3/2B protease. 121

Table 4.3. The hydrogen bonds between substrate residues of aprotinin or cyclic peptides and DENV3 protease. 130

Table A.1. Genotype 1a HCV NS3/4A protease substrate sequences. 158

Table A.2. Crystallographic statistics for HCV protease with Boceprevir co-crystal structures. 166

Table A.3. Drug susceptibilities against wild-type and resistant HCV clones and inhibitory activities against NS3/4A proteases. 168

Table A.4. Intermolecular hydrogen bonds between the inhibitor and protease active site residues..... 169

Table A.5. Intramolecular salt bridges forming a network at the NS3/4A active site surface in crystal structures, and stabilities assessed by MD simulations. .. 170

Table B.1. The expression, purification and crystallization trials of dengue NS3/2B protease constructs or NS2B/full length NS3 constructs. 180

Table C.1. Selected results of fragment library screening against DENV3 WT protease. 185

Table E.1. The inhibition constants of linear P site inhibitors against dengue NS3/2B protease. 195

List of Figures

Figure 1.1. HIV-1 Gag and GagPol polyproteins. 6

Figure 1.2. HIV-1 life cycle. 7

Figure 1.3. Structure of HIV-1 protease in complex with substrate. 10

Figure 1.4. Substrate and Inhibitor Envelopes of HIV-1 Protease. 12

Figure 1.5. Chemical structures of FDA approved HIV-1 protease inhibitors. 15

Figure 1.6. HIV-1 protease mutations selected by different classes of protease inhibitors. 16

Figure 1.7. The life cycle of dengue virus. 21

Figure 1.8. Crystal structure of dengue NS3/2B protease. 24

Figure 1.9. Dengue NS3/2B protease cleavage site sequences. Residues are colored based on the properties of side chains. 28

Figure 2.1. HIV-1 protease and p1-p6 cleavage site co-evolution with I50V primary drug resistance mutation. 40

Figure 2.2. Distance-difference maps showing the effect of protease-substrate co-evolution on internal Ca-Ca distances with respect to the wild-type complex. 44

Figure 2.3. Superposed HIV-1 protease substrate conformations. 45

Figure 2.4. The fit of HIV-1 protease substrates within the substrate envelope and the packing of the substrate within the protease active site. 46

Figure 2.5. The changes in van der Waals contacts of protease residues in monomer (A) and (B) relative to WT complex. 51

Figure 2.6. Distal effects of substrate mutations on protease interactions. 52

Figure 2.7. The hydrogen bonds between p1-p6 substrate and HIV-1 protease. 54

Figure 2.8. The distance (in Å) distribution between T80–T80' across the active site during MD simulations of substrate–protease complexes..... 57

Figure 3.1. Design of aprotinin constructs mimicking dengue protease substrates 77

Figure 3.2. The inhibition constants (K_i) of WT-AP and aprotinin constructs against DENV3 WT protease ordered from tightest to weakest binder. 78

Figure 3.3. The secondary structure is conserved across all aprotinin constructs used in this study. 79

Figure 3.4. The binding of WT-AP and aprotinin constructs to DENV3 protease is entropy driven. 84

Figure 3.5. The relative enthalpic contribution to the binding free energy of aprotinin constructs versus WT aprotinin to DENV3 protease. 85

Figure 3.6. Molecular dynamics (MD) of binding loops bound to DENV protease active site. 87

Figure 3.7. Aprotinin constructs' binding loop fluctuations correlate with affinity. 88

Figure 3.8. The interaction of substrate residues with DENV3 protease ($NS3_{BL}$-AP binding to DENV3 protease is shown as an example). 91

Figure 3.9. The overall binding loop vdW contacts with DENV3 protease calculated from MD simulations of WT-AP and aprotinin constructs..... 93

Figure 4.1. Design cyclic peptides derived from aprotinin as dengue NS3/2B protease inhibitor. 116

Figure 4.2. Liquid chromatography-mass spectrometry (LC-MS) analysis of Apro9 with (red) or without (blue) pre-incubation with DENV3 protease. 117

Figure 4.3. Liquid chromatography-mass spectrometry (LC-MS) analysis of Apro10 with (red) or without (blue) pre-incubation with DENV3 protease. 118

Figure 4.4. Molecular dynamics (MD) of aprotinin or apro10 bound to dengue protease active site.....127

Figure 4.5. The binding loop vdW contacts during MD simulations of aprotinin or apro10 against DENV3 protease. 129

Figure A.1. Schematic representation of HCV NS3/4A protease inhibitors. 159

Figure A.2. Topology of hepatitis C viral serine protease, NS3/4A. 167

Figure A.3. The dynamic distance distribution between protease residue pairs sampled during MD simulations. 171

Figure A.4. The distribution of Ca–Ca distance during MD simulations for F43 with K136, R/K155, A156 and A157.....172

Figure B.1. Selected crystal packing surface of DENV2 protease apo-structure. 177

Figure B.2. Building disulfide bonds between NS3 and NS2B domains of DENV2 protease.....178

Figure B.3. Applying surface entropy reduction to DENV2 protease for crystallization purpose. 179

Figure C.1. Examples of docking result of hit obtained from fragment based screening with different target residues. 184

Figure D.1. Surface comparison of serine proteases' active sites 187

Figure D.2. Modeling of P site macrocyclic inhibitors onto dengue and human serine proteases. 188

Figure D.3. Chemical structures of P site cyclic peptidomimetic inhibitors with P1–P4 macro-cyclization. 189

Figure D.4. Chemical structures of P site cyclic peptidomimetic inhibitors with P1–N terminus capping macro-cyclization. 190

Figure E.1. Inactive derivatives of tri-peptide compounds. 193

Figure E.2. Chemical structures of linear P site inhibitors without c-terminal capping group. 194

List of Third Party Copyrighted Materials

The Following were adapted from journals where permission is required		
Chapter / Figure (Table)	**Publisher**	**License number**
Chapter 1, Figure 1.4	Elsevier	3912640699896
Chapter 1, Figure 1.7	Nature Publishing Group	3913170527164

The Following were adapted from journals where no permission required	
Chapter / Figure (Table)	**Publisher**
Chapter 1, Figure 1.2	National Institute of Allergy and Infectious Diseases
Chapter 1, Figure 1.5	Viruses
Chapter 1, Figure 1.6	IAS–USA Drug Resistance Mutations Group
Chapter 1, Table 1.2	American Chemical Society

List of Abbreviations

AIDS	Acquired immunodeficiency syndrome
AP	Aprotinin
APV	Amprenavir
Apro	Aprotinin derived cyclic peptide
BL	Binding loop
CD	Circular Dichroism
DAA	Direct acting antiviral
DENV	Dengue virus
DRV	Darunavir
ER	Endoplasmic Reticulum
FDA	Food and Drug Administration
FRET	Förster resonance energy transfer
HAART	Highly Active Antiretroviral Therapy
HCV	Hepatitis C virus
HIV	Human immunodeficiency virus
IC50	Cellular half-maximal inhibition constants
ITC	Isothermal titration calorimetry
K_d	Dissociation constants
K_i	Inhibition constants
MD	Molecular Dynamics
NS	Nonstructural protein
PDB	Protein Data Bank
PI	Protease inhibitor
RMSD	Root mean square deviation
RMSF	Root mean square fluctuations
vdW	van der Waals
WT	Wild type
ZIKV	Zika virus

Chapter I

Introduction

1.1 Viral protease as drug target

Viral proteases play an essential role in the maturation of the viral particles and have been shown to be effective targets of anti-viral therapies for human immunodeficiency virus (HIV) and hepatitis C virus (HCV). There are currently nine HIV-1 protease inhibitors (PIs) and five HCV protease inhibitors that are FDA-approved for clinical use, and all are competitive inhibitors binding at the active site. Currently, there is no approved drug against dengue virus; however, the lessons we learned from HIV-1 and HCV should be applicable to dengue virus inhibitor design.

Under the selective pressure of PI-including therapy regimens, viral variants carrying mutations in the protease gene and occasionally in the cleavage sites on the polyprotein impair the inhibitor efficacy. While the protease inhibitors become weak binders, the substrates are still hydrolyzed, skewing the balance between inhibitor binding and substrate processing in favor of the latter and causing drug resistance. Therefore, understanding the mechanism of how drug resistant protease interacts with its substrates and inhibitors is crucial to understand how to avoid drug resistance.

1.2 Human immunodeficiency virus

Acquired immunodeficiency syndrome (AIDS) is a global problem caused by the infection of Human Immunodeficiency Virus 1 (HIV-1) (Barre-Sinoussi et al., 1983), with an estimated 35 million people living with HIV/AIDS and 1.5 million people dying early because of AIDS-related illness (UNAIDS, 2016). HIV-1 primarily infects $CD4^+$ T cells, macrophages and dendritic cells, and the progressive failure of cell-mediated immunity makes infected individuals susceptible to opportunistic infections and eventually causes death.

Currently, Highly Active Antiretroviral Therapy (HAART) is the main therapeutic strategy against HIV-1 infection. HAART consists of at least three drugs from at least two classes: protease and reverse transcriptase inhibitors (Yeni et al., 2002). HAART was designed to reduce the likelihood of developing viral resistance, and has been shown to successfully decrease the mortality and morbidity rate of HIV-1 infected patients. However, drug resistance is still observed in patients under treatment and diminishing drug susceptibility. Further understanding at the molecular level of how resistance happens would be beneficial to the development of more efficient therapies.

1.2.1 HIV-1 viral proteins and viral lifecycle

HIV-1 virus belongs to the family *Retroviridae* and genus *Lentivirus*, the virus has two copies of positive sense 9.2 kb RNA genome (Chiu et al., 1985). The gag gene encodes structural proteins (MA, CA, SP1, NC, SP2, p6), the env gene encodes envelope proteins (gp120, gp41), and the pol gene encodes the enzymes (reverse transcriptase, protease, integrase, Ribonuclease H) (Riemann and Kohler, 1988). Six additional proteins, Vif, Vpr, Nef, Tat, Rev and Vpu, are also encoded by the HIV-1 genome, and the first three proteins are found in the viral particle, but the functions of these proteins are not fully understood (Fig. 1.1).

The life cycle of HIV-1 virus starts with the interaction between the viral envelope complex (gp160) and CD4 receptor of the target cell (Fig. 1.2). The gp160 protein has transmembrane domain gp41 and surface domain gp120. The binding of gp120 to CD4 receptor requires the viral particle to interact with one of the co-receptors (CCR5 or CXCR4) at the same time (Clapham and Weiss, 1997). This binding interaction causes structural rearrangements that allow gp41 to penetrate the cell membrane and induces the fusion of viral and cell membranes (Gallo et al., 2001), followed by genome uncoating in the host cell cytoplasm.

The reverse transcription of the viral RNA genome into double-stranded DNA is mediated by the viral reverse transcriptase. This viral DNA is then transported to the nucleus as a pre-integration complex including the integrase, which inserts the viral DNA into the host cell DNA (Broder et al., 1985). Once the viral DNA is inserted into the host genome, this DNA is called provirus, which can remain inactive for years (Koup, 2001).

The provirus uses host cell RNA polymerase to synthesize viral mRNA with the help of Tat, which is a viral transcriptional transactivator protein (Parada and Roeder, 1996). Viral mRNAs are then exported to the cytoplasm where they are translated into viral Gag and GagPol polyproteins. Two copies of viral RNA genome are encapsidated and packaged with Gag proteins. Viral envelope protein is cleaved by host cell proteases into gp120 and gp41 during transport through the Golgi apparatus (Earl et al., 1991, Chan et al., 1997). These proteins are then incorporated into viral particles while interacting with matrix domain of Gag, which is responsible for plasma membrane targeting (Frankel and Young, 1998, Facke et al., 1993, Bryant and Ratner, 1990).

Upon budding, the newly assembled viral RNA genome, structure proteins and enzymes emerge on the cell surface as an immature viral particle. The Gag and GagPol polyproteins are then processed by the viral protease into mature structure proteins and functional enzymes during this process (Kohl et al., 1988), leading to the maturation of viral particle and next round of infection. Since these cleavage steps are required for viral maturation, HIV-1 protease has been a promising target for drug development.

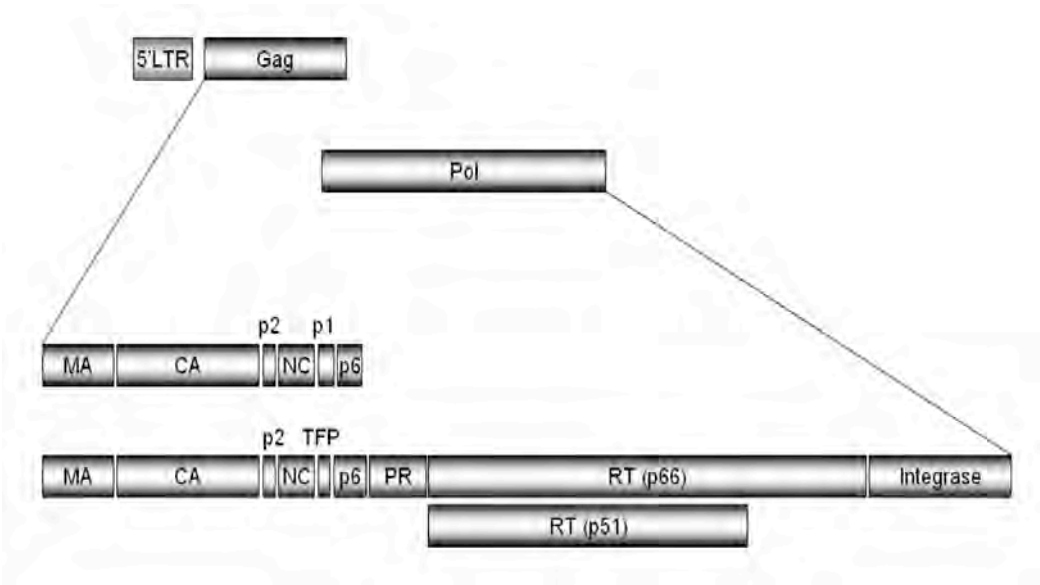


Figure 1.1. HIV-1 Gag and GagPol polyproteins.

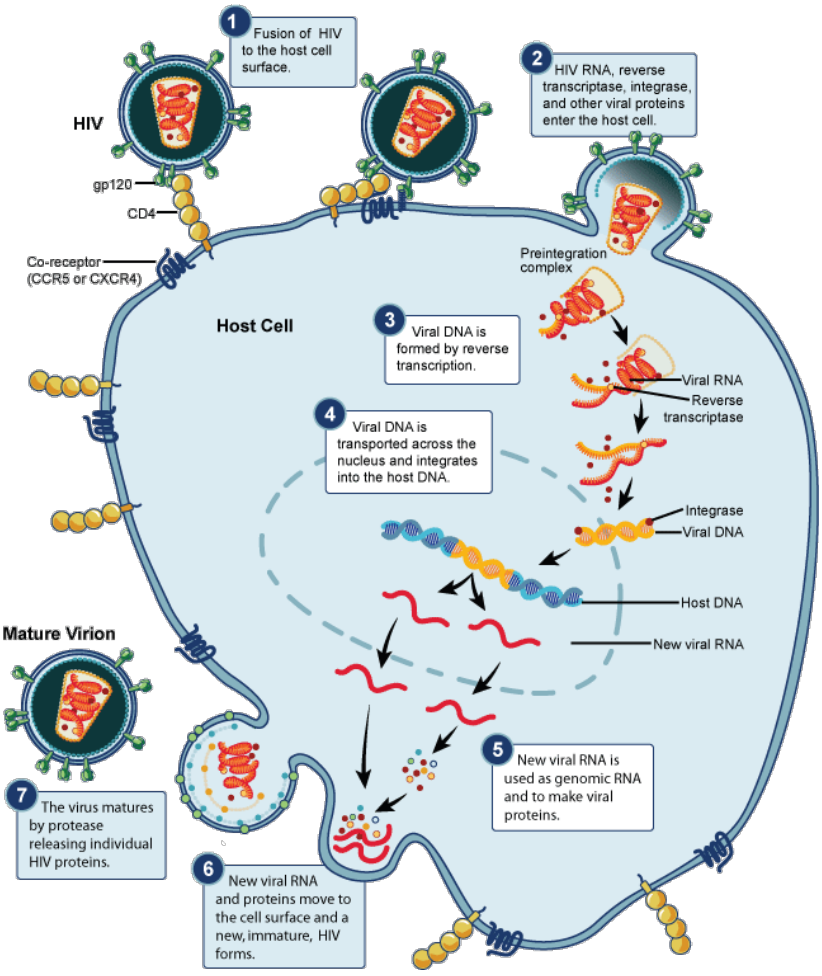


Figure 1.2. HIV-1 life cycle. The steps of HIV-1 viral infection include: binding, fusion, reverse transcription, integration, transcription and translation, assembly, budding and maturation.

Image courtesy: National Institute of Allergy and Infectious Diseases

1.2.2 HIV-1 protease and substrate recognition

HIV-1 protease is an aspartyl protease with a conserved Asp-Thr-Gly triad as observed in other retroviral proteases (Seelmeier et al., 1988). HIV-1 protease has 99 residues and functions as a homodimer (Fig. 1.3), and the triad from each monomer contributes to the formation of the active site (Navia et al., 1989). There are four main structural regions of HIV-1 protease, the active site (residues 23-28, 81-84), the dimer interface (1-10, 90-99), the flap region (40-60) and the core domain of each monomer (residues 11-22, 29-39, 61-80, 85-90). The flap regions of the protease control the entry of ligands and also participate in the substrate binding. HIV-1 protease has the conserved Asp-Thr-Gly triad as observed in other aspartic proteases. The triad from each monomer contributes to form the active site. Aspartic acid 25 is the catalytic residue, and the protease uses an activated water molecule to attack the carbonyl group of the substrate's scissile bond (Navia et al., 1989).

HIV-1 protease cleaves at least ten cleavage sites at Gag and GagPol polyproteins and gives rise to the release of structure proteins, matrix (MA), capsid (CA), nucleocapsid (NC), p6; two spacer peptides p1 and p2; and the enzymes, protease (PR), reverse transcriptase (RT) and integrase (IN) (Kohl et al., 1988) (Fig. 1.1).

HIV-1 protease's substrate residues spanning from P4 to P4' positions have been shown to be required for proper substrate recognition (Weber et al., 1989, Tozser et al., 1991). However, the substrate sequences of different cleavages sites are very diverse (Table. 1.1). How the protease is able to recognize these variable sequences was revealed by crystal structures solved in the Schiffer lab. HIV-1 protease recognizes these

substrates through a three dimensional consensus volume occupied by different substrates when bound to the protease, spanning P4 to P4' substrate positions, which was named the substrate envelope (Fig. 1.4) (Prabu-Jeyabalan et al., 2002). HIV-1 inhibitors protruding out the substrate envelope were found to be more susceptible to drug resistance mutations since protease residues that interact with protruding inhibitor atoms are more important for inhibitor binding than substrate binding (Fig. 1.4) (King et al., 2004, Prabu-Jeyabalan et al., 2006).

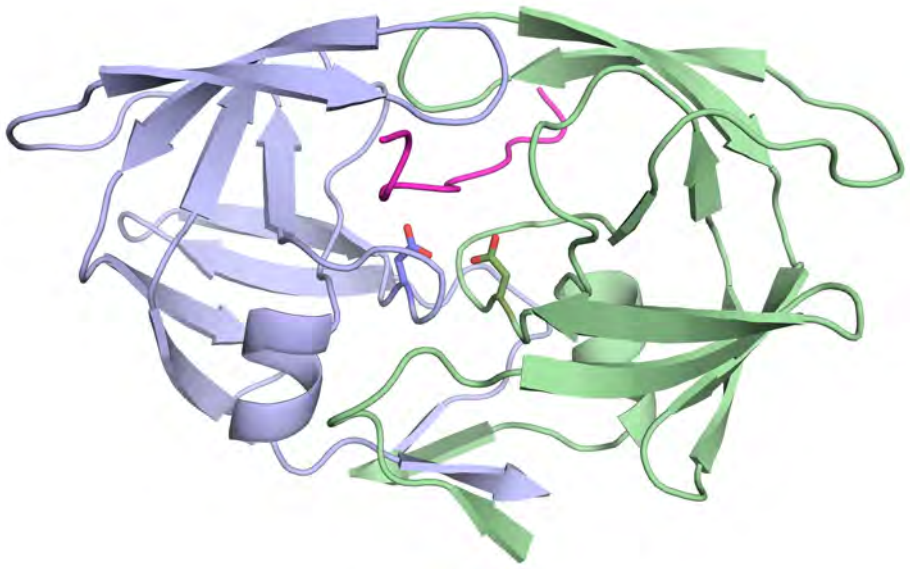


Figure 1.3. Structure of HIV-1 protease in complex with substrate. The two monomers are shown in purple and green, and the substrate in magenta. Catalytic aspartates are shown as sticks.

Table 1.1. HIV-1 protease cleavage site sequences based on the HXB2 sequence.

Gag	**Amino acid sequence**									
	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'
MA/CA	V	S	Q	N	Y	P	I	V	Q	N
CA/p2	K	A	R	V	L	A	E	A	M	S
P2/NC	S	A	T	I	M	M	Q	R	G	N
NC/p1	E	R	Q	A	N	F	L	G	K	I
P1/p6^{gag}	R	P	G	N	F	L	Q	S	R	P

Pol	**Amino acid sequence**									
	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'
NC/TFP	E	R	Q	A	N	F	F	R	E	D
TFP/p6^{pol}	E	D	L	A	F	L	Q	G	K	A
p6^{pol}/PR	V	S	F	N	F	P	Q	I	T	L
PR/RT	C	T	L	N	F	P	I	S	P	I
RT(p51)/RT(p15)	G	A	E	T	F	Y	V	D	G	A
RT/IN	I	R	K	V	L	F	L	D	G	I

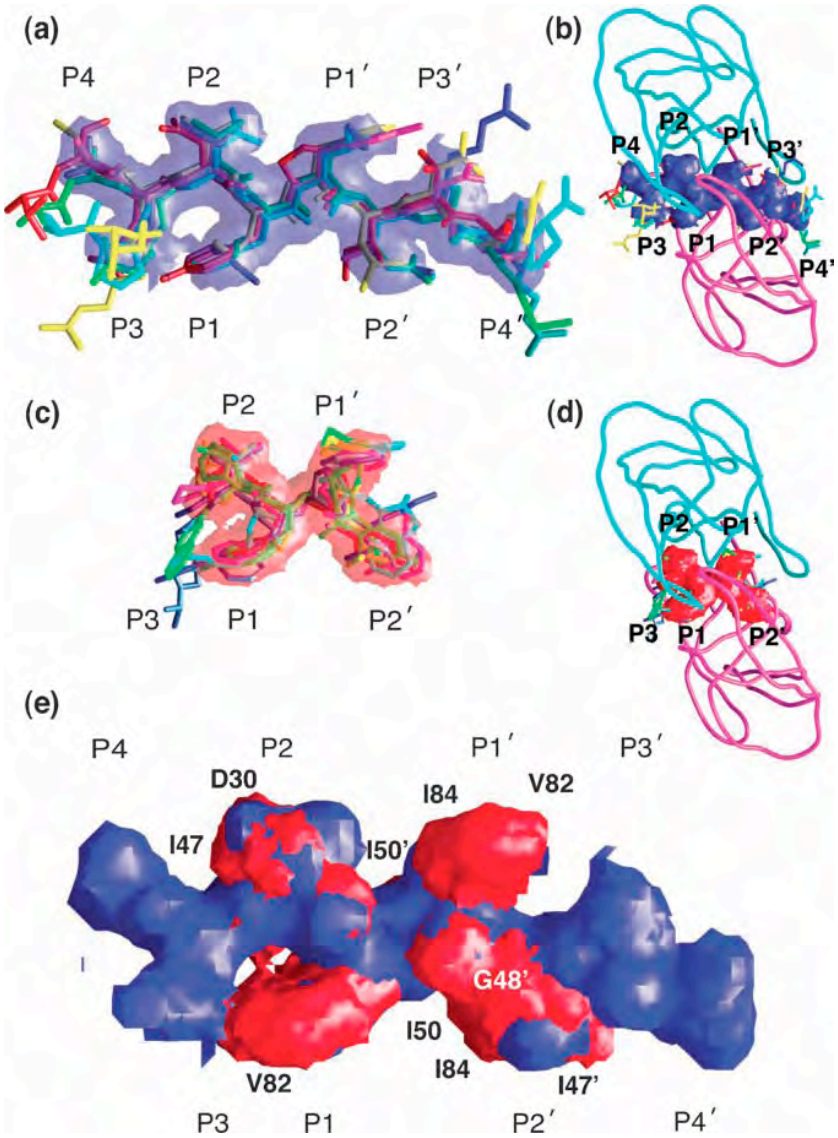


Figure 1.4. Substrate and Inhibitor Envelopes of HIV-1 Protease. (A) Substrate envelope in blue with superposed substrate peptides. (B) Top view of the substrate envelope and the active site of the protease. (C) Inhibitor envelope in red, calculated from the consensus volume occupied by five or more of the inhibitor complexes with active HIV-1 protease. (D) Top view of the inhibitor envelope and the active site of the protease. (E) Superposition of the substrate envelope and inhibitor envelope. Protease residues contacted by the inhibitor that mutate to cause drug resistance are labeled.

1.2.3 HIV-1 protease inhibitors as antivirals and drug resistance

There are currently nine protease inhibitors (PIs) that are FDA-approved for clinical use: saquinavir (SQV), indinavir (IDV), ritonavir (RTV), nelfinavir (NFV), amprenavir (APV), lopinavir (LPV), atazanavir (ATV), tipranavir (TPV) and darunavir (DRV) (Vacca et al., 1994, Turner et al., 1998, Sham et al., 1998, Robinson et al., 2000, Patick et al., 1996, Partaledis et al., 1995, Kempf et al., 1995, Craig et al., 1991) (Fig. 1.5). These drugs are included in highly active antiretroviral therapy (HAART) (Yeni et al., 2002), and all are competitive inhibitors binding at the active site. These inhibitors are all peptidomimetic inhibitors except TPV.

HAART containing protease inhibitors have been shown to successfully decrease viral replication. However, drug resistance mutations decrease the efficacy of the antiviral therapy. Because of the high replication rate of the virus and lack of proofreading function of the viral reverse transcriptase, viral mutations are quickly selected under the drug pressure to confer protease inhibitor resistance(Kaplan et al., 1993, Kaplan et al., 1994, Condra et al., 1995, Molla et al., 1996, Rhee et al., 2003, Shafer, 2006) .

Major primary drug resistance mutations include D30N, V32I, M46I, G48V, I50V/L, V82A/F/I, I84V, which are commonly observed in patients failing antiviral treatment (Gulnik et al., 1995), and certain mutations are selected under specific drug treatment (Fig. 1.6) (Wensing et al., 2014). For example, I50V is commonly observed in patients failing therapy with PIs amprenavir and darunavir (Mahalingam et al., 1999, Colonno et al., 2004), and D30N is observed in patients failing nelfinavir treatment

(Patick et al., 1998). As primary protease mutations may decrease protease activity, secondary mutations are usually selected to partially rescue protease function, such as A71V mutation, which rescues I50V protease's catalytic efficiency (Clemente et al., 2003). The mechanisms of how these protease resistance mutations affect drug binding have been investigated before, e.g., I50V protease makes less vdW interactions with the sulfonyl moiety in APV/DRV compared to those in the WT complexes (Mittal et al., 2013). Beside mutations of the viral protease, drug resistance mutations at the substrate cleavage sites were observed to co-exist with protease mutations, indicating protease–substrate co-evolution [Kolli, 2009 #100].

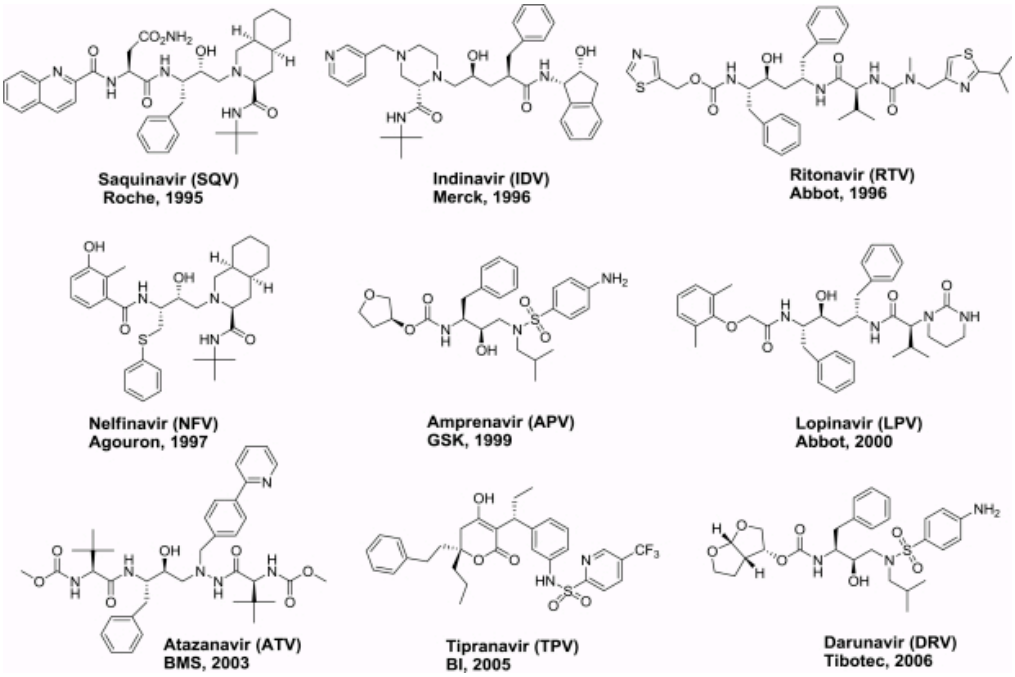


Figure 1.5. Chemical structures of FDA approved HIV-1 protease inhibitors.

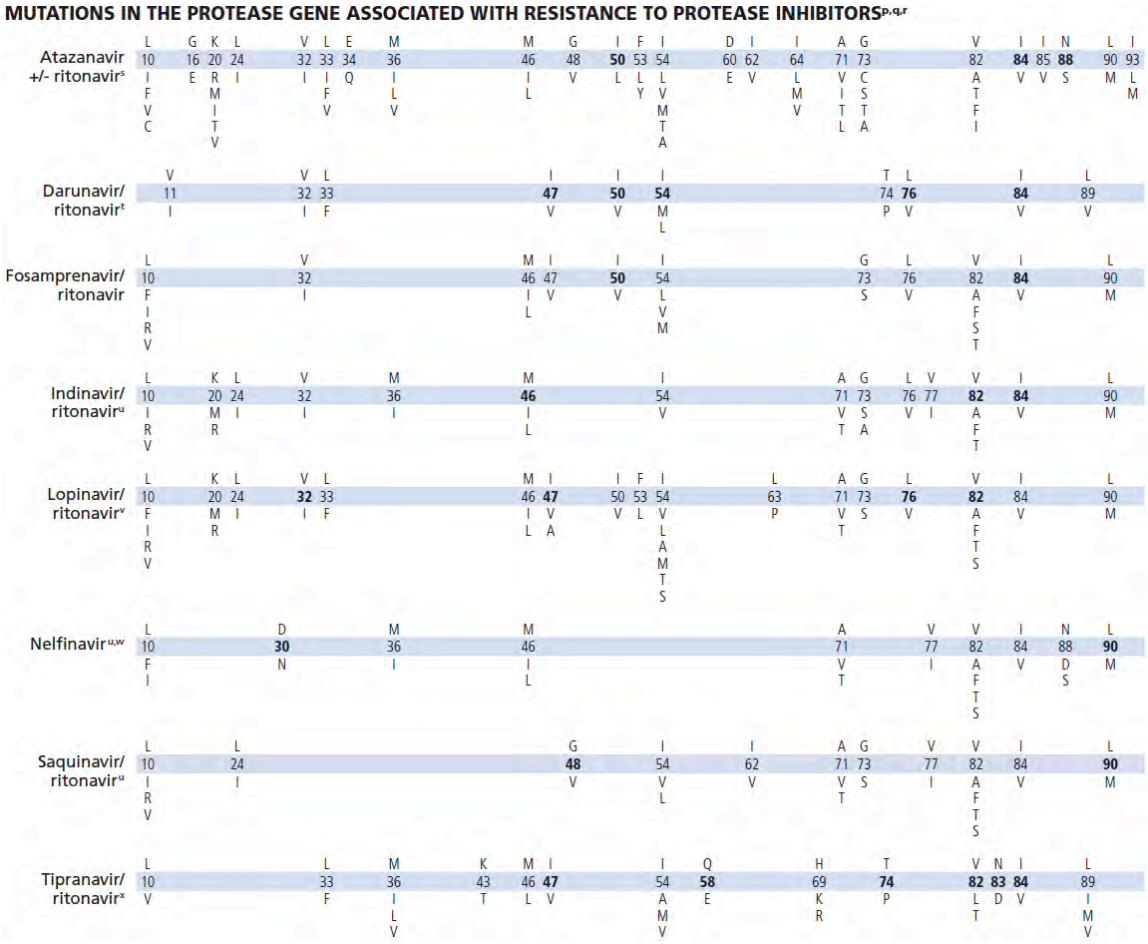


Figure 1.6. HIV-1 protease mutations selected by different classes of protease inhibitors.

Reprinted with permission from Wensing, A. M., Calvez, V., Gunthard, H. F., Johnson, V. A., Paredes, R., Pillay, D., Shafer, R. W. & Richman, D. D., 2014. 2014 Update of the drug resistance mutations in HIV-1. *Top Antivir Med,* 22**,** 642-50. Copyright 2014 IAS–USA Drug Resistance Mutations Group.

1.2.4 HIV-1 protease-substrate co-evolution

Several mutations in Gag cleavage sites co-evolve with primary protease mutations, and contribute to resistance. Based on the viral sequences retrieved from patients, mutations at the p1-p6 cleavage site are statistically associated with I50V or D30N/N88D protease mutation (Kolli et al., 2009, Kolli et al., 2006), and mutations within NC-p1 cleavage site correlate with V82A protease resistance mutation (Zhang et al., 1997, Maguire et al., 2002).

These co-evolved substrates have been reported to rescue protease cleavage activity. In particular, for the I50V protease and p1-p6 substrate co-evolution, Gag L449F mutation rescues the protease activity by 10 fold, while P453L, despite being distal from the catalytic site, causes a 23-fold enhancement (Maguire et al., 2002). The molecular basis for the selection advantage of these correlated mutations and the mechanism by which the compensatory mutations restore substrate recognition in drug resistance was not clear.

By studying the molecular mechanism of protease-substrate co-evolution, I aimed to understand how HIV-1 mutant protease is able to recognize and cleave its substrates while the inhibitor binding is impeded. The results help us re-evaluate the idea of substrate envelope and design inhibitors less likely to be susceptible to the selection of resistance mutations. This project will be discussed in Chapter II.

1.3 Dengue virus

Dengue fever, caused by infection with dengue virus (DENV), is a worldwide infectious disease endemic in more than 110 countries. Approximately 390 million people are infected yearly, with 96 million of infected developing disease symptoms and about 20,000 annual deaths (WHO, 2009, Bhatt et al., 2013).

Mosquito *Aedes aegypti* is the major vector of dengue virus, and due to the narrow temperature tolerance of *Aedes*, DENV is transmitted predominantly in tropical and subtropical region. Eighty percent of dengue infections are asymptomatic, and for those who are symptomatic, symptoms range from undifferentiated fever to dengue hemorrhagic fever and dengue chock syndrome [WHO, 1997 #813] (or severe dengue for the last two) [WHO, 2009 #103], with severe symptoms including plasma leakage, haemoconcentration, abnormalities in homeostasis and shock.

There are four serotypes of dengue virus (DENV 1-4), and each serotype shares 65-70% sequence identity of the genome (Rico-Hesse, 1990). Ninety percent of dengue hemorrhagic fever cases were observed in patients with secondary infections of heterologous serotypes (Green and Rothman, 2006).

1.3.1 DENV lifecycle and viral proteins

Dengue virus (DENV), a member of the family *Flaviviridae* (which includes other viruses like hepatitis C virus, yellow fever virus, Zika virus), is an enveloped virus with a positive single-strand 11,000 nucleotide RNA genome. Dengue RNA genome has one open reading frame that encodes a single polyprotein of three structural proteins (C, prM, and E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Chambers et al., 1990a). The role of the structural proteins include: the C protein which is required for the formation of nucleocapsid, prM inhibits the pre-maturation of viral particle, and E protein is required for the receptor binding and fusion activity (Zhang et al., 2004, Kuhn et al., 2002). The nonstructural proteins are mainly enzymes and include: N terminus of NS3 protein is the protease domain and C terminus is the ATP-dependent helicase domain. NS2B is the cofactor of NS3 protease and is required for proper protease function (Yusof et al., 2000, Chambers et al., 1990a, Bera et al., 2007). NS5 has the functions of methyltransferase and RNA-dependent RNA polymerase, and NS4B has been shown to inhibit interferon signal transduction. The functions of NS1, NS2A, and NS4A are incompletely understood (Ross, 2010). Due to the known functions of NS3 and NS5, these two proteins have been targeted for inhibitor design (Lim et al., 2013).

The structures of dengue viral particle (3J27) (Zhang et al., 2013), envelope protein E (1TG8) (Zhang et al., 2004), pre-membrane protein prM (3C5X, 3C6E) (Li et al., 2008), capsid protein C (1R6R) (Ma et al., 2004), NS3/2B protease (2FOM) (Erbel et al., 2006) (3L6P) (Chandramouli et al., 2010) (3U1I) (Noble et al., 2012), NS2B with full

length NS3 (2VBC) (Luo et al., 2008), NS3 helicase/NTPase (2BHR, 2BMF) (Xu et al., 2005), methyltransferase (2P1D, 2P40) (Egloff et al., 2002, Egloff et al., 2007), and RNA-dependent RNA polymerase (2J7U, 2J7W) (Yap et al., 2007) have been determined.

DENV enters the target cell through clathrin-mediated endocytosis, then the viral particle is delivered to Rab5-positive endosome, which will further mature to Rab7-positive endosome (Fig. 1.7). Acidification of the endosome induces the fusion of viral and vesicular membranes, after which the nucleocapsid is released and uncoated (Clyde et al., 2006). Dengue RNA genome is translated into a single polyprotein on the endoplasmic reticulum membrane, where this polyprotein gets processed at the cytoplasmic side of ER membrane by dengue NS2B-NS3 protease and at the lumen side by the host cell peptidase (Chambers et al., 1990a), giving rise to functional viral proteins.

Non-structural proteins initiate the synthesis of the RNA genome, and then the genome is packaged by C protein to form the nucleocapsid. The prM/E heterodimeric protein forms trimers in the lumen of ER membrane and encapsulates the nucleocapsid to form the immature viral particle. This immature viral particle then travels through the Golgi apparatus where the dropping of pH induces the dissociation of prM/E heterodimer, which allows human furin protease to cleave between M protein and pr peptide. This cleavage process allows the maturation of viral particle (Rodenhuis-Zybert et al., 2010).

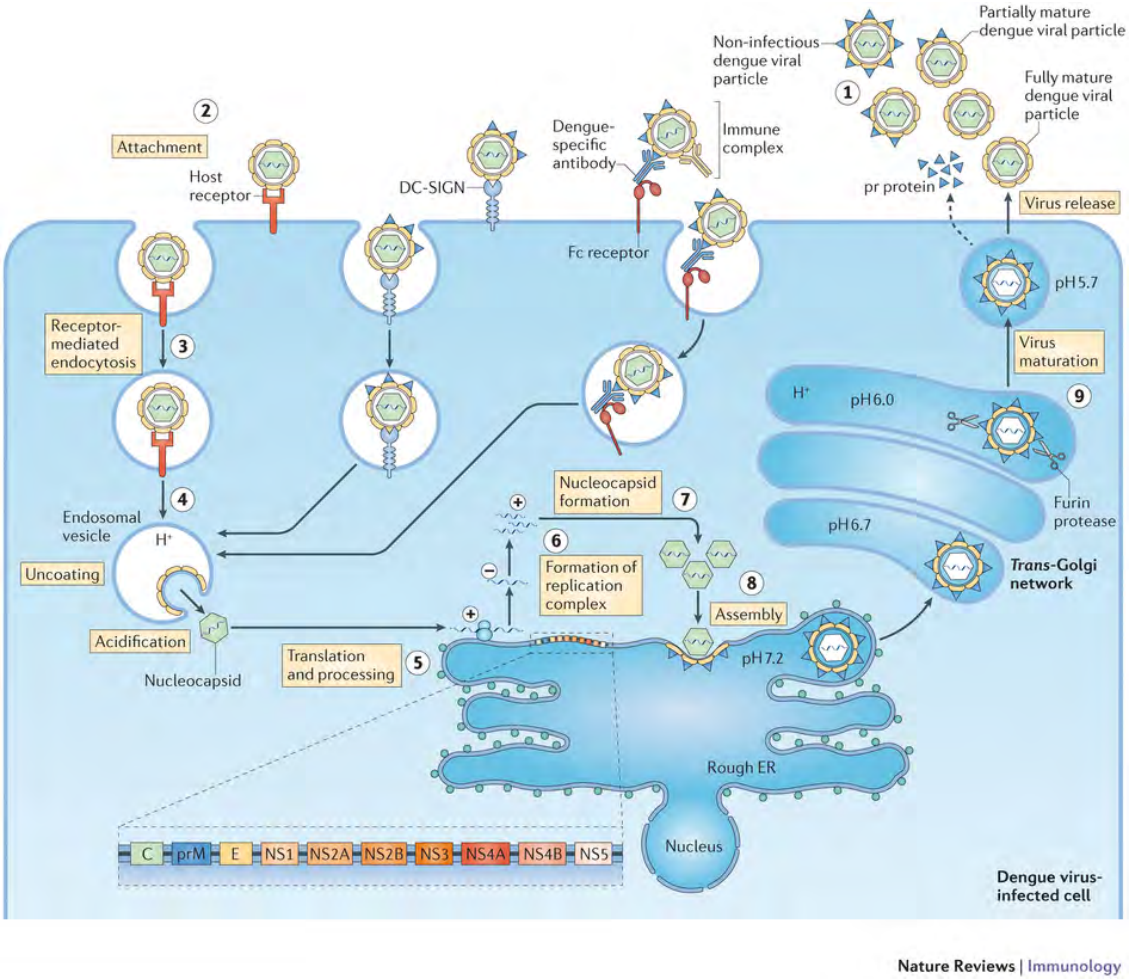


Figure 1.7. The life cycle of dengue virus. The steps of dengue viral infection include: attachment, endocytosis, uncoating, translation and processing, genome replication, assembly, budding and maturation.

Reprinted with permission from Screaton, G., Mongkolsapaya, J., Yacoub, S. and Roberts, C., 2015. New insights into the immunopathology and control of dengue virus infection. *Nature Reviews Immunology*, *15*(12), pp.745-759. Copyright 2015 Nature Publishing Group.

1.3.2 Therapeutic strategies against dengue virus

The first dengue vaccine CYD-TDV (Dengvaxia) (Sanofi Pasteur) was approved in Mexico in December 2015. This vaccine is a formulation of four chimeric yellow fever 17D vaccine viruses, which was designed to express the membrane and envelope proteins of all four serotypes of DENV (Simmons, 2015).

Unlike HIV-1 and HCV, there is no approved drug (direct acting antiviral) against dengue virus. The main strategy to reduce the spread of dengue virus is vector control. Inhibitors targeting dengue protease, helicase, polymerase and methyl transferase have been investigated (Lim et al., 2013), but there are no drugs in development or clinical trials. My research was focused on the dengue NS3/2B protease substrate recognition and inhibitor design.

1.3.3 Dengue NS3/2B protease and structural features

Dengue NS2B-NS3 protease is a serine protease (Ser135 is the catalytic residue), which belongs to the chymotrypsin family with a classic Ser-His-Asp catalytic triad (His51-Asp75-Ser135) (Bera et al., 2007). The hydrophilic core of NS2B cofactor (cNS2B; amino acids 1394-1440) is required for the proper function of NS3 protease (NS3pro185; amino acids 1476-1660) and participates in substrate recognition (Yusof et al., 2000, Noble et al., 2012).

Crystal structures of serotype 1 (PDB: 3L6P) (Chandramouli et al., 2010) and serotype 2 (PDB: 2FOM) (Erbel et al., 2006) dengue NS3/2B protease, and serotype 4 NS2B with full length NS3 protein (PDB: 2VBC) (Luo et al., 2008) have been determined all in apo form. For serotype 3, there is one peptidomimetic inhibitor bound (PDB: 3U1I) and one aprotinin bound (PDB: 3U1J) NS3/2B protease complex structure (Noble et al., 2012). Apo NS3/2B protease adopts an opened conformation, which is the inactive form (Fig. 1.8). In the ligand-bound DENV3 NS3/2B protease structure (closed, active form), the co-factor 2B loops around the NS3 protease domain and participates in the formation of S2 and S3 pockets of the active site (Fig. 1.8). The NS3 protease domain adopts a chymotrypsin-like conformation with two β-barrels and the catalytic triad located in between. Asp129 in S1 pocket and Asp75 in S2 pocket may confer the protease's preferences of basic residues at these two positions (Erbel et al., 2006).

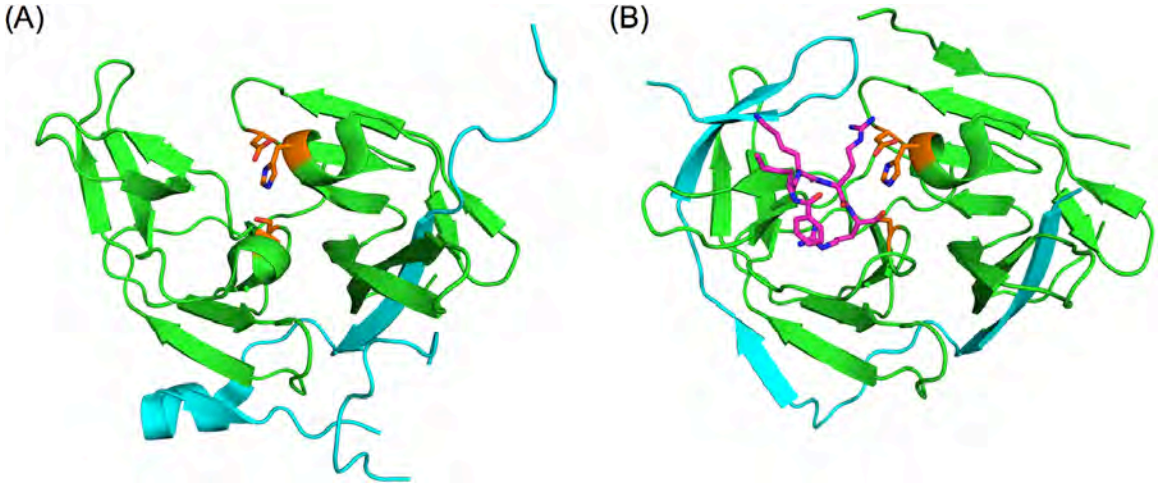


Figure 1.8. Crystal structure of dengue NS3/2B protease. (A) Apo-structure of DENV2 protease (PDB: 2FOM). (B) Peptidiomimetic inhibitor bound DENV3 protease structure (PDB: 3U1I). NS3 is in green, NS2B in cyan, catalytic triads in orange and ligand in magenta.

1.3.4 Dengue NS3/2B protease substrate specificity

Dengue protease is responsible for processing eight of the thirteen polyprotein cleavage sites (C, NS2A, NS2A-NS2B, NS2B-NS3, NS3, NS3-NS4A, NS4A, NS4B-NS5) (Falgout et al., 1991), and these cleavage steps are required for the maturation of the viral particle, making dengue NS2B-NS3 protease a promising target for drug development.

Previously, two basic residues at P2 and P1 sites and a small and polar residue at P1' site were found as optimal for substrate recognition and processing across all four serotypes (Chambers et al., 1990b). However, there are non-basic residues at P2 position and residues at P5 to P3 and P' sites are quite diverse (Fig. 1.9) (Strain name: DENV1: Nauru/West Pac/1974, DENV2: Australia TSV01/1993, DENV3: Singapore 8120/1995, DENV4: Taiwan 2K0713/2000). How dengue protease is able to recognize these diverse substrates is not fully understood. Meanwhile, dengue protease has similar substrate sequence preferences as human serine proteases (Furin RXRR, thrombin P1 R, trypsin P1 R), making the design of dengue protease inhibitor more challenging. In contrast, dengue protease does not have similar P' site substrate sequence preferences with other serine proteases. Thus revealing P' site protease–substrate interactions would facilitate the design of more specific P' site inhibitors.

Linear peptides corresponding to dengue NS3/2B protease cleavage sites have been used to investigate the residue preference at each position, however, the results were quite different depending on experimental approaches. For example, acidic residues (Asp, Glu) were found to be favorable at P2' position when screened with peptides nKRR-

XOXX (O is fixed amino acid and X stands for amino acid mixtures while cysteine is excluded and methionine is replace by isostere norleucine) (Li et al., 2005a), but Glu was found to be the least favorable when screened with peptides GLKRGOAK (O is fixed P2' amino acid) (Shiryaev et al., 2007). This discrepancy suggests that background sequence plays a key role in determining the contribution of a specific residue: using amino acid mixtures as a background can mask the effect of specific residues at a given position, such as due to intra-molecular interactions, while using a fixed background sequence may miss extra interdependencies.. Further, these tested sequences may not exist in nature. Linear substrate peptides spanning P4 to P4' have been used to investigate substrate cleavage (Chanprapaph et al., 2005). However, with diverse residues at P site (both basic and non-basic P2 residues, distinct P3 and P4 residues), to distinguish the contributions of P or P' sites is difficult.

General serine protease inhibitor aprotinin, which was used in cardiac surgery to reduce bleeding (Mahdy and Webster, 2004), is a potent inhibitor against DENV2 protease (26 nanomolar K_i) (Mueller et al., 2007). Unlike P site inhibitors, aprotinin occupies the dengue protease active site from P3 to P4' positions (Noble et al., 2012). Based on the published aprotinin–dengue protease complex structure (3U1J), the P' site substrate's binding mode in the protease active site was revealed, making this structure a useful tool to investigate the P' site inhibitors.

I took advantage of the high-affinity binding and structural availability of aprotinin (3U1J) and screened the binding loop with corresponding P1 to P4' substrate sequences (the sequences are conserved among DENV3 genotypes: China/80-2/1980,

Philippines/H87/1956, Singapore/8120/1995, SriLanka/1266/2000 and Martinique/1243/1999) to investigate how diverse P’ site substrate sequences affect binding interactions with the protease. This project will be discussed in Chapter III.

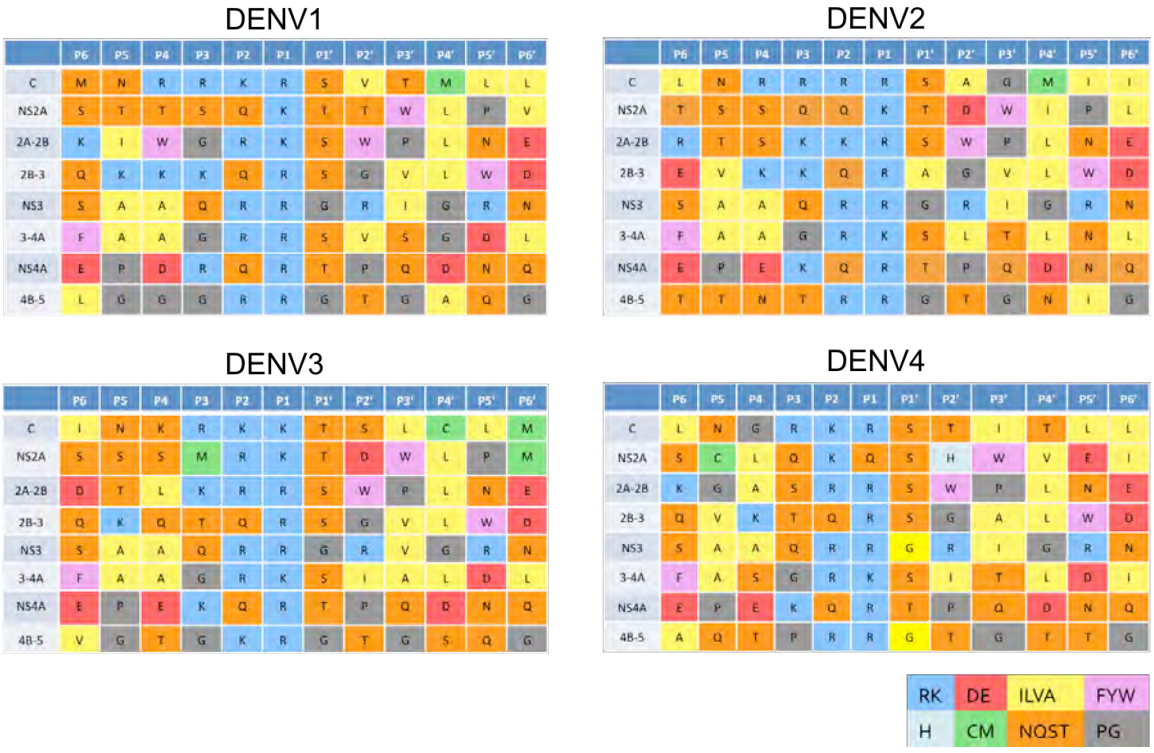


Figure 1.9. Dengue NS3/2B protease cleavage site sequences. Residues are colored based on the properties of side chains.

1.3.5 Inhibitors targeting dengue NS3/2B protease

Peptidomimetic dengue protease inhibitors have been designed mainly based on P site substrate sequences, which usually have basic P2 and P1 residues (Yin et al., 2006a, Yin et al., 2006b, Nitsche et al., 2011, Nitsche et al., 2012, Bastos Lima et al., 2015) (Table 1.2). These inhibitors were improved to nanomolar level binding affinity according to a recently published study (Behnam et al., 2015). Since dengue protease has similar substrate sequence preferences as human serine proteases do (furin RXRR, thrombin P1 R, trypsin P1 R), these inhibitors may not be specific to dengue protease. Therefore, further optimization of these inhibitors is required.

Besides peptide mimicking strategy, small molecule library screening has been applied to targeting dengue NS3/2B protease, and several hits with micromolar affinity were found (Ganesh et al., 2005, Cregar-Hernandez et al., 2011, Boonyasuppayakorn et al., 2014, Yang et al., 2014). Due to the difficulty of getting complex crystal structures of these inhibitors with dengue protease, how these molecules bind to the protease is not clear.

Fragment based drug design was also applied to investigate dengue NS3/2B protease inhibitors, and compounds with micromolar potencies were discovered (Knehans et al., 2011). These compounds resemble dengue protease substrates' basic P2 and P1 residues, therefore, further optimization is required to increase the specificity of these compounds against dengue protease.

Other than competitive dengue protease inhibitors, allosteric inhibitors were shown to be able to affect the proper structural arrangement of co-factor NS2B (Yildiz et al., 2013), which is required for the function of NS3 domain (Yusof et al., 2000).

Overall, further optimizations are required to increase P site inhibitors' specificity to dengue protease. Since dengue protease does not share similar P' site substrate sequence preferences with other serine proteases, targeting S' pocket is a strategy to design more specific dengue protease inhibitors. I have designed cyclic peptides as dengue NS3/2B protease inhibitors based on the favorable interactions I identified at this pocket, and this project will be discussed in Chapter IV.

Table 1.2. Selected peptidiomimetic dengue NS3/2B protease inhibitors.

number	structure	inhibition potential (DENV-2)
6	Bz-nKRR-H	K_i = 5.8 μM
11	Bz-nKR(4-Me)F-H	K_i = 6.0 μM
12	Bz-nKRW-H	K_i = 7.5 μM
13	Bz-nKR(4-guanidinyl)F-H	K_i = 2.8 μM
14	Bz-KRR-H	K_i = 1.5 μM
15	4-Phenylphenylacetyl-KKR-H	K_i = 12.2 μM
16	2-Phenylacetyl-KRR-H	K_i = 6.7 μM
17	Itz-CTRRGLPVCGSEE GPTNCGRRS	K_i = 1.4 μM
18	Itz-CRRSWPVCTRRS GRRSEESGCVPLG	K_i = 3.0 μM
Tyrothricin	mixture of peptides	K_i = 3.0 μM
Protegrin-1	RGGRLCYCRRRFCVCVGR	K_i = 5.85 μM
Retrocyclin-1	GICRCICGRGICRCICGR	IC_{50} = 46.1 μM (20 °C) IC_{50} = 14.1 μM (40 °C)
Plectasin	linear, 40 amino acids	K_i = 5.03 μM
Protegrin-1-MAP30-Plectasin	fusion peptide, 37.7 kDa	IC_{50} = 0.5 μM
19	CGYKLC	K_i = 12.5 μM
20	CGKRKLC	K_i = 2.5 μM
21	CGKRKSC	K_i = 1.4 μM
22	CAGKRKSG	K_i = 2.2 μM
23	Ac-RRRRHWCW-NH_2	K_i = 0.3 μM
24	WYCW-NH_2	K_i = 4.8 μM

Reprinted with permission from Nitsche, C., Holloway, S., Schirmeister, T. and Klein, C.D., 2014. Biochemistry and medicinal chemistry of the dengue virus protease. *Chemical reviews*, *114*(22), pp.11348-11381. Copyright 2014 American Chemical Society.

1.4 Thesis scope

This thesis attempts to fill the gaps in molecular level understanding of how drug resistance happens in viral proteases and how protease−substrate recognition can be applied to drug design, more specifically how HIV-1 protease−substrate co-evolution contributes to drug resistance, how dengue protease recognizes diverse P' substrate sequences, and how this knowledge can be used in inhibitor design targeting dengue protease.

First, I demonstrate that co-evolved p1-p6 substrates rescue the HIV-1 I50V protease's binding activity by forming more vdW contacts and hydrogen bonds, and that co-evolution restores the dynamics at the active site for all three mutant substrates (Chapter II).

To apply the principles and strategies we learned from HIV-1 structural studies to dengue protease, I elucidated the dengue protease's P' site substrate recognition patterns and designed inhibitors based on the favorable interactions at this pocket. First, I used aprotinin as a platform to investigate protease–substrate interactions. I presented that P' side residues significantly modulate substrate affinity to protease, and the contributions by distinct substrate residues to binding (Chapter III). Next, I investigate P' site inhibitors derived from aprotinin binding loop, through optimizing the length and sequence of the peptides; the tightest cyclic peptide reached low micromolar potency against DENV3 protease (Chapter IV).

How protease's substrate recognition pattern and molecular dynamics simulations can benefit drug design will be discussed in chapter V. Also, how to improve dengue

protease inhibitor and the implications for evolving virus based on dengue research will be covered in this chapter.

Chapter II

Structural Basis and Distal Effects of Gag Substrate Co-evolution in Drug Resistance to HIV-1 Protease

PREFACE

Chapter II has been previously published as:

Aysegül Özen[1], **Kuan-Hung Lin**[1], Nese Kurt Yilmaz[1], and Celia A. Schiffer. **Structural Basis and Distal Effects of Gag Substrate Co-evolution in Drug Resistance to HIV-1 Protease.** Proc Natl Acad Sci U S A. 2014 Nov 11;111(45): 15993-8. doi: 10.1073/pnas.1414063111.

[1]Equal contributors

Author contributions: A.Ö., N.K.Y., and C.A.S. designed research; **K.-H.L.** and A.Ö. performed research; **K.-H.L.**, A.Ö., and N.K.Y. analyzed data; and **K.-H.L.**, A.Ö., N.K.Y., and C.A.S. wrote the paper.

Contributions from Kuan-Hung Lin: I performed protein expression and purification and solved six crystal structures for this study with the guidance of my mentor Celia Schiffer and my thesis research committee. I performed structural analysis and I created figures and tables for the manuscript. Aysegül Özen, Nese Kurt-Yilmaz and Celia Schiffer guided interpretation of the data. Aysegül Özen, Nese Kurt-Yilmaz and I wrote the manuscript, and Celia Schiffer provided editorial assistance.

2.1 Abstract

Drug resistance mutations in response to HIV-1 protease inhibitors are selected not only in the drug target but elsewhere in the viral genome, especially at the protease cleavage sites in the Gag polyprotein. To understand the molecular basis of this protease–substrate coevolution, we solved the crystal structures of drug resistant I50V/A71V HIV-1 protease with p1-p6 substrates bearing coevolved mutations. Analyses of the protease–substrate interactions reveal that compensatory co-evolved mutations in the substrate do not restore interactions lost due to protease mutations, but instead establish novel interactions that are not restricted to the site of mutation. Mutation of a substrate residue has distal effects on other residues' interactions as well, including through the induction of a conformational change in the protease. Additionally, molecular dynamics simulations suggest that restoration of active site dynamics is an additional constraint in the selection of coevolved mutations. Hence, protease–substrate co-evolution permits mutational, structural, and dynamic changes via molecular mechanisms that involve distal effects contributing to drug resistance.

2.2 Introduction

Resistant pathogens evolve under the selective pressure of drug therapies, commonly by acquiring mutations in the drug target (Yun et al., 2008, Janne et al., 2009, Ali et al., 2010, Theuretzbacher and Mouton, 2011). Most of these mutations cluster around the drug binding site and alter key interactions between the drug and its target. Strikingly, mutations in other off-target proteins have also been reported to contribute to drug resistance (Kern et al., 2000, Kolli et al., 2006, Dam et al., 2009, Li and Nikaido, 2009) where the mechanism of resistance is not as straightforward to rationalize. In the case of HIV-1, mutations both in the target protease gene and elsewhere in Gag confer resistance to protease inhibitors (PIs). There are currently nine protease inhibitors (PIs) that are FDA-approved for clinical use including in highly active antiretroviral therapy (HAART) (Debouck, 1992), and all are competitive inhibitors binding at the active site.

HIV-1 protease is a key antiviral drug target due to its essential function of processing Gag and Gag-Pol viral polyproteins in viral maturation (Chou et al., 1996, Pettit et al., 1998, Sadler and Stein, 2002). Under the selective pressure of PI-including therapy regimens, viral variants carrying mutations in the protease gene and occasionally in the cleavage sites on the Gag polyprotein are populated impairing the inhibitor efficacy. While the PIs become weaker binders of these resistant protease variants, the substrates are still hydrolyzed (Kantor et al., 2002, Rhee et al., 2003), skewing the balance between inhibitor binding and substrate processing in favor of the latter. Earlier work from our group revealed the molecular determinants of this fine balance and formulated the substrate envelope hypothesis to effectively explain the molecular

mechanism of resistance due to primary active site mutations (King et al., 2004). Among primary protease mutations, I50V is commonly observed in patients failing therapy with PIs APV and DRV (Partaledis et al., 1995, Van Marck H, 2007, Vermeiren et al., 2007). Residue 50 is located at the flap tip of the flexible loop (50s loop) that controls the access of substrates and competitive inhibitors to the protease active site. In addition to conferring resistance to PIs, I50V mutation also impairs substrate processing (Maguire et al., 2002). The loss of catalytic efficiency with I50V is compensated by secondary mutations, in particular A71V (Nijhuis et al., 1999), which is observed in more than 50% of patient sequences bearing I50V (Rhee et al., 2003).

Several mutations in Gag cleavage sites co-evolve with primary protease mutations, and contribute to resistance (Doyon et al., 1996, Zhang et al., 1997, Mammano et al., 1998, Bally et al., 2000, Feher et al., 2002, Kolli et al., 2006, Dam et al., 2009). Particularly, mutations in the p1-p6 cleavage site are statistically associated with I50V protease mutation in the viral sequences retrieved from patients (Fig. 2.1) (Kolli et al., 2009). Based on catalytic efficiency (kcat/Km), Gag L449F mutation rescues the protease activity by 10 fold, while P453L, despite being distal from the catalytic site, causes a 23-fold enhancement (Maguire et al., 2002). However, the molecular basis for the selection advantage of these correlated mutations and the mechanism by which the compensatory mutations restore substrate recognition in drug resistance is not clear. In this study, we report the structural basis for the co-evolution of I50V/A71V protease with the p1-p6 substrate. Through a series of co-crystal structures the Gag mutations L449F and P453L were shown to enhance van der Waals (vdW) interactions between the substrate and

mutant protease, while R452S results in an additional hydrogen bond. Unexpectedly, the P453L substrate mutation causes a conformational change in the protease flap loop, revealing the molecular mechanism by which this distal substrate mutation is able to enhance substrate–protease interactions. In addition, molecular dynamics simulations suggest that co-evolution restores the dynamics at the active site, a key aspect of substrate recognition and turnover that is largely uncharacterized.

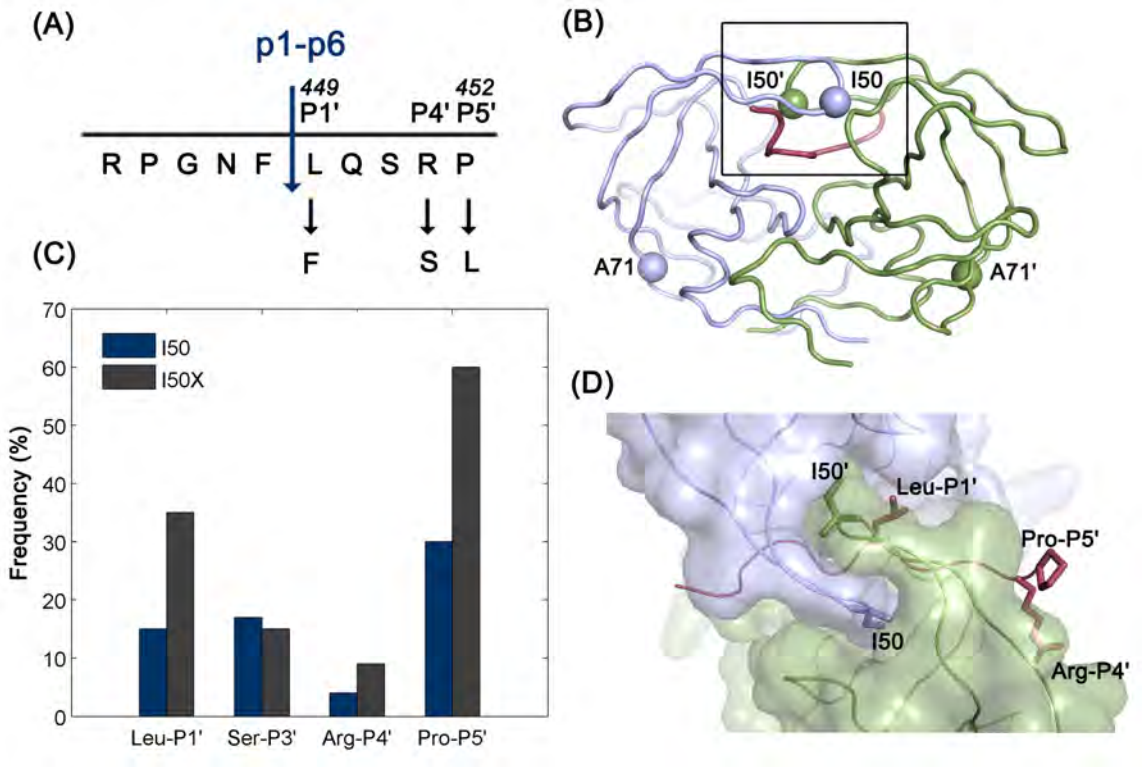


Figure 2.1. HIV-1 protease and p1-p6 cleavage site co-evolution with I50V primary drug resistance mutation. (A) p1-p6 cleavage site sequence and the most common co-evolution mutations at P1', P4', and P5' sites. Gag numbering 449 corresponds to P1' site. (B) Residues 50 and 71 are indicated as spheres on the homodimeric HIV-1 protease structure; two protease monomers and the bound substrate are colored light purple, green, and red, respectively. (C) Frequency of mutations in the p1-p6 cleavage site without (dark blue) and with (gray) any mutations at position 50 of the protease. The difference is statistically significant for LP1', RP4', and PP5'. (D) Side chains of the substrate residues LP1', RP4', and PP5' and the protease residue I50 are shown as sticks. Monomers of HIV-1 protease are represented as light purple and green surfaces and the substrate is depicted as a red ribbon.

2.3 Results

To understand how HIV-1 protease–substrate co-evolution alters binding interactions, a series of co-crystal structures of WT and I50V/A71V protease with p1-p6 substrate variant peptides (WT, L449F, R452S, P453L) were determined. All structures were solved to resolution 1.50–2.14 Å (Table 2.1). In addition to an overall structural comparison, the alterations in co-evolved substrates' fit within the substrate envelope, and details of molecular interactions between the protease and substrate were analyzed. Finally, MD simulations were carried out to reveal any dynamic changes in co-evolved complexes relative to WT.

2.3.1 The overall structure and substrate envelope is conserved in co-evolved complexes

The overall backbone conformation of substrate–protease complexes is conserved in all co-evolved structures. When the structures are superposed onto the WT_{WT} complex (based on the structurally conserved regions, residues 24-26 and 85-90) (Ozen et al., 2011), the RMSD of Cα atoms are within 0.48 Å. Minor structural changes in some co-evolved structures are located mainly at crystal contact surfaces (Fig. 2.2). One notable exception is in the $I50V_{PP5'L}$ structure, where a peptide bond flips to change the flap loop conformation relative to the WT structure in one monomer. Similar to the protease, all substrate residues from P4 to P4' overlap well, with the exception of P4' arginine having an altered orientation in WT_{WT} complex (Fig. 2.3). The distal P5 and P5' residues are

more flexible, P5 is often disordered, while P5’ has altered conformations in both LP1'F and PP5'L substrates as detailed below.

The protease and substrate mutations were evaluated for their effect on the substrate’s fit to the substrate envelope, which is key to recognition by the protease (Fig. 2.4A,B). The substrate volume within the envelope, V_{in}, is the largest in the WT_{WT} complex, indicating that the WT substrate fills the substrate envelope better than the other substrate variants. The I50V/A71V protease mutations worsen the fit of the WT substrate within the substrate envelope resulting in a 53 $Å^3$ (5.0%) reduction in V_{in}. The LP1'F substrate partially restores this loss in V_{in} and the coevolved substrate better fills the substrate envelope by 25 $Å^3$ (2.5%). Overall, the co-evolved substrates maintain a comparable fit within the substrate envelope regardless of whether the protease carries the I50V/A71V mutations or not, supporting that substrate envelope is the recognition motif, and coevolved mutations at the cleavage site do not drastically alter the fit within the substrate envelope.

Table 2.1. Crystallographic statistics for HIV-1 protease with p1-p6 substrate co-crystal structures.

	WT_{WT}	$WT_{LP1'F}$	$WT_{RP4'S}$	$WT_{PP5'L}$	$I50V_{WT}$	$I50V_{LP1'F}$	$I50V_{RP4'S}$	$I50V_{PP5'L}$
PDB code	1KJF	4OBH	4QJ9	4QJA	4QJ2	4QJ6	4QJ7	4QJ8
resolution (A°)	2.00	1.85	1.84	1.54	2.14	1.50	1.67	2.00
space group	P212121	P21	P212121	P212121	P21	P21	P21	P21
a (A°)	51.3	51.1	50.9	51.0	52.2	51.4	51.6	51.3
b (A°)	59.1	62.8	57.9	58.2	63.3	59.8	58.2	62.7
c (A°)	61.8	61.5	61.8	61.8	58.8	60.6	61.8	57.5
Rmerge (%)	6.7	7.3	10.4	4.6	3.3	5.6	4.7	7.6
completeness (%)	93.4	99.7	98.4	97.1	94.5	97.5	94.6	94
total no. of reflections	41786	113050	127633	195712	69773	220012	119749	64360
no. of unique reflections	12376	28931	16417	27160	20171	56731	40242	23065
no. of molecules	1	2	1	1	2	2	2	2
Rfree (%)	25.1	21.28	20.81	17.78	23.84	21.41	22.97	23.53
Rfactor (%)	20.3	16.3	16.01	15.17	18.35	17.40	17.87	17.03
rms in bond length (A°)	0.006	0.009	0.0093	0.0088	0.0097	0.0090	0.0089	0.0094
rms angle (deg)	1.3	1.423	1.492	1.362	1.462	1.341	1.402	1.463
no. of waters	101	215	161	198	279	421	404	214

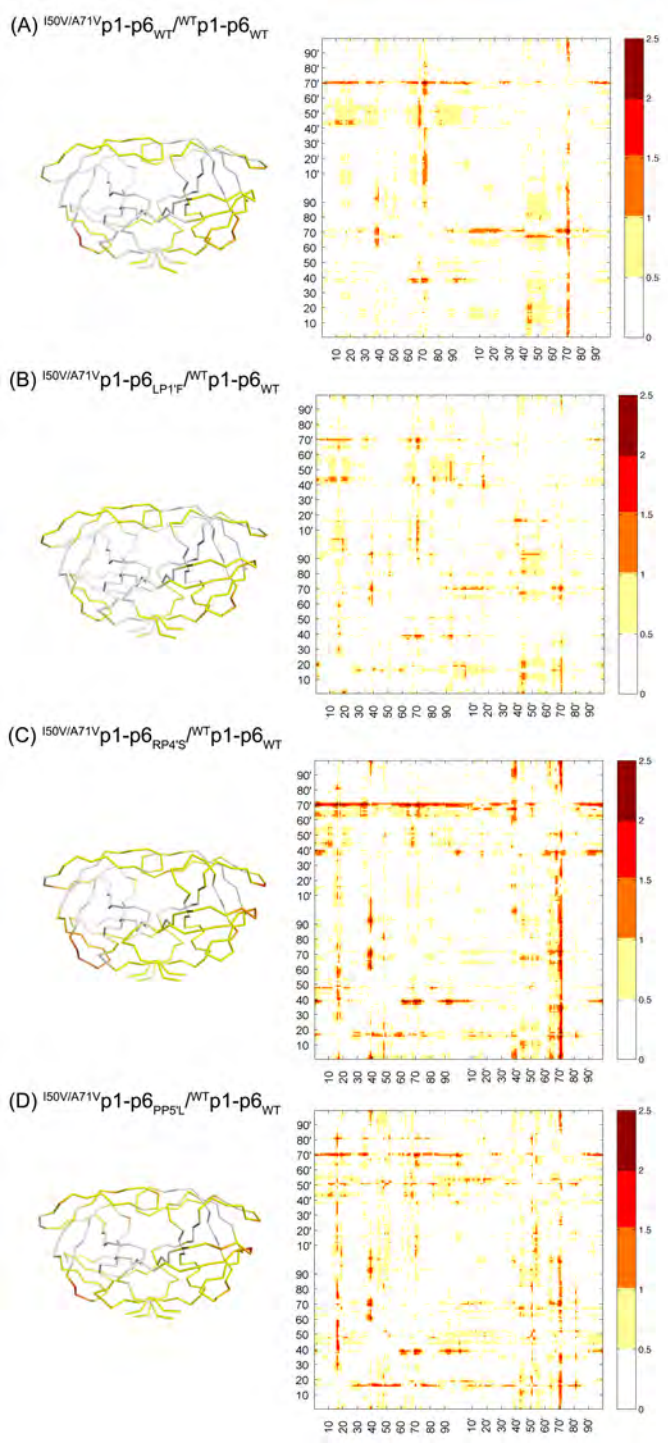


Figure 2.2. Distance-difference maps showing the effect of protease-substrate co-evolution on internal Ca-Ca distances with respect to the wild-type complex.

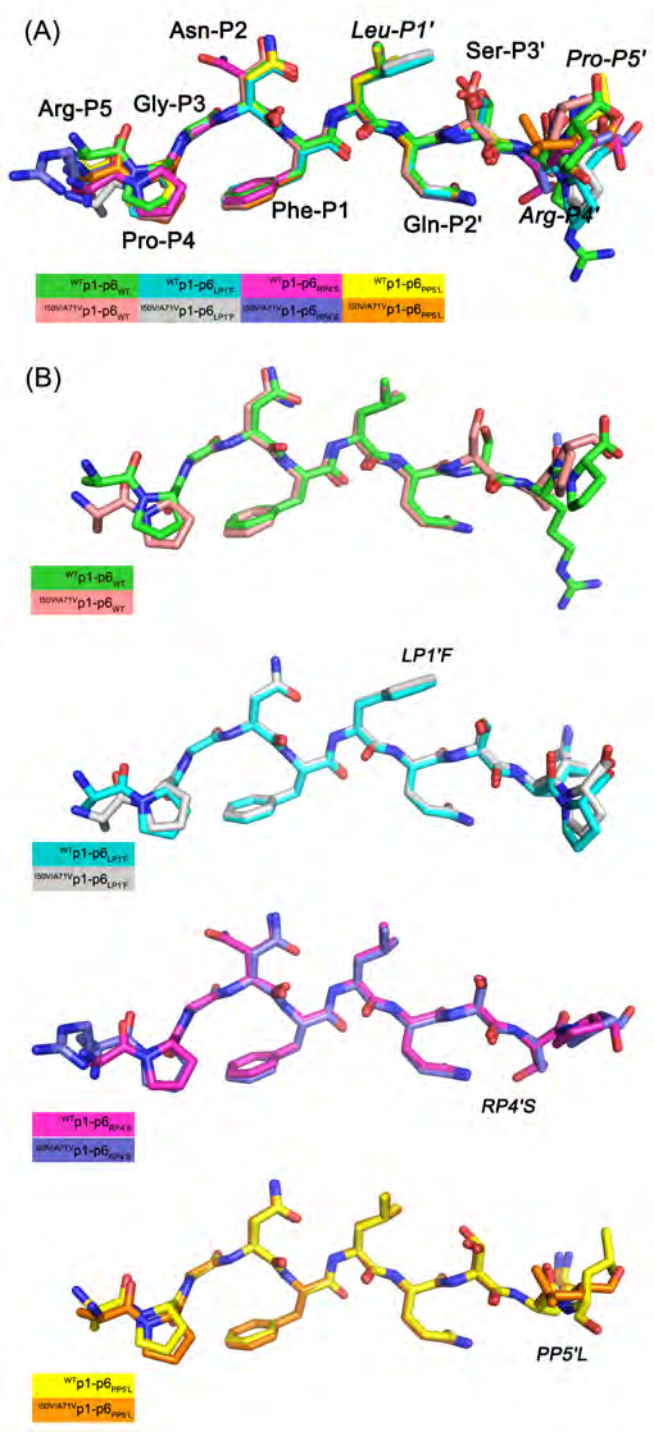


Figure 2.3. Superposed HIV-1 protease substrate conformations. (A) Superposed substrate conformations from eight protease-substrate co-crystal structures. (B) Pairwise superposed substrates bound to wild-type and I50V/A71V protease variants.

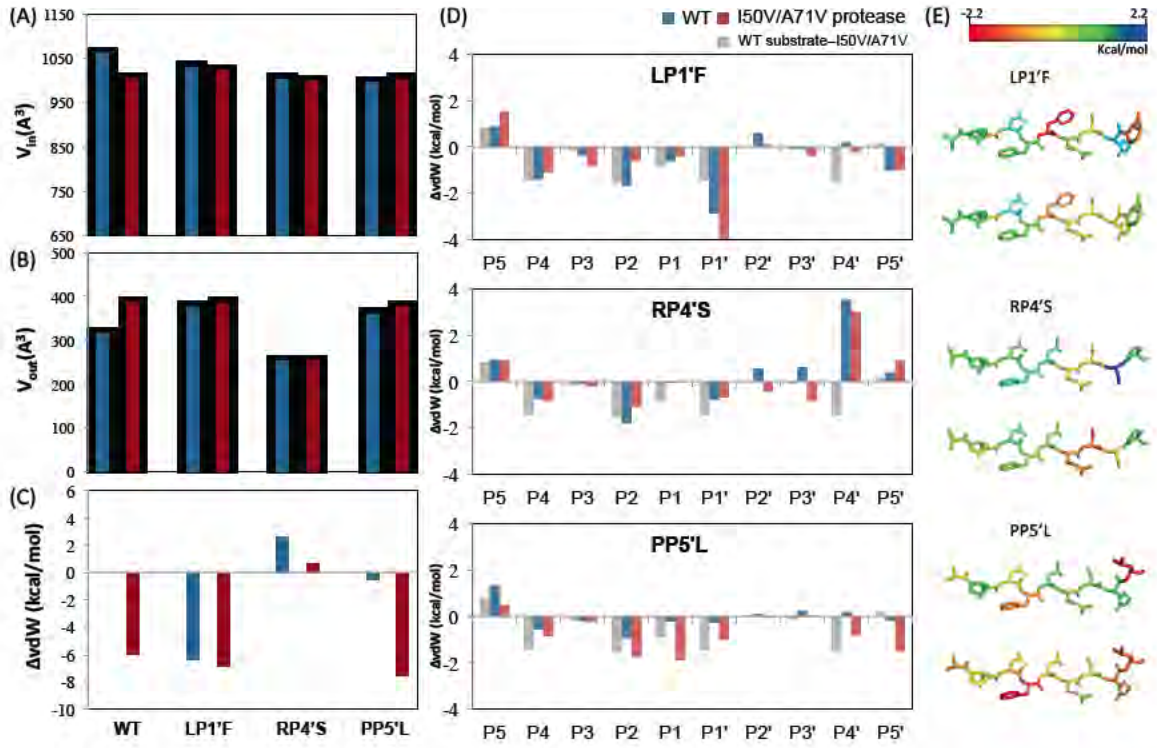


Figure 2.4. The fit of HIV-1 protease substrates within the substrate envelope and the packing of the substrate within the protease active site. (A) Volume within the envelope (V_{in}) (B) Volume protruding outside the envelope (V_{out}). (C) The packing of the substrate within the protease active site in the various crystal structures changes relative to the wild type complex, as reflected in the changes in total van der Waals contacts between the substrate and protease. Negative values indicate enhanced interactions. (D) The change in vdW contacts of substrate residues (in kcal/mol) in mutant HIV-1 protease–substrate cocrystal structures relative to the wild type complex. Negative values indicate enhanced contacts. (E) The difference between vdW contact changes depicted on the substrate structure. Upper structure is for the changes in the co-evolved complex relative to I50V/A71V_{WT} (difference between the red and gray bars in panel D) and the lower structure is relative to WT protease with the corresponding mutant substrate (red and blue bars in panel D).

2.3.2 Substrate mutations have distal effects and enhance packing at the active site

To investigate the alterations in substrate packing at the active site, the vdW interactions were quantified between the protease and substrate. The changes in substrate packing relative to WT_{WT} structure are displayed in Fig. 2.4C, where negative values indicate enhanced packing. The LP1'F substrate has more vdW contacts with I50V/A71V protease compared to those in either $I50V/A71V_{WT}$ or $WT_{LP1'F}$. This is also the case for the PP5'L substrate, where the co-evolved $I50V/A71V_{PP5'L}$ complex makes more vdW contacts than either the $I50V/A71V_{WT}$ or $WT_{PP5'L}$ complexes. Hence, the co-evolution of HIV-1 protease and p1-p6 substrate may rescue the loss of binding interactions caused by mutations on either protease or substrate alone. The enhanced interactions of $I50V/A71V_{LP1'F}$ was partially due to the LP1'F mutation (Fig. 2.4D), that has more interactions with residues that surround the S1' pocket Arg8, Leu23, Pro81 and Val82 compared to $I50V/A71V_{WT}$, and more interaction with Val82 compared to $WT_{LP1'F}$ (Fig. 2.5). In addition, the LP1'F mutation causes a distal change at the substrate P5' proline (Fig. 2.4D,E and Fig. 2.6A) that is in an alternative position in the WT complex (Fig. 2.3). This change increases the P5' proline's vdW contacts, specifically with Lys45' and Met46' instead of Phe53'. Hence, the effect of LP1'F mutation is not localized solely to the site of mutation. This alteration of the structure and vdW contacts of residues at a distal position illustrates the adaptability and interdependency of interactions when HIV-1 protease recognizes substrates.

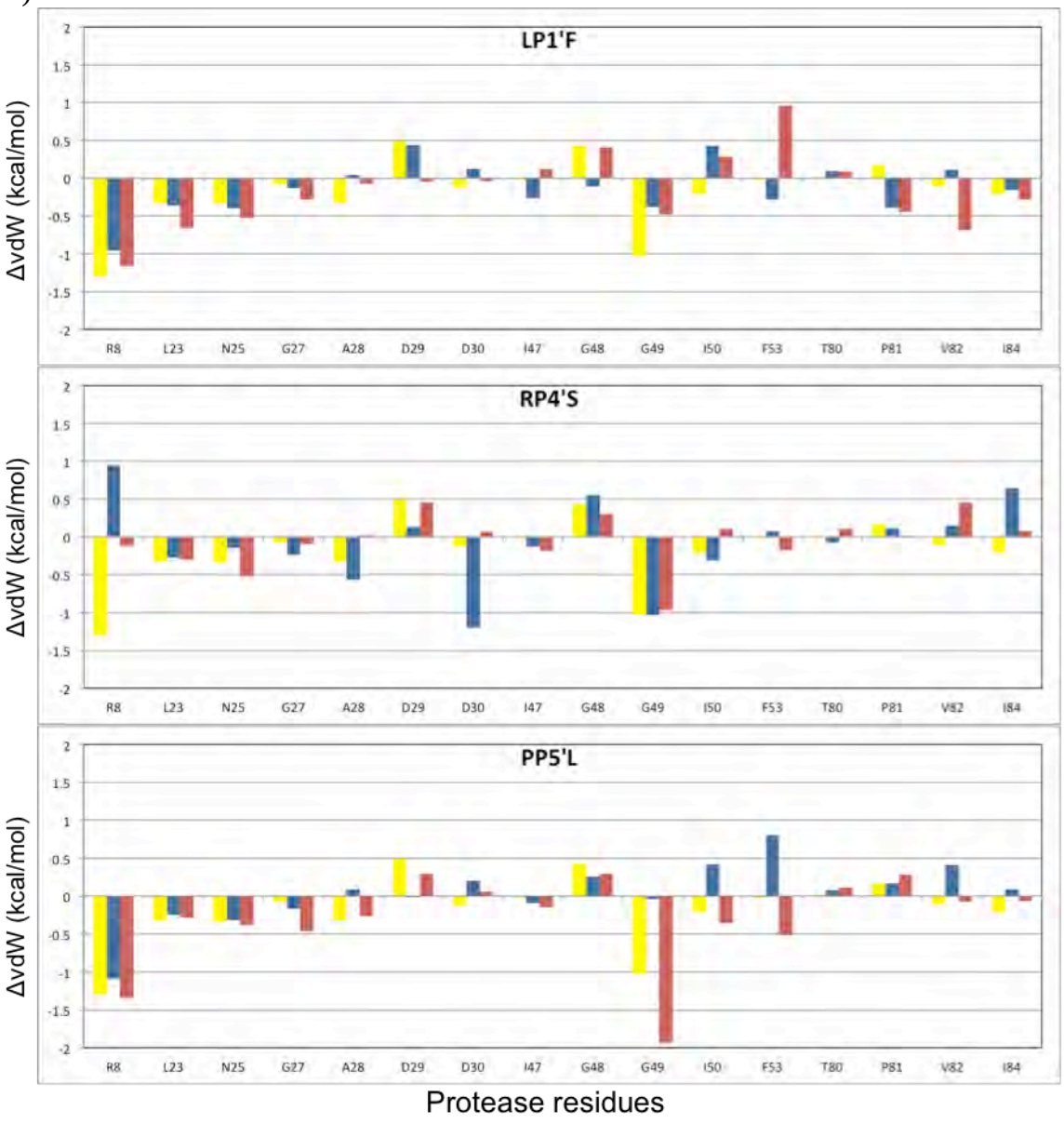
The RP4'S substrate has more overall vdW contacts in the co-evolved $I50V/A71V_{RP4'S}$ complex relative to $WT_{RP4'S}$ but less than the WT substrate in

I50V/A71V$_{WT}$. Since P4’ serine is a smaller residue compared to arginine, the mutated serine makes less vdW contacts when bound to either WT or I50V/A71V protease compared to arginine with the corresponding protease (Fig. 2.4C). However, the mutation at P4' residue actually influences the interactions at other positions: P3' Ser and P2' Gln has enhanced vdW contacts in I50V/A71V$_{RP4'S}$ compared to either I50V/A71V$_{WT}$ or WT$_{RP4'S}$. Specifically, P3' Ser makes more contacts with Asp 29' compared to I50V/A71V$_{WT}$, while P2' Gln–Asp30' and P3’ Ser–Arg8/Ile47' interactions are enhanced compared to WT$_{RP4'S}$ (Fig. 2.5). As with the LP1'F mutation, the RP4'S mutation impacts the interactions of other substrate residues with the protease.

The PP5'L mutation increases the P5' residue’s interactions with I50V/A71V protease as the leucine packs closer to the 50s loop in one monomer, while influencing the residues closer in the active site, P2 Asn and P1 Leu, to also form more extensive contacts. In addition, the co-evolved structure makes more contacts at P1' residue compared to WT$_{PP5'L}$. Interestingly, the major structural change due to this substrate mutation is observed within the protease (Fig. 2.6B). The peptide bond between Gly51 and Gly52 in the I50V/A71V$_{PP5'L}$ structure is flipped compared to the other structures, and this flipped peptide bond pushes the 50s loop towards the substrate, causing increased vdW contacts between protease residues Gly48, Gly49, Ile50, Phe53 and the substrate. Hence, the co-evolved site not only impacts its own fit within protease active site, but also alters the interactions of distal residues in the substrate by stabilizing alternative conformations of the protease.

In conclusion, the detailed analysis of vdW contacts between the protease and substrates show interdependent distal effects in binding interactions where the alterations are not localized at the mutated residue itself, but also occur at other residues. These distal alterations are caused by structural changes in the protease, the substrate, or both.

(A)


LP1'F
RP4'S
PP5'L
ΔvdW (kcal/mol)
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
R8
L23
N25
G27
A28
D29
D30
I47
G48
G49
I50
F53
T80
P81
V82
I84
Protease residues

(B)

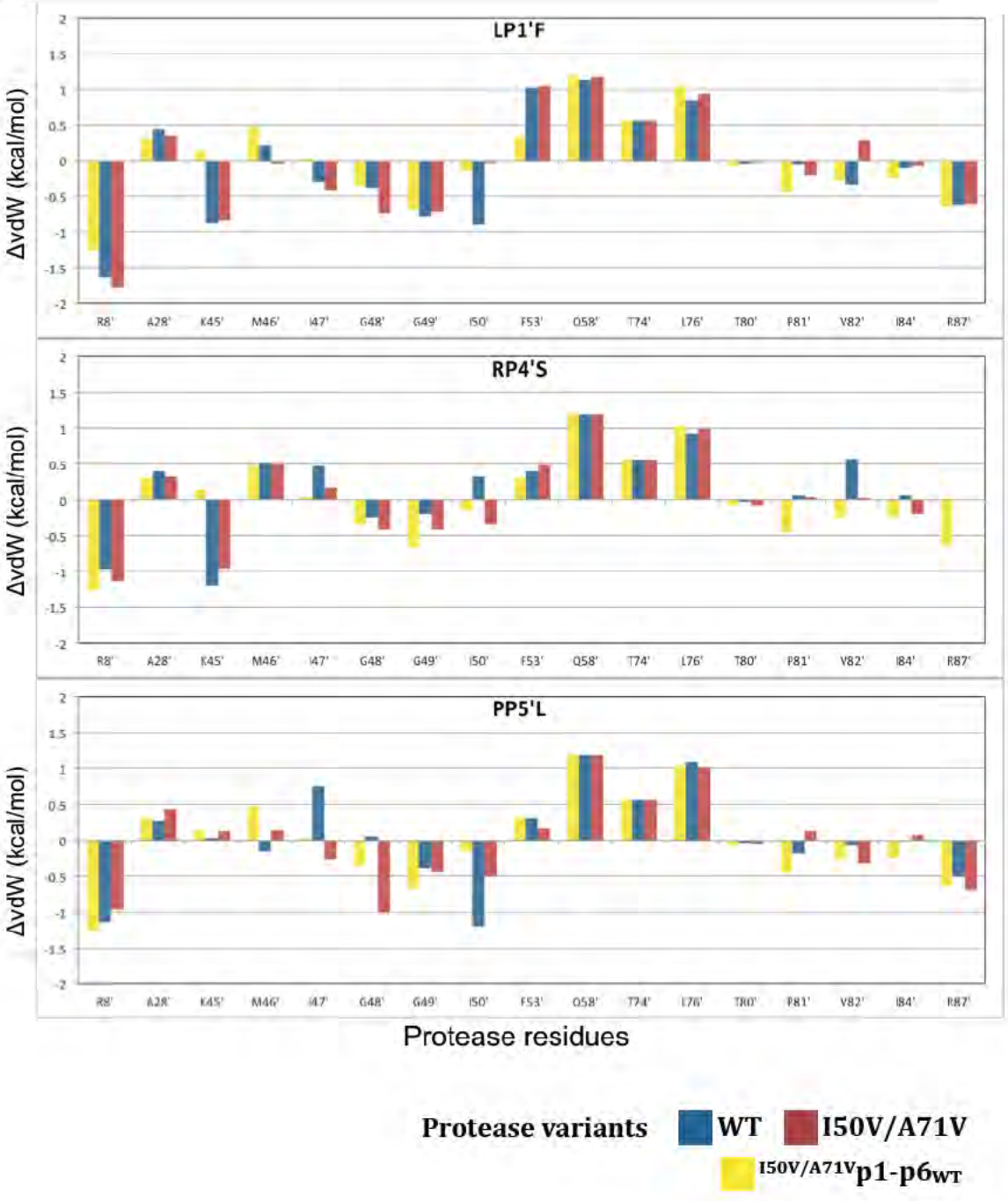


Figure 2.5. The changes in van der Waals contacts of protease residues in monomer (A) and (B) relative to WT complex. Negative values indicate enhanced interactions.

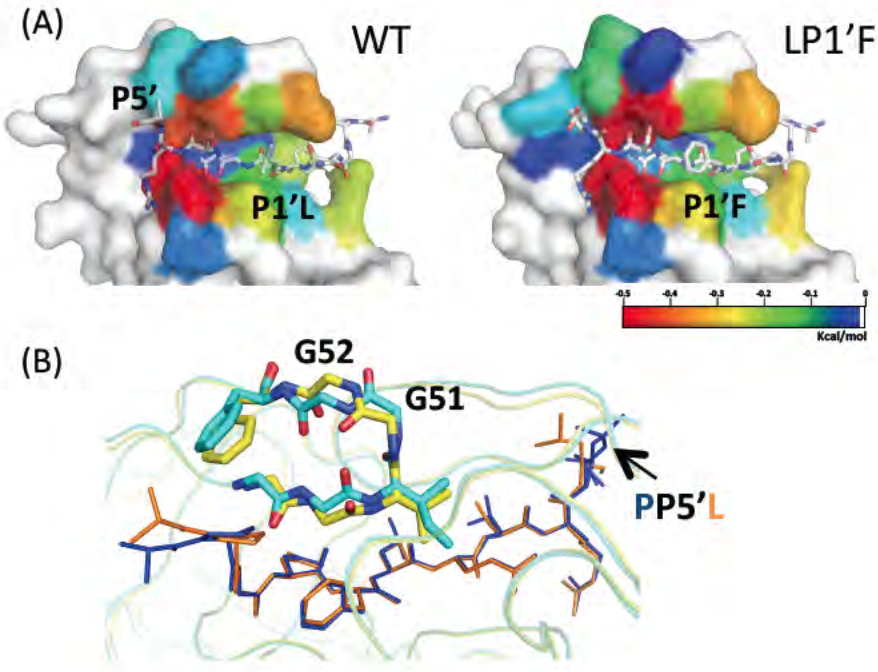


Figure 2.6. Distal effects of substrate mutations on protease interactions. (A) The vdW contacts of residues in HIV-1 protease–substrate co-crystal structures colored blue to red for increasing contacts. The substrate mutation at P1' position enhances contacts at P5'. (B) The distal substrate mutation PP5'L causes a conformational change in the protease flap and alters substrate–protease interactions.

2.3.3 The RP4'S substrate has less packing interactions but an additional hydrogen bond

The intramolecular hydrogen bonds between the protease and the substrates are conserved across all complexes, with two exceptions (Fig. 2.7, Table 2.2): First, the P4' arginine in the WTWT structure orients in the opposite direction compared to the other complexes making a hydrogen bond with the side chain of Asp30’ instead of Asp29'. Second, the RP4'S substrate forms an additional hydrogen bond with both WT and I50V/A71V protease through the P4' serine side chain (Fig. 2.7). This extra hydrogen bond may compensate for the loss of vdW contacts due to the smaller size of serine in these two complexes.

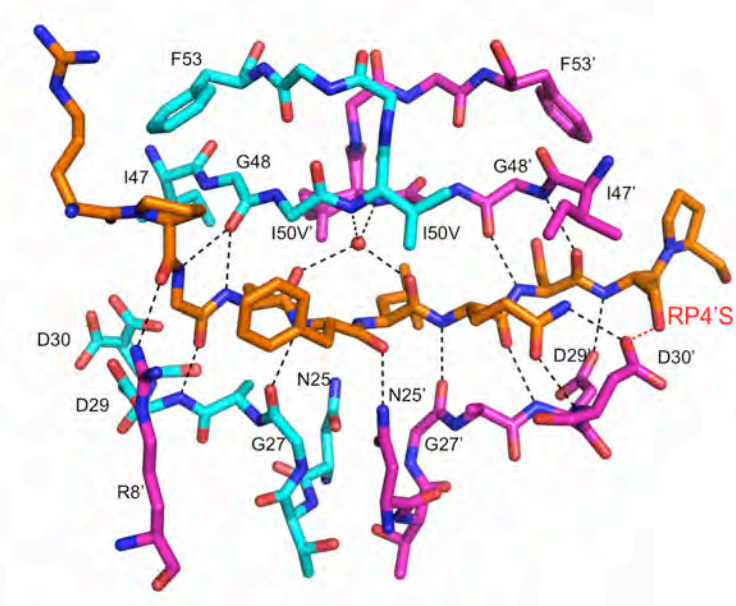


Figure 2.7. The hydrogen bonds between p1-p6 substrate (orange) and HIV-1 protease (cyan and magenta monomers), including those mediated by a conserved water molecule (red sphere). Bonds shared in all substrate variants are in black, while the additional hydrogen bond formed by RP4'S variant is indicated in red.

Table 2.2. Hydrogen bonds between the p1-p6 substrate and HIV-1 protease in cocrystal structures. Bonds with acceptor–donor distances less and more than 3 Å are indicated by red and yellow, respectively. Distances greater than 3.8 Å are not considered bonded but listed for comparison, in green.

Protease	Substrate	WT_{WT}	$I50V_{WT}$	$I50V_{LP1'F}$	$WT_{LP1'F}$	$I50V_{RP4'S}$	$WT_{RP4'S}$	$I50V_{PP5'L}$	$WT_{PP5'L}$
Arg8' NH2	P4 Pro O	4.0	3.01	2.72	2.73	2.94	2.97	2.89	2.94
Gly48 O	P3 Gly N	3.01	2.97	3.03	2.77	2.90	3.03	2.88	3.03
Asp29 N	P3 Gly O	2.79	2.87	2.82	2.86	2.86	2.92	2.82	2.80
Gly48 O	P2 Asn N	3.05	2.86	2.88	2.74	2.85	2.88	2.96	2.95
Gly27 O	P1 Leu/Phe N	2.93	2.90	2.81	2.88	2.90	2.80	2.78	2.88
Asn 25' ND2	P1 Leu/Phe O	2.7	4.0/2.7	2.72	2.74	3.9/2.9	3.7/2.7	2.70	2.76
Gly27 O	P2' Gln N	2.95	2.94	2.97	3.07	2.91	2.83	2.95	2.98
Asp 30' N	P2' Gln OE1	2.74	2.72	2.89	2.86	2.81	2.82	2.83	2.77
Asp 30' OD2	P2' Gln NE2	2.95	2.73	2.79	3.02	2.81	4.6/2.8	2.98	2.88
Asp 29 N	P2' Gln O	3.07	3.07	3.12	3.15	3.04	3.01	3.06	3.10
Gly 48' O	P3' Ser N	3.00	2.58	2.83	2.89	2.81	2.89	2.84	2.85
Gly 48' N	P3' Ser O	2.78	2.91	3.07	3.26	2.87	3.01	3.02	3.02
Asp 29' OD2	P4' Arg/Ser N	3.8	2.95	2.85	2.95	2.96	3.06	2.84	2.94
Asp 30' OD2	P4' Arg NE/Ser OG	3.02				2.59	2.6/2.6		

2.3.4 Active site dynamics is restored by coevolution

Resistance mutations, in addition to altering the molecular interactions, affect the conformational dynamics at the active site. Specifically, the distance between the two 80s loops, which reflect the overall size of the protease active site, varies between protease–substrate complexes during the MD simulations. In crystal structures, the 80-80' distance is similar in all the structures and varies between 17.1 and 18.0 Å (Fig. 2.8). In the dynamic conformational ensemble of the WT_{WT} structure, the distance between these two loops is around 17.5 Å, and expands to 19–19.5 Å with mutations in either the protease or substrate. Strikingly, in all three cases, co-evolution brings this distance back to 17.5–18.0 Å, which is similar to the WT inter-loop distance (Fig. 2.8). Hence, mutation of either the protease or the substrate alone disturbs the dynamics of the protease active site, while co-evolution of both restores the active site dynamics and possibly the protease activity.

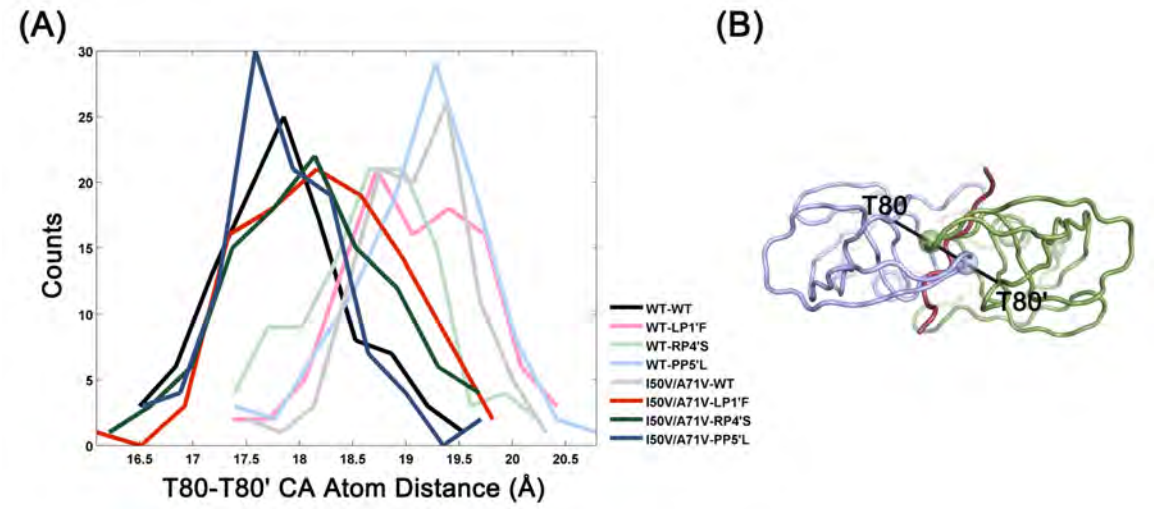


Figure 2.8. The distance (in Å) distribution between T80–T80' across the active site during MD simulations of substrate–protease complexes. (A) Mutations in only protease or only substrate increase the distance sampled (lighter shades) while coevolution of both together (darker shades) brings the distance back to that in the WT–WT complex (black). (B) The T80–T80' distance across the active site depicted on the protease with a view from the top of the flaps. The colors are the same as in Figure 2.1.

2.4 Discussion

HIV-1 I50V/A71V protease is commonly observed in patients failing therapy with APV and DRV, and substrate mutations in Gag cleavage sites coevolve with these primary protease mutations to contribute to inhibitor resistance. Gag L449F mutation rescues the protease activity by 10 fold, while P453L, although located distal from the catalytic site, causes a 23-fold enhancement (Maguire et al., 2002). The mutated substrates are cleaved more efficiently than the wild-type substrate by I50V/A71V protease (Maguire et al., 2002). Interestingly, WT protease also processes the wild-type substrate less efficiently than the mutant substrates. This sub-optimal cleavage efficiency at p1-p6 site should be critical for temporal regulation of Gag processing preventing premature viral maturation (Pettit et al., 1998, Pettit et al., 2005). Under drug pressure, the resistance mutations I50V/A71V are populated, and the impaired protease activity on the wild-type Gag may interfere with the ordered processing of Gag. Co-evolution of substrates possibly restores proper Gag processing by getting more efficiently cleaved by the protease.

The molecular basis of such protease–substrate co-evolution was investigated in this study by solving the crystal structures of the complexes of WT and I50V/A71V protease with p1-p6 substrate variants combined with molecular dynamics simulations. While all complexes have similar overall backbone structures, vdW contacts and hydrogen bonds between the protease and substrate are altered. Coevolved complex structures display enhanced overall substrate interactions, due to either more vdW

contacts or hydrogen bonds, compared to complexes with mutations in either the protease or the substrate alone.

Coevolving mutations in the substrate are not selected to restore the specific interactions lost due to drug resistance mutations but instead enhance substrate–protease interactions through a variety of molecular mechanisms. P1' and P5' mutations enhance substrate packing at the active site, while P4' contributed an additional hydrogen bond with the protease. A similar compensation of interactions is observed in substrate coevolution with nelfinavir resistant D30N/N88D protease, where lost interactions were compensated for by new interactions particularly at the site of substrate coevolution mutation (Ozen et al., 2012, Kolli et al., 2014).

In I50V/A71V protease co-evolution, the effects of substrate mutations are not local, but propagate to distal parts of both the substrate and the protease. The mutated P4' serine affects the interactions at other positions, particularly at P3' Ser and P2' Gln, which has enhanced vdW contacts in I50V/A71V$_{RP4's}$ compared to I50V/A71V$_{WT}$ and WT$_{RP4's}$. Substrate positions P1' and P5' mutually influence each other's interactions with the protease. Mutation at either residues results in enhancement of vdW contacts at both sites: In the presence of LP1'F mutation the P5' residue, and in the presence of PP5'L mutation P1' residue packing is altered. Apart from these synergistic effects within the substrate, the P5' mutation in the substrate stabilized an unexpected structural change within the protease. The protease structure accommodated this mutation flipping the peptide bond between Gly51 and Gly52 in one of the flaps. This backbone flip brings the flap closer to the substrate and increases vdW contacts at P2, P1 and P1' positions. Hence,

HIV-1 protease adopts a conformational change to favor the substrate's binding in the co-evolved complex. This ability of mutations to have distal effects explains why a co-evolution mutation at P5' position, which is away from the core region of the substrate at the active site, may be selected and how this distal mutation is able to alter substrate–protease interactions.

HIV-1 protease is a highly dynamic protein, and conformational dynamics especially around the active site is crucial to substrate binding and enzymatic activity (Ishima et al., 1999, Freedberg et al., 2002, Perryman et al., 2004, Foulkes-Murzycki et al., 2007, Mittal et al., 2012). Although crystal structures provide key insights, alteration of dynamic behavior, not captured by static structures, is emerging as an additional contribution to mechanisms of drug resistance (Cai et al., 2012, de Vera et al., 2013). We found that drug resistance mutations in the protease or in the native substrate disturbed the active site dynamics, which was restored in all co-evolved complexes bearing complementary mutations in both the protease and the substrate. These results suggest that, in addition to the specific three-dimensional shape adopted and shared by all substrates when bound to HIV-1 protease, as defined by the substrate envelope, a conserved dynamic behavior around the active site may be an additional substrate recognition and selection constraint. This dynamic constraint may contribute to the selection of these specific substrate co-evolution mutations in response to the disturbed dynamics in mutated drug resistant protease.

Previously, we showed that I50V/A71V protease has decreased vdW interactions with the protease inhibitors APV and DRV compared to WT (0.61 and 1.98 kcal/mol

respectively), mainly due to the loss of a methyl group interacting with the sulfonyl moiety in APV/DRV (Mittal et al., 2013). The coevolved I50V/A71V$_{LP1'F}$ and I50V/A71V$_{PP5'L}$ structures have more vdW contacts, and I50V/A71V$_{RP4'S}$ more hydrogen bonds compared to WT$_{WT}$ complex. Unlike substrates, the inhibitors cannot adapt to conformational changes in the drug resistant protease, such as the peptide bond flip in the flap of I50V/A71V$_{PP5'L}$ protease. Hence, the structural adaptability of the protease–substrate system allows drug resistance to evolve by selecting mutations in the protease that decrease inhibitor affinity, and additional compensatory mutations in the substrate to enhance any inadvertently lost substrate interactions through various molecular mechanisms, including propagating distal effects.

Co-evolution cause distal changes both in the substrate and the protease, and adaptability of the complex permits mutational, structural, and dynamic plasticity to confer drug resistance. Therefore, the resistance mechanism is an interdependent process whereby multiple residues act in concert on both sides. The molecular rationale, reported here, for the distal effects of mutations in the non-target polyprotein under HIV-1 PI treatment should provide insights into allosteric events in a wider range of co-evolving systems where function is maintained by complex interdependent protein interactions.

2.5 Conclusion

In summary, different co-evolutions adopt varying ways to rescue the interactions. Gag mutations L449F and P453L enhance van der Waals (vdW) interactions between the substrate and mutant I50A/A71V protease, and R452S results in an additional hydrogen bond. The distal effect of protease-substrate co-evolution between P1’ and P5’ positions enhance substrate–protease interactions. In addition, the active site dynamics is restored by coevolution, a key aspect of substrate recognition and turnover that is largely uncharacterized.

2.6 Materials and Methods

2.6.1 Nomenclature

HIV-1 protease (WT or I50V/A71V) complexes with different p1-p6 substrate variants (WT, L449F (LP1'F), R452S (RP4'S) and P453L (PP5'L)) are distinguished with subscripts. For example, WT protease in complex with LP1'F p1-p6 substrate is denoted by $WT_{LP1'F}$ and I50V/A71V protease in complex with WT p1-p6 substrate is denoted by $I50V/A71V_{WT}$. HIV-1 protease functions as a homodimer, and residues in monomer A are simply indicated by residue number while residues in monomer B are marked with residue number with an apostrophe. For example, arginine 8 in monomer A is Arg8, and valine 50 in monomer B is Val50'.

2.6.2 Substrate peptides

Substrate peptides of the p1-p6 processing site within the Gag polyprotein (amino acids 444–453) and its variants were purchased from Quality Controlled Biochemicals. The substrate sequences of WT and co-evolved substrates are: (1) $p1\text{-}p6_{WT}$ – RPGNFLQSRP (2) $p1\text{-}p6_{LP1'F}$ – RPGNFFQSRP (3) $p1\text{-}p6_{RP4'S}$ –RPGNFLQSSP and (4) $p1\text{-}p6_{PP5'L}$ – RPGNFLQSRL.

2.6.3 Protease gene construction

Synthetic protease gene was constructed using codon optimization for protein expression in Escherichia coli, and Q7K mutation was introduced to prevent autoproteolysis (Rose et al., 1993). For protease-substrate co-crystallization purposes, the

D25N mutation was introduced to prevent substrate cleavage; this mutation has negligible impact on protease structure (Sayer et al., 2008). I50V/A71V protease mutations were introduced sequentially by using the QuikChange site-directed mutagenesis kit (Stratagene).

2.6.4 Protein expression, purification and crystallization

The gene encoding HIV protease was sub-cloned into the heat-inducible pXC35 expression vector (ATCC), and transformed into E. coli TAP-106 cells. Protein expression and purification were performed as previously described (King et al., 2002). Protease purified from size exclusion column (equilibrated with gel filtration buffer containing 0.05 M sodium acetate at pH 5.5, 5% ethylene glycol, 10% glycerol, and 5 mM DTT) was concentrated to 2 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore) for crystallization. The concentrated samples were incubated with 10 molar excess of substrates overnight at 4 °C. Concentrated protein solution was then mixed with either precipitant solution (126 mM sodium phosphate buffer (pH 6.2), 63 mM sodium citrate, and 20% to 32% ammonium sulfate, or 0.1 M citrate phosphate (pH 5.5), 0.5-3.0 M ammonium sulphate) at a 1:1 ratio in 24-well VDX hanging-drop trays at room temperature. Diffractionquality crystals were obtained within a week.

2.6.5 Data collection and structure solution

Diffraction-quality crystals were flash-frozen in liquid nitrogen for storage. Constant cryostream was applied when mounting the crystal, and X-ray diffraction data

were collected at Advanced Photon Source LS-CAT 21-ID-F or at our in-house Rigaku_Saturn 944 X-ray system. The substrate complexes' diffraction intensities were indexed, integrated, and scaled using the program HKL2000 (Zbyszek Otwinowski, 1997). Number of molecules in the asymmetric unit was determined by Matthews coefficient calculation. The structure solutions were generated using simple isomorphous molecular replacement with PHASER (McCoy et al., 2007). WT protease–DRV co-crystal structure was used as the starting model (1T3R). Initial refinement was carried out in the absence of modeled substrate, which was subsequently built in during later stages of refinement. Upon obtaining the correct molecular replacement solutions, ARP/wARP or Phenix (Adams et al., 2010) was applied to improve the phases by building solvent molecules (Morris et al., 2002a). Crystallographic refinement was carried out using the CCP4 program suite or PHENIX with iterative rounds of TLS and restrained refinement until convergence was achieved (Collaborative Computational Project, 1994). MolProbity (Davis et al., 2007) was applied to evaluate the final structures prior to deposition in the Protein Data Bank. Five percent of the data was reserved for the free R-value calculation to limit the possibility of model bias throughout the refinement process (Brunger, 1992). Interactive model building and electron density viewing were carried out with COOT (Emsley and Cowtan, 2004).

2.6.6 Structural analysis

Hydrogen bonds were determined using Maestro (Suite 2012: Maestro, version 9.3, Schrödinger). A hydrogen bond was defined by a distance between the donor and acceptor of less than 3.5 Å and a donor–hydrogen–acceptor angle of greater than 120°.

The vdW contacts between the protease and substrate were estimated using a simplified Lennard-Jones potential $V(r) = 4\varepsilon [(\sigma/r)_{12} – (\sigma/r)_6]$, with the well depth (ε) and hard sphere diameter (σ) for each protease–substrate atom pair. V(r) for all protease–substrate atom pairs was computed within 6 Å, and when the distance between nonbonded pairs was less than ε, V(r) was considered equal to –ε. The rationale for this modification to the original 6-12 Lennard-Jones potential was previously described in detail (27). Using this simplified potential for each nonbonded pair, ΣV(r) was then computed for the protease–substrate complex.

The HIV-1 protease substrate envelope was defined using a three-dimensional grid, and the fit of a substrate within this substrate envelope for a given cocrystal structure was evaluated by V_{in} and V_{out} (volumes of the substrate within and outside the substrate envelope, respectively), as previously described in detail (27). Only the P4 to P4' residues of the substrates were modeled in the substrate envelope, since the substrate residues beyond these positions do not share a significant consensus volume.

2.6.7 Molecular dynamics simulations

The crystal structures were prepared for simulations by keeping the crystallographic waters within 4.0 Å of any protease or substrate atom but removing the

buffer salts from the coordinate file. The structures were further processed with the Protein Preparation Tool from Schrodinger by adding hydrogen atoms, building side chains with missing atoms, and determining the optimal protonation states for the ionizable side chains. The hydrogen bonding network of the initial structures was optimized by flipping the terminal chi angle of Asn, Gln, and His residues and sampling hydroxyl/thiol polar hydrogens with the exhaustive/water orientational sampling options. Before solvation, the structures were minimized in vacuum with restraints on heavy atoms using the Impact refinement module with the OPLS2005 force field until the root-mean-square deviation (RMSD) reached 0.3 Å allowing the hydrogens to be freely minimized while relaxing the strained bonds, angles and potential clashes. The prepared systems were solvated in a truncated octahedron solvent box with the SPC water model extending 10 Å beyond the protein in all directions, using the System Builder utility. The overall charge was neutralized

by adding the necessary number of counter ions (Na^+ or Cl^-).

Desmond was used in all simulations with OPLS2005 force field. Each system was first relaxed using a protocol consisting of an initial minimization restraining the solute heavy atoms with a force constant of 1000 kcal mol^{-1} $Å^{-2}$ for 10 steps with steepest descent and with LBFSG method up to 2000 total steps with a convergence criterion of 50.0 kcal mol^{-1} $Å^{-2}$. The system was further minimized by restraining only the backbone and allowing the free motion of the side chains. At this stage, the restraint on the backbone was gradually reduced from 1000 to 1.0 kcal mol^{-1} $Å^{-2}$ in 5000 steps (250 steepest descent plus 4750 LBFSG) for each value of force constant (1000, 500, 250, 100,

50, 10, 1.0 kcal mol^{-1} $Å^{-2}$) and finally an unrestrained energy minimization was performed.

After energy minimization, each system was equilibrated by running a series of short MD steps. First, a 10 ps MD simulation at 10 K was performed with a 50 kcal mol^{-1} $Å^{-2}$ restraint on solute heavy atoms and using Berendsen thermostat in the NVT ensemble. MD steps were integrated using a two time-step algorithm, with 1 fs steps for bonded and shortrange interactions within the 9 Å cutoff and 3 fs for long-range electrostatic interactions, which were treated with the smooth particle-mesh Ewald (PME) method. Time steps were kept shorter at this first MD stage to reduce numerical issues associated with large initial forces before the system equilibrates. This was followed by another restrained MD simulation for 10 ps at 10 K with a 2 fs inner and 6 fs outer time step in NPT ensemble. The temperature of the system was slowly increased from 10 K to 300 K over 10 ps retaining the restraint on the system and 10 ps MD was performed without the harmonic restraints. Production MD simulations were carried out at 300 K and 1 bar for 20 ns using the NPT ensemble, Nose-Hoover thermostat, and Martyna-Tuckerman-Klein barostat. The long-range electrostatic interactions were computed using a smooth particle mesh Ewald (PME) approximation with a cutoff radius of 9 Å for the transition between the particle-particle and particle-grid calculations and van der Waals (vdW) interactions were truncated at 9 Å. The coordinates and energies were recorded every 5 ps.

Chapter III

Dengue Protease Substrate Recognition: Binding of the Prime Side

PREFACE

Chapter III has been submitted to *ACS Infectious Diseases* as:

Kuan-Hung Lin, Ellen A. Nalivaika, Kristina L. Prachanronarong, Nese Kurt Yilmaz, Celia A. Schiffer. **Dengue Protease Substrate Recognition: Binding of the Prime Side.**

Author contributions: **K.-H.L.** and C.A.S. designed research; **K.-H.L.** performed research; **K.-H.L.** and N.K.Y. analyzed data; **K.-H.L.**, N.K.Y., and C.A.S. wrote the paper.

Contributions from Kuan-Hung Lin: With the guidance of my mentor Celia Schiffer and my thesis research committee, I performed all the experiments including: protein engineering, protein expression and purification, circular dichroism spectroscopy experiment, enzyme inhibition assay, isothermal titration calorimetry experiment (guided by Ellen A. Nalivaika), MD simulations (guided by Kristina L. Prachanronarong), and structure analysis. Nese Kurt-Yilmaz and Celia Schiffer guided interpretation of the data. I created all the figures and tables for the manuscript. I wrote the manuscript, and Nese Kurt-Yilmaz and Celia Schiffer provided editorial assistance.

3.1 Abstract

Dengue virus (DENV), transmitted predominantly in tropical and subtropical regions by the mosquito *Aedes aegypti,* infects millions of people and leads to dengue fever and thousands of deaths each year. There are no direct-acting antivirals to combat DENV, and a scarcity of molecular and structural knowledge required to develop such compounds. The dengue NS2B/NS3 protease is a promising target for direct-acting antivirals, as viral polyprotein cleavage during replication is required for the maturation of the viral particle. The NS2B/NS3 protease processes eight of the thirteen viral polyprotein cleavage sites to allow viral maturation. Although these sites share little sequence homology beyond the P1 and P2 positions, most are well conserved among the serotypes. How the other substrate residues, especially at the P' side, affect substrate recognition remains unclear. We exploited the tight-binding general serine protease inhibitor aprotinin to investigate protease–substrate interactions at the molecular level. We engineered aprotinin's binding loop with sequences mimicking P' side of DENV substrates. P' residues significantly modulate substrate affinity to protease, with inhibition constants varying from nanomolar to sub-millimolar. Structural and dynamic analysis revealed the molecular basis of this modulation, and allowed identifying optimal residues for each of the P' positions. In addition, isothermal titration calorimetry showed binding to be solely entropy driven for all constructs. Potential flaviviral P' side inhibitors could benefit from mimicking the optimal residues at P' positions, and incorporate hydrophobicity and rigidity to maintain entropic advantage for potency.

3.2 Introduction

Dengue virus (DENV), a member of the family *Flaviviridae*, is an enveloped virus with a positive single-strand RNA genome. There are four different serotypes of dengue virus (DENV 1–4), and each serotype shares 65-70% sequence identity of the viral genome (Rico-Hesse, 1990). Dengue fever, which is caused by DENV, is a worldwide infectious disease endemic in more than 110 countries. Approximately 390 million people are infected yearly, with 96 million of those infected developing disease symptoms and about 20,000 annual deaths (Monath, 1994, WHO, 2009, Bhatt et al., 2013). The mosquito *Aedes aegypti* is the major vector of dengue virus, and due to the narrow temperature tolerance of *Aedes*, DENV is transmitted predominantly in tropical and subtropical regions. No vaccine or antiviral drug to combat dengue infections is available, except for the first-ever dengue vaccine approved for use in Mexico in December 2015 (Simmons, 2015). DENV is in the same flaviviral family as the Zika virus, and is also transmitted by *Aedes aegypti* to cause major outbreaks in the tropical and sub-tropical regions, with no specific treatment. A better understanding of DENV will contribute to the development of a targeted therapy against this global health threat.

The dengue RNA genome has one open reading frame which encodes a single polyprotein including three structural proteins (C, prM, and E) and seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) (Chambers et al., 1990a) (Fig. 3.1A). The viral polyprotein gets processed at the lumen side of the rough endoplasmic reticulum membrane by the host cell peptidase, while dengue NS2B/NS3 protease cleaves the polyprotein at the cytoplasmic side (Chambers et al., 1990a). Dengue

NS2B/NS3 protease is a serine protease (Ser 135 is the catalytic residue) of the chymotrypsin family with a classic Ser-His-Asp catalytic triad (Bera et al., 2007). The hydrophilic core of NS2B cofactor (cNS2B; amino acids 1394-1440) is required for the proper function of NS3 protease (NS3pro185; amino acids 1476-1660) (Yusof et al., 2000) and participates in substrate recognition (Noble et al., 2012). Dengue protease is responsible for processing 8 of the 13 polyprotein cleavage sites (C, NS2A, NS2A-NS2B, NS2B-NS3, NS3, NS3-NS4A, NS4A, NS4B-NS5) (Fig. 3.1A) (Falgout et al., 1991). Polyprotein processing is required for the maturation of the viral particle.

The polyprotein cleavage site sequences that DENV protease processes share little homology (Fig. 3.1B). Two basic residues at P2 and P1 positions and a small polar residue at P1' are preferred in general by flaviviral NS3 proteases (Chambers et al., 1990b). However, in DENV, some substrates have non-basic residues at the P2 position and residues at P3–P5 and P' positions are quite diverse (Fig. 3.1B). Although diverse between different cleavage sites, the sites themselves are well conserved across all serotypes (and even in some cases with Zika), in particular the NS3, NS4A, and NS2B-NS3 (Table 3.1), implying that these sequences may be required for regulating the temporal processing of the polyprotein. However how dengue protease recognizes these diverse substrates is not well understood.

Dengue protease does not share P' substrate sequence preference with other serine proteases in humans, but does so at P1 and P2 positions (furin RXRR, thrombin P1 R, trypsin P1 R). Thus, the peptidomimetic dengue protease inhibitors designed based on the conserved P2 and P1 positions of substrate sequences alone may not be specific (Yin et

al., 2006a, Yin et al., 2006b, Nitsche et al., 2011). Incorporating P' moieties to current inhibitors could improve specificity. P' residues at each position have been screened to investigate the favored amino acids (Li et al., 2005a) but elucidation of the interdependence of these residues and key physical features of P' recognition is still lacking.

In this study, we exploited the high-affinity binding and structural availability of bound aprotinin to investigate P' side substrate interactions with dengue protease. The binding loop of aprotinin shares close homology with the DENV NS3 cleavage site (Fig. 3.1B). The homology makes this structure a useful template to investigate how different P' native substrate residues may interact with the protease. We replaced the aprotinin binding loop with corresponding P1 to P4' substrate sequences of the eight DENV3 protease cleavage sites (C, NS2A, NS2A/B, NS2B/3, NS3, NS3/4A, NS4A, NS4B/5) (Fig. 3.1B). To elucidate how P' substrate sequence affects binding affinity, we measured the inhibition constant for each aprotinin construct in enzymatic assays. These were complemented with molecular dynamics simulations based on molecular models of the aprotinin–DENV3 complex. Isothermal titration calorimetry revealed that binding is solely entropy driven, and enthalpy contributions are unfavorable. These constructs varied in binding affinity by five orders of magnitude, and their dynamic behavior implicates the recognition patterns of various cleavage sites. Thus P' side interactions can significantly affect ligand binding and incorporating these interactions could help achieve additional specificity for dengue protease inhibition.

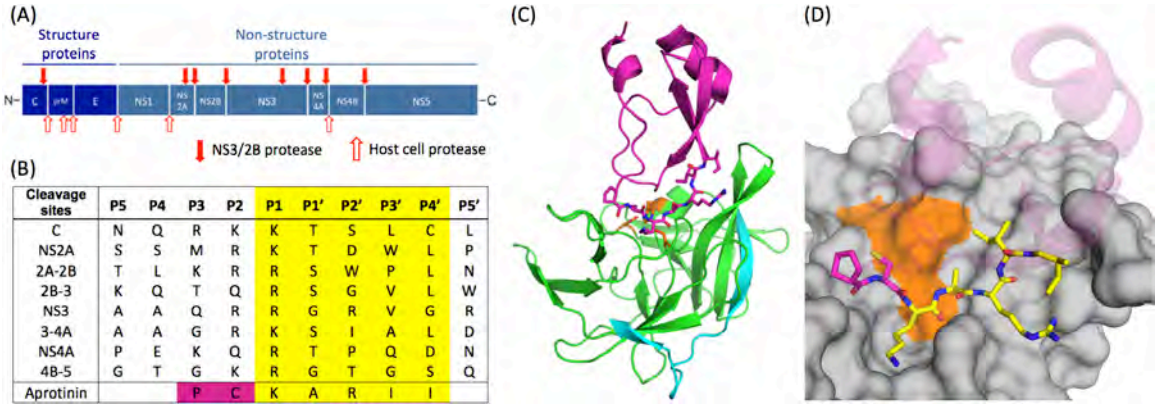
3.3 Results

3.3.1 Design of aprotinin constructs mimicking dengue protease substrates

Dengue protease recognizes substrates with diverse sequences, and favors two basic residues at P1 and P2 positions (Fig. 3.1B) but little is known about how the protease recognizes the entire substrate sequences which are well conserved between serotypes. To investigate the role of the prime side in conferring specificity to DENV protease we used the serine protease inhibitor aprotinin, whose binding loop shares high homology with the DENV NS3 cleavage site sequence (Fig. 3.1B). Aprotinin has previously been shown to bind DENV2 protease with an inhibition constant (K_i) of 26 nM (Mueller et al., 2007), and we measured a K_i of 3.7 nM to DENV3 protease (Fig. 3.2). The crystal structure of the aprotinin–dengue protease complex (PDB: 3u1j) (Noble et al., 2012) (Fig. 3.1C) reveals how P' positions bind at the active site extending from P3 to P4' position (Fig. 3.1D) (Noble et al., 2012), enabling engineering of constructs mimicking P' side of substrates.

In this study the binding loop (BL) residues in aprotinin (AP) were replaced with P1 to P4' substrate sequences of eight protease cleavage sites (C, NS2A, NS2A/B, NS2B/3, NS3, NS3/4A, NS4A, NS4B/5 (Fig. 3.1). These cleavage site residues will be referred to as $(\text{cleavage site})_{BL}$-AP for clarity, and DENV protease residues will be distinguished using single-letter amino acid codes. Since the disulfide bond between Cys14 (P2 residue) and Cys38 is required for the proper folding of aprotinin (Bode and Huber, 1992), residues at P2 and P3 position were kept as the aprotinin sequence in all constructs. The eight engineered aprotinin constructs as well as the wildtype aprotinin

were expressed and purified. To test whether the constructs maintained the aprotinin fold, protein secondary structure was confirmed by circular dichroism spectroscopy (Fig. 3.3). Seven of the eight engineered constructs, except 2A-2B, maintained the aprotinin-like fold and were used for all further experiments, with wildtype aprotinin as a control.



Cleavage sites	P5	P4	P3	P2	P1	P1'	P2'	P3'	P4'	P5'
C	N	Q	R	K	K	T	S	L	C	L
NS2A	S	S	M	R	K	T	D	W	L	P
2A-2B	T	L	K	R	R	S	W	P	L	N
2B-3	K	Q	T	Q	R	S	G	V	L	W
NS3	A	A	Q	R	R	G	R	V	G	R
3-4A	A	A	G	R	K	S	I	A	L	D
NS4A	P	E	K	Q	R	T	P	Q	D	N
4B-5	G	T	G	K	R	G	T	G	S	Q
Aprotinin			P	C	K	A	R	I	I	

Figure 3.1. Design of aprotinin constructs mimicking dengue protease substrates. (A) Dengue virus polyprotein cleavage sites. (B) Polyprotein cleavage site sequences of DENV3 protease. (C) Aprotinin-DENV3 protease complex structure (3u1j). NS3 protease domain is in green, NS2B co-factor cyan and aprotinin purple. (D) The binding loop of aprotinin is displayed as sticks and the residues screened with corresponding P1 to P4' substrate sequences are colored yellow.

Aprotinin	Ki (nM)
WT	3.7 ± 1.2
NS3	5.9 ± 1.6
3-4A	13.8 ± 2.0
2B-3	44.6 ± 6.1
C	192.4 ± 16.6
4B-5	711.6 ± 76.7
NS2A	$2.7x10^3 \pm 0.3x10^3$
NS4A	$184.4x10^3 \pm 30.5x10^3$

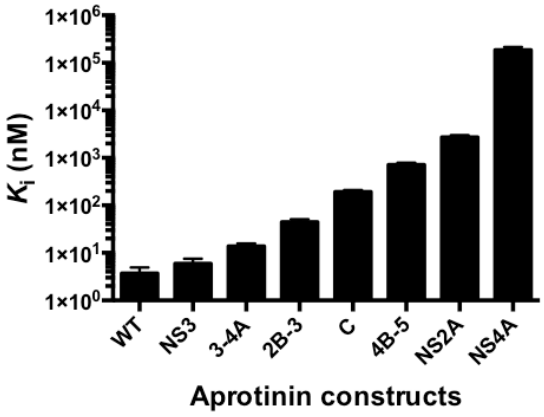


Figure 3.2. The inhibition constants (K_i) of WT-AP and aprotinin constructs against DENV3 WT protease ordered from tightest to weakest binder. Each K_i value and standard error was calculated from three independent determinations.

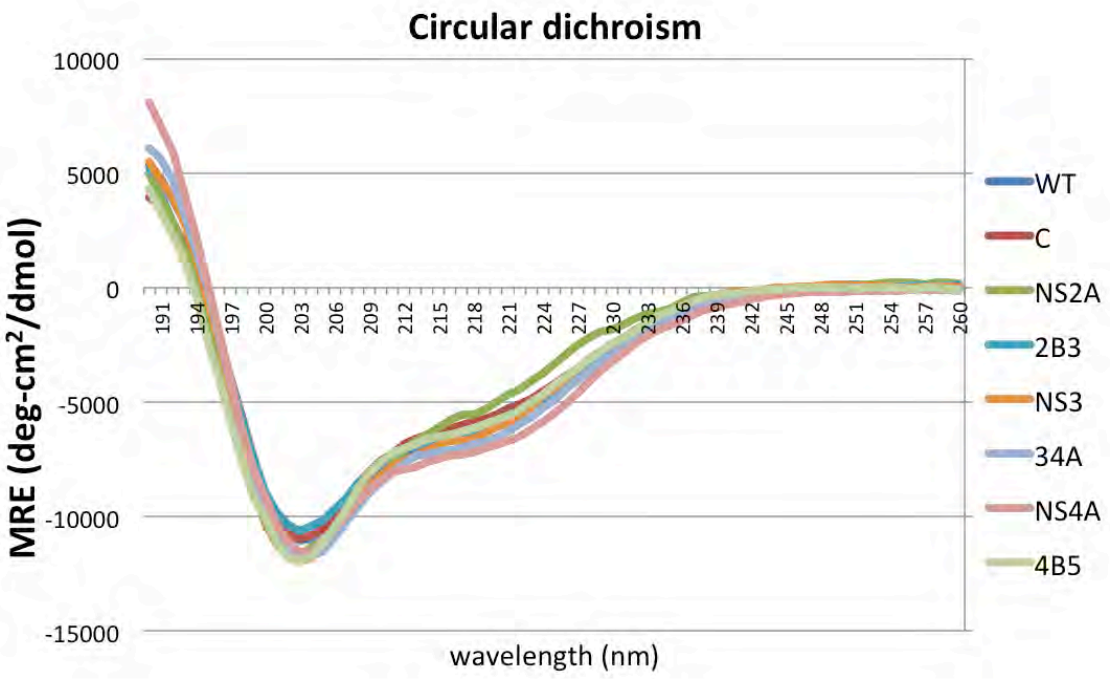


Figure 3.3. The secondary structure is conserved across all aprotinin constructs used in this study.

3.3.2 Substrate sequences have widely varying affinity to DENV3 protease

To investigate how diverse P' substrate sequences affect the binding to dengue protease, the inhibition constants of wild type aprotinin and engineered constructs against DENV3 WT protease were determined using a FRET-based enzymatic assay. While WT aprotinin has a K_i of 3.7 nM against DENV3 protease, the aprotinin constructs have a wide variety of K_i values ranging from 5.9 nM ($NS3_{BL}$-AP) to 184 μM ($NS4A_{BL}$-AP) (Fig. 3.2). The five orders of magnitude difference in binding affinity strongly suggests that these sequences modulate the specificity of DENV3 protease. The sequences are distributed throughout the range of affinities with three sequences ($NS3_{BL}$-AP; NS3-$4A_{BL}$-AP; NS2B-3_{BL}-AP) binding in the low to mid nanomolar; three sequences (C-AP; NS4B-$5A_{BL}$-AP; NS2A-AP) binding in the sub to low micromolar and finally $NS4A_{BL}$-AP in the sub millimolar affinity. The tightest binding $NS3_{BL}$-AP is very similar to the corresponding cleavage site in Zika (Table 3.1). Overall, different P' residues significantly change the binding interactions between DENV3 protease and the substrate, over several orders of magnitude, and this variation in affinity likely provides some specificity for the regulation of the viral polyprotein processing.

Table 3.1. The sequence alignments of dengue and Zika proteases' polyprotein cleavage sites.

NS3

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	A	A	Q	R	R		G	R	I	G	R
DENV2	A	A	Q	R	R		G	R	I	G	R
DENV3	**A**	**A**	**Q**	**R**	**R**		**G**	**R**	**V**	**G**	**R**
DENV4	A	A	Q	R	R		G	R	I	G	R
Zika	A	A	Q	R	R		G	R	I	G	R

3-4A

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	A	A	G	R	R		S	V	S	G	D
DENV2	A	A	G	R	K		S	L	T	L	N
DENV3	**A**	**A**	**G**	**R**	**K**		**S**	**I**	**A**	**L**	**D**
DENV4	A	S	G	R	K		S	I	T	L	D
Zika	A	A	G	K	R		G	A	A	F	G

2B-3

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	K	K	K	Q	R		S	G	V	L	W
DENV2	V	K	K	Q	R		A	G	V	L	W
DENV3	**K**	**Q**	**T**	**Q**	**R**		**S**	**G**	**V**	**L**	**W**
DENV4	V	K	T	Q	R		S	G	A	L	W
Zika	K	T	G	K	R		S	G	A	L	W

C

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	N	R	R	K	R		S	V	T	M	L
DENV2	N	R	R	R	R		S	A	G	M	I
DENV3	**N**	**K**	**R**	**K**	**K**		**T**	**S**	**L**	**C**	**L**
DENV4	N	G	R	K	R		S	T	I	T	L
Zika	E	K	K	R	R		G	A	D	T	S

4B-5

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	G	G	G	R	R		G	T	G	A	Q
DENV2	T	N	T	R	R		G	T	G	N	I
DENV3	**G**	**T**	**G**	**K**	**R**		**G**	**T**	**G**	**S**	**Q**
DENV4	Q	T	P	R	R		G	T	G	T	T
Zika	L	V	K	R	R		G	G	G	T	G

NS2A

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	T	T	S	Q	K		T	T	W	L	P
DENV2	S	S	Q	Q	K		T	D	W	I	P
DENV3	**S**	**S**	**M**	**R**	**K**		**T**	**D**	**W**	**L**	**P**
DENV4	C	L	Q	K	Q		S	H	W	V	E
Zika	G	S	V	K	K		N	L	P	F	V

NS4A

	P5	P4	P3	P2	P1		P1'	P2'	P3'	P4'	P5'
DENV1	P	D	R	Q	R		T	P	Q	D	N
DENV2	P	E	K	Q	R		T	P	Q	D	N
DENV3	**P**	**E**	**K**	**Q**	**R**		**T**	**P**	**Q**	**D**	**N**
DENV4	P	E	K	Q	R		T	P	Q	D	N
Zika	P	E	K	Q	R		S	P	Q	D	N

3.3.3 The binding interactions are solely entropy driven

Binding thermodynamics of the constructs to DENV3 protease was evaluated using ITC experiments. ITC measures the contributions of enthalpy and entropy to the binding free energy, which dictates the binding affinity. As expected, the equilibrium dissociation constants (K_d) obtained from ITC experiments (Table 3.2) are in very good agreement with the inhibition constants (K_i) from enzymatic assays (Pearson's r: 0.99, R squared: 0.99) (Fig. 3.2). The contributions of enthalpy and entropy to the binding free energy of each construct (Fig. 3.4) indicates that the entropic contributions ($-T\Delta S$) are negative (favorable) while the enthalpy of binding is positive (unfavorable) for all constructs. Thus, binding is solely entropy driven and, with the exception of 4B-5_{BL}-AP, the binding entropy strictly follows the trend of inhibition constants. Although still having an overall unfavorable binding enthalpy, the constructs of the natural DENV substrate sequences are less unfavorable than the WT-AP (Fig. 3.5 / Table 3.2) by ~2–4.5 kcal suggesting that these natural substrate constructs are more complementary to the specific binding surface of DENV protease.

Table 3.2. The values derived from ITC experiments of WT-AP and aprotinin constructs against DENV3 protease. Each value and standard error was calculated from at least two independent determinations except NS4A (n=1).

Aprotinin	K_d(nM)	ΔG	ΔH	−TΔS
WT	1.7 ± 0.7	-11.8 ± 0.3	7.2 ± 0.2	-18.9
NS3	2.2 ± 0.5	-11.6 ± 0.1	3.5 ± 1.1	-15.1
3-4A	29.5 ± 6.5	-10.1 ± 0.1	4.4 ± 0.1	-14.5
2B-3	238.6 ± 21.0	-8.9 ± 0.1	2.8 ± 0.0	-11.7
C	$6.8x10^3 \pm 2.\ 10^3$	-6.9 ± 0.2	3.6 ± 0.3	-10.6
4B-5	816.9 ± 56.9	-8.2 ± 0.0	5.3 ± 0.0	-13.5
NS2A	$9.5x10^3 \pm 3.0x10^3$	-6.7 ± 0.2	3.7 ± 1.1	-10.4
NS4A	$161.8x10^3 \pm 37.3x10^3$	-5.1 ± 0.0	5.3 ± 1.1	-10.4

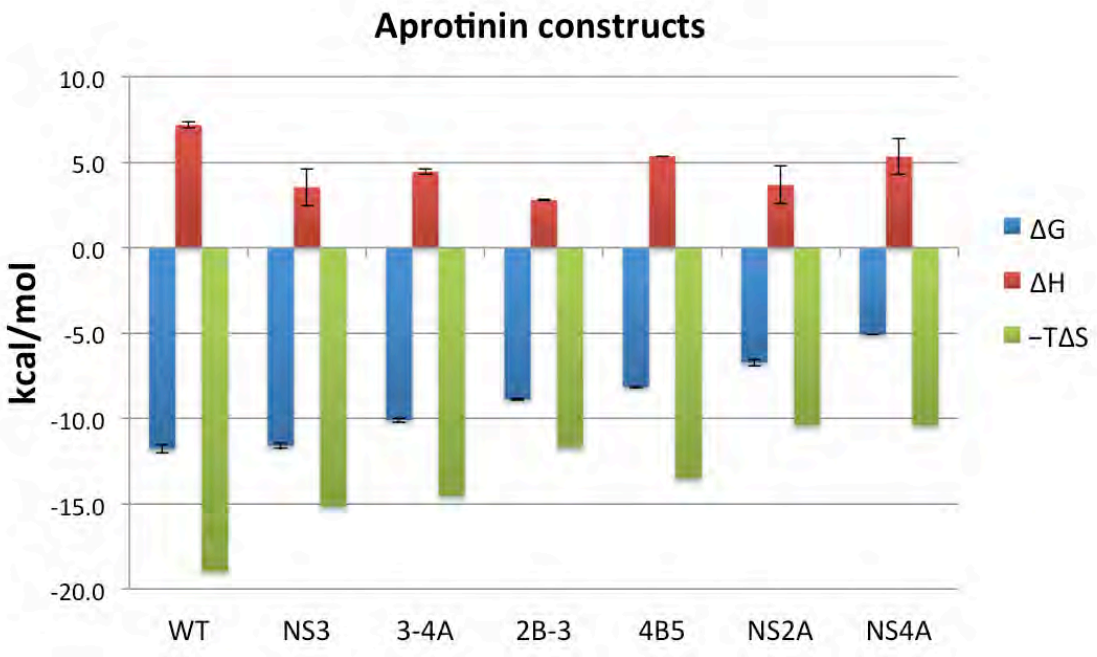


Figure 3.4. The binding of WT-AP and aprotinin constructs to DENV3 protease is entropy driven. Changes of Gibbs free energy (ΔG), enthalpy (ΔH) and entropy (-TΔS) upon binding are displayed in blue, red and green respectively.

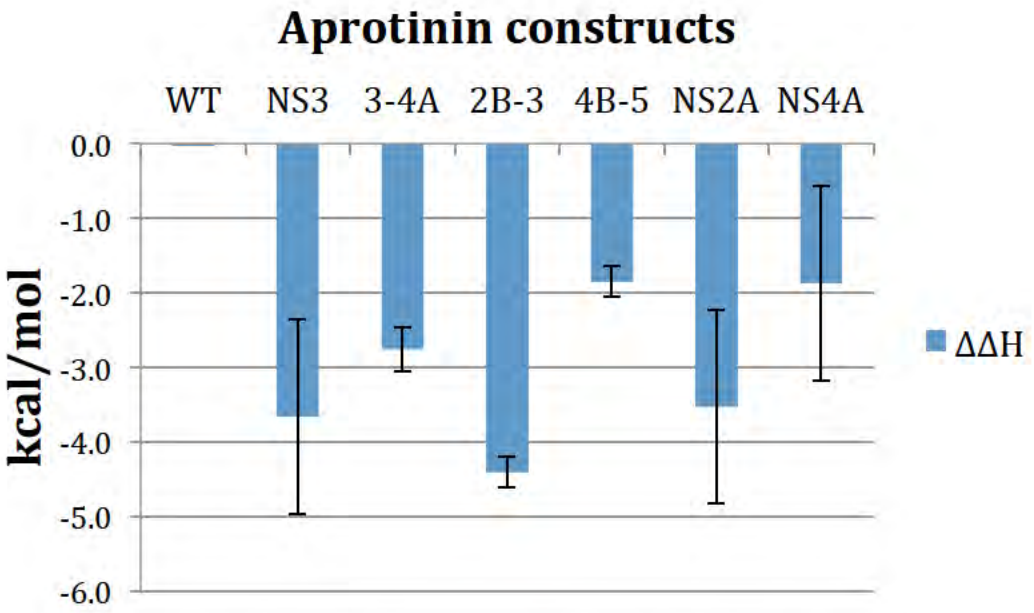


Figure 3.5. The relative enthalpic contribution to the binding free energy of aprotinin constructs versus WT aprotinin to DENV3 protease. Each value and standard error was calculated from at least two independent determinations except NS4A (n=1).

3.3.4 Binding loop fluctuations correlate with affinity

To investigate the molecular interactions between different substrate residues and DENV3 protease, extensive fully-solvated MD simulations were performed starting from structural models generated based on the complex crystal structure (PDB: 3ulj), where the binding loop was replaced by corresponding substrate sequences *in silico* (Schrödinger, 2015). The binding loop of wild type aprotinin was stable over the simulation, with well-overlapping snapshots and relatively low root-mean-square fluctuation (RMSF) values around ~1 Å (Fig. 3.6A, Fig. 3.7).

Aprotinin constructs with binding loops corresponding to the various substrate sequences had varying levels of loop stability. The extent of the RMSF values corresponded closely with the affinities of the various substrate sequences, where the $NS3_{BL}$-AP had the lowest RMSF of 1.2 Å and the tightest experimental binding affinity, and $NS4A_{BL}$-AP the highest RMSF value of 4.3 Å and the weakest binding to DENV3 protease. In fact, the interactions between DENV3 protease and $NS4A_{BL}$-AP were not strong enough to hold the binding loop in the starting pose, and by the end of the simulation only P1 Arg of $NS4A_{BL}$-AP and D129 of the protease were still forming ionic interactions. Overall, fluctuations (RMSF) of residues corresponding to substrate sequences in each construct correlated well with the experimental K_i values (Pearson's r: 0.88, R squared: 0.78). The Pearson's r and R squared values for ΔG versus RMSF are 0.88 and 0.78 respectively.

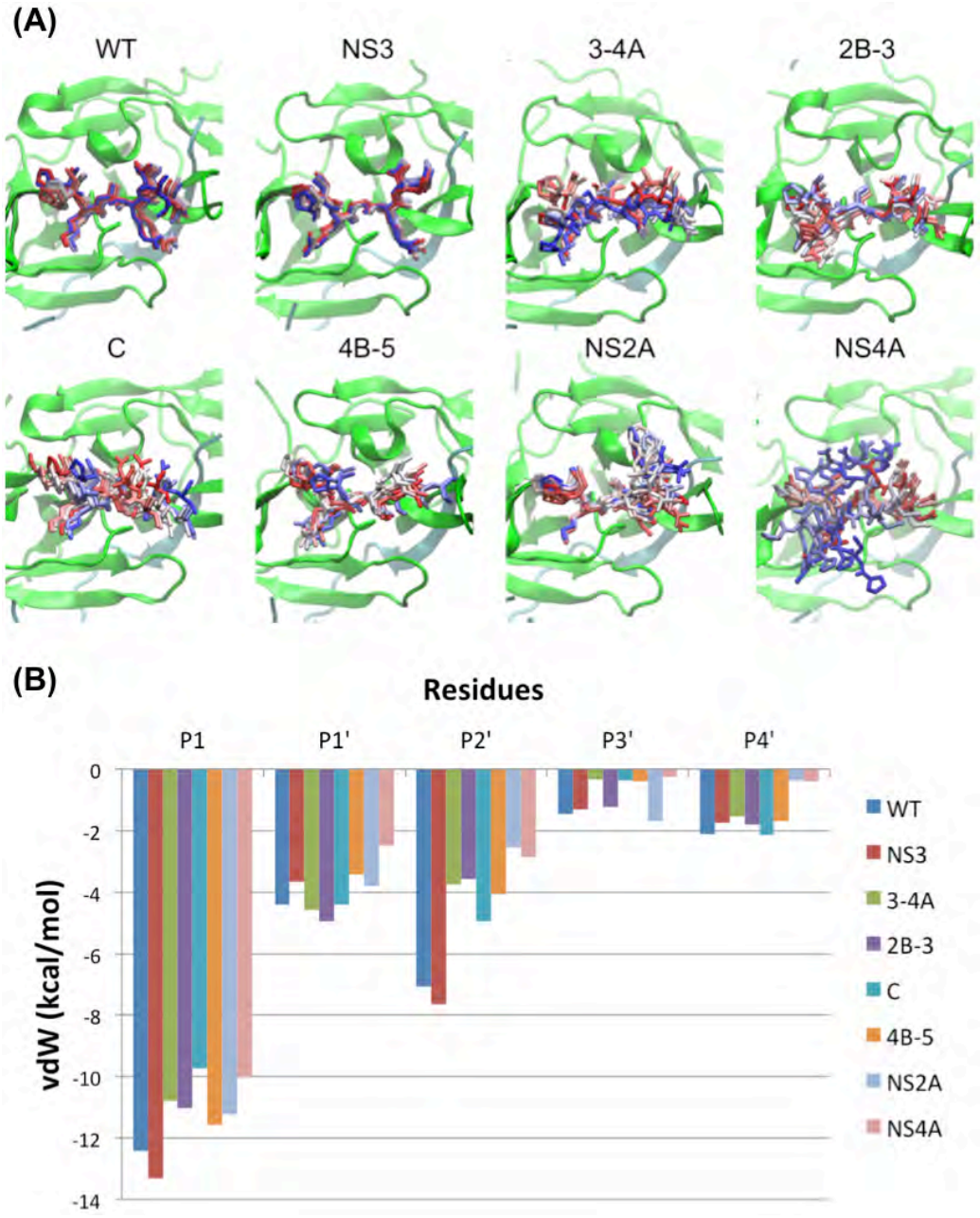
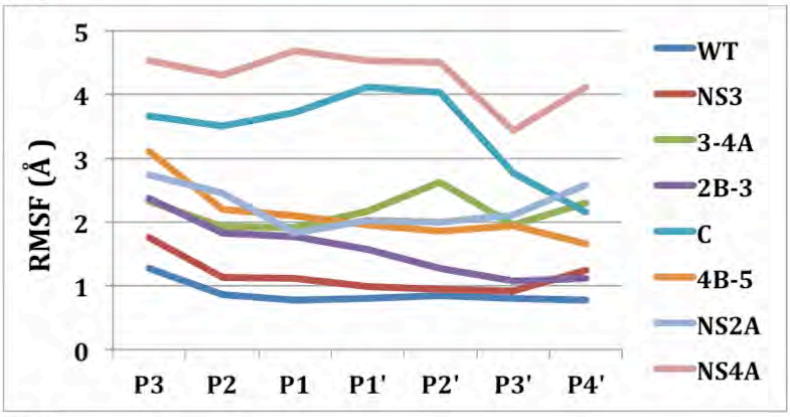


Figure 3.6. Molecular dynamics (MD) of binding loops bound to DENV protease active site. (A) Snapshots from MD simulation trajectories of WT-AP and aprotinin constructs bound to DENV3 protease. The protease is shown in cartoon representation and the aprotinin binding loops as sticks from start (in red) to end (in blue) of the trajectory. (the simulations were performed with full-length aprotinin, but only residues corresponding to native substrates are shown for clarity). (B) The binding loop vdW contacts during MD simulations of WT-AP and aprotinin constructs against DENV3 protease.

(A)



(B)

Aprotinin constructs	WT	NS3	3-4A	2B-3	C	4B-5	NS2A	NS4A
Binding loop RMSF (Å)	**0.9**	**1.2**	**2.2**	**1.6**	**3.4**	**2.1**	**2.3**	**4.3**

Figure 3.7. Aprotinin constructs' binding loop fluctuations correlate with affinity. (A) WT-AP and aprotinin constructs' RMSF values by residue. (B) The overall RMSF values of WT-AP and aprotinin constructs. (Snapshots were taken every ten nanoseconds throughout the simulation trajectory for the different sequences)

3.3.5 Substrate sequences with higher affinity have better packing at the protease active site

To investigate how packing of substrate residues at the protease active site contribute to protease binding, the average vdW contact energy was calculated over the simulation time and mapped onto the protease surface (Fig. 3.8B). The catalytic H51 packs against Cys14 and Cys38 (outside the binding loop of aprotinin), and F31 of protease packs against Thr32 of aprotinin located outside the binding loop. To determine the interactions that dictate substrate specificity, the overall vdW contact energy (P1 to P4') of residues in the binding loop of different aprotinin constructs was calculated. The two constructs with the lowest (most favorable) vdW interactions are the two tightest binders (Table 3.3, Fig. 3.9), WT-AP and $NS3_{BL}$-AP (-27.5 and -27.7 kcal/mol respectively). The $NS3_{BL}$-AP substrate residues pack against protease residues G151, S135, T134, G133, P132 and V36, with vdW contact energies ranging from -1.3 (G151) to -4.0 (P132) kcal/mol. Other residues with significant amount of contact energy include packing of Y161 (-4.0 kcal/mol) against P1 residue, P132 (-4.0 kcal/mol) against P1, P2' residues, V36 (-3.2 kcal/mol) against P3' residue, and I30 (-3.5 kcal/mol) against P4' residue. The binding loops of the other constructs ($2B3_{BL}$-AP, C_{BL}-AP, $4B\text{-}5_{BL}$-AP, $3\text{-}4A_{BL}$-AP and $NS2A_{BL}$-AP) have similar levels of vdW energies (-22.6 to -19.6 kcal/mol), and the weakest binder $NS4A_{BL}$-AP has the highest (least favorable) contact energy (-16.0 kcal/mol). With the exception of $3\text{-}4A_{BL}$-AP, the overall contact energies of the various constructs reflect the differences observed in the experimental K_i values, where

stronger binders have more contacts (lower energy) and hence pack better against protease active site residues.

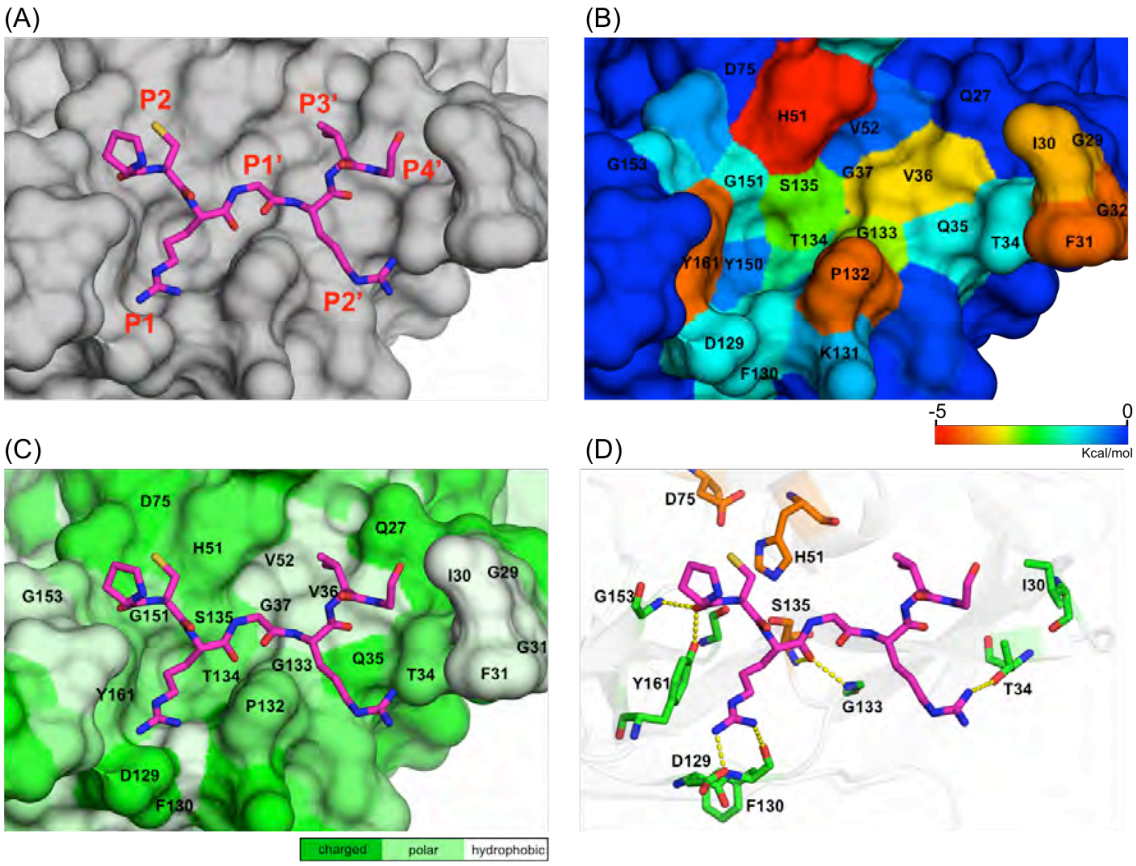


Figure 3.8. The interaction of substrate residues with DENV3 protease ($NS3_{BL}$-AP binding to DENV3 protease is shown as an example). (A) $NS3_{BL}$ is displayed as sticks and DENV3 protease is as surface. (B) DENV3 protease's surface colored according to the extent of vdW contacts with $NS3_{BL}$-AP. (C) DENV3 protease's surface is colored based on polarity. (D) Hydrogen bonds between $NS3_{BL}$-AP and DENV3 protease are displayed as yellow dashes.

Table 3.3. The values of binding loop van der Waals contacts calculated from MD simulations of WT-AP and aprotinin constructs against DENV3 protease.

Residue	WT	NS3	3-4A	2B-3	C	4B-5	NS2A	NS4A
P1	-12.4	-13.3	-10.8	-11.0	-9.7	-11.6	-11.2	-10.0
P1'	-4.4	-3.7	-4.6	-5.0	-4.4	-3.4	-3.8	-2.5
P2'	-7.1	-7.6	-3.7	-3.6	-4.9	-4.1	-2.5	-2.9
P3'	-1.5	-1.3	-0.3	-1.2	-0.4	-0.4	-1.7	-0.3
P4'	-2.1	-1.7	-1.5	-1.8	-2.2	-1.7	-0.4	-0.4
Overall	-27.5	-27.7	-21.0	-22.6	-21.6	-21.2	-19.6	-16.0

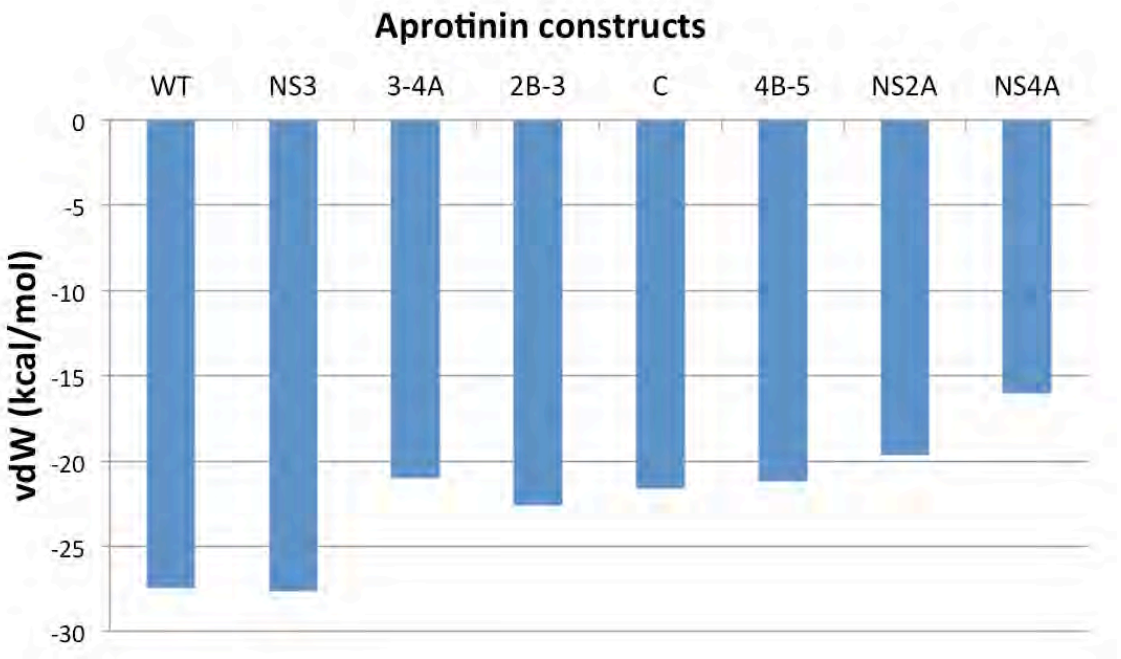


Figure 3.9. The overall binding loop vdW contacts with DENV3 protease calculated from MD simulations of WT-AP and aprotinin constructs.

3.3.6 Tightest binders have more hydrogen bonds

To quantify the contribution of hydrogen bonds to the binding of aprotinin constructs to DENV3 protease, the percentage of time each hydrogen bond existed during the MD trajectories were calculated (Table 3.4). The most highly populated hydrogen bonds are formed with P1 and P2'. The binding loop of WT-AP had 5 hydrogen bonds (3 strong and 2 weak ones) with DENV3 protease that existed more than 40% of the simulation time. $NS3_{BL}$-AP had four hydrogen bonds (3 strong and 1 weak) (Fig. 3.8D, Table 3.4). All other constructs had 3 hydrogen bonds with DENV3 protease, except 4B-5_{BL}-AP, which had only 2 hydrogen bonds. Thus the number of hydrogen bonds overall correlates with binding affinity.

The hydrogen bond between the P1 side chain and protease D129 side chain is the strongest interaction, and is conserved in all constructs (Fig. 3.8D). This hydrogen bond existed more than 70% of the simulation time for all constructs (Table 3.4). P1 forms additional side chain and backbone hydrogen bonds with the backbone atoms of F130 and G133 in the protease, respectively. Overall, the P1 residue contributes the most hydrogen bonds compared to other residues in the binding loop, acting as an anchor to ensure binding.

The major difference in hydrogen bonds of different aprotinin constructs was at the P2' position. There is one extra hydrogen bond between WT-AP or $NS3_{BL}$-AP's P2' Arg side chain and T34 side chain of DENV3 protease (Fig. 3.8D), which existed 54.3% and 61.3% of simulation time respectively (Table 3.4). This extra hydrogen bond stabilizes the P2' residue in WT-AP and $NS3_{BL}$-AP, resulting in more favorable vdW

contacts compared to P2’ residues in other constructs (Fig. 3.6B). Thus, Arg is more favorable than other amino acids in the P2’ position of substrate sequences.

Table 3.4. The hydrogen bonds between substrate residues at the binding loop and DENV3 protease for the various aprotinin constructs, and the percentage of time (yellow > 40%, red >60%) the hydrogen bonds existed during the MD simulations.

Aprotinin	Protease	WT	NS3	34A	2B3	C	4B5	NS2A	NS4A
P13-main	Y161-side	54.7%	21.7%	13.4%	40.2%		23.1%	33.3%	33.3%
K15-side	D129-side	86.4%	73.1%	86.8%	110.9%	79.8%	73.6%	87.7%	126.6%
K15-side	F130-main	71.4%	45.9%	73.5%	37.9%	63.7%	37.2%	62.8%	40.2%
K15-main	G133-main	78.5%	66.2%	49.9%	54.3%	61.9%	76.2%	80.2%	45.9%
R17-side	T34-side	54.3%	61.3%						

3.3.7 Interactions at individual positions reveal amino acids for optimal packing

The contribution of various residues to intermolecular interactions at individual positions of the binding loop was further examined (Fig. 3.6B, Table 3.3). The P1 position has the best vdW contacts across all constructs, with the P1 arginine of $NS3_{BL}$-AP both forming hydrogen bonds and having the lowest vdW contact energy (-13.3 kcal/mol), lower by 3.5 to 1.7 kcal/mol compared to the other constructs. However in all constructs this side chain is either Arg or Lys, so the relative level of interaction is reflective of the interdependence of the rest of the binding site recognition.

At the P1' position, serine in both the 2B-3_{BL}-AP and 3-$4A_{BL}$-AP constructs has slightly better contact energy compared to other P1' residues. This serine can form a weak hydrogen bond with the backbone of protease H51 or V36 (12.1% to 32.6% of simulation time respectively), and appears to be the most favored side chain at this position. $NS3_{BL}$-AP and 4B-5_{BL}-AP have a Gly at P1' position, resulting in contact energy below -4 kcal/mol. The worst contact energy at the P1' position is of $NS4A_{BL}$-AP's Thr; however this reflects interdependent loss of contacts and high loop fluctuations mentioned above (also observed for P2' to P4' positions) during the simulation.

At the P2' position, vdW contact energy has major variations among substrates, with WT-AP's and $NS3_{BL}$-AP's P2' Arg having significantly lower contact energy than other constructs' P2' residues. In addition to being the largest side chain with the most contact surface, the extra hydrogen bond between P2' Arg and T34 of the protease also stabilizes this residue (Table 1). C_{BL}-AP has the third most contacts, with the P2' serine also forming a hydrogen bond with Q35 backbone, but this hydrogen bond (24.0% of the

time simulated) is not as strong as the one formed by Arg. Serine has smaller contact surface than arginine, resulting in less favorable contact energy compared to WT-AP and $NS3_{BL}$-AP. Hence, Arg followed by Ser is favored for vdW contacts and hydrogen bonds at the P2' position.

At the P3' position, $NS2A_{BL}$-AP's Trp has the lowest contact energy. Tryptophan has a much larger side chain compared to the P3' residues in other constructs, and the extra surface area contributes to more vdW interactions. However, Trp is too big to optimally fit in the S3' pocket, and cannot maintain stable interactions with the protease throughout the simulation time (Fig. 3.6A). Poor fit and stability of this Trp causes the neighboring P2' and P4' residues to drift away from the correlated biding pockets, resulting in the worst vdW contacts at these two positions among all the constructs. The hydrophobic P3' residues of WT-AP, 2B-3_{BL}-AP and $NS3_{BL}$-AP (Ile, Val and Val respectively) have lower contact energy (-1.22 to -1.46 kcal/mol) compared to the remaining 4 constructs (-0.26 to -0.41 kcal/mol). Hence, P3' side chains smaller than Trp better fill the hydrophobic S3' pocket, by packing against V36 in the protease active site (Fig. 3.8C).

Finally, for the P4' position, glutamate in the $NS4A_{BL}$-AP has the least vdW contacts with the protease as this residue cannot pack against protease's hydrophobic loop (G29, I30, F31, G32) which defines the S4' pocket. This unfavorable interaction caused P4' residue to drift away from the active site, and $NS4A_{BL}$-AP was bound to DENV3 protease only through the interaction between P1 Arg and D129 of protease at

the end of the simulation. Because of the hydrophobic nature of the S4’ surface, a hydrophobic P4’ residue would be favorable at this position.

Overall, the intermolecular contacts contributed by different aprotinin constructs correlate well with the K_i values (Fig. 3.2, Table 3.3) and reveal which combination of residues might be optimal.

3.4 Discussion

In this study, the binding interactions of P' substrate residues to dengue protease were investigated by exploiting the high affinity serine protease inhibitor aprotinin. Substrate affinity was strongly modulated by different P' sequences with the inhibition constant varying over five orders of magnitude (Fig. 3.2), which indicates protease's affinity and processing may highly vary, potentially in both rate and temporal sequence, among various cleavage sites. The biological function of the internal cleavage at NS3 (tightest binder) is not clear (Arias et al., 1993), however, processing of 3-4A and 2B-3 (2^{nd} and 3^{rd} tightest binders respectively) is required to release mature NS3 protein (Zhang and Padmanabhan, 1993), which can then cleave the other cleavage sites. Hence, P' sequences likely play a key role in regulating the specificity of polyprotein processing.

Previously, substrate peptides have been used to investigate residue preference at each position, with contradictory results. For example, acidic residues Glu and Asp were found to be favored at P2' position in the context of nKRR-X(P2')XX (X: amino acid mixture except Cys, and isostere norleucine instead of Met) (Li et al., 2005a), but Glu was found to be the least favorable when screened within GLKR-G(P2')AK (Shiryaev et al., 2007). Using amino acid mixtures as a background can be misleading and mask the effect of specific residues at a given position, such as due to intramolecular interactions, while using a fixed background sequence may miss other key interdependencies. Linear substrate peptides corresponding to natural sequences spanning P4 to P4' have been used to investigate substrate cleavage (Shiryaev et al., 2007). However, with diverse residues

at the P positions (both basic and non-basic P2 residues, and distinct P3 and P4 residues) distinguishing the contributions of individual P' sites was very challenging.

Therefore, instead of changing P' substrate residues one by one or the whole cleavage site at once, our results reveal the effect of P1 to P4' positions as a group. Within this context, we identified the favored amino acids at each substrate position: basic residues Arg and Lys at P1, Ser at P1', Arg at P2' and hydrophobic residues at P3' and P4' sites. Compared to previous peptide-based screening results (Li et al., 2005a), the main discrepancies are at P2' and P3', which suggested acidic residues and serine to be favorable at P2' and P3' respectively. For the P2' pocket, hydrogen bonds and vdW contacts are likely the dominant forces, and potentially an acidic residue may form hydrogen bonds with the same T34 residue that Arg interacts with (Fig. 3.8D). Regarding P3' position, Ser may interact with Q27 through a hydrogen bond. However, we found that hydrophobic substrate residues pack well against V36 of the protease. Hydrophobic protease residues 30 and 31, which define the S4' pocket, were proposed to interact with the membrane (Li et al., 2005a, Chappell et al., 2008, Assenberg et al., 2009) and further studies are required to investigate the effect of hydrophobic interactions at this pocket. Thus, Arg/Lys-Ser-Arg-Ile/Val-Leu is the optimal sequence for P1–P4' positions of the dengue protease substrates. None of the natural sequences contain this combination (Fig. 3.1B) potentially suggesting an evolutionary advantage not to have the highest affinity combination for a particular cleavage site. This sequence may be a potent pan-flaviviral protease binder, based on the high level of homology among cleavage sites in the dengue viral serotypes and other flaviviruses such as the recently pandemic Zika virus.

Thus we suggest extending inhibitors to P' sites to enhance both affinity and specificity against dengue protease. Previous peptidomimetic dengue protease inhibitors have been designed mainly based on only P site substrate sequences (Yin et al., 2006a, Yin et al., 2006b, Nitsche et al., 2011), and usually contain basic P2 and P1 residues. Since dengue protease has similar substrate sequence preferences (basic P1 or P1/P2 residues) as human serine proteases (thrombin, trypsin and furin), further optimization of these inhibitors are required to increase specificity. In this study, we found that P' amino acids can significantly affect the binding affinity. Dengue protease does not share P' site sequence preference with other human serine proteases, therefore inhibitor design extending to or based on P' sites could increase both affinity and specificity.

Designing potent DENV protease inhibitors will likely require using a combination of strategies. Inhibitors spanning P1-P1' positions needed to mimic the transition state of the peptide cleavage reaction. Transition state mimicking compounds have proved successful in targeting HCV proteases (Malcolm et al., 2006), and enzymes have evolved to bind strongest to the transition state, rather than substrate or product. Therefore, incorporating chemical moieties into inhibitors to mimic the transition state could increase inhibitor affinity. In fact, the peptidomimetic inhibitor Bz-Nle-Lys-Arg-Arg-B(OH)2, reported to bind DENV2 protease tightly (K_i value 43 nM) (Yin et al., 2006b), uses a boronic acid as a warhead for the catalytic serine with the hydroxyl group mimicking the transition state. Based on our current study, we also suggest incorporating moieties that target the P' side of the active site, and possibly designing a relatively rigid macrocyclic inhibitor that could leverage the entropically driven interactions. Future

studies will need to explore and efficiently combine a series of these potential strategies for the design of DENV, or other flavivirus such as Zika, protease inhibitors to ensure both potency and specificity.

3.5 Conclusion

In Summary, different P’ residues significantly change the binding interactions between DENV3 protease and the substrate, over five orders of magnitude. The favored amino acids at each substrate position were revealed: basic residues Arg and Lys at P1, Ser at P1’, Arg at P2’ and hydrophobic residues at P3’ and P4’ sites. In addition, the binding interactions of aprotinin constructs to DENV3 protease are solely entropy driven. These properties should be considered when extending inhibitors to P’ sites to enhance both affinity and specificity against dengue protease.

3.6 Materials and Methods

3.6.1 Protease gene construction, protein expression and purification

Synthetic DENV3 protease gene (cDNA encoding NS2B cofactor (cNS2B; aa 1394-1440) and NS3 protease (NS3pro185; aa 1476-1660) with a G_4SG4 in between) was constructed using codon optimization for protein expression in *Escherichia coli*. Protease gene was constructed between BamH1 and Xho1 sites of pGEX6p1 (GE Healthcare). The plasmid was transformed into BL21(DE3) for protein expression. We followed the published protein expression and purification protocols (Li et al., 2005a).

3.6.2 Aprotinin gene construction, protein expression and purification

The cDNA encoding SUMO and aprotinin with His6 tag at N terminus was constructed between Nde1 and Sac1 of pET28a. The plasmid was transformed into BL21(DE3) cells for protein expression. We followed the published aprotinin expression and purification protocols (Sun et al., 2009).

3.6.3 Enzyme inhibition assay

FRET-based enzymatic cleavage assay was used to measure inhibition constants of aprotinin constructs against DENV3 WT protease. 100 nM of DENV3 NS2B/NS3 protease was incubated with varying concentrations of aprotinin for 60 min in 50 mM Tris assay buffer (20% glycerol, 1 mM CHAPS, pH 8.5) (Leung et al., 2001). Proteolysis reactions were initiated by adding 5 μM protease substrate [Ac-[D-EDANS]KRRSWP[K-DABCYL]-AMIDE] (21^{st} Century Biochemicals) and monitored using the EnVision

plate reader (Perkin Elmer) at excitation and emission wavelengths of 340 nm and 490 nm, respectively. The initial cleavage reaction velocities were determined using nonlinear fit to one-phase association of the whole progress curves (Salykin et al., 2013). Apparent inhibition constants (K_i) were obtained by nonlinear regression fitting of initial velocity versus inhibitor concentration to the Morrison equation using Prism 6 (GraphPad Software). Data were collected in triplicate and processed globally to calculate the shared inhibition constant and standard deviation.

3.6.4 Molecular dynamic simulations

All molecular dynamics simulations were performed in triplicate following previously published protocols (Ozen et al., 2014). Briefly, aprotinin constructs were modeled on DENV3 protease binding pocket based on the aprotinin and DENV3 WT protease complex structure (PDB: 3u1j). The modeled structures were prepared for simulations by keeping the crystallographic waters within 4.0 Å of any protease or aprotinin atom but removing the buffer salts from the coordinate file. By using Protein Preparation Tool from Schrodinger {Sastry, 2013 #32}, hydrogen atoms were added to the structures and the optimal protonation states for the ionizable side chains were determined. The hydrogen-bonding network of the initial structures was optimized and the structures were minimized in vacuum using the Impact refinement module with the OPLS2005 force field {Shivakumar, 2010 #33}. The prepared systems were solvated in a truncated octahedron solvent box with the TIP3P water model extending 10 Å beyond the protein in all directions, using the System Builder utility. The overall charge was

neutralized by adding the necessary number of counter ions (Na^+ or Cl^-). Desmond was used in all simulations with OPLS2005 force field. Each simulation was carried out at 300 K and 1 bar for 100 ns and the coordinates and energies were recorded every 5 ps.

3.6.5 MD results analysis

Hydrogen bonds were determined using VMD (Humphrey et al., 1996). A hydrogen bond was defined by a distance between the donor and acceptor of less than 3 Å and a donor–hydrogen–acceptor angle of greater than 150°.

The vdW contacts between the protease and aprotinin binding loop were calculated using a simplified Lennard-Jones potential $V(r) = 4\varepsilon\,[(\sigma/r)^{12} - (\sigma/r)^{6}]$, with the well depth ($\varepsilon$) and hard sphere diameter (σ) for each protease–aprotinin atom pair. The $V(r)$ for all protease–aprotinin atom pairs was computed within 6 Å, and when the distance between nonbonded pairs was less than e, $V(r)$ was considered equal to $-e$. More details of the rationale for this modification to the original 6-12 Lennard-Jones potential were described before (Ozen et al., 2011). The total $\Sigma V(r)$ of the protease–aprotinin complex was computed using this simplified potential for each nonbonded pair.

3.6.6 Isothermal titration calorimetry experiment

The ITC experiments were performed using Microcal ITC200 (Malvern) at 20 °C with the protease in the sample cell and aprotinin constructs in the syringe. The assay buffer was 20 mM Tris pH 8.5. By fitting the data with non-linear regression using Origin® 7.0, the change in enthalpy (ΔH) and corresponding dissociation constant (K_d)

were determined. The change in Gibbs free energy (ΔG) and entropy (ΔS) were derived from these values.

Chapter IV

Inhibitors of Dengue Protease: Exploiting the Primed Side

PREFACE

Manuscript of Chapter IV is in preparation for *ACS Chemical Biology* as:

Kuan-Hung Lin, Akbar Ali, Djade I. Soumana, Nese Kurt Yilmaz, Celia A. Schiffer. **Inhibitors of Dengue Protease: Exploiting the Primed Side.**

Author contributions: **K.-H.L.**, A. A., and C.A.S. designed research; **K.-H.L.** performed research; **K.-H.L.** and N.K.Y. analyzed data; **K.-H.L.**, N.K.Y., and C.A.S. wrote the paper.

Contributions from Kuan-Hung Lin: With the guidance of my mentor Celia Schiffer and my thesis research committee, I performed all the experiments including: cyclic peptide design, protein expression and purification, enzyme inhibition assay, MD simulations, and structure analysis. Akbar Ali and Djade I. Soumana contributed to the design of cyclic design. Nese Kurt-Yilmaz and Celia Schiffer guided interpretation of the data. I created all the figures and tables for the manuscript. I wrote the manuscript, and Nese Kurt-Yilmaz and Celia Schiffer provided editorial assistance.

4.1 Abstract

The mosquito-transmitted dengue virus (DENV) infects millions of people in tropical and sub-tropical regions. Maturation of DENV particles requires proper cleavage of the viral polyprotein, including processing 8 of the 13 substrate cleavage sites by dengue NS2B/NS3 protease. With no available direct-acting antiviral targeting DENV, dengue protease is a promising target for inhibitor design. We have previously found that the prime side of cleavage sites significantly modulates dengue protease binding affinity over several orders of magnitude, and identified optimal amino acids for each position. In this study, we designed a series of cyclic peptides based on the prime side of substrate cleavage (P') as inhibitors of dengue protease. By optimizing the length and amino acid sequence, the tightest cyclic peptide achieved a K_i value of 2.9 μM against DENV3 WT protease. These peptides mainly interact with the protease from P1 to P3'/P4' sites, and since dengue protease does not share P' site substrate sequence preference with other human serine protease, these cyclic peptides could be used as leads to design inhibitors with higher specificity.

4.2 Introduction

Dengue virus (DENV), the causative agent of disease dengue fever, is endemic in more than 110 countries with approximately 390 million people infected yearly leading to about 20,000 deaths (Monath, 1994, WHO, 2009, Bhatt et al., 2013). Currently, no direct-acting drugs are available either in clinic or development to combat dengue infections. Thus, a better understanding of the causative virus is needed to develop effective therapies.

DENV is a member of the family *Flaviviridae*, an enveloped virus with a positive single-strand RNA genome. There are four different serotypes (DENV 1–4), and each serotype shares 65-70% sequence identity of the genome (Rico-Hesse, 1990). Dengue RNA genome encodes a single polyprotein, which needs to get processed at the cytoplasmic side of host cell rough endoplasmic reticulum membrane by dengue NS2B/NS3 protease, and by the host cell peptidase at the lumen side (Chambers et al., 1990a). Dengue NS2B/NS3 protease is a serine protease that belongs to the chymotrypsin family with a classic Ser-His-Asp catalytic triad (Bera et al., 2007). The cofactor NS2B (cNS2B; amino acids 1394–1440) is required for the proper function of NS3 protease (NS3pro185; amino acids 1476–1660) (Yusof et al., 2000) and participates in substrate recognition (Noble et al., 2012). Dengue protease is responsible for the cleavage at 8 of the 13 polyprotein cleavage sites (Falgout et al., 1991). These cleavage steps are required for maturation of the viral particle, making dengue NS2B-NS3 protease a promising target for drug development.

Inhibitors targeting dengue protease have largely been based on the P side of substrate cleavage (Yin et al., 2006a, Yin et al., 2006b, Nitsche et al., 2011, Nitsche et al., 2012, Bastos Lima et al., 2015), and these inhibitors were improved to nanomolar level binding affinity according to a recently published study (Behnam et al., 2015). P side of substrates has been explored as P1 and P2 positions are rather conserved (basic amino acids) while the rest of the cleavage sequences are diverse. Since dengue protease has similar substrate sequence preferences as human serine proteases (Furin RXRR, thrombin P1 R, trypsin P1 R), these linear peptide-based inhibitors are not designed to be specific to the viral protease. General serine protease inhibitor aprotinin inhibits DENV2 protease with a K_i value 26 nM (Mueller et al., 2007). The binding loop of aprotinin is highly analogous in sequence to the native NS3 cleavage site and spans from P3 to P4' position at the active site of DENV protease (Noble et al., 2012). In chapter III, by engineering the binding loop of aprotinin, I recently identified the optimal amino acids for each of the P' positions. I determined that forming specific intermolecular interactions (such as hydrogen bonds contributed by P1 and P2' residues, hydrophobic packing of P3' and P4' residues), and maintaining the overall structure of the aprotinin's binding loop is critical for binding affinity.

I now take advantage of this knowledge, and design a series of cyclic peptides where we optimize the sequence and linker as inhibitors of dengue protease. To help maintain the loop structure in aprotinin, the binding loop (Pro13 to Ile18/Ile19) was cyclized with a second loop (Tyr35/Gly36 to Arg39) through a linkage between Ile18/Ile19 and Tyr35/Gly36. Optimizing the length and sequence of the cyclic peptide

led to an inhibitor with a K_i value 2.9 μM against DENV3 WT protease (Apro10). In addition, Arg54 on the protease was targeted to form more interactions (hydrogen bond and electrostatic interactions). This arginine is conserved through all four dengue serotypes but does not exist in human serine proteases (furin, thrombin and trypsin), and further optimization of interactions with this residue may increase the specificity of these peptides against dengue protease. The cyclic peptides we designed may be used as lead compounds in the design of DENV inhibitors with good specificity and potency.

4.3 Results

Previously, by using aprotinin as a platform and replacing the binding loop with corresponding native substrate sequences, we showed that P' side of substrates can significantly affect binding interactions with dengue protease and certain interactions are critical for binding (Chapter III). Taking advantage of these known interactions, we have designed cyclic peptides targeting DENV protease pocket from S3 to S4' position (Fig. 4.1). As we identified entropy to be the major driving force for binding, a second aprotinin loop was incorporated into the design with the aim of maintaining binding loop structure and rigidity. The binding loop of aprotinin (Pro13 to Ile18/Ile19) and a second loop (Tyr35/ Gly36 to Arg39) were linked together with or without glycine spacers between Ile18/Ile19 and Tyr35/Gly36. The disulfide bond between Cys14 and Cys38 already in aprotinin was kept to cyclize the peptide. The inhibition constants of cyclic peptides against dengue protease were measured by FRET-based enzymatic assays, and molecular dynamics simulations were applied to investigate the molecular interactions between cyclic peptides and the protease active site residues. To determine whether these cyclic peptides get cleaved by DENV3 protease during the FRET-based enzymatic assay, we performed the assay with or without protease and then analyzed these samples using liquid chromatography-mass spectrometry (LC/MS). Based on the results, these cyclic peptides did not get cleaved by the protease during the assay (Apro9 and Apro10 as examples) (Fig. 4.2, Fig. 4.3, Table 4.2).

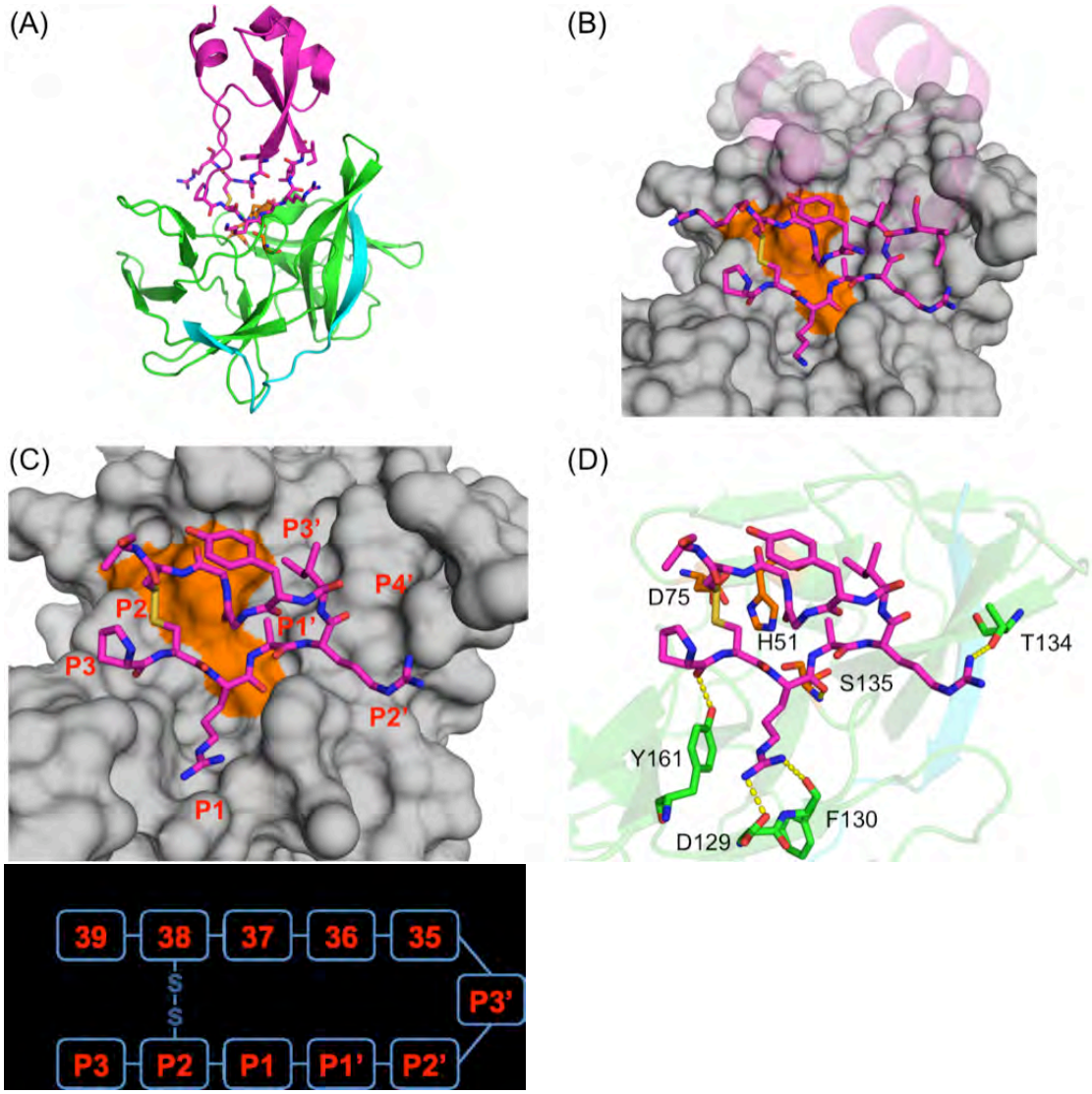


Figure 4.1. Design cyclic peptides derived from aprotinin as dengue NS3/2B protease inhibitor. (A) Aprotinin-DENV3 protease complex structure (3u1j). NS3 protease domain is in green, NS2B co-factor cyan and aprotinin purple. (B) The binding loop and 2nd loop of aprotinin are displayed as sticks. Dengue NS3/2B protease is shown as surface and the catalytic triad is in orange. (C) Cyclic peptide derived from aprotinin is displayed as sticks. The residues are numbered based on substrate residue position. (D) Hydrogen bonds between cyclic peptide and dengue NS3/2B protease are displayed as yellow dashes. (E) Naming of residues in cyclic peptide.

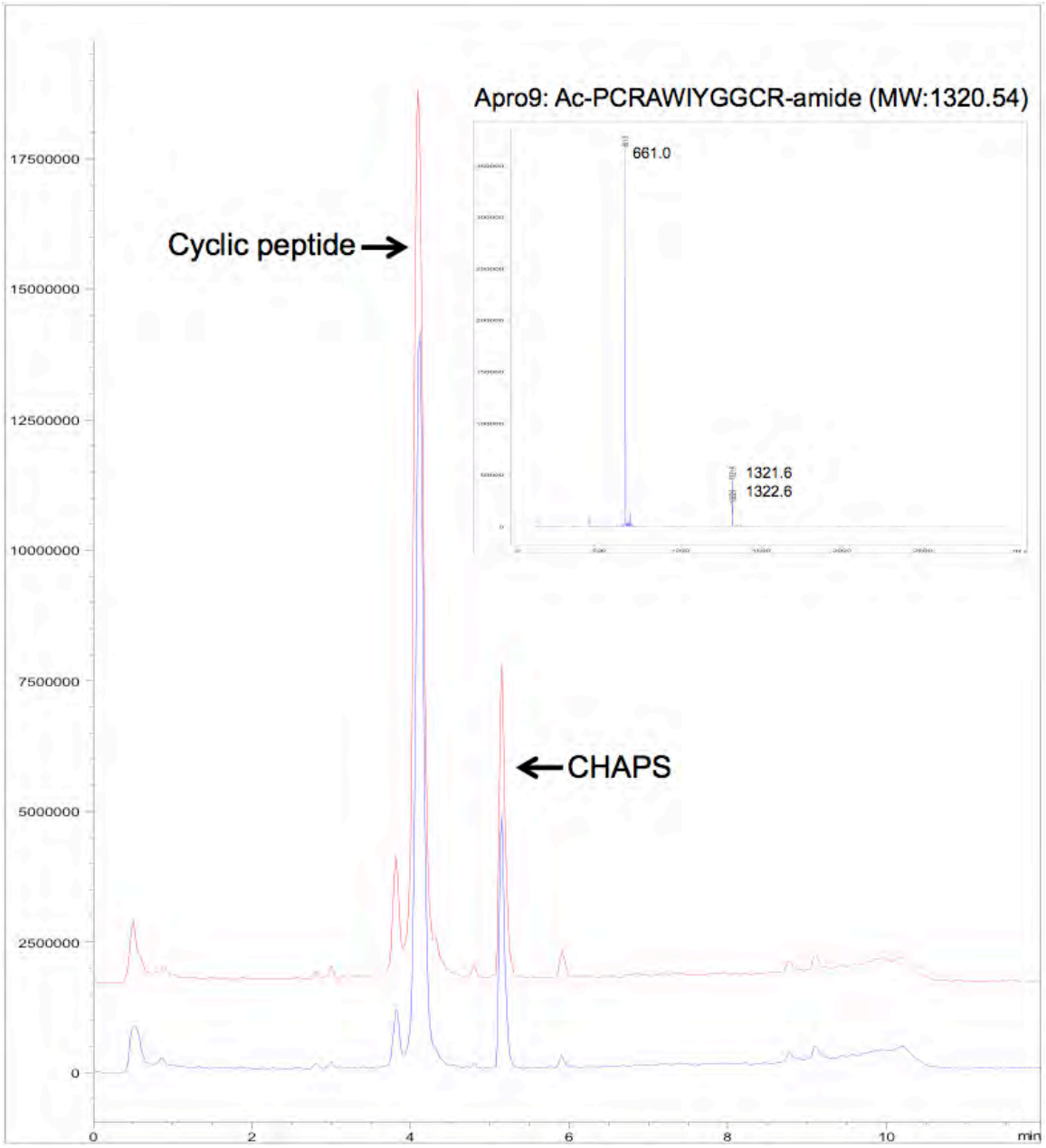


Figure 4.2. Liquid chromatography-mass spectrometry (LC-MS) analysis of Apro9 with (red) or without (blue) pre-incubation with DENV3 protease.

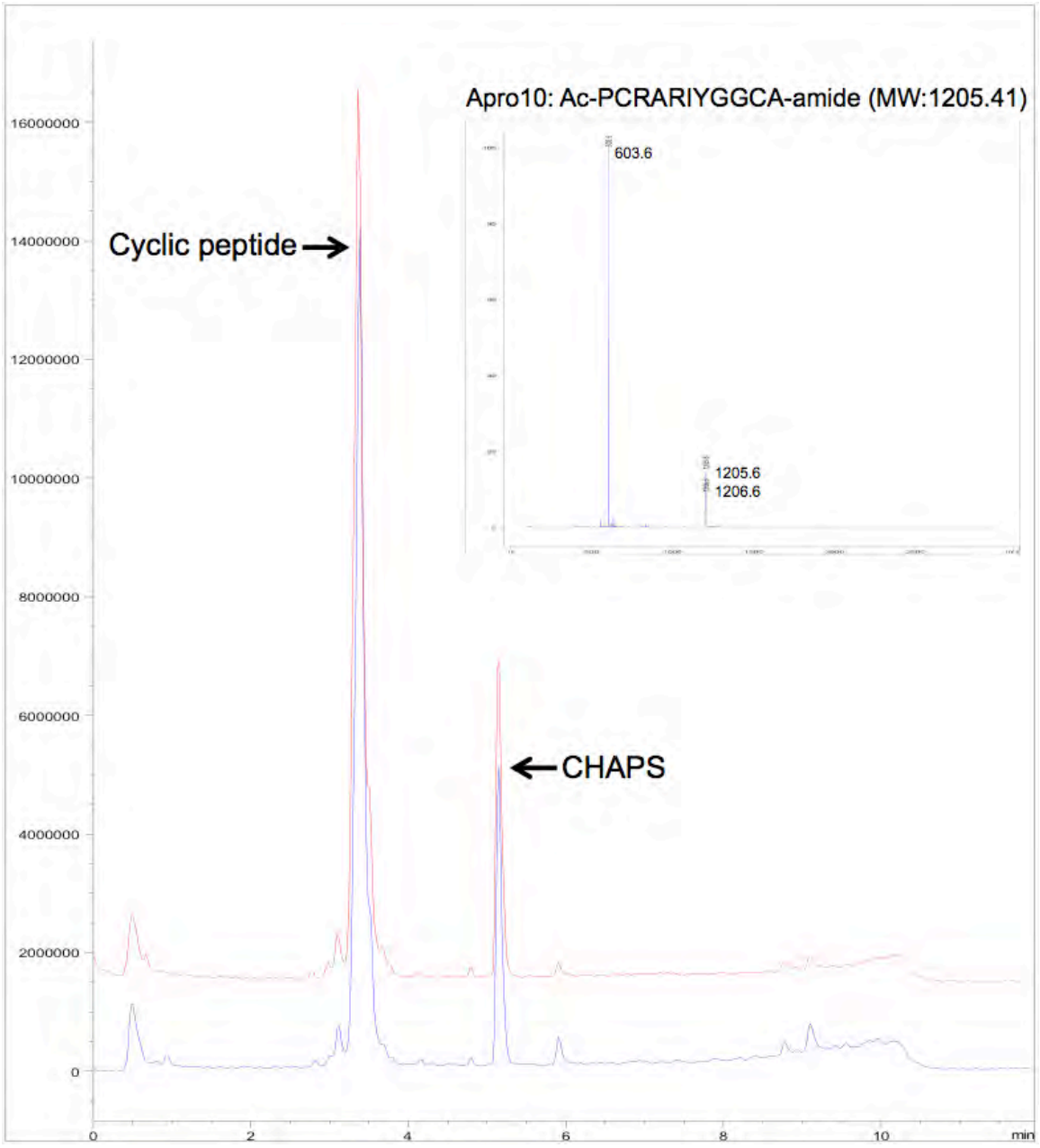


Figure 4.3. Liquid chromatography-mass spectrometry (LC-MS) analysis of Apro10 with (red) or without (blue) pre-incubation with DENV3 protease.

4.3.1 Optimizing the linker and peptide length

The inhibition constants of cyclic peptides against DENV3 protease vary from 2.9 μM (Apro10) to 780.3 μM (Apro14) (Table 4.1, Table 4.2). To determine the optimal length of cyclic peptide, peptides were designed with varying lengths from the longest Apro 3 (residues P3 to P4', two glycine linkers and residues 35 to 39) to the shortest Apro8 (residues P3 to P3', and with no linker and residues 37 to 39).

First, the linker between the binding loop and second loop was optimized for length. Comparing Apro1 to Apro5, or Apro2 to Apro10, no glycine linker is required between Ile18 and Tyr35. Without glycine linker, the K_i of Apro1 decreases from 376.8 μM to 14.5 μM (Apro5), and the K_i of Apro2 decreases from 966.1 μM to 2.9 μM (Apro10). Comparing Apro3 (K_i value 14.5 μM) to Apro5 (K_i value 678.3 μM), P4' isoleucine and two glycine residues as a linker is not favorable. Thus, no linker is required to connect the two loops and shorter cyclic peptides have better inhibition against dengue protease.

Next, we investigated length of the second loop, by comparing Apro10 to peptides with shorter second loops: For Apro8 (without Tyr35 and Gly36), Apro13 (without Gly36) and Apro14 (without Tyr35), the K_i values increase from 2.9 μM to 101.2 μM, 467.9 μM and 780.3 μM respectively. Therefore, the optimal number of residues for the second loop of the cyclic peptide between P3' Ile and Cys38 is three residues.

Table 4.1. The K_i values of different length of cyclic peptides against DENV3 NS3/2B protease.

Peptide name	Peptide sequence														K_i (μM)	Fold change
	P3	P2	P1	P1'	P2'	P3'	P4'			35	36	37	38	39		
Apro4	P	C	K	A	R	I	I	G	G	Y	G	G	C	R	232.3 ± 69.9	1.00
Apro3	P	C	K	A	R	I	I	G	G	Y	G	G	C	A	678.3 ± 244.1	2.92
Apro1	P	C	K	A	R	I		G		Y	G	G	C	A	376.8 ± 204.1	1.62
Apro2	P	C	R	A	R	I		G		Y	G	G	C	A	966.1 ± 328.8	4.16
Apro5	P	C	K	A	R	I				Y	G	G	C	A	14.5 ± 6.7	0.06
Apro6	P	C	R	A	R	I				Y	G	G	C	R	187.3 ± 43.2	0.81
Apro10	P	C	R	A	R	I				Y	G	G	C	A	2.9 ± 0.8	0.01
Apro13	P	C	R	A	R	I				Y		G	C	A	467.9 ± 68.7	2.01
Apro14	P	C	R	A	R	I					G	G	C	A	780.3 ± 195.5	3.36
Apro7	P	C	R	A	R	I					G	G	C	R	29.7 ± 5.8	0.13
Apro11		C	R	A	R	I				Y	G	G	C	R	145.1 ± 27.9	0.62
Apro8	P	C	R	A	R	I						G	C	A	101.2 ± 24.8	0.44

Table 4.2. The K_i values of cyclic peptides with sequence optimization against DENV3 NS3/2B protease.

Peptide name	Peptide sequence											K_i (μM)	Fold change
	P3	P2	P1	P1’	P2’	P3’	35	36	37	38	39		
Apro10	P	C	R	A	R	I	Y	G	G	C	A	2.9 ± 0.8	1.00
Apro16	Bz	C	R	A	R	I	Y	G	G	C	A	33.2 ± 6.5	11.45
Apro12	P	C	R	A	Q	I	Y	G	G	C	A	NB	NB
Apro9	P	C	R	A	W	I	Y	G	G	C	R	19.7 ± 8.8	6.79
Apro21	P	C	R	A	W	I	Y	G	G	C	A	144.8 ± 54.4	49.93
Apro17	P	C	R	A	R	I	D	G	G	C	A	69.3 ± 24.4	23.90
Apro18	P	C	R	V	R	I	D	G	G	C	A	35.8 ± 16.4	12.34
Apro19	P	C	R	V	R	I	Y	G	G	C	A	25.6 ± 9.9	8.83

4.3.2 Optimizing residues for P1 and P2' positions of the peptides

Dengue protease prefers basic residues lysine and arginine at the P1 position, and whether one of these residues is preferred over the other was tested. Apro1, with lysine at P1, has lower K_i (376.8 μM) compared to Apro2, with arginine at P1 (966.1 μM), but Apro10, with arginine at P1, has lower K_i (2.9 μM) compared to Apro5, with lysine at P1 (14.5 μM) (Table 4.1). Thus, the preference between the two basic amino acids is subtle and context dependent. While Arg can form one more hydrogen bond compared to Lys, both residues can form electrostatic interactions with the negatively charged residue D129 on the protease. The charge interaction might be the dominant force in this pocket, explaining no significant difference between these two amino acids.

Previously, I have shown that for different substrate sequences, arginine is favored at P2' position and can form both a hydrogen bond and more vdW contacts compared to other residues. In this study, arginine (big side chain, hydrogen bond partner), glutamine (smaller side chain, hydrogen partner) and tryptophan (big side chain with more contribution to vdW contacts) at the P2' position were compared: Apro10, Apro12 and Apro21 have K_i values 2.9, 1273 and 145 respectively (Table 4.2). Glutamine of Apro12, which can also form a hydrogen bond, is less favorable at this position compared to arginine (Apro10), and tryptophan of Apro21, which was designed to interact with protease through extensive vdW contacts, is also less favorable compared to Apro10. These results reflect that both hydrogen bonding and vdW contacts contributed by P2' residue are critical for the binding interactions in this pocket.

4.3.3 Optimal C terminal residue depends on the peptide length

To decrease the flexibility at the C terminus of the cyclic peptides, Arg39 (present in native aprotinin sequence) was replaced with alanine, resulting in Apro10. For longer peptides with linkers between the binding and second loop, Apro3, with an arginine (678.3 μM), gains potency in Apro4, with an alanine, with a lower K_i (232.3 μM). Similarly comparing Apro10, with an arginine, with Apro6, with an alanine, the affinity improves with the K_i value dropping from 187.3 μM to 2.9 μM (Table 4.1). Thus, decreased flexibility at the peptide C terminus is beneficial for tighter binding interactions. However, in contrast for shorter peptides where Tyr 35 has been also removed, Apro7, with an arginine, has a lower K_i value (29.7 μM) compared to Apro14, with an alanine (780.3 μM). Potentially in these smaller peptides with Tyr 35 removed the binding conformation is altered and the terminal arginine interacts with the S2' pocket, specifically with Asp75, resulting in the observed lower K_i value. Overall, the length of the peptide likely determines the position of the terminal residue when bound to dengue protease active site, resulting in peptides with different lengths favoring different amino acid at this position.

4.3.4 N terminal capping

To investigate the contribution of P3 residue (N terminus), the proline at this position was replaced with a smaller benzoyl capping group, which can mimic interactions contributed by the ring. Apro16 with the capping group has a higher K_i value (33.2 μM) compared Apro10 (2.9 μM), indicating proline is more favorable compared to benzoyl group (Table 4.2). Further, to investigate the requirement of a P3 residue, we designed Apro11 which does not have P3 residue. Comparing Apro6 and Apro11, removal of the whole proline residue does not significantly change the K_i values, which suggests that P3 proline is not required (Table 4.1).

4.3.5 Targeting Arg54 on the protease

Arg54 of dengue protease is conserved throughout all four serotypes of dengue virus but not in human furin, thrombin or trypsin. To build interactions with this particular residue and thus potentially enhance specificity, various residues on the second loop of the cyclic peptides were replaced with acidic residues. Replacing Tyr 35 with aspartate, Apro17 and Apro18 have K_i values 69.3 μM and 35.8 μM respectively (Table 4.2). While P1' alanine is more favorable over valine comparing Apro10 to Apro19, there is no significant difference between alanine and valine when Tyr35 is replaced with aspartate. Since the S1' pocket is relatively small, alanine may fit better compared to valine. However, the P1' valine may push the second loop closer to Arg54 on the protease, stabilizing the interaction between cyclic peptide's Asp35 and Arg54 of the protease. The interactions I aimed to build between these two residues were reflected by an extra hydrogen bond between this cyclic peptide-protease pair as discussing in later paragraph.

4.3.6 Cyclic peptide is more flexible compared to binding loop in aprotinin

Based on the K_i values of the cyclic peptides designed and assayed, both sequence and length of the cyclic peptide significantly affect binding to dengue protease. However, the K_i values are still higher compared to the same peptide sequence within the binding loop of aprotinin. I found before that entropy is the driving force for binding and fluctuations in MD simulations correlate with binding affinity for aprotinin constructs. To investigate how the dynamics of molecular interactions compare to those within the context of aprotinin constructs, we carried out MD simulations of the highest affinity cyclic peptide Apro10 bound to dengue protease (Apro10 was modeled into the binding pocket by aligning to aprotinin-DENV3 protease complex structure (PDB: 3u1j)). The dynamics were compared to aprotinin-protease complex structure (PDB: 3u1j) to examine whether those interactions critical for aprotinin binding were still maintained.

The snapshots of binding loop within aprotinin (the simulation included the whole aprotinin, but only the binding loop and second loop are displayed) and the corresponding cyclic peptide Apro10 are compared in Figure 4.4. Both the superimposed snapshots and RMSF values calculated indicate that the binding loop and second loop are more flexible in Apro10, with 0.7 Å and 1.6 Å higher RMSF values compared to the corresponding loops within the context of aprotinin, respectively. The second loop of Apro10 was incorporated into the design with the aim of sustaining the binding loop structure and rigidity as observed in aprotinin, however, these results illustrate that the cyclic peptide is relatively more flexible. The decreased rigidity of the cyclic peptide compared to aprotinin may be responsible for the increased K_i values (decreased affinity).

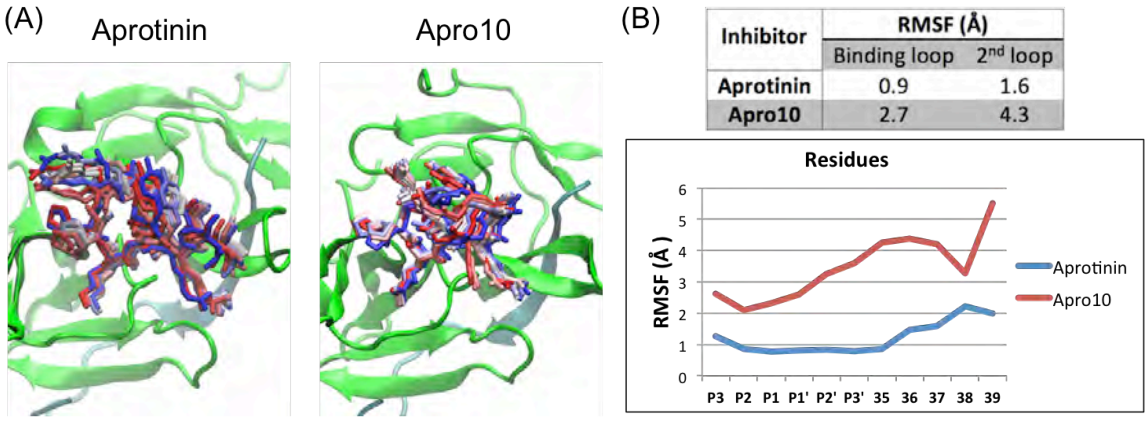


Figure 4.4. Molecular dynamics (MD) of aprotinin or apro10 bound to dengue protease active site. (A) Snapshots from MD simulation trajectories of aprotinin or apro10 bound to DENV3 protease. The protease is shown in cartoon representation and the aprotinin binding/2nd loops or apro10 as sticks from start (in red) to end (in blue) of the trajectory. (the simulations were performed with full-length aprotinin, but only residues corresponding to native substrates are shown for clarity). (B) The RMSF values during MD simulations are displayed as average of binding loop, 2nd loop, or individual residues.

4.3.7 Binding loop residues lose interactions in cyclic peptide compared to aprotinin

To investigate how each residue packs against the protease active site within the cyclic peptide, the average contact energy of each binding loop residue was calculated throughout the simulation time (Figure 4.5). Similar to the contacts of same positions within aprotinin binding loop, P1 Lys/Arg and P1' Arg contribute the most contact energies. Overall, residues in aprotinin have better packing compared to Apro10 at each corresponding position. As these binding loops within aprotinin and within the cyclic peptide have the same sequence except P1 residue (lysine or arginine), the differences may reflect how stable these residues stay in the binding pockets. Thus, residues lose favorable packing contacts within the cyclic peptide compared to aprotinin.

The intermolecular hydrogen bonds formed by the binding loop residues were also analyzed and compared to those in aprotinin (Table 4.3). The hydrogen bonds that are critical for aprotinin binding to the protease are conserved in the cyclic peptides; however, these hydrogen bonds are more stable in aprotinin. These hydrogen bonds formed in lower percentage of time, possibly due to the higher flexibility of the cyclic peptide.

For the peptide Apro17, which was designed to target Arg54 of the protease, there is one extra hydrogen bond formed between Asp35 of Apro17 and Arg54 of the protease, which is promising. As Arg54 on the protease is conserved within all four serotypes of dengue proteases but not in human serine proteases, further optimization of interactions with this residue is a potential strategy to increase the specificity of these peptides against dengue protease.

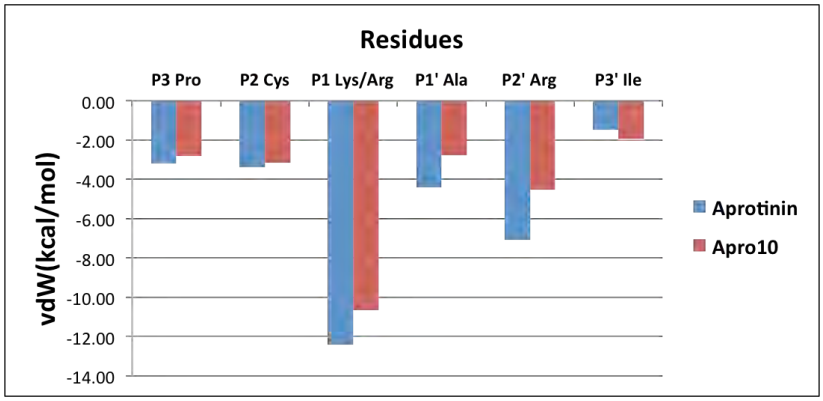


Figure 4.5. The binding loop vdW contacts during MD simulations of aprotinin or apro10 against DENV3 protease.

Table 4.3. The hydrogen bonds between substrate residues of aprotinin or cyclic peptides and DENV3 protease, and the percentage of time (yellow > 40%, red >60%) the hydrogen bonds existed during the MD simulations.

Ligand	Protease	Aprotinin	Apro10	Apro17
P13-Main	Y161-Side	55%	21%	10%
K/R15-Side	D129-Side	86%	76%	86%
K/R15-Side	F130-Main	71%	46%	22%
K/R15-Main	G133-Main	79%	58%	50%
R17-Side	T34-Side	54%	32%	48%
D35-Side	R54-Side			70%

4.4 Discussion

In this study, we designed cyclic peptide inhibitors targeting both P and P' sites of dengue protease, and optimized the sequence and length of these peptides. The best binder (Apro10) has a 2.9 μM K_i value against DENV3 WT protease. These inhibitors were derived based on the binding loop of aprotinin, and as intended, favorable interactions for aprotinin binding were preserved, specifically the hydrogen bonds contributed by P1 and P2' residues and vdW contacts of each residue position. However, these interactions in the cyclic peptides were not as strong as those in aprotinin upon binding to dengue protease. The RMSF values reflect that even though a second loop was included in the design to support the structure and stability of the binding loop, the cyclic peptide is still more flexible compared to the corresponding loop in aprotinin. As the binding between aprotinin and the protease is entropy driven (Chapter III), the increased flexibility may account for the decreased affinity between the cyclic peptides and protease, which explains why these cyclic peptides have higher K_i values against DENV3 WT protease compared to aprotinin.

To make sure the entropy of binding is favorable, forming hydrophobic interactions at P3' and P4' positions is also critical. Hydrophobic protease residues 30 and 31, which define the S4' pocket, were proposed to interact with the membrane (Chappell et al., 2008, Assenberg et al., 2009) and further studies are required to investigate the effect of hydrophobic interactions at this pocket.

Since most of P site inhibitors are highly charged and polar, design inhibitors forming hydrophobic interactions with prime site pocket can adjust the hydrophilicity of these compounds, potentially affects their cell permeability.

These cyclic peptides will serve as leads for future design. For enhancement of affinity and specificity of dengue inhibitors. To increase the affinity of the cyclic peptide against dengue protease, the rigidity needs to be increased, such as by using a carbon linker instead of disulfide bond to connect the binding loop to the second loop, using non-rotatable backbone, and further decreasing the overall size of the peptide. While there is no linkage between these two loops in cyclic peptides beside the ends, incorporating heterocycles between binding loop and 2^{nd} loop of cyclic peptides at different positions could be another approach to rigidify the peptides. To ensure specificity, Arg54 on the protease should be considered when designing P' site inhibitors, as this residue is conserved between all serotypes of dengue protease, but not in human serine proteases. Further optimization of the interactions with this residue could be a strategy to increase the specificity of these peptide inhibitors.

The strategies of optimizing cyclic peptides will be further discussed in chapter 5.3.

4.5 Conclusion

Overall, we leverage the potency of aprotinin and our detailed analysis of the specificity for dengue protease to design the scaffold for specific inhibitors. We design a series of cyclic peptides as DENV3 NS3/2B protease inhibitors, and the tightest cyclic peptide achieved a K_i value of 2.9 μM. Since dengue NS3/2B protease does not share substrate preference with human serine proteases and certain residue we targeted (Arg54) is conserved through all four dengue serotypes, the cyclic peptides we designed may be used as lead compounds in the design of DENV inhibitors with good specificity and potency.

4.6 Material and Methods

4.6.1 Protein and peptides

DENV3 protease gene (cDNA encoding NS2B cofactor [cNS2B; amino acids 1394–1440] and NS3 protease [NS3pro185; amino acids 1476–1660] with a G_4SG4 linker in between) was constructed for protein expression in *Escherichia coli*, as described previously (Noble et al., 2012). The protease gene was constructed between BamH1 and Xho1 sites of pGEX6p1 plasmid (GE Healthcare) and BL21(DE3) cells were used for protein expression. Protein expression and purification protocols published previously (Li et al., 2005a) were used. All cyclic peptides were purchased from 21st Century Biochemicals.

4.6.2 Enzyme inhibition assay

Fluorescence Resonance Energy Transfer (FRET) based enzymatic assay was used to determine the inhibition constants of cyclic peptides against DENV3 WT protease. 100 nM of the DENV3 protease was incubated with varying concentrations of cyclic peptides (5 mM to 500 nM) for 60 min in 50 mM Tris assay buffer (20% glycerol, 1 mM CHAPS, pH 8.5) (Leung et al., 2001). 5 μM of protease substrate [Ac-[D-EDANS]KRRSWP[K-DABCYL]-AMIDE] (21st Century Biochemicals) were added to initiate proteolysis reactions and monitored using the EnVision plate reader (Perkin Elmer) at excitation and emission wavelengths of 340 nm and 490 nm, respectively. One-phase association to nonlinear fit the whole progress curves was applied to determine the initial cleavage velocities (Salykin et al., 2013). By nonlinear regression fitting to the

Morrison equation of initial velocity versus inhibitor concentration using Prism 6 (GraphPad Software), apparent inhibition constants (K_i) were obtained. Data were collected in triplicate and processed independently to calculate the shared inhibition constant and standard deviation.

4.6.3 Molecular dynamics simulations

All molecular dynamics simulations were performed in triplicate following previously published protocols (Ozen et al., 2014). To prepare the modeled complex structures, aprotinin-bound DENV3 WT protease structure (3u1j) was used as a starting point. Aprotinin residues except the binding loop (Pro13 to Ile18/ Ile19) and the second loop (Tyr35/ Gly36 to Arg39) were removed from the structure, and these two loops were linked together with a peptide bond using Maestro (Suite 2012: Maestro, version 9.3, Schrödinger). The modeled structures were prepared for simulation by keeping the crystallographic waters within 4.0 Å of any protein atom but removing the buffer salts from the coordinate file. Next, hydrogen atoms were added to the structure using Protein Preparation Tool from Schrodinger (Sastry et al., 2013), and the optimal protonation states for the ionizable side chains were determined. Hydrogen bonding network of the initial structures was optimized, and by using the Impact refinement module with the OPLS2005 force field, the structures were minimized in vacuum (Shivakumar et al., 2010). Using the System Builder utility, these prepared systems were then solvated with the TIP3P water model extending 10 Å beyond the protein in all directions in a truncated octahedron solvent box. Overall charge of the systems were neutralized by adding

counter ions (Na^+ or Cl^-). Desmond was used in all simulations with the OPLS2005 force field. Each simulation was carried out at 300 K and 1 bar for 100 ns and the coordinates and energies were recorded every 5 ps.

4.6.4 Structural analysis

The hydrogen bonds were determined using VMD (Humphrey et al., 1996). A hydrogen bond was defined by the donor and acceptor distance less than 3 Å and a donor–hydrogen–acceptor angle of greater than 150°. To calculate the vdW contacts between cyclic peptides and protease, a simplified Lennard-Jones potential was applied, as described before (Ozen et al., 2011).

Chapter V

Discussion

5.1 Viral protease substrate recognition and drug design

HIV-1 protease inhibitors (PIs) are the only antiretroviral drugs used as a mono-therapy (Perez-Valero and Arribas, 2011), and they also have the highest intrinsic antiviral activity among HIV-1 drugs (Shen et al., 2008, Jilek et al., 2012), making PIs one of the most effective antiretroviral drugs (Rabi et al., 2013). Therefore, even with the occurrence of PI drug resistance and the availabilities of inhibitors against other HIV-1 viral protein targets, a deeper understanding of how protease recognize its ligands and optimizing PIs is still required for the development of antiviral therapy.

Viral proteases are able to recognize diverse substrate sequences. HIV-1 protease substrates have heterogeneity at all positions (Table 1.1). HCV protease substrate sequences are similar at P6 (acidic residues), P1 (mainly cysteine) and P1' (mainly serine) positions; however, residues from P5 to P2 and P2' to P4' are diverse (Table A.1). Dengue protease substrate sequences show homology at P2 and P1 (basic residues with exceptions at P2), and P1' (small/polar residues with exceptions), but no similarity at other positions (Figure 1.9). Since most FDA approved HIV-1 and HCV protease inhibitors are peptidomimetics (except HIV-1 protease inhibitor TPV), applying the knowledge of how viral proteases recognize diverse sequences is critical to drug design and development.

For both HIV-1 and HCV proteases, the recognition motif is not a consensus substrate sequence but a consensus volume shared by substrates. Previous studies in the Schiffer lab on HIV-1 protease substrate recognition showed that the protease recognizes the substrates through a three dimensional consensus volume occupied by different

substrates, and this volume spans from P4 to P4' substrate positions (Prabu-Jeyabalan et al., 2002). The same phenomenon was also observed in HCV: HCV NS3/4A protease recognizes N-terminal of substrates through a three-dimensional consensus volume occupied by different substrates, and this consensus volume was defined as the substrate envelope as in HIV-1 protease (Romano et al., 2010). Viral proteases are less likely to develop drug resistance mutations to inhibitors that fit within the substrate envelope because protease mutations that affect these inhibitors' binding would simultaneously affect substrate binding.

Inhibitors that protrude out the substrate envelope were found to be susceptible to resistance mutations. HIV-1 I50V protease, for example, was observed in patients failing amprenavir and darunavir treatment; these protease inhibitors protrude out of the HIV-1 protease substrate envelope and interact with residue I50 (Partaledis et al., 1995, Van Marck H, 2007, Vermeiren et al., 2007). The I50V removes this key van der Waal contact with the inhibitor. HCV protease drug resistance mutation R155K is often observed in patients treated with protease inhibitors with a big aromatic P2 moieties. In danoprevir (Lim et al., 2012) the P2 moiety protrudes out substrate envelope and forms a pi-stacking interaction against arginine 155 that is lost with the R155K. Furthermore, HIV-1 protease inhibitors which were designed by using substrate envelope as a constraint were shown to retain robust binding to multi-drug resistance protease variants (Nalam et al., 2013).

Protease–substrate co-evolution was observed in patients who received therapy that included HIV-1 protease inhibitors. For example, mutations at the p1-p6 cleavage site are statistically associated with protease mutation I50V (Kolli et al., 2009), and these

co-evolved substrates rescue mutated protease's activity. In Chapter II, I showed that co-evolved substrates can rescue binding interactions by forming more vdW contacts and hydrogen bonds, or by restoring the dynamic fluctuations of protease binding site. Meanwhile, co-evolved substrates maintain a comparable fit within the substrate envelope regardless of whether the protease carries the I50V/A71V mutations or not, supporting that substrate envelope is the recognition motif.

Since protease inhibitors are more rigid than the substrates and can not mutate to accommodate the lost binding interactions with the mutant protease, the protease-substrate co-evolution would skew the balance between inhibitor binding and substrate processing in favor of the latter and cause drug resistance. Therefore, this substrate envelope should be used as a constraint in protease inhibitor design to avoid drug resistance.

As dengue has a large amount of variation between serotypes, there is a high probability that resistance might emerge for any direct acting antiviral. Since the substrate envelope has been proven to be a critical constraint for inhibitor design targeting HIV-1 or HCV proteases, identifying the dengue NS3/2B protease substrate envelope would be beneficial to designing a Dengue Protease inhibitor. I have tried to co-crystallize different aprotinin constructs with DENV3 protease, however, the crystallization trials were not successful. Methods like co-crystallizing inactive DENV3 NS3/2B protease with substrates or co-crystallizing DENV3 NS3/2B protease with non-cleavable substrates (or products) are potential ways to get substrate complexed structures, which will help us define dengue NS3/2B protease's substrate envelope.

Along with DENV3 protease, I have also tried to co-crystallize active or inactive DENV2 proteases with substrates or substrate products, and these trials will be discussed in Appendix B.

5.2 Dynamics of protease substrate/inhibitor interactions

Protein complex structures enable investigating protein-protein or protein-ligand interactions at the molecular level; however, the atomic fluctuations, which are critical for intra- and inter- molecular interactions, are not always captured in the static crystal structure. NMR experiments are performed in solution and allow us to capture the flexibility of proteins or the dynamics of protein-ligand interactions; nevertheless, this method cannot fully elucidate the conformational flexibility in atomic scale. To fill this gap, molecular dynamic simulations have been successfully applied to address the dynamics of protein-protein interactions and protein folding at atomic resolution (Cho et al., 2008, Pan et al., 2013).

Throughout my thesis I used molecular dynamics (MD) to capture some of these motions. In Chapter II, I show that co-evolved p1-p6 substrates rescue the HIV-1 I50V protease's binding activity by forming more vdW contacts and hydrogen bonds (Fig. 2.4, Table 2.2), and molecular dynamics simulations suggest that co-evolution restores the dynamics at the active site (Fig. 2.8). Studying static structures alone cannot elucidate these phenomena.

In Chapter III, the MD analysis of aprotinin constructs bound to DENV3 protease revealed that the binding loop fluctuations, vdW contacts and hydrogen bonds all correlate well with the experimental K_i values and also highlight the favorable binding interactions (Fig. 3.2, Fig. 3.6, Fig. 3.9, Table 3.4). In Chapter IV, dynamic analysis helped us understand why cyclic peptides derived from aprotinin experience a significant potency loss against DENV3 protease (Fig. 3.2, Table 4.1, Table 4.2, Fig. 4.4). The

RMSF values derived from MD simulations show that even though we incorporated the second loop in cyclic peptide to sustain the binding loop structure as observed in aprotinin, both loops are more flexible compared to corresponding loops in aprotinin (Fig. 4.4). Since the entropy is the main driving force for binding (Fig. 3.4), the loss of rigidity causes the cyclic peptides to lose affinity against DENV3 protease.

In Appendix A, I present that the impact of the distal V36M mutation on the HCV NS3/4A protease-inhibitor complex crystal structures is subtle, with little change in the inhibitor binding mode, intermolecular hydrogen bonds and salt bridges (Table A.4, Table A.5). However, MD simulations results reveal that the impact of V36M mutation propagates to the binding site through B1 β-strand, and this distal modulation decreases the size of the active site (Fig. A.3, Fig. A.4). Since HCV protease inhibitors typically protrude beyond the dynamic substrate envelope (Ozen et al., 2013), this shrinking is expected to impair inhibitor binding more than substrate recognition, and contribute to drug resistance.

Traditional structure-based drug design approaches allow us to design inhibitor targeting certain static state of a protein of interest. However, static structures are not enough to explain the existence of drug resistance in several cases (Fig. 2.8, Fig. A.3, Fig. A.4). MD simulations enabled investigating the dynamic interactions between protein and ligand, which could not be captured in the static crystal structures. Since these protein-protein and protein-ligand dynamics are crucial in determining inhibitor binding and drug resistance, incorporating the mutational

and conformational flexibility of target protein into drug design would be critical to avoid resistance.

5.3 Optimization of dengue protease inhibitors

I designed a series of aprotinin derived cyclic peptides as DENV3 NS3/2B protease inhibitors, and the tightest cyclic peptide achieved a K_i value of 2.9 μM (Table 4.1). However, these cyclic peptides are still weak DENV3 protease binders compared to aprotinin (3.7 nM) (Fig. 3.2). Entropy is the main driving force for binding of all aprotinin constructs to dengue protease (Fig. 3.4), and compared to the binding loop of aprotinin, aprotinin-derived cyclic peptides are much more flexible upon binding to DENV3 protease (higher RMSF values) (Fig. 4.4). Therefore, to make these cyclic peptides better binders, their rigidity needs to be increased. Certain strategies to address this problem including incorporating non-reducible bond, non-rotatable bond, peptide bioisosteres, heterocycles and macrocycles among side chains to the peptide design, and shortening the length of cyclic peptide.

The two loops of cyclic peptide inhibitor (Apro10 as an example) are linked together through a peptide bond between Ile18 and Tyr35, and a disulfide bond between Cys14 and Cys38 (Fig. 4.1, Table 4.1, Table 4.2). Nevertheless, this disulfide bond can be broken under reducing conditions, making it unstable. Using a non-reducible bond instead is a strategy to increase the stability of the cyclic peptide, or incorporating heterocycles at the same position to increase both stability and rigidity.

Since the peptide bond between P1 and P1' residues is still cleavable, replacing with a non-cleavable bond would be beneficial to the integrity of the cyclic peptide. The carbonyl group of P1 residue forms a hydrogen bond with protease's Ser135 backbone amine groups (Fig. 4.1D, Table 4.3); therefore, replacing this peptide bond with peptide

bioisosteres such as ester, thioester, ketoethylene and ketone, this hydrogen bond can be maintained while making the bond non-cleavable.

Decreasing the length of cyclic peptide is a feasible way to increase the overall conformational entropy. Since the length of second loop is already optimized, the overall length of the peptide can be decreased by cutting residues at both termini. Further, even though the overall length of the second loop is optimized, the rigidity of this loop can be increased by incorporating non-peptide-like rigid chemical groups. Meanwhile, all backbones within the cyclic peptides are rotatable, giving the cyclic peptides great flexibility. Hence, replacing backbone bonds flanking the peptide bonds (phi and psi) with non-rotatable bonds would also increase rigidity. This design should be done carefully, investigating the dynamics of corresponding bonds in the aprotinin-DENV3 complex structure.

Macrocyclic inhibitors were shown to successfully inhibit HCV NS3/4A protease compared to their linear analogs [Summa, 2012 #658;Jiang, 2014 #233;Rosenquist, 2014 #231]. These inhibitors were cyclized through links between side chains (Fig. A.1), unlike through-backbone bonds in dengue cyclic peptides. Since dengue protease's S2' and S4' pockets are solvent exposed and overlapping (Fig. 4.1C), similar to the S4 and S2 pockets (or S3 and S1 pockets) of HCV NS3/4A protease, introducing a P2'-P4' macrocycle to these cyclic peptides can potentially increase both peptide rigidity and also the packing against residues in these two pockets.

Besides the prime site cyclic peptides, cyclizing dengue P site peptides is another strategy to increase both rigidity and specificity, which will be discussed in Chapter 5.4 and Appendix D.

5.4 Design of dengue inhibitors with higher specificity

Dengue protease shares similar substrate sequence preferences with human serine proteases (furin RXRR, thrombin P1 R, trypsin P1 R) (Lim et al., 2013). Therefore, increasing specificity of dengue protease inhibitor is a necessary but challenging endeavor. With the lack of investigation into increasing the specificity of dengue protease inhibitor, we can apply the lesson we learned from the HCV protease inhibitor design (targeting conserved residue) to address this problem. Also, we can take advantage of dengue NS3/2B protease's unique structural feature at the active site and design inhibitor with higher specificity.

Residue Cys159 of HCV NS3/4A protease is conserved across all HCV genotypes and subtypes, and this residue is structurally unique to HCV protease (Hagel et al., 2011). Covalent inhibitors designed to interact with this cysteine were shown to have high specificity against HCV protease (Hagel et al., 2011). Therefore, targeting an active site residue that is conserved among various genotypes (or serotypes) of virus is a feasible strategy to increase inhibitor specificity. Residue Arg54 in the protease domain of dengue NS3/2B protease is conserved across all serotypes of dengue virus. In addition human serine protease furin, thrombin and trypsin do not have arginine at the analogous position. Thus targeting this arginine by building electrostatic interactions or hydrogen bonds can increase inhibitor specificity against dengue protease.

Even though dengue protease shares similar substrate sequence preferences with human serine proteases, the active site structure of dengue protease is configured very differently from these human proteases. Dengue protease has a flat and solvent exposed

S1 pocket, but the shapes of these human proteases' S1 pockets are relatively narrow (Fig. D.1). Therefore, I expect that cyclized P site peptides (tetra-peptide with cyclization between N terminus capping and modified P1 side chain as an example) (Fig. D.2) will still bind to dengue protease, but might be too big to fit into human serine protease active sites. The macrocylic inhibitor design targeting this pocket will be further discussed in Appendix D.

5.5 Implications for Zika virus

In addition to the challenges we face in drug design against dengue virus, there is no effective therapy against Zika virus either. Zika virus (ZIKV), first isolated from febrile sentinel rhesus monkey in 1947 (Dick et al., 1952), had a major outbreak in Brazil recently, with an estimated 500,000-1,500,000 cases in 2015 (Bogoch et al., 2016). Zika virus, a mosquito-borne virus that belongs to *Flaviviridae*, *Flavivirus*, is an enveloped virus with a positive single-stranded 10,794 nucleotide RNA genome, which encodes a single polyprotein. Similar to dengue virus, the ZIKV polyprotein is further processed into three structural and seven non-structural proteins by the viral protease (Kuno and Chang, 2007, Faye et al., 2014).

DENV protease and ZIKV protease structures are highly similar; DENV3 protease (PDB: 3U1I) and ZIKV protease (PDB: 5LC0) structures have a RMSD value of 0.6 Å. ZIKV and DENV proteases active site residues are highly conserved (72% identity, 88% similarity), and the tightest prime site substrate NS3 is also highly conserved among different serotypes of DENV and ZIKV (Table 3.1). Therefore, the lessons I learned from inhibitor design targeting DENV protease should be applicable to ZIKV, and the cyclic peptide (P1 to P3') I designed could be a starting scaffold for ZIKV inhibitor design. The main difference between DENV and ZIKV proteases is at the NS2B cofactor domain. ZIKV protease has an aspartic acid at position 81 (glycine in dengue protease), located between S3 and S2 pockets. How an acidic residue at this position would affect substrate or inhibitor binding is not clear. Further research is required to study this specific residue's contribution.

Appendix A: Distal Mutation V36M Allosterically Modulates the Active Site to Accentuate Drug Resistance in HCV NS3/4A Protease

A.1 Preface

Manuscript of Appendix A is in preparation as:

Ayşegül Özen, Kuan-Hung Lin, Keith P Romano, Davide Tavella, Alicia Newton, Christos J. Petropoulos, Wei Kuang, Cihan Aydin, and Celia A. Schiffer. **Resistance from Afar: Distal Mutation V36M Allosterically Modulates the Active Site to Accentuate Drug Resistance in HCV NS3/4A Protease.**

Author contributions: A.Ö. and C.A.S. designed research; **K.-H.L.,** A.Ö., K.P.R., D.T. and C.A. performed research; **K.-H.L.,** A.Ö., D.T., and N.K.Y. analyzed the data.

Contributions from Kuan-Hung Lin: I performed protein expression and purification and solved three crystal structures for this study. I created figures for the manuscript. Nese Kurt-Yilmaz and Celia Schiffer guided interpretation of the data.

Former lab member Keith Romano solved the crystal of WT, R155K, R155K/V36M NS3/4A protease in complex with telaprevir and Danoprevir. These structures were analyzed by former lab member Ayşegül Özen and also Davide Tavella from Dr.

Francesca Massi's lab. Aysegul and Davide performed molecular dynamics simulations to investigate the dynamic interactions between HCV NS3/4A protease and the inhibitors. The enzyme inhibition assay was performed by former lab member Cihan Aydin.

A.1.1 Hepatitis C virus

Hepatitis C virus is a worldwide health burden with an estimated 170 million individuals chronically infected, and the infection of HCV causes severe liver diseases include fibrosis, cirrhosis and hepatocellular carcinoma (WHO, 2014). There are six genotypes of HCV (genotypes 1-6) with more than 50 subtypes, genotype 1 is the most common one globally (46%), followed by genotype 3 (22%), and genotype 3 and 4 (13% each) (Simmonds et al., 2005, Gower et al., 2014). Each genotype is further subcategorized in to different subtypes (a, b, c, etc.).

HCV is RNA virus belongs to the family *Flaviviridae* and genus *Hepacivirus*. The 9.6 kb RNA genome of HCV has a open reading frame (ORF) flanked by 5' and 3' untranslated regions (UTRs), this ORF is translated via internal ribosome entry site (IRES). The genome gives rise to a single polyprotein, which is further processed to structural proteins (core, E1, E2 and p7) and nonstructural proteins (NS2, NS3, NS4A, NS4B, NS5A and NS5B) (Major and Feinstone, 1997).

A.1.2 HCV NS3/4A protease and substrate recognition

HCV genome is translated at ER membrane by host replication machinery into a single polyprotein. The polyprotein then be processed by cellular proteases (signalase and signal peptide peptidase) (Moradpour et al., 2007) and viral proteases (NS2/NS3 protease and NS3/4A protease) (Egger et al., 2002, Chang et al., 2003) and give rise to structural proteins (core, E1 and E2) and nonstructural proteins (p7, NS2, NS3, NS4A, NS4B, NS5A and NS5B). NS3/4A protease cleaves at four viral polyprotein cleavage sites (3-

4A, 4A-4B, 4B-5A and 5A-5B), and these cleavages are required for the replication and maturation of viral particle, making the protease a promising drug target. NS3/4A protease also cleaves human cellular targets TRIF and MAVS (Kawai et al., 2005, Li et al., 2005b, Seth et al., 2005). These substrate sequences show homologies at P6 (acidic residues), P1 (mainly cysteine) and P1' (mainly serine), however, residues from P5 to P2 and P2' to P4' show heterogeneity (Table A.1).

HCV NS3 protein is a 631 amino-acid bi-functional protein, with a serine protease domain at the N-terminus followed by C-terminal NTPase/ RNA helicase domain. NS3/4A protease, a chymotrypsin-like protein with two β-barrel domains, has a catalytic triad (H57, D81, S139) located in a cleft separating the two barrels. NS4A, a 54-amino acid peptide, is required as a cofactor for optimal proteolytic activity. The central 11 amino acids of the cofactor inserts as a β-strand to the N-terminal β-barrel of NS3 to form the active enzyme (Yao et al., 1999).

The study published by Schiffer lab in 2010 showed that the NS3/4A protease recognizes N-terminal of substrates through a three dimensional consensus volume occupied by different substrates, this consensus volume was defined as substrate envelope as in HIV protease (Prabu-Jeyabalan et al., 2002, Romano et al., 2010). Inhibitors that fit within the substrate envelope were proposed less likely to be susceptible to drug-resistant mutations, since protease mutations impacting such inhibitors would simultaneously impact the processing of substrates. For example, protease inhibitors, ITMN-191 (Jiang et al., 2014), TMC435 (Rosenquist et al., 2014) and boceprevir (Malcolm et al., 2006) all have P2 moieties protrude out the substrate

envelope and contact protease's residues A156 and R155, and these residues were reported to mutate and cause drug resistance (Kieffer et al., 2007, Sarrazin et al., 2007b, He et al., 2008, Tong et al., 2008, Yi et al., 2009, Lenz et al., 2010).

A.1.3 HCV NS3/4A protease inhibitors as antivirals

Protease inhibitors Telaprevir (Perni et al., 2006, Kwong et al., 2011) and Boceprevir (Malcolm et al., 2006) were approved in 2011 and used as part of the direct-acting antivirals (DAAs) for HCV genotype 1 infections. The cure rate increased from 45% to 70% when these PIs were added to standard PEG-IFN/ribavirin treatment (Jacobson et al., 2011, Poordad et al., 2011).

HCV NS3/4A peptidomimetic inhibitors are mainly spanning from P3 to P1 positions, with different moieties or chemical groups at P4, P2 and P1 positions (Fig. A.1). There are three main classes of inhibitors: linear α-ketoamide covalent inhibitors (telaprevir, boceprevir and narlaprevir), linear non-covalent inhibitors (asunaprevir, sovaprevir, GS-9451 and Faldaprevir) and macrocyclic inhibitor (BILN-2061, Simeprevir, Paritaprevir (ABT-450), Danoprevir (ITMN-191), Vaniprevir (MK-7009) and Grazoprevir (MK-5172)) (Fig. A.1).

A.1.4 Drug resistance mutation selected by HCV NS3/4A protease's inhibitor usage

Because of the high error rate of HCV RNA polymerase (10^{-4} substitutions per base per year, 10 fold higher than HIV reverse transcriptase's) (Ogata et al., 1991, Mansky and Temin, 1995) and high virion production rate (10^{12} new virions per day, 100

fold higher than HIV), drug resistant strains may preexist and quickly be selected under drug pressure (Perelson et al., 1996, Neumann et al., 1998).

Mutations arise in the NS3/4A protease depending on the therapeutic regime; A156 mutations were observed in patients treated with linear ketoamide inhibitors (Kieffer et al., 2007, Sarrazin et al., 2007a, Susser et al., 2009, Halfon and Sarrazin, 2012, Svarovskaia et al., 2012, Vermehren and Sarrazin, 2012, Welsch and Zeuzem, 2012) and macrocyclic PIs mainly select for D168A and R155K variants (Manns et al., 2011a, Manns et al., 2011b, Lim et al., 2012). Mutations at V36, T54, and V36/R155 were initially reported to be associated with resistance to ketoamide inhibitors (Kieffer et al., 2007, Sarrazin et al., 2007a, Susser et al., 2009). In particular, V36M/R155K mutations were observed in patients receiving boceprevir (Howe et al., 2015), and these double mutations were also selected in patients received mericitabine/danoprevir combination therapy (Gane, 2012).

Even though the HCV protease inhibitors are in nanomolar or subnanomolar level, the existence of protease mutations still make inhibitors weak binders while the substrates are hydrolyzed, skewing the balance between inhibitor binding and substrate processing in favor of the latter and cause drug resistance. While resistance via active site mutations in the viral NS3/4A protease has been well characterized, e.g., mutations R155K and D168A disrupt favorable cation-π stacking interactions with R155 and confer danoprevir and vaniprevir resistance (Romano et al., 2012), the mechanism of resistance caused by non-active site mutations is unclear. In particular, protease mutations R155K and V36M often co-evolve, and while R155K alters the electrostatic network at the

binding site, V36M is >13 Å away. In the absence of crystal structures, molecular modeling led to the hypothesis that the distal mutations at V36 and T54 can impair interaction with telaprevir's cyclopropyl group (Welsch et al., 2008). However, no experimental data exists on the structural changes due to V36M and how the effects of this distal mutation may affect PI binding at the active site.

In this study the mechanism by which V36M confers resistance, in the context of R155K, is elucidated with drug susceptibility assays, crystal structures, and molecular dynamics (MD) simulations for three protease inhibitors: telaprevir, boceprevir and danoprevir.

Table A.1. Genotype 1a HCV NS3/4A protease substrate 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
TRIF	P	S	S	T	P	C	S	A	H	L
MAVS	E	R	E	V	P	C	H	R	P	S

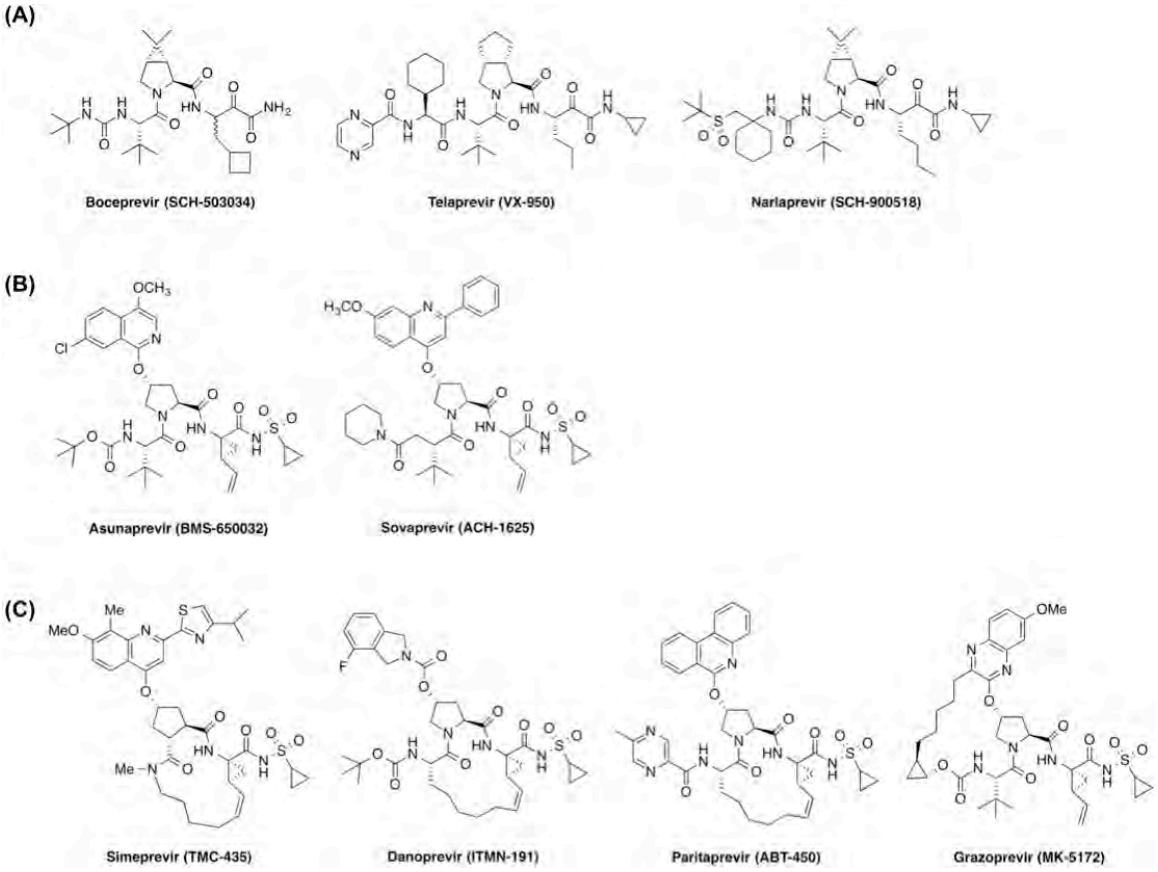


Figure A.1. Schematic representation of HCV NS3/4A protease inhibitors. (A) Linear α-ketoamide covalent inhibitors. (B) linear non-covalent inhibitors in pink. (C) Macrocyclic inhibitors.

A.2 Methods

A.2.1 Protein Expression and purification

NS3/4A protease expression and purification were carried out as described previously (Gallinari et al., 1998, Wittekind et al., 2002). Transformed Escherichia coli BL21(DE3) cells were grown at 37°C until OD600 reached 0.6 and induced by adding 1 mM IPTG. Cells were harvested after overnight expression at 4°C and pelleted. Cell pellets were resuspended in 5 mL/g of resuspension buffer (50 mM phosphate buffer at pH 7.5, 500 mM NaCl, 10% glycerol, 2 mM β-mercaptoethanol [β-ME]), and lysed with a cell disruptor. The soluble fraction was applied to a nickel column (Qiagen), washed with resuspension buffer supplemented with 20 mM imidazole, and eluted with resuspension buffer supplemented with 200 mM imidazole. The eluant was dialyzed overnight (molecular mass cutoff, 10 kDa) against resuspension buffer to remove the imidazole, thrombin treatment was applied simultaneously to remove the His tag. The nickel-purified protein was then flash-frozen with liquid nitrogen and stored at -80°C for future use.

A.2.2 Crystallization

Danoprevir was prepared in-house using our convergent reaction sequence as described previously (Romano et al., 2010); boceprevir was provided by Merck & Co., Inc; telaprevir was purchased from A ChemTek, Inc. (Worcester, MA). For crystallization, the protein solution was thawed and loaded on a HiLoad Superdex75 16/60 column equilibrated with gel filtration buffer (25 mM morpholineethanesulfonic

acid [MES] at pH 6.5, 500mMNaCl, 10% glycerol, 30 μM zinc chloride, and 2 mM dithiothreitol [DTT]). The protease fractions were pooled and concentrated to 25 mg/mL using an Amicon Ultra-15 10-kDa device (Millipore). The concentrated samples were incubated 1 h with 2 to 20 molar excess of protease inhibitors. Concentrated protein solutions were then mixed with precipitant solution (20 to 26% polyethylene glycol [PEG] 3350, 0.1 M sodium MES buffer at pH 6.5, and 4% ammonium sulfate) at a 1:1 ratio in 24-well VDX hanging-drop trays and diffraction-quality crystals were obtained overnight.

A.2.3 Data collection and structure solution

Crystals large enough for data collection were flash-frozen in liquid nitrogen for storage. Constant cryostream was applied when mounting crystal, and X-ray diffraction data were collected at Advanced Photon Source LS-CAT 21-ID-F and our in-house Rigaku_Saturn 944 X-ray system, respectively. The product complexes diffraction intensities were indexed, integrated, and scaled using the program HKL2000 (Otwinowski and Minor, 1997). All structure solutions were generated using simple isomorphous molecular replacement with PHASER (McCoy et al., 2007). The model of viral substrate N-terminal product 5A-5B (3M5O) (Romano et al., 2010) 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, ARP/wARP or Phenix (Adams et al., 2010) were applied to improve the phases by building solvent molecules

(Morris et al., 2002b). Crystallographic refinement was carried out within the CCP4 program suite or PHENIX with iterative rounds of TLS and restrained refinement until convergence was achieved (Collaborative-Computational-Project, 1994). The final structures were evaluated with MolProbity (Davis et al., 2007) prior to deposition in the Protein Data Bank. Five percent of the data was reserved for the free R-value calculation to prevent model bias throughout the refinement process (Brunger, 1992). Manual model building and electron density viewing were carried out using the program COOT (Emsley and Cowtan, 2004).

Drug susceptibility assay, enzyme inhibition assay, distance-difference maps, hydrogen bonds, salt bridges, van der Waals interactions, molecular dynamics simulations and molecular modeling were performed by other members working on this project, the methods for these sections will not be discussed here.

A.3 Results

To understand the molecular basis of the selection of V36M distal mutation under the pressure of PI including regimens, wild-type (WT) and resistant protease variants carrying R155K and R155K/V36M mutations were compared for binding telaprevir, boceprevir, and danoprevir. WT apo protease was published before (). Apo R155K and R155K/V36M variants' homology models were generated using crystal structures of the danoprevir-bound R155K and R155K/V36M variants. Together, these twelve (10 crystal and 2 models) resulting structures and their subsequent extensive MD simulations (100ns in triplicate) were analyzed to investigate the effects of V36M.

A.3.1 V36M Further Decreases Susceptibility of R155K variants to PIs

The activity of the three PIs against R155 and R155/V36M variants was assessed by enzymatic (K_i) and the cellular half-maximal inhibition constants (IC50) (Table A.3). While the double mutant has varying degrees of susceptibility to the three PIs, in all cases, V36M enhances resistance against both linear and macrocyclic compounds compared to R155K alone. Replicon-based cellular inhibition results correlate well with the enzyme inhibition constants. For all three PIs, the loss of antiviral activity (fold-change in IC50 relative to wild-type) against the R155K/V36M clones is substantial compared to the HCV clones carrying the R155K single mutation. In conclusion, the non-active site mutation V36M reduces the activity of all three PIs both on molecular and cellular levels in the presence of the active site mutation R155K.

A.3.2 Resistance Mutations Cause Changes in the Active Site Electrostatic Network

The PIs make hydrogen bonds mainly with the backbone donors /acceptors in the binding site. As a result, the impact of R155K on the hydrogen-bonding network is minimal (Table A.4).

The electrostatic network at the binding site of NS3/4A, involving residues D81–R155–D168–R123, is critical for substrate recognition and inhibitor binding (Ali et al., 2013, Romano et al., 2011). Disruption of this network can arise as a mechanism of resistance by weakening inhibitor binding. Therefore, sensitivity of the salt bridges this network to resistance mutations was assessed (Table A.5). R155 can make two salt bridges, one with D81 and the other with D168 in all PI-bound structures. While R155K mutation favors D81–K155 over K155–D168 in telaprevir and danoprevir complexes, this mutation favors K155–D168 over D81–K155 in boceprevir complexes (Table A.5). R155K mutation disrupted the electrostatic network in all three PIs' binding mode, since this network was shown critical for inhibitor binding, the fluctuations in the salt bridges contribute to the decreased Susceptibility of R155K variants to PIs. The effects of V36M were not observed in this electrostatic network.

A.3.3 The Distal V36M Mutation Changes the Active Site via F43

To better capture the dynamic changes at the binding site due to resistance mutations, inter-residue Ca-Ca distance during MD simulations was tracked for pairs of protease residues. V36M mutation changed the dynamics and distance sampled for some key residue pairs across and away from the active site (Fig. A.3, Fig. A.4). In the

presence of V36M, residue 36 and F43 become closer to each other compared to WT and R155K complexes (Figure A.3A). F43 interacts with the bound inhibitor and is located in B1 strand, which bridges the distal V36M mutation on the A1 strand to the binding site (Fig. A.2). We found that four residues, K136 in α2 helix and residues 155–157 on E2 β-strand, also become closer to F43 with V36M mutation (Fig. A.3, Fig. A.4). The fact that E2 and B1 β-strands and the α2 helix are closer to each other suggests a slight shrinking in the binding site across the B1–E2 strands direction.

Structural and dynamic reorganization in the binding site involving the α2 helix and B1 and E2 strands also impacts the catalytic triad (Fig. A.3B). In telaprevir and danoprevir complexes, R155K mutation causes H57 and D81 Cα atoms to get farther away from each other, which is further aggravated by V36M in danoprevir-bound protease. When farther apart, the sidechains of these two catalytic residues are not oriented properly for strong salt bridging (Table A.5). In contrast, the distances and the H57-D81 salt bridge are robust in complexes of boceprevir, which is the least susceptible of the three inhibitors to R155K/V36M resistance mutations (Table A.3).

In summary, detailed investigation of co-crystal structures and conformational ensembles coupled with drug susceptibility assays show that the distal site mutation V36M, in the presence of the binding site mutation R155K, alters the binding site shape through changes in conformational dynamics.

Table A.2. Crystallographic statistics for HCV protease with Boceprevir co-crystal structures.

Drug	**Boceprevir**		
Protease	**WT**	**R155K**	**R155K/V36M**
Resolution	1.7	1.8	1.7
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
a (Å)	55.1	55.4	55.4
b (Å)	58.8	59.0	58.9
c (Å)	60.1	59.8	60.0
Molecules in	1	1	1
R_{merge} (%)[b]	4.5	3.2	3.4
Completeness	95.8	90.8	90.8
I/ σ	22.8	15.4	11.6
Measured	96791	48641	41927
Unique	21118	15757	19992
Redundancy	4.1	3.1	2.1
RMSD[c] in:			
Bonds (Å)	0.012	0.006	0.007
Angles (°)	1.492	1.117	1.128
R_{free} (%)	19.3	18.8	20.0
R_{factor} (%)	15.6	15.2	16.3
No. of waters	247	186	212
PDB Code	5EBQ	5EBR	5EBS

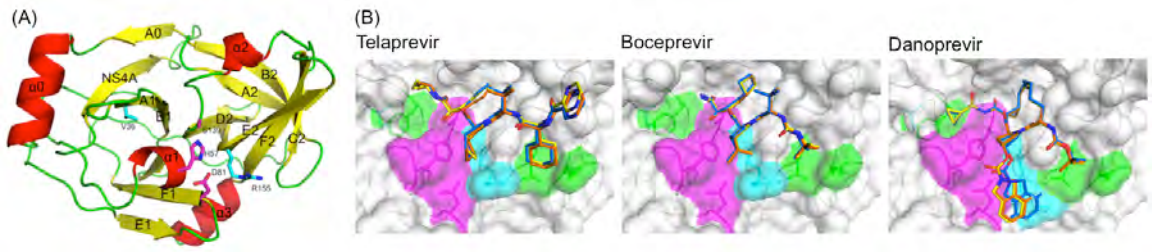


Figure A.2. Topology of hepatitis C viral serine protease, NS3/4A. Protease is colored by the secondary structures; α-helices, strands, and loops colored in red, yellow, and green, respectively. Side chains of the catalytic triad are in magenta and the mutation sites, R155K and V36 are in cyan (A). Crystallographic binding modes of telaprevir, boceprevir, and danoprevir in the wild-type, R155K and R155K/V36M protease complexes. Protease and inhibitors are represented as surface and sticks, respectively. Side chains of key residues are also shown as sticks; the drug resistance mutation sites R155K and V36M (cyan), the catalytic triad H57-D81-S139 (magenta), and other binding site residues (green). Inhibitors in complex with wild-type, R155K and R155k/V36M proteases are shown in blue, orange and yellow respectively.

Table A.3. Drug susceptibilities against wild-type and resistant HCV clones and inhibitory activities against NS3/4A proteases.

Full-length NS3/4A Enzyme - K_i (nM)[a]			
	WT	***R155K***	***R155K/V36M***
Telaprevir	40.9 ± 3.7	824.0 ± 75.1 (20)	>10,000 (>244)
Boceprevir	34.7 ± 2.9	390.8 ± 43.0 (11)	1018.0 ± 192.3 (29)
Danoprevir	1.2 ± 0.1	132.0 ± 18.0 (111)	292.9 ± 38.6 (246)
Protease domain K_i (nM)[a]			
	WT	***R155K***	***R155K/V36M***
Telaprevir	33.3 ± 4.0	803.7 ± 89.9 (24)	7342 ± 1281 (220)
Boceprevir	35.4 ± 3.3	236.7 ± 44.3 (7)	1097.0 ± 120.4 (31)
Danoprevir	1.0 ± 0.1	157.9 ± 20.5 (158)	295.5 ± 34.3 (295)
Replicon IC_{50} (nM)[a]			
	WT	***R155K***	***R155K/V36M***
Telaprevir	1349	4740 (3.5)	15759 (12)
Boceprevir	971	2788 (3)	3941 (4)
Danoprevir	0.24	>100 (>416)	>100 (>416)

[a]Numbers in parentheses reflect fold-change relative to wild-type; > indicates IC_{50} and K_i values higher than the maximum drug concentration tested in the assay.

Table A.4. Intermolecular hydrogen bonds between the inhibitor and protease active site residues.

Crystal structures - Distance between donor and acceptor (Å)

		Telaprevir				Boceprevir				Danoprevir		
		WT	R155K	R155K V36M		WT	R155K	R155K V36M		WT	R155K	R155K V36M
H57-NE2	OBR	2.6	2.5	2.5	O4	2.6	2.7	2.8	N3	3.1	3.1	3.1
G137-N	OBS	2.7	2.7	2.7	O5	2.8	2.8	2.8	O9	2.9	2.9	2.9
G137-N	-	-	-	-	-	-	-	-	O6	3.1	3.1	3.0
S139-N	OBS	3.0	3.0	3.0	O5	3.0	3.1	2.9	-	-	-	-
R155-O	NAE	3.0	3.1	3.1	N4	3.0	3.1	3.2	N2	3.0	3.0	2.9
A157-O	NAC	2.9	2.9	2.8	N2	2.9	2.9	2.9	N1	2.9	2.9	2.9
A157-O	-	-	-	-	N3	3.0	3.0	2.9	-	-	-	-
A157-N	OBT	3.0	2.9	3.0	O1	3.0	3.0	3.0	O3	2.9	2.9	2.9
S159-N	OBW	3.0	2.8	2.7	-	-	-	-	-	-	-	-
S159-OG	OBW	2.9	2.9	2.8	-	-	-	-	-	-	-	-

MD simulations - %Time

		Telaprevir				Boceprevir				Danoprevir		
		WT	R155K	R155K V36M		WT	R155K	R155K V36M		WT	R155K	R155K V36M
H57-NE2	OBR	19	2	1	O4	59	60	34	N3	2	1	6
G137-N	OBS	74	69	82	O5	61	43	35	O9	16	26	21
G137-N	-	-	-	-	-	-	-	-	O6	47	-	18
S139-N	OBS	62	62	46	O5	44	38	45	O4	41	-	7
R155-O	NAE	84	81	68	N4	93	94	86	N2	83	91	83
A157-O	NAC	90	90	91	N2	48	46	47	N1	74	50	53
A157-O	-	-	-	-	N3	49	47	50	-	-	-	-
A157-N	OBT	87	85	86	O1	71	74	65	O3	88	78	81
S159-N	OBW	70	77	77	-	-	-	-	-	-	-	-
S159-OG	OBW	12	12	12	-	-	-	-	N3	57.4		12.2

Table A.5. Intramolecular salt bridges forming a network at the NS3/4A active site surface in crystal structures, and stabilities assessed by MD simulations.

		APO			Telaprevir			Boceprevir			Danoprevir		
		WT	R155K	R155K V36M	WT	R155K	R155K V36M	WT	R155K	R155K V36M	WT	R155K	R155K V36M
Crystal Structures (Distance between the side chain oxygen and nitrogen in Å)													
H57	D81	3.9			3.9	3.8	3.8	3.9	3.9	3.9	3.9	3.9	3.9
D81	R155	6.4			4.5	-	-	4.5	3.7	3.5	-	-	-
R155	D168	4.8			4.8	4.8	5.2	4.8	-	-	3.7	-	-
D168	R123	3.6			4.8	4.8	4.3	4.6	3.5	3.6	5.1	3.5	3.5
D25	R11	4.5			4.5	4.4	4.4	4.5	4.5	4.4	4.4	4.4	4.4
E173	E118	5.7			5.7	5.6	5.5	5.5	5.6	5.5	5.4	5.4	5.4
E32	R92[a]	3.5			-	-	-	4.1	-	3.3	-	-	-
E30	R997	4.5			4.7	4.6	4.6	4.8	4.7	4.5	4.9	4.7	4.8
D103	R117	4.5			4.6	4.5	4.5	4.6	4.6	4.6	4.6	4.6	4.7
D121[b]	R118[b]	-			-	-	-	-	-	-	6.7	-	-
D112[b]	R130[b]	-			-	-	-	-	-	-	-	6.8	6.5
MD Simulations (% of the time the salt bridge criterion was satistifed													
H57	D81	100	99.4	99.9	93.9	29.1	34.6	100	100	100	100	22.7	30.5
D81	R155	10.3	12.2	16.4	95.8	89.3	83.3	85.5	2.6	3.9	0	83.9	78.4
R155	D168	98.3	5.6	3.1	86.2	23.8	23.9	79.6	65.0	67.3	100	22.2	21.3
D168	R123	21.7	96.2	99.7	88.5	96.0	95.1	88.7	91.7	84.9	22.3	69.5	90.8

[a]In only three structures, unambiguous electron density was observed for R92.

[b]All structures had relatively high B-factors for these residues.

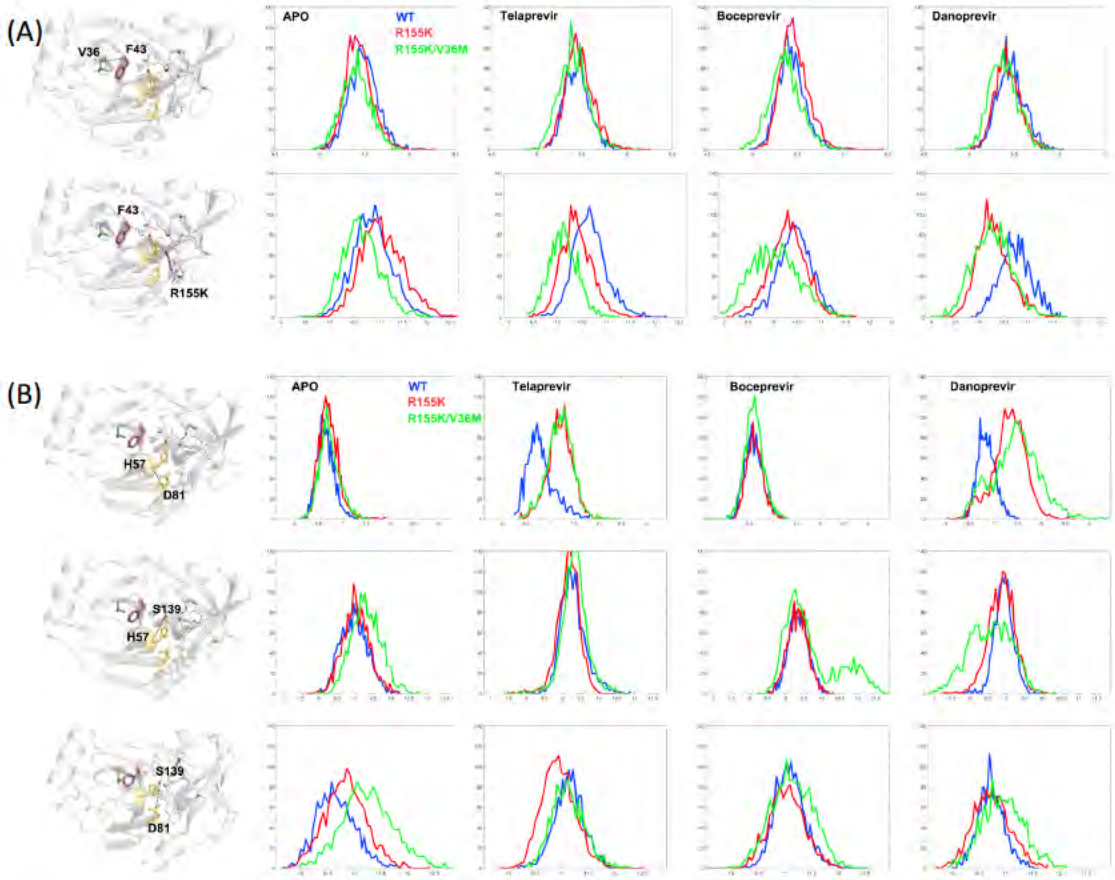


Figure A.3. The dynamic distance distribution between protease residue pairs sampled during MD simulations. The distance between (A) F43 and the sites of mutation (V/M36–F43, F43–R/K155), and (B) the catalytic triad residues (H57–D81, H57–S139, D81–S139). The distance (in Å) between the Ca atoms of the two residues is plotted against the percent time that distance was observed.

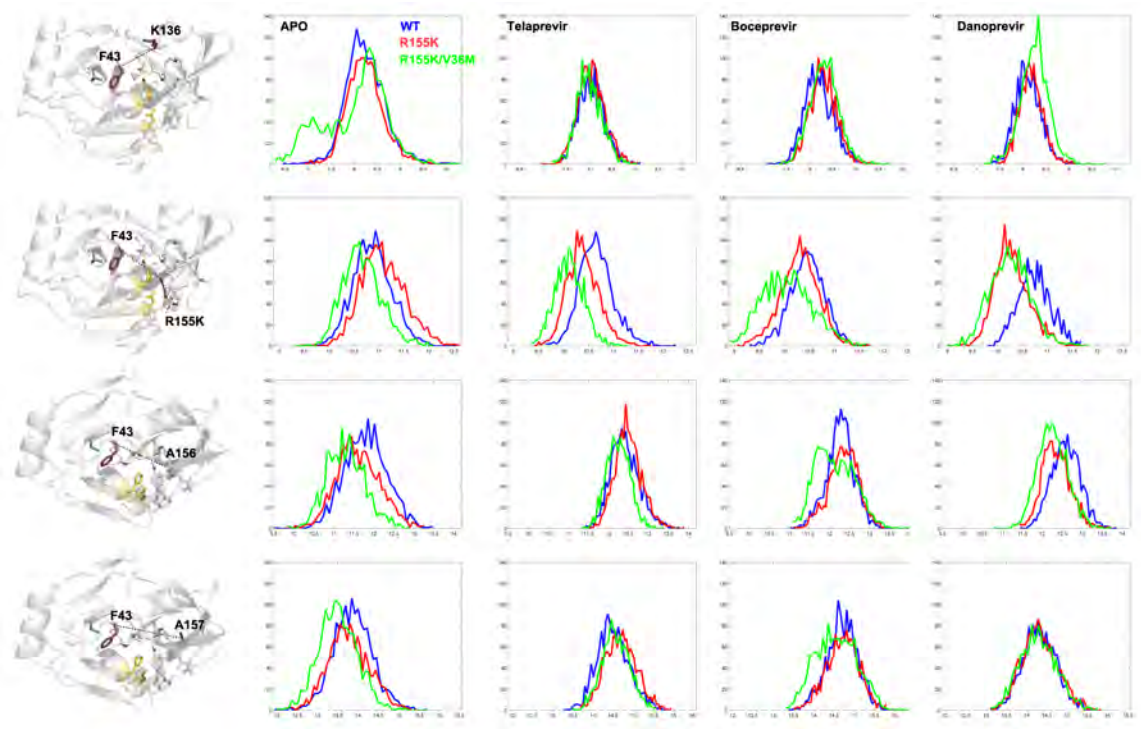


Figure A.4. The distribution of Ca–Ca distance during MD simulations for F43 with K136, R/K155, A156 and A157.

Appendix B: Engineering DENV2 NS2B/3 protease for crystallization

B.1 Preface

Due to the lack of information about how dengue NS3/2B protease can recognize diverse substrate sequences, I aimed to solve the complex structures of DENV2 protease with either substrates or substrate products to address this problem. However, the S1 pocket of the DENV2 protease active site is blocked in the apo structure because of crystal packing, making this crystallization condition not optimal for co-crystallization purpose (PDB: 2fom). As shown in Figure B.1, leucine 31 from the NS3 domain of the bottom monomer blocks the S1 pocket of the top monomer and vice versa. To address this problem, I did protein engineering to disrupt the interactions formed by this leucine residue.

The co-factor NS2B was observed to adapt an opened conformation in the apo-DENV2 protease structure (PDB: 2fom), and the electron density beyond residue 76 of NS2B domain was not observed, which may reflect multiple conformations in this region (Erbel et al., 2006). In contrast, the NS2B domain of DENV3 protease (PDB: 3u1i) (Noble et al., 2012) adapts a closed active conformation, and the c terminus of NS2B was observed to contribute to the formation of protease binding pocket (Fig. 1.3.2). Therefore, I introduced disulfide bond between DENV2 protease's NS2B and NS3 domains to induce the folding of NS2B, which may decrease the flexibility of NS2B, facilitate the formation of protease active conformation, and increase the possibility of getting a complex crystal.

Certain regions on DENV2 protease surface are highly flexible, e.g., ^{90}KEGE^{93} and ^{142}KKGK^{145}, which may decrease the possibility for this protease to crystallize. These two highly charged regions are entropically unfavorable once participating in packing interactions. Previously, surface entropy reduction has been shown successfully applied to the crystallization of HCV protease (Hagel et al., 2011). Therefore, to facilitate crystallization of DENV2 protease, I proposed to apply this method and build more favorable packing surface on the protease.

B.2 Methods and Results

To make the DENV2 protease active site available for ligand binding, I mutated leucine 31 to phenylalanine or tryptophan to disrupt the interactions formed by this residue upon crystal packing (Fig. B.1). Since DENV3 protease has phenylalanine at this position and the S1 pocket of DENV3 protease is available for inhibitor binding, this mutation was expected to break the packing interactions formed by residue 31. Mutation L31W was also expected to break the packing because the S1 pocket is not big enough to accommodate large tryptophan side chain. After the mutagenesis, I could not grow crystals of these mutant constructs' using same crystallization condition, which may reflect the success of breaking packing surface formed by L31. Therefore, I performed crystallization screening of DENV2 L31F or L31W protease with substrates or substrate products. However, no hits were obtained from the crystallization trials.

To facilitate the folding of NS2B cofactor of DENV2 protease, I introduced disulfide bond between NS2B and NS3 domains: S70C−S127C or S71C−G114C. These residue pairs were picked using Disulfide by Design (Craig and Dombkowski, 2013) based on distances between Cα carbons (Fig. B.2). Next, I combined these two sets of mutations with the L31 mutations and then performed crystallization screening. However, no hit was found.

To facilitate the crystallization of DENV2 protease, I applied surface entropy reduction method and mutated two highly charged and flexible regions on protease surface: ^{90}KEGE^{93} and ^{142}KKGK^{145}. These two regions were mutated to residues AAGA, which were expected to significantly decrease the entropy penalty upon packing (Fig.

B.3). These mutations were then combined with L31 mutations and disulfide mutations (Table B.1), and crystallization screening were applied. However, no hit was obtained. Further construct engineering and crystallization trials are required to get DENV2 protease-ligand complex structures.

Along with DENV2 NS3/2B protease constructs, I also performed crystallizations of DENV3 WT NS3/2B protease, unlinked DENV2 NS2B NS3 protease, and NS2B/FL NS3 protein (DENV2, DENV3). All these constructs were co-crystallized with either substrates or substrate products in the crystallization trials. To identify new crystallization condition, co-crystallizing these protease constructs with the cyclic peptides (Chapter IV) or P site cyclic inhibitors (Appendix D) should be pursued.

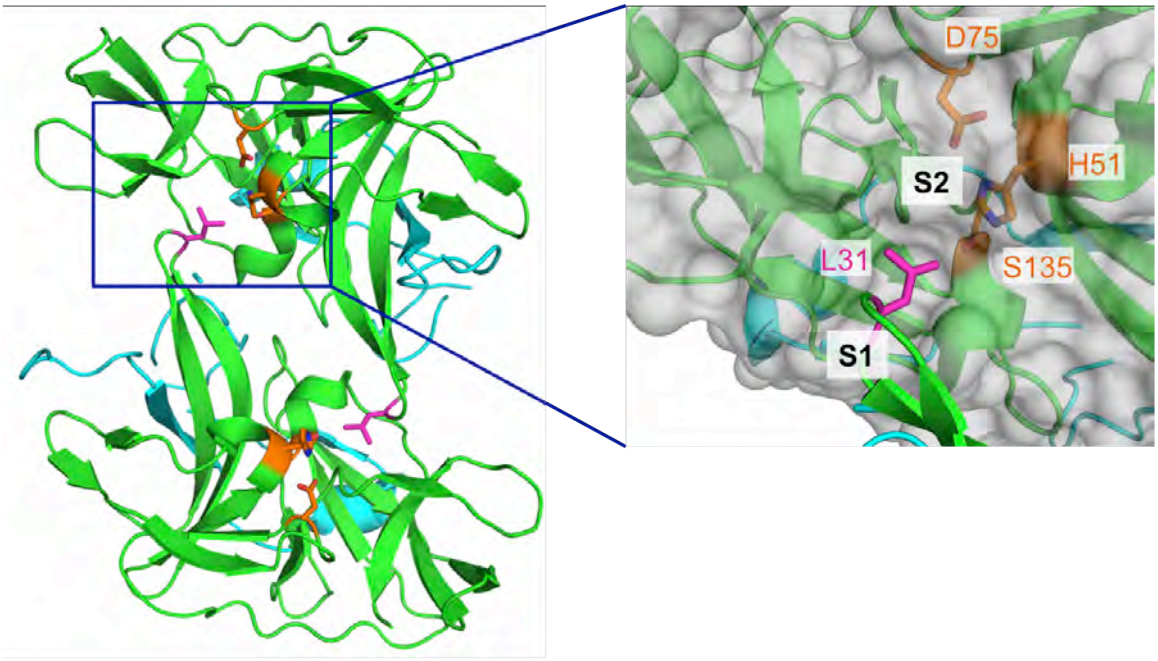


Figure B.1. Selected crystal packing surface of DENV2 protease apo-structure. NS3 is shown in green, NS2B in cyan, catalytic triads in orange and leucine 31 in magenta.

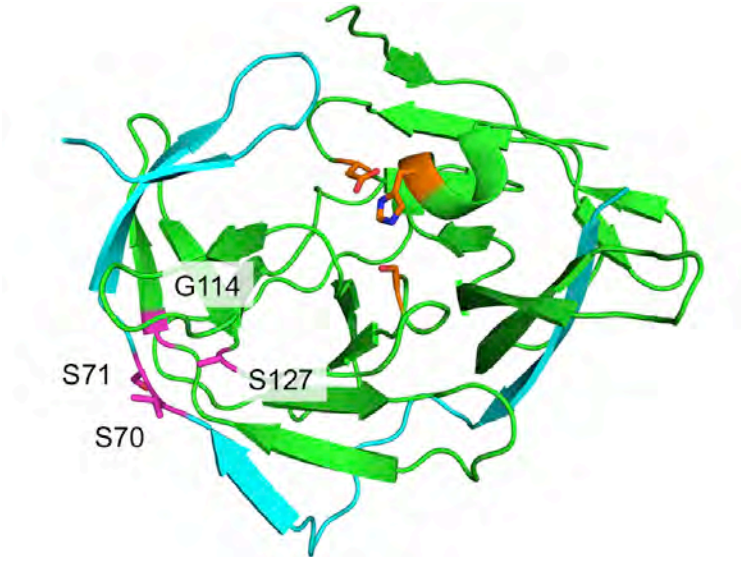


Figure B.2. Building disulfide bonds between NS3 and NS2B domains of DENV2 protease. DENV3 protease structure (3u1i) is shown as an example of where the disulfide bonds are, and the ligand in this structure is removed for visualization purpose. Residues are labeled based on DENV2 protease sequence. NS3 is shown in green, NS2B in cyan, catalytic triads in orange and mutated residues in magenta.

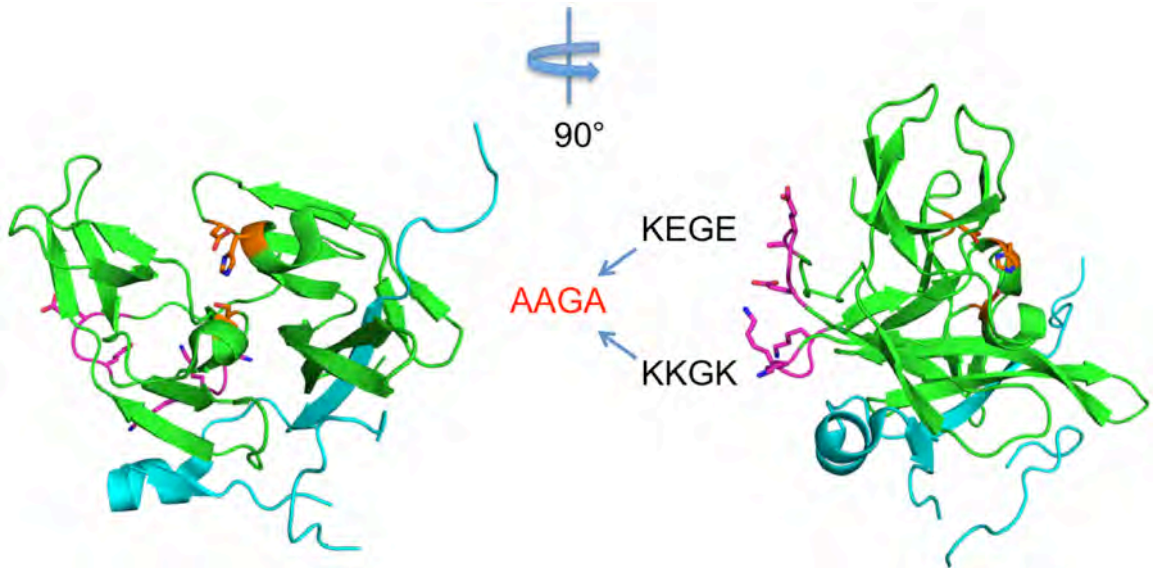


Figure B.3. Applying surface entropy reduction to DENV2 protease for crystallization purpose. Two flexible and charged loop regions are highlighted as sticks. NS3 is shown in green, NS2B in cyan, catalytic triads in orange and mutated residues in magenta.

Table B.1. The expression, purification and crystallization trials of dengue NS3/2B protease constructs or NS2B/full length NS3 constructs.

Protease construct	Expression	Purification	Crystallization
DENV2 WT	✓	✓	✓
DENV2 L31F	✓	✓	✓
DENV2 L31W	✓	✓	✓
DENV2 S135A	✓	✓	✓
DENV2 ^{90}AAGA^{93}	✓	✓	✓
DENV2 70-127 disulfide bond	✓	✓	✓
DENV2 71-114 disulfide bond	✓	✓	✓
DENV2 S135A / ^{90}AAGA^{93}	✓	✓	✓
DENV2 S135A / 70-127 disulfide bond	✓	✓	✓
DENV2 S135A / 71-114 disulfide bond	✓	✓	✓
DENV2 L31F / K90A	✓	✓	✓
DENV2 L31F / ^{90}AAGA^{93}	✓	✓	✓
DENV2 L31F / K143A			
DENV2 L31F / 70-127 disulfide bond	✓	✓	✓
DENV2 L31F / 71-114 disulfide bond	✓	✓	✓
DENV2 L31F / S135A	✓	✓	✓
DENV2 L31F / S135A / K90A	✓	✓	✓
DENV2 L31F / S135A / ^{90}AAGA^{93}	✓	✓	✓
DENV2 L31F / S135A / K143A			
DENV2 L31F / S135A / 70-127 disulfide bond	✓	✓	✓
DENV3 WT	✓	✓	✓
DENV2 unlinked NS2B NS3	✓	✓	✓
DENV2 unlinked NS2B NS3 (K90A)	✓	✓	✓
DENV2 unlinked NS2B NS3 (^{90}AAGA^{93})	✓	✓	✓
DENV2 WT NS2B + FL NS3	✓		✓
DENV3 WT NS2B + FL NS3	✓		✓

Appendix C: Virtual fragment screening against DENV3 NS3/2B protease

C.1 Preface

Fragment based drug design has become a powerful method to target individual protein active site pocket, and several compounds have been successfully built by either fragment linking or growing methods (Kumar et al., 2012).

Previously, fragment based drug design was applied to discover inhibitors targeting the S pocket of dengue protease active site (Knehans et al., 2011). However, as mentioned in chapter 1.3.5, these inhibitors may not be specific to dengue protease. To address this issue, I proposed to screen fragments against conserved residues at protease active site, which would help us design inhibitors more specific to DENV.

C.2 Methods

The fragment library (ChemBridge Corp.) was prepared using canvas (Duan et al., 2010) and ligand preparation (LigPrep, version 2.3, Schrödinger, LLC). Target protein DENV3 protease (PDB:3u1i) was prepared using protein preparation wizard (Sastry et al., 2013), and the grid was generated using receptor grid generation tool in maestro. These grids were centered at residues conserved among different serotypes of dengue proteases. Glide (Halgren et al., 2004, Friesner et al., 2004) was then applied for the docking of fragment libraries onto DENV3 protease.

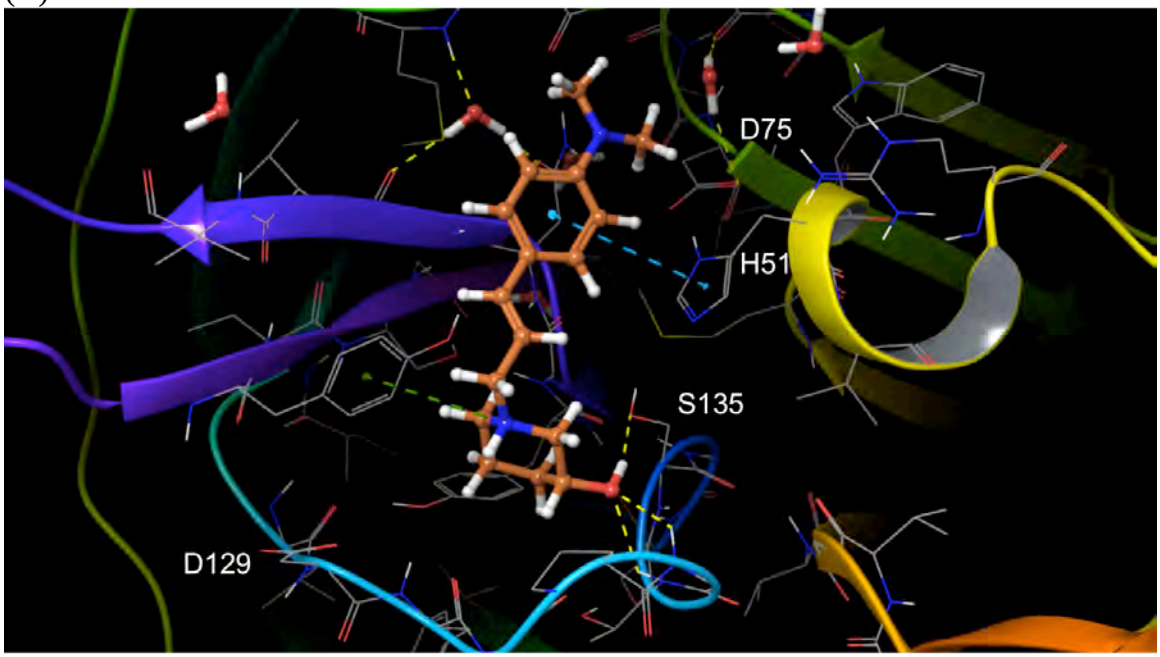
Top 10 percent of docking hits were further filtered by MM-GBSA free energy calculation (Prime, version 2.1, Schrödinger, LLC), and the top ranked fragments with favorable docking poses were chose for further experimental verification. Enzyme inhibition assay was applied to measure the inhibition constants of these fragments against DENV3 protease. The method for this assay is the same as described in chapter 3.5.3..

Fragment hits verified by enzyme inhibition assay would then be developed using fragment linking or fragment growing methods, which will be performed by Jacqueto Zephyr in our lab.

C.3 Summary of findings

I obtained fragment hits with mircomolar to low millimolar K_i values against DENV3 NS3/2B protease (Table C.1). These fragment hits were categorized based on their targeting protease residues, e.g., R54, D75, and S135 (Table C.1). The hit against R54 allows us to build compound outside the S pocket of dengue protease (Fig. C.1B). As mentioned in chapter 4.3.5, this arginine residue is conserved among different serotypes of dengue viruses. Further fragment growing based on the hit against this residue could help us obtain compound with higher specificity against dengue protease.

(A)



(B)

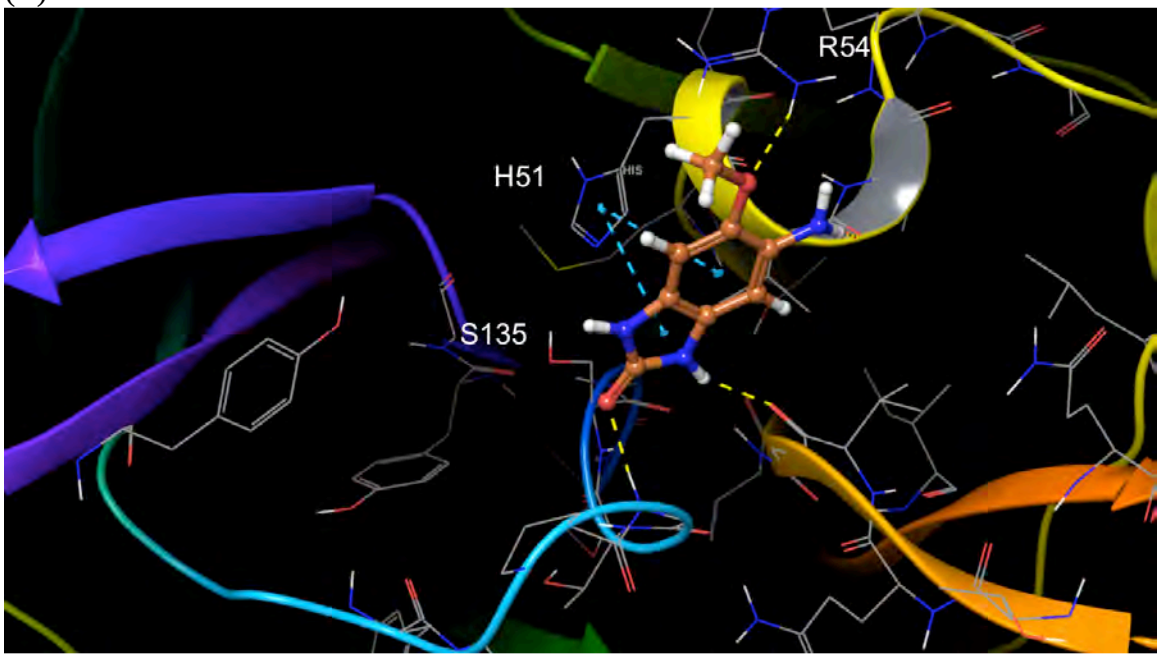
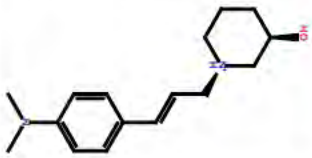
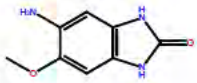
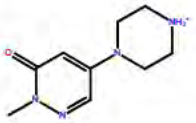
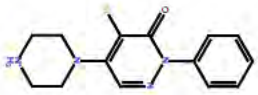


Figure C.1. Examples of docking result of hit obtained from fragment based screening with different target residues. (A) Fragment targeting S135. (B) Fragment targeting R54.

Table C.1. Selected results of fragment library screening against DENV3 WT protease.

Compound ID	Chemical structure	K_i (μM)	Target residue
5426254		588 ± 169	S135
4028462		1698 ± 316	R54
4030938		1304 ± 283	D75
7773406		1176 ± 221	S135

Appendix D: Design of P site cyclic peptide inhibitors against dengue NS3/2B protease

D.1 Preface

Dengue protease shares similar P site substrate sequence preferences with human serine protease, however, the dengue protease active site structures are different from these human proteases'. Dengue protease has a flat and solvent exposed S1 pocket, however, the shapes of these human proteases' S1 pockets are relatively narrow (Fig. D.1). Therefore, by cyclizing P site peptide (tetra-peptide with cyclization among N terminus capping and modified P1 side chain as an example) (Fig. D.2), I expect this P site macro-cyclic inhibitor to bind to dengue protease but too big to fit into human serine proteases' active sites.

D.2 Methods and Results

P site cyclic inhibitors were designed by connecting P1 and P4 side chains, or P1 side chain and N terminal capping group. The linkers include five member rings and six member rings (Fig. D.3, Fig. D.4). These designed compounds were modeled onto DENV3 protease active site and followed by energy minimization using protein preparation wizard in Maestro (Sastry et al., 2013). Compounds with favorable binding mode were chose, and chemists Dr. Linah Rusere and Jacqueto Zephyr are working on the synthesis.

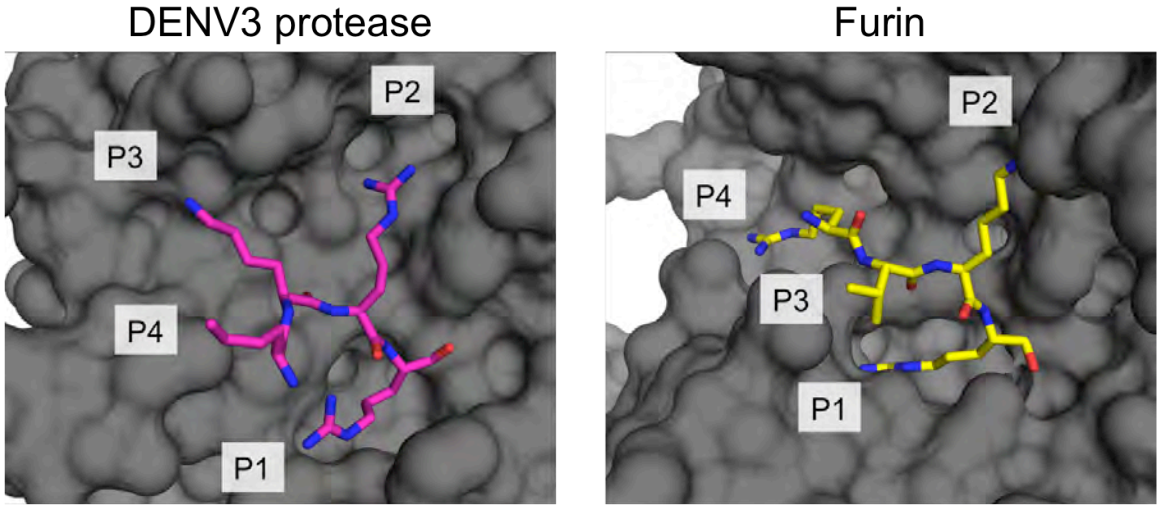


Figure D.1. Surface comparison of serine proteases' active sites. DENV3 protease's inhibitor is labeled based on corresponding substrate residue positions. Furin substrate residues are labeled. The n terminus beyond P4 position of both ligands are removed for visualization purpose.

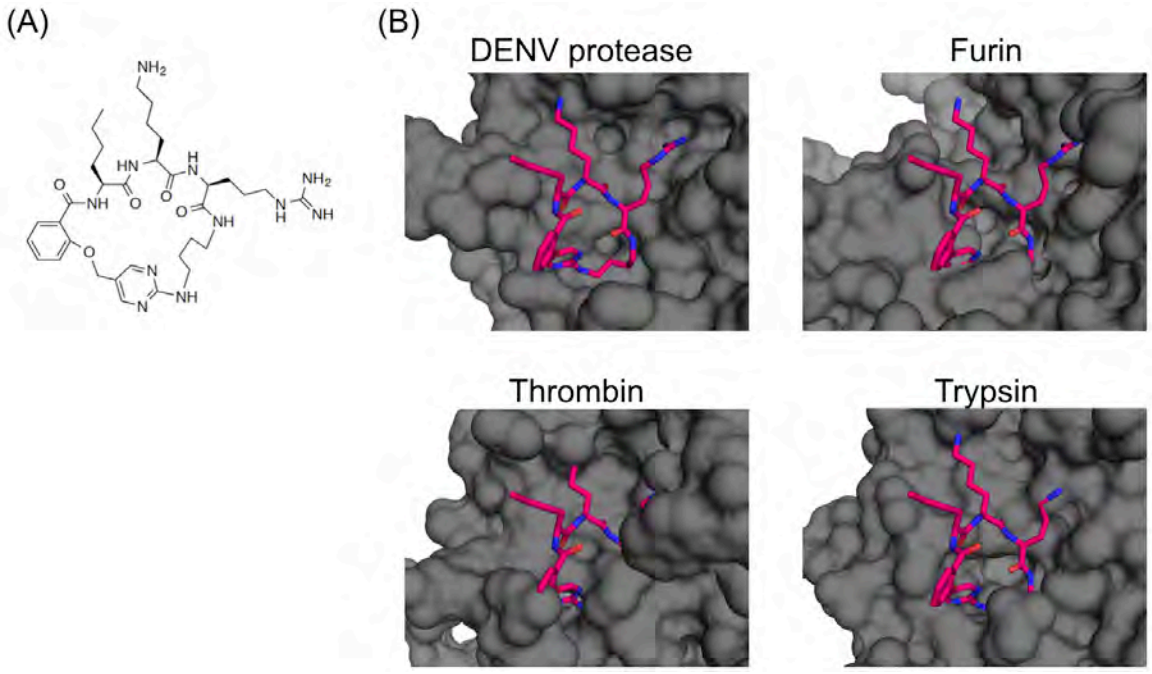


Figure D.2. Modeling of P site macrocyclic inhibitors onto dengue and human serine proteases. (A) Chemical structure of selected protease inhibitor. (B) Modeling of inhibitor onto DENV3 protease (3u1i), furin (1p8j), thrombin (2afq) and trypsin (1trn).

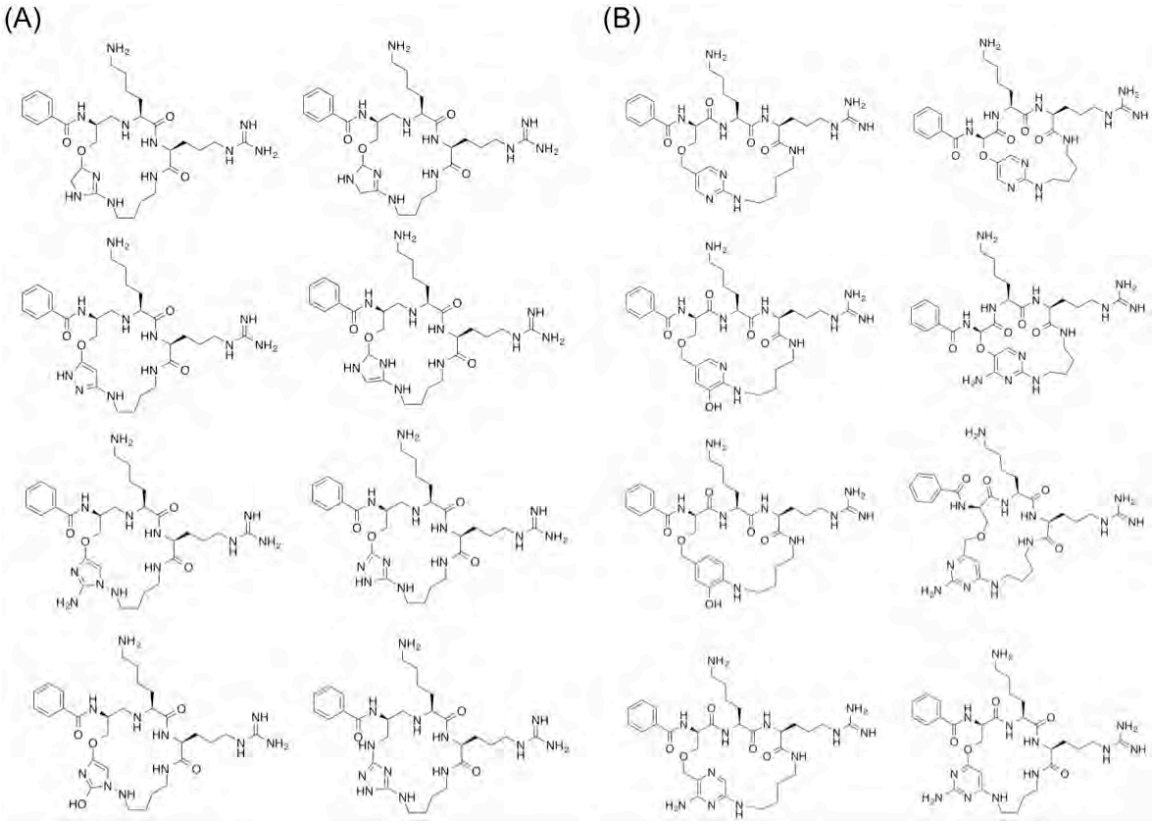


Figure D.3. Chemical structures of P site cyclic peptidomimetic inhibitors with P1−P4 macro-cyclization.

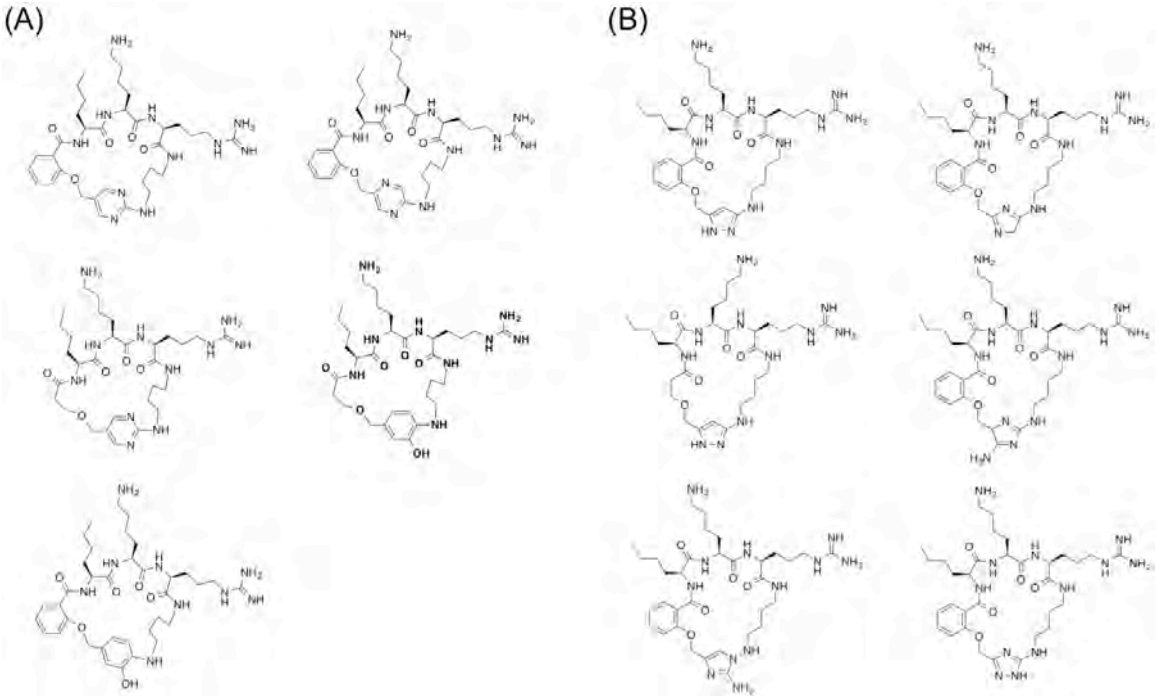


Figure D.4. Chemical structures of P site cyclic peptidomimetic inhibitors with P1−N terminus capping macro-cyclization.

Appendix E: Design of linear P site inhibitors against dengue NS3/2B protease

E.1 Preface

A potential reason I did not get any dengue NS3/2B protease apo- or ligand bond crystal was that I did not have a tighter binder to stabilize the flexible NS2B co-factor domain. Beside introducing a disulfide bond between NS2B and NS3 domains as described in appendix B, I designed a series of linear P site inhibitors to interact and stabilize the co-factor. Low micromolar P site inhibitors with c-terminus aldehyde capping group have been investigated, and the structural information was available (Yin et al., 2006a, Yin et al., 2006b, Noble et al., 2012). However, due to the cyclization and hydration issues, we could not get pure aldehyde compound (Fig. E.1). Beside aldehyde compounds, linear P site inhibitors without c-terminus capping group have also been shown to inhibit dengue NS3/2B protease with low- to sub-micromolar K_i values (Hammamy et al., 2013). Therefore, I proposed to use P site peptidomimetic inhibitors without c-terminus capping to stabilize the NS2B co-factor and help us identify favorable crystallization conditions for dengue NS3/2B protease.

E.2 Methods and results

Working with Dr. Akbar Ali in our lab, we designed a series of P site peptidomimetic inhibitors without c-terminus capping. Akbar taught me how to used solution-phase methods to synthesize these compounds.

These inhibitors were tested against dengue NS3/2B protease using assay described in chapter 3.5.3., and the K_i values are in micromolar range (Table E.1). The crystallization trials of these inhibitors against dengue NS3/2B protease were performed using commercial crystallization screening kits, however, no hit was obtained. Further compound optimization is required to identify tighter binders for crystallization purpose.

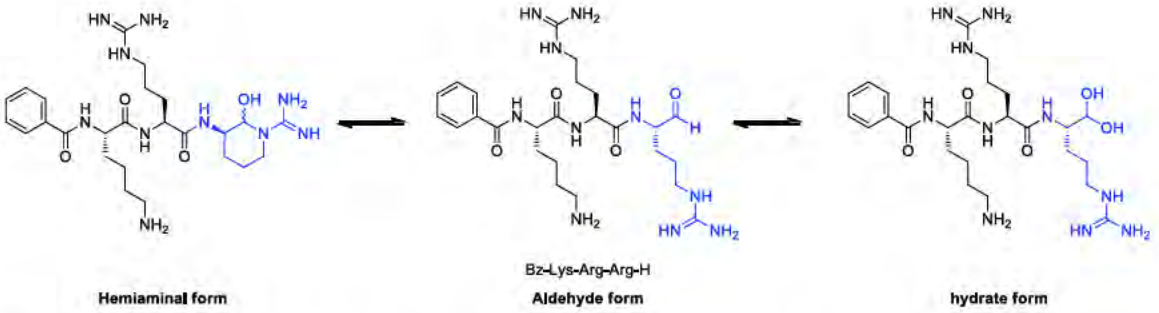


Figure E.1. Inactive derivatives of tri-peptide compounds.

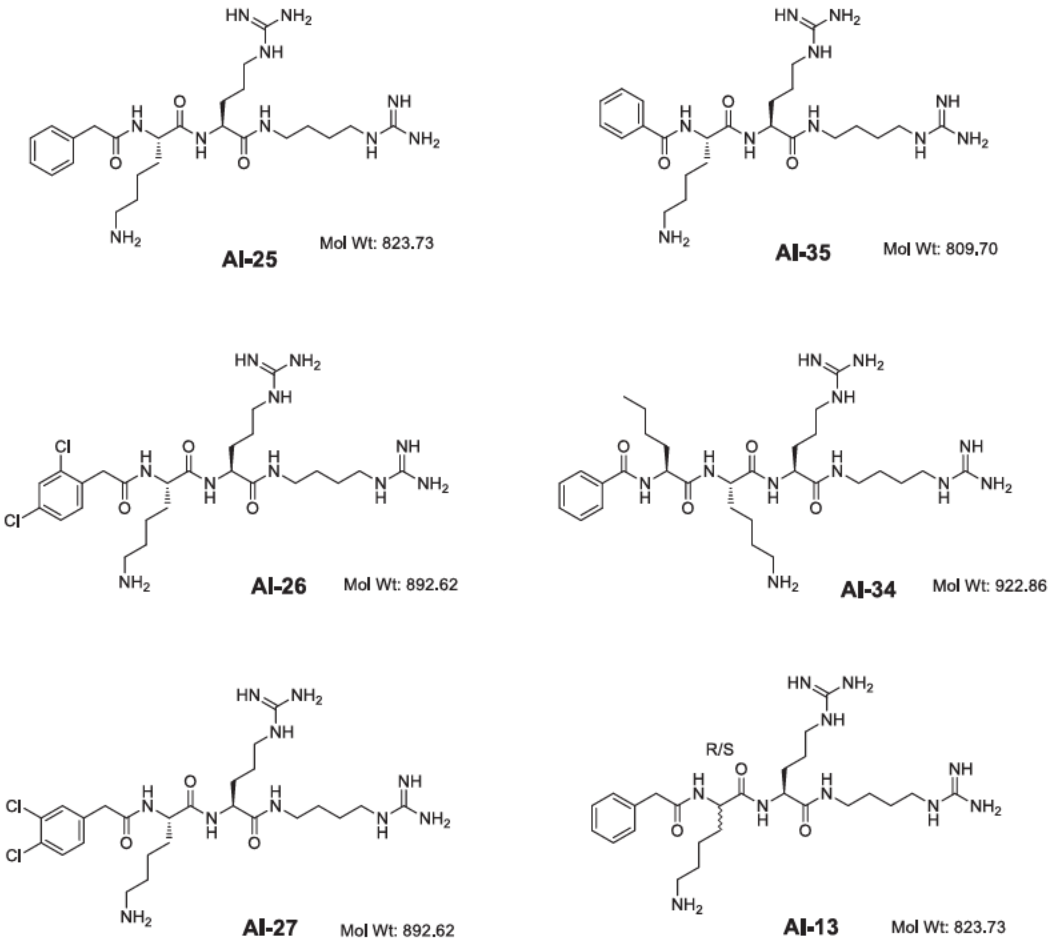


Figure E.2. Chemical structures of linear P site inhibitors without c-terminal capping group.

Table E.1. The inhibition constants of linear P site inhibitors against dengue NS3/2B protease.

Compound	K_i value (μM)
AI-13	237.5 ± 23.9
AI-25	166.5 ± 10.4
AI-26	109.7 ± 11.2
AI-27	142.3 ± 30.5
AI-34	137.1 ± 8.3
AI-35	336.6 ± 85.0

References

ADAMS, P. D., AFONINE, P. V., BUNKOCZI, G., CHEN, V. B., DAVIS, I. W., ECHOLS, N., HEADD, J. J., HUNG, L. W., KAPRAL, G. J., GROSSE-KUNSTLEVE, R. W., MCCOY, A. J., MORIARTY, N. W., OEFFNER, R., READ, R. J., RICHARDSON, D. C., RICHARDSON, J. S., TERWILLIGER, T. C. & ZWART, P. H. 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr,* 66**,** 213-21.

ALI, A., AYDIN, C., GILDEMEISTER, R., ROMANO, K. P., CAO, H., OZEN, A., SOUMANA, D., NEWTON, A., PETROPOULOS, C. J., HUANG, W. & SCHIFFER, C. A. 2013. Evaluating the Role of Macrocycles in the Susceptibility of Hepatitis C Virus NS3/4A Protease Inhibitors to Drug Resistance. *ACS Chem Biol*.

ALI, A., BANDARANAYAKE, R. M., CAI, Y., KING, N. M., KOLLI, M., MITTAL, S., MURZYCKI, J. F., NALAM, M. N., NALIVAIKA, E. A., OZEN, A., PRABU-JEYABALAN, M. M., THAYER, K. & SCHIFFER, C. A. 2010. Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses,* 2**,** 2509-35.

ARIAS, C. F., PREUGSCHAT, F. & STRAUSS, J. H. 1993. Dengue 2 virus NS2B and NS3 form a stable complex that can cleave NS3 within the helicase domain. *Virology,* 193**,** 888-99.

ASSENBERG, R., MASTRANGELO, E., WALTER, T. S., VERMA, A., MILANI, M., OWENS, R. J., STUART, D. I., GRIMES, J. M. & MANCINI, E. J. 2009. Crystal structure of a novel conformational state of the flavivirus NS3 protein: implications for polyprotein processing and viral replication. *J Virol,* 83**,** 12895-906.

BALLY, F., MARTINEZ, R., PETERS, S., SUDRE, P. & TELENTI, A. 2000. Polymorphism of HIV type 1 gag p7/p1 and p1/p6 cleavage sites: clinical significance and implications for resistance to protease inhibitors. *AIDS Res Hum Retroviruses,* 16**,** 1209-13.

BARRE-SINOUSSI, F., CHERMANN, J. C., REY, F., NUGEYRE, M. T., CHAMARET, S., GRUEST, J., DAUGUET, C., AXLER-BLIN, C., VEZINET-BRUN, F., ROUZIOUX, C., ROZENBAUM, W. & MONTAGNIER, L. 1983. Isolation of a T-lymphotropic retrovirus from a patient at risk for acquired immune deficiency syndrome (AIDS). *Science,* 220**,** 868-71.

BASTOS LIMA, A., BEHNAM, M. A., EL SHERIF, Y., NITSCHE, C., VECHI, S. M. & KLEIN, C. D. 2015. Dual inhibitors of the dengue and West Nile virus NS2B-NS3 proteases: Synthesis, biological evaluation and docking studies of novel peptide-hybrids. *Bioorg Med Chem,* 23**,** 5748-55.

BEHNAM, M. A., GRAF, D., BARTENSCHLAGER, R., ZLOTOS, D. P. & KLEIN, C. D. 2015. Discovery of Nanomolar Dengue and West Nile Virus Protease Inhibitors Containing a 4-Benzyloxyphenylglycine Residue. *J Med Chem,* 58**,** 9354-70.

BERA, A. K., KUHN, R. J. & SMITH, J. L. 2007. Functional characterization of cis and trans activity of the Flavivirus NS2B-NS3 protease. *J Biol Chem,* 282**,** 12883-92.

BHATT, S., GETHING, P. W., BRADY, O. J., MESSINA, J. P., FARLOW, A. W., MOYES, C. L., DRAKE, J. M., BROWNSTEIN, J. S., HOEN, A. G., SANKOH, O., MYERS, M. F., GEORGE, D. B., JAENISCH, T., WINT, G. R., SIMMONS, C. P., SCOTT, T. W., FARRAR, J. J. & HAY, S. I. 2013. The global distribution and burden of dengue. *Nature,* 496**,** 504-7.

BODE, W. & HUBER, R. 1992. Natural protein proteinase inhibitors and their interaction with proteinases. *Eur J Biochem,* 204**,** 433-51.

BOGOCH, II, BRADY, O. J., KRAEMER, M. U., GERMAN, M., CREATORE, M. I., KULKARNI, M. A., BROWNSTEIN, J. S., MEKARU, S. R., HAY, S. I., GROOT, E., WATTS, A. & KHAN, K. 2016. Anticipating the international spread of Zika virus from Brazil. *Lancet,* 387**,** 335-6.

BOONYASUPPAYAKORN, S., REICHERT, E. D., MANZANO, M., NAGARAJAN, K. & PADMANABHAN, R. 2014. Amodiaquine, an antimalarial drug, inhibits dengue virus type 2 replication and infectivity. *Antiviral Res,* 106**,** 125-34.

BRODER, S., YARCHOAN, R., COLLINS, J. M., LANE, H. C., MARKHAM, P. D., KLECKER, R. W., REDFIELD, R. R., MITSUYA, H., HOTH, D. F., GELMANN, E. & ET AL. 1985. Effects of suramin on HTLV-III/LAV infection presenting as Kaposi's sarcoma or AIDS-related complex: clinical pharmacology and suppression of virus replication in vivo. *Lancet,* 2**,** 627-30.

BRUNGER, A. T. 1992. Free R value: a novel statistical quantity for assessing the accuracy of crystal structures. *Nature,* 355**,** 472-5.

BRYANT, M. & RATNER, L. 1990. Myristoylation-dependent replication and assembly of human immunodeficiency virus 1. *Proc Natl Acad Sci U S A,* 87**,** 523-7.

CAI, Y., YILMAZ, N. K., MYINT, W., ISHIMA, R. & SCHIFFER, C. A. 2012. Differential Flap Dynamics in Wild-type and a Drug Resistant Variant of HIV-1 Protease Revealed by Molecular Dynamics and NMR Relaxation. *J Chem Theory Comput,* 8**,** 3452-3462.

CHAMBERS, T. J., HAHN, C. S., GALLER, R. & RICE, C. M. 1990a. Flavivirus genome organization, expression, and replication. *Annu Rev Microbiol,* 44**,** 649-88.

CHAMBERS, T. J., WEIR, R. C., GRAKOUI, A., MCCOURT, D. W., BAZAN, J. F., FLETTERICK, R. J. & RICE, C. M. 1990b. Evidence that the N-terminal domain of nonstructural protein NS3 from yellow fever virus is a serine protease responsible for site-specific cleavages in the viral polyprotein. *Proc Natl Acad Sci U S A,* 87**,** 8898-902.

CHAN, D. C., FASS, D., BERGER, J. M. & KIM, P. S. 1997. Core structure of gp41 from the HIV envelope glycoprotein. *Cell,* 89**,** 263-73.

CHANDRAMOULI, S., JOSEPH, J. S., DAUDENARDE, S., GATCHALIAN, J., CORNILLEZ-TY, C. & KUHN, P. 2010. Serotype-specific structural differences in the protease-cofactor complexes of the dengue virus family. *J Virol,* 84**,** 3059-67.

CHANG, M., WILLIAMS, O., MITTLER, J., QUINTANILLA, A., CARITHERS, R. L., JR., PERKINS, J., COREY, L. & GRETCH, D. R. 2003. Dynamics of hepatitis C virus replication in human liver. *Am J Pathol,* 163**,** 433-44.

CHANPRAPAPH, S., SAPARPAKORN, P., SANGMA, C., NIYOMRATTANAKIT, P., HANNONGBUA, S., ANGSUTHANASOMBAT, C. & KATZENMEIER, G. 2005. Competitive inhibition of the dengue virus NS3 serine protease by synthetic peptides representing polyprotein cleavage sites. *Biochem Biophys Res Commun,* 330**,** 1237-46.

CHAPPELL, K. J., STOERMER, M. J., FAIRLIE, D. P. & YOUNG, P. R. 2008. West Nile Virus NS2B/NS3 protease as an antiviral target. *Curr Med Chem,* 15**,** 2771-84.

CHIU, I. M., YANIV, A., DAHLBERG, J. E., GAZIT, A., SKUNTZ, S. F., TRONICK, S. R. & AARONSON, S. A. 1985. Nucleotide sequence evidence for relationship of AIDS retrovirus to lentiviruses. *Nature,* 317**,** 366-8.

CHO, S. S., WEINKAM, P. & WOLYNES, P. G. 2008. Origins of barriers and barrierless folding in BBL. *Proc Natl Acad Sci U S A,* 105**,** 118-23.

CHOU, K. C., TOMASSELLI, A. G., REARDON, I. M. & HEINRIKSON, R. L. 1996. Predicting human immunodeficiency virus protease cleavage sites in proteins by a discriminant function method. *Proteins,* 24**,** 51-72.

CLAPHAM, P. R. & WEISS, R. A. 1997. Immunodeficiency viruses. Spoilt for choice of co-receptors. *Nature,* 388**,** 230-1.

CLEMENTE, J. C., HEMRAJANI, R., BLUM, L. E., GOODENOW, M. M. & DUNN, B. M. 2003. Secondary mutations M36I and A71V in the human immunodeficiency virus type 1 protease can provide an advantage for the emergence of the primary mutation D30N. *Biochemistry,* 42**,** 15029-35.

CLYDE, K., KYLE, J. L. & HARRIS, E. 2006. Recent advances in deciphering viral and host determinants of dengue virus replication and pathogenesis. *J Virol,* 80**,** 11418-31.

COLLABORATIVE COMPUTATIONAL PROJECT, N. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr,* 50**,** 760-3.

COLLABORATIVE-COMPUTATIONAL-PROJECT, N. 1994. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr,* 50**,** 760-763.

COLONNO, R., ROSE, R., MCLAREN, C., THIRY, A., PARKIN, N. & FRIBORG, J. 2004. Identification of I50L as the signature atazanavir (ATV)-resistance mutation in treatment-naive HIV-1-infected patients receiving ATV-containing regimens. *J Infect Dis,* 189**,** 1802-10.

CONDRA, J. H., SCHLEIF, W. A., BLAHY, O. M., GABRYELSKI, L. J., GRAHAM, D. J., QUINTERO, J. C., RHODES, A., ROBBINS, H. L., ROTH, E., SHIVAPRAKASH, M. & ET AL. 1995. In vivo emergence of HIV-1 variants resistant to multiple protease inhibitors. *Nature,* 374**,** 569-71.

CRAIG, D. B. & DOMBKOWSKI, A. A. 2013. Disulfide by Design 2.0: a web-based tool for disulfide engineering in proteins. *BMC Bioinformatics,* 14**,** 346.

CRAIG, J. C., DUNCAN, I. B., HOCKLEY, D., GRIEF, C., ROBERTS, N. A. & MILLS, J. S. 1991. Antiviral properties of Ro 31-8959, an inhibitor of human immunodeficiency virus (HIV) proteinase. *Antiviral Res,* 16**,** 295-305.

CREGAR-HERNANDEZ, L., JIAO, G. S., JOHNSON, A. T., LEHRER, A. T., WONG, T. A. & MARGOSIAK, S. A. 2011. Small molecule pan-dengue and West Nile virus NS3 protease inhibitors. *Antivir Chem Chemother,* 21**,** 209-17.

DAM, E., QUERCIA, R., GLASS, B., DESCAMPS, D., LAUNAY, O., DUVAL, X., KRAUSSLICH, H. G., HANCE, A. J., CLAVEL, F. & GROUP, A. S. 2009. Gag mutations strongly contribute to HIV-1 resistance to protease inhibitors in highly drug-experienced patients besides compensating for fitness loss. *PLoS Pathog,* 5**,** e1000345.

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. & RICHARDSON, D. C. 2007. MolProbity: all-atom contacts and structure validation for proteins and nucleic acids. *Nucleic Acids Res,* 35**,** W375-83.

DE VERA, I. M., SMITH, A. N., DANCEL, M. C., HUANG, X., DUNN, B. M. & FANUCCI, G. E. 2013. Elucidating a relationship between conformational sampling and drug resistance in HIV-1 protease. *Biochemistry,* 52**,** 3278-88.

DEBOUCK, C. 1992. The HIV-1 protease as a therapeutic target for AIDS. *AIDS Res Hum Retroviruses,* 8**,** 153-64.

DICK, G. W., KITCHEN, S. F. & HADDOW, A. J. 1952. Zika virus. I. Isolations and serological specificity. *Trans R Soc Trop Med Hyg,* 46**,** 509-20.

DOYON, L., CROTEAU, G., THIBEAULT, D., POULIN, F., PILOTE, L. & LAMARRE, D. 1996. Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors. *J Virol,* 70**,** 3763-9.

DUAN, J., DIXON, S. L., LOWRIE, J. F. & SHERMAN, W. 2010. Analysis and comparison of 2D fingerprints: insights into database screening performance using eight fingerprint methods. *J Mol Graph Model,* 29**,** 157-70.

EARL, P. L., MOSS, B. & DOMS, R. W. 1991. Folding, interaction with GRP78-BiP, assembly, and transport of the human immunodeficiency virus type 1 envelope protein. *J Virol,* 65**,** 2047-55.

EGGER, D., WOLK, B., GOSERT, R., BIANCHI, L., BLUM, H. E., MORADPOUR, D. & BIENZ, K. 2002. Expression of hepatitis C virus proteins induces distinct membrane alterations including a candidate viral replication complex. *J Virol,* 76**,** 5974-84.

EGLOFF, M. P., BENARROCH, D., SELISKO, B., ROMETTE, J. L. & CANARD, B. 2002. An RNA cap (nucleoside-2'-O-)-methyltransferase in the flavivirus RNA polymerase NS5: crystal structure and functional characterization. *EMBO J,* 21**,** 2757-68.

EGLOFF, M. P., DECROLY, E., MALET, H., SELISKO, B., BENARROCH, D., FERRON, F. & CANARD, B. 2007. Structural and functional analysis of methylation and 5'-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J Mol Biol,* 372**,** 723-36.

EMSLEY, P. & COWTAN, K. 2004. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr,* 60**,** 2126-32.

ERBEL, P., SCHIERING, N., D'ARCY, A., RENATUS, M., KROEMER, M., LIM, S. P., YIN, Z., KELLER, T. H., VASUDEVAN, S. G. & HOMMEL, U. 2006. Structural basis for the activation of flaviviral NS3 proteases from dengue and West Nile virus. *Nat Struct Mol Biol,* 13**,** 372-3.

FACKE, M., JANETZKO, A., SHOEMAN, R. L. & KRAUSSLICH, H. G. 1993. A large deletion in the matrix domain of the human immunodeficiency virus gag gene redirects virus particle assembly from the plasma membrane to the endoplasmic reticulum. *J Virol,* 67**,** 4972-80.

FALGOUT, B., PETHEL, M., ZHANG, Y. M. & LAI, C. J. 1991. Both nonstructural proteins NS2B and NS3 are required for the proteolytic processing of dengue virus nonstructural proteins. *J Virol,* 65**,** 2467-75.

FAYE, O., FREIRE, C. C., IAMARINO, A., FAYE, O., DE OLIVEIRA, J. V., DIALLO, M., ZANOTTO, P. M. & SALL, A. A. 2014. Molecular evolution of Zika virus during its emergence in the 20(th) century. *PLoS Negl Trop Dis,* 8**,** e2636.

FEHER, A., WEBER, I. T., BAGOSSI, P., BOROSS, P., MAHALINGAM, B., LOUIS, J. M., COPELAND, T. D., TORSHIN, I. Y., HARRISON, R. W. & TOZSER, J. 2002. Effect of sequence polymorphism and drug resistance on two HIV-1 Gag processing sites. *Eur J Biochem,* 269**,** 4114-20.

FOULKES-MURZYCKI, J. E., SCOTT, W. R. & SCHIFFER, C. A. 2007. Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease. *Structure,* 15**,** 225-33.

FRANKEL, A. D. & YOUNG, J. A. 1998. HIV-1: fifteen proteins and an RNA. *Annu Rev Biochem,* 67**,** 1-25.

FREEDBERG, D. I., ISHIMA, R., JACOB, J., WANG, Y. X., KUSTANOVICH, I., LOUIS, J. M. & TORCHIA, D. A. 2002. Rapid structural fluctuations of the free HIV protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations. *Protein Sci,* 11**,** 221-32.

FRIESNER, R. A., BANKS, J. L., MURPHY, R. B., HALGREN, T. A., KLICIC, J. J., MAINZ, D. T., REPASKY, M. P., KNOLL, E. H., SHELLEY, M., PERRY, J. K., SHAW, D. E., FRANCIS, P. & SHENKIN, P. S. 2004. Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem,* 47**,** 1739-49.

GALLINARI, P., BRENNAN, D., NARDI, C., BRUNETTI, M., TOMEI, L., STEINKUHLER, C. & DE FRANCESCO, R. 1998. Multiple enzymatic activities associated with recombinant NS3 protein of hepatitis C virus. *J Virol,* 72**,** 6758-69.

GALLO, S. A., PURI, A. & BLUMENTHAL, R. 2001. HIV-1 gp41 six-helix bundle formation occurs rapidly after the engagement of gp120 by CXCR4 in the HIV-1 Env-mediated fusion process. *Biochemistry,* 40**,** 12231-6.

GANE, E. J., P. POCKROS, ET AL. . Interferon-Free Treatment With A Combination Of Mericitabine And Danoprevir/R With Or Without Ribavirin In Treatment-Naive Hcv Genotype 1-Infected Patients. 63rd Annual Meeting of the American Association for the Study of Liver Diseases., 2012.

GANESH, V. K., MULLER, N., JUDGE, K., LUAN, C. H., PADMANABHAN, R. & MURTHY, K. H. 2005. Identification and characterization of nonsubstrate based inhibitors of the essential dengue and West Nile virus proteases. *Bioorg Med Chem,* 13**,** 257-64.

GOWER, E., ESTES, C., BLACH, S., RAZAVI-SHEARER, K. & RAZAVI, H. 2014. Global epidemiology and genotype distribution of the hepatitis C virus infection. *J Hepatol,* 61**,** S45-57.

GREEN, S. & ROTHMAN, A. 2006. Immunopathological mechanisms in dengue and dengue hemorrhagic fever. *Curr Opin Infect Dis,* 19**,** 429-36.

GULNIK, S. V., SUVOROV, L. I., LIU, B., YU, B., ANDERSON, B., MITSUYA, H. & ERICKSON, J. W. 1995. Kinetic characterization and cross-resistance patterns of HIV-1 protease mutants selected under drug pressure. *Biochemistry,* 34**,** 9282-7.

HAGEL, M., NIU, D., ST MARTIN, T., SHEETS, M. P., QIAO, L., BERNARD, H., KARP, R. M., ZHU, Z., LABENSKI, M. T., CHATURVEDI, P., NACHT, M., WESTLIN, W. F., PETTER, R. C. & SINGH, J. 2011. Selective irreversible inhibition of a protease by targeting a noncatalytic cysteine. *Nat Chem Biol,* 7**,** 22-4.

HALFON, P. & SARRAZIN, C. 2012. Future treatment of chronic hepatitis C with direct acting antivirals: is resistance important? *Liver Int,* 32 Suppl 1**,** 79-87.

HALGREN, T. A., MURPHY, R. B., FRIESNER, R. A., BEARD, H. S., FRYE, L. L., POLLARD, W. T. & BANKS, J. L. 2004. Glide: a new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening. *J Med Chem,* 47**,** 1750-9.

HAMMAMY, M. Z., HAASE, C., HAMMAMI, M., HILGENFELD, R. & STEINMETZER, T. 2013. Development and characterization of new peptidomimetic inhibitors of the West Nile virus NS2B-NS3 protease. *ChemMedChem,* 8**,** 231-41.

HE, Y., KING, M. S., KEMPF, D. J., LU, L., LIM, H. B., KRISHNAN, P., KATI, W., MIDDLETON, T. & 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**,** 1101-10.

HOWE, A. Y., LONG, J., NICKLE, D., BARNARD, R., THOMPSON, S., HOWE, J., ALVES, K. & WAHL, J. 2015. Long-term follow-up of patients receiving boceprevir for treatment of chronic hepatitis C. *Antiviral Res,* 113**,** 71-8.

HUMPHREY, W., DALKE, A. & SCHULTEN, K. 1996. VMD: visual molecular dynamics. *J Mol Graph,* 14**,** 33-8, 27-8.

ISHIMA, R., FREEDBERG, D. I., WANG, Y. X., LOUIS, J. M. & TORCHIA, D. A. 1999. Flap opening and dimer-interface flexibility in the free and inhibitor-bound HIV protease, and their implications for function. *Structure,* 7**,** 1047-55.

JACOBSON, I. M., MCHUTCHISON, J. G., DUSHEIKO, G., DI BISCEGLIE, A. M., REDDY, K. R., BZOWEJ, N. H., MARCELLIN, P., MUIR, A. J., FERENCI, P., FLISIAK, R., GEORGE, J., RIZZETTO, M., SHOUVAL, D., SOLA, R., TERG, R. A., YOSHIDA, E. M., ADDA, N., BENGTSSON, L., SANKOH, A. J.,

KIEFFER, T. L., GEORGE, S., KAUFFMAN, R. S., ZEUZEM, S. & TEAM, A. S. 2011. Telaprevir for previously untreated chronic hepatitis C virus infection. *N Engl J Med,* 364**,** 2405-16.

JANNE, P. A., GRAY, N. & SETTLEMAN, J. 2009. Factors underlying sensitivity of cancers to small-molecule kinase inhibitors. *Nat Rev Drug Discov,* 8**,** 709-23.

JIANG, Y., ANDREWS, S. W., CONDROSKI, K. R., BUCKMAN, B., SEREBRYANY, V., WENGLOWSKY, S., KENNEDY, A. L., MADDURU, M. R., WANG, B., LYON, M., DOHERTY, G. A., WOODARD, B. T., LEMIEUX, C., GECK DO, M., ZHANG, H., BALLARD, J., VIGERS, G., BRANDHUBER, B. J., STENGEL, P., JOSEY, J. A., BEIGELMAN, L., BLATT, L. & SEIWERT, S. D. 2014. Discovery of danoprevir (ITMN-191/R7227), a highly selective and potent inhibitor of hepatitis C virus (HCV) NS3/4A protease. *J Med Chem,* 57**,** 1753-69.

JILEK, B. L., ZARR, M., SAMPAH, M. E., RABI, S. A., BULLEN, C. K., LAI, J., SHEN, L. & SILICIANO, R. F. 2012. A quantitative basis for antiretroviral therapy for HIV-1 infection. *Nat Med,* 18**,** 446-51.

KANTOR, R., FESSEL, W. J., ZOLOPA, A. R., ISRAELSKI, D., SHULMAN, N., MONTOYA, J. G., HARBOUR, M., SCHAPIRO, J. M. & SHAFER, R. W. 2002. Evolution of primary protease inhibitor resistance mutations during protease inhibitor salvage therapy. *Antimicrob Agents Chemother,* 46**,** 1086-92.

KAPLAN, A. H., MICHAEL, S. F., WEHBIE, R. S., KNIGGE, M. F., PAUL, D. A., EVERITT, L., KEMPF, D. J., NORBECK, D. W., ERICKSON, J. W. & SWANSTROM, R. 1994. Selection of multiple human immunodeficiency virus type 1 variants that encode viral proteases with decreased sensitivity to an inhibitor of the viral protease. *Proc Natl Acad Sci U S A,* 91**,** 5597-601.

KAPLAN, A. H., ZACK, J. A., KNIGGE, M., PAUL, D. A., KEMPF, D. J., NORBECK, D. W. & SWANSTROM, R. 1993. Partial inhibition of the human immunodeficiency virus type 1 protease results in aberrant virus assembly and the formation of noninfectious particles. *J Virol,* 67**,** 4050-5.

KAWAI, T., TAKAHASHI, K., SATO, S., COBAN, C., KUMAR, H., KATO, H., ISHII, K. J., TAKEUCHI, O. & AKIRA, S. 2005. IPS-1, an adaptor triggering RIG-I- and Mda5-mediated type I interferon induction. *Nat Immunol,* 6**,** 981-8.

KEMPF, D. J., MARSH, K. C., DENISSEN, J. F., MCDONALD, E., VASAVANONDA, S., FLENTGE, C. A., GREEN, B. E., FINO, L., PARK, C. H., KONG, X. P. & ET AL. 1995. ABT-538 is a potent inhibitor of human immunodeficiency virus protease and has high oral bioavailability in humans. *Proc Natl Acad Sci U S A,* 92**,** 2484-8.

KERN, W. V., OETHINGER, M., JELLEN-RITTER, A. S. & LEVY, S. B. 2000. Non-target gene mutations in the development of fluoroquinolone resistance in Escherichia coli. *Antimicrob Agents Chemother,* 44**,** 814-20.

KIEFFER, T. L., SARRAZIN, C., MILLER, J. S., WELKER, M. W., FORESTIER, N., REESINK, H. W., KWONG, A. D. & ZEUZEM, S. 2007. Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis C virus replication in patients. *Hepatology,* 46**,** 631-9.

KING, N. M., MELNICK, L., PRABU-JEYABALAN, M., NALIVAIKA, E. A., YANG, S. S., GAO, Y., NIE, X., ZEPP, C., HEEFNER, D. L. & SCHIFFER, C. A. 2002. Lack of synergy for inhibitors targeting a multi-drug-resistant HIV-1 protease. *Protein Sci,* 11**,** 418-29.

KING, N. M., PRABU-JEYABALAN, M., NALIVAIKA, E. A. & SCHIFFER, C. A. 2004. Combating susceptibility to drug resistance: lessons from HIV-1 protease. *Chem Biol,* 11**,** 1333-8.

KNEHANS, T., SCHULLER, A., DOAN, D. N., NACRO, K., HILL, J., GUNTERT, P., MADHUSUDHAN, M. S., WEIL, T. & VASUDEVAN, S. G. 2011. Structure-guided fragment-based in silico drug design of dengue protease inhibitors. *J Comput Aided Mol Des,* 25**,** 263-74.

KOHL, N. E., EMINI, E. A., SCHLEIF, W. A., DAVIS, L. J., HEIMBACH, J. C., DIXON, R. A., SCOLNICK, E. M. & SIGAL, I. S. 1988. Active human immunodeficiency virus protease is required for viral infectivity. *Proc Natl Acad Sci U S A,* 85**,** 4686-90.

KOLLI, M., LASTERE, S. & SCHIFFER, C. A. 2006. Co-evolution of nelfinavir-resistant HIV-1 protease and the p1-p6 substrate. *Virology,* 347**,** 405-9.

KOLLI, M., OZEN, A., KURT-YILMAZ, N. & SCHIFFER, C. A. 2014. HIV-1 protease-substrate coevolution in nelfinavir resistance. *J Virol,* 88**,** 7145-54.

KOLLI, M., STAWISKI, E., CHAPPEY, C. & SCHIFFER, C. A. 2009. Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance. *J Virol,* 83**,** 11027-42.

KOUP, R. A. 2001. A new latent HIV reservoir. *Nat Med,* 7**,** 404-5.

KUHN, R. J., ZHANG, W., ROSSMANN, M. G., PLETNEV, S. V., CORVER, J., LENCHES, E., JONES, C. T., MUKHOPADHYAY, S., CHIPMAN, P. R., STRAUSS, E. G., BAKER, T. S. & STRAUSS, J. H. 2002. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell,* 108**,** 717-25.

KUMAR, A., VOET, A. & ZHANG, K. Y. 2012. Fragment based drug design: from experimental to computational approaches. *Curr Med Chem,* 19**,** 5128-47.

KUNO, G. & CHANG, G. J. 2007. Full-length sequencing and genomic characterization of Bagaza, Kedougou, and Zika viruses. *Arch Virol,* 152**,** 687-96.

KWONG, A. D., KAUFFMAN, R. S., HURTER, P. & MUELLER, P. 2011. Discovery and development of telaprevir: an NS3-4A protease inhibitor for treating genotype 1 chronic hepatitis C virus. *Nat Biotechnol,* 29**,** 993-1003.

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. & SIMMEN, K. A. 2010. In vitro resistance profile of the hepatitis C virus NS3/4A protease inhibitor TMC435. *Antimicrob Agents Chemother,* 54**,** 1878-87.

LEUNG, D., SCHRODER, K., WHITE, H., FANG, N. X., STOERMER, M. J., ABBENANTE, G., MARTIN, J. L., YOUNG, P. R. & FAIRLIE, D. P. 2001. Activity of recombinant dengue 2 virus NS3 protease in the presence of a

truncated NS2B co-factor, small peptide substrates, and inhibitors. *J Biol Chem,* 276**,** 45762-71.

LI, J., LIM, S. P., BEER, D., PATEL, V., WEN, D., TUMANUT, C., TULLY, D. C., WILLIAMS, J. A., JIRICEK, J., PRIESTLE, J. P., HARRIS, J. L. & VASUDEVAN, S. G. 2005a. Functional profiling of recombinant NS3 proteases from all four serotypes of dengue virus using tetrapeptide and octapeptide substrate libraries. *J Biol Chem,* 280**,** 28766-74.

LI, K., FOY, E., FERREON, J. C., NAKAMURA, M., FERREON, A. C., IKEDA, M., RAY, S. C., GALE, M., JR. & LEMON, S. M. 2005b. 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**,** 2992-7.

LI, L., LOK, S. M., YU, I. M., ZHANG, Y., KUHN, R. J., CHEN, J. & ROSSMANN, M. G. 2008. The flavivirus precursor membrane-envelope protein complex: structure and maturation. *Science,* 319**,** 1830-4.

LI, X. Z. & NIKAIDO, H. 2009. Efflux-mediated drug resistance in bacteria: an update. *Drugs,* 69**,** 1555-623.

LIM, S. P., WANG, Q. Y., NOBLE, C. G., CHEN, Y. L., DONG, H., ZOU, B., YOKOKAWA, F., NILAR, S., SMITH, P., BEER, D., LESCAR, J. & SHI, P. Y. 2013. Ten years of dengue drug discovery: progress and prospects. *Antiviral Res,* 100**,** 500-19.

LIM, S. R., QIN, X., SUSSER, S., NICHOLAS, J. B., LANGE, C., HERRMANN, E., HONG, J., ARFSTEN, A., HOOI, L., BRADFORD, W., NAJERA, I., SMITH, P., ZEUZEM, S., KOSSEN, K., SARRAZIN, C. & SEIWERT, S. D. 2012. Virologic escape during danoprevir (ITMN-191/RG7227) monotherapy is hepatitis C virus subtype dependent and associated with R155K substitution. *Antimicrob Agents Chemother,* 56**,** 271-9.

LUO, D., XU, T., HUNKE, C., GRUBER, G., VASUDEVAN, S. G. & LESCAR, J. 2008. Crystal structure of the NS3 protease-helicase from dengue virus. *J Virol,* 82**,** 173-83.

MA, L., JONES, C. T., GROESCH, T. D., KUHN, R. J. & POST, C. B. 2004. Solution structure of dengue virus capsid protein reveals another fold. *Proc Natl Acad Sci U S A,* 101**,** 3414-9.

MAGUIRE, M. F., GUINEA, R., GRIFFIN, P., MACMANUS, S., ELSTON, R. C., WOLFRAM, J., RICHARDS, N., HANLON, M. H., PORTER, D. J., WRIN, T., PARKIN, N., TISDALE, M., FURFINE, E., PETROPOULOS, C., SNOWDEN, B. W. & KLEIM, J. P. 2002. Changes in human immunodeficiency virus type 1 Gag at positions L449 and P453 are linked to I50V protease mutants in vivo and cause reduction of sensitivity to amprenavir and improved viral fitness in vitro. *J Virol,* 76**,** 7398-406.

MAHALINGAM, B., LOUIS, J. M., REED, C. C., ADOMAT, J. M., KROUSE, J., WANG, Y. F., HARRISON, R. W. & WEBER, I. T. 1999. Structural and kinetic analysis of drug resistant mutants of HIV-1 protease. *Eur J Biochem,* 263**,** 238-45.

MAHDY, A. M. & WEBSTER, N. R. 2004. Perioperative systemic haemostatic agents. *Br J Anaesth,* 93**,** 842-58.

MAJOR, M. E. & FEINSTONE, S. M. 1997. The molecular virology of hepatitis C. *Hepatology,* 25**,** 1527-38.

MALCOLM, B. A., LIU, R., LAHSER, F., AGRAWAL, S., BELANGER, B., BUTKIEWICZ, N., CHASE, R., GHEYAS, F., HART, A., HESK, D., INGRAVALLO, P., JIANG, C., KONG, R., LU, J., PICHARDO, J., PRONGAY, A., SKELTON, A., TONG, X., VENKATRAMAN, S., XIA, E., GIRIJAVALLABHAN, V. & NJOROGE, F. G. 2006. SCH 503034, a mechanism-based inhibitor of hepatitis C virus NS3 protease, suppresses polyprotein maturation and enhances the antiviral activity of alpha interferon in replicon cells. *Antimicrob Agents Chemother,* 50**,** 1013-20.

MAMMANO, F., PETIT, C. & CLAVEL, F. 1998. Resistance-associated loss of viral fitness in human immunodeficiency virus type 1: phenotypic analysis of protease and gag coevolution in protease inhibitor-treated patients. *J Virol,* 72**,** 7632-7.

MANNS, M., REESINK, H., BERG, T., DUSHEIKO, G., FLISIAK, R., MARCELLIN, P., MORENO, C., LENZ, O., MEYVISCH, P., PEETERS, M., SEKAR, V., SIMMEN, K. & VERLOES, R. 2011a. Rapid viral response of once-daily TMC435 plus pegylated interferon/ribavirin in hepatitis C genotype-1 patients: a randomized trial. *Antivir Ther,* 16**,** 1021-33.

MANNS, M. P., BOURLIERE, M., BENHAMOU, Y., POL, S., BONACINI, M., TREPO, C., WRIGHT, D., BERG, T., CALLEJA, J. L., WHITE, P. W., STERN, J. O., STEINMANN, G., YONG, C. L., KUKOLJ, G., SCHERER, J. & BOECHER, W. O. 2011b. Potency, safety, and pharmacokinetics of the NS3/4A protease inhibitor BI201335 in patients with chronic HCV genotype-1 infection. *J Hepatol,* 54**,** 1114-22.

MANSKY, L. M. & TEMIN, H. M. 1995. Lower in vivo mutation rate of human immunodeficiency virus type 1 than that predicted from the fidelity of purified reverse transcriptase. *J Virol,* 69**,** 5087-94.

MCCOY, A. J., GROSSE-KUNSTLEVE, R. W., ADAMS, P. D., WINN, M. D., STORONI, L. C. & READ, R. J. 2007. Phaser crystallographic software. *J Appl Crystallogr,* 40**,** 658-674.

MITTAL, S., BANDARANAYAKE, R. M., KING, N. M., PRABU-JEYABALAN, M., NALAM, M. N., NALIVAIKA, E. A., YILMAZ, N. K. & SCHIFFER, C. A. 2013. Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in HIV-1 protease with mutations at residue 50. *J Virol,* 87**,** 4176-84.

MITTAL, S., CAI, Y., NALAM, M. N., BOLON, D. N. & SCHIFFER, C. A. 2012. Hydrophobic core flexibility modulates enzyme activity in HIV-1 protease. *J Am Chem Soc,* 134**,** 4163-8.

MOLLA, A., KORNEYEVA, M., GAO, Q., VASAVANONDA, S., SCHIPPER, P. J., MO, H. M., MARKOWITZ, M., CHERNYAVSKIY, T., NIU, P., LYONS, N., HSU, A., GRANNEMAN, G. R., HO, D. D., BOUCHER, C. A., LEONARD, J. M., NORBECK, D. W. & KEMPF, D. J. 1996. Ordered accumulation of mutations in HIV protease confers resistance to ritonavir. *Nat Med,* 2**,** 760-6.

MONATH, T. P. 1994. Dengue: the risk to developed and developing countries. *Proc Natl Acad Sci U S A,* 91**,** 2395-400.

MORADPOUR, D., PENIN, F. & RICE, C. M. 2007. Replication of hepatitis C virus. *Nat Rev Microbiol,* 5**,** 453-63.
MORRIS, R. J., PERRAKIS, A. & LAMZIN, V. S. 2002a. ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr D Biol Crystallogr,* 58**,** 968-75.
MORRIS, R. J., PERRAKIS, A. & LAMZIN, V. S. 2002b. ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr D Biol Crystallogr,* D58**,** 968-975.
MUELLER, N. H., YON, C., GANESH, V. K. & PADMANABHAN, R. 2007. Characterization of the West Nile virus protease substrate specificity and inhibitors. *Int J Biochem Cell Biol,* 39**,** 606-14.
NALAM, M. N., ALI, A., REDDY, G. S., CAO, H., ANJUM, S. G., ALTMAN, M. D., YILMAZ, N. K., TIDOR, B., RANA, T. M. & SCHIFFER, C. A. 2013. Substrate envelope-designed potent HIV-1 protease inhibitors to avoid drug resistance. *Chem Biol,* 20**,** 1116-24.
NAVIA, M. A., FITZGERALD, P. M., MCKEEVER, B. M., LEU, C. T., HEIMBACH, J. C., HERBER, W. K., SIGAL, I. S., DARKE, P. L. & SPRINGER, J. P. 1989. Three-dimensional structure of aspartyl protease from human immunodeficiency virus HIV-1. *Nature,* 337**,** 615-20.
NEUMANN, A. U., LAM, N. P., DAHARI, H., GRETCH, D. R., WILEY, T. E., LAYDEN, T. J. & PERELSON, A. S. 1998. Hepatitis C viral dynamics in vivo and the antiviral efficacy of interferon-alpha therapy. *Science,* 282**,** 103-7.
NIJHUIS, M., SCHUURMAN, R., DE JONG, D., ERICKSON, J., GUSTCHINA, E., ALBERT, J., SCHIPPER, P., GULNIK, S. & BOUCHER, C. A. 1999. Increased fitness of drug resistant HIV-1 protease as a result of acquisition of compensatory mutations during suboptimal therapy. *AIDS,* 13**,** 2349-59.
NITSCHE, C., BEHNAM, M. A., STEUER, C. & KLEIN, C. D. 2012. Retro peptide-hybrids as selective inhibitors of the Dengue virus NS2B-NS3 protease. *Antiviral Res,* 94**,** 72-9.
NITSCHE, C., STEUER, C. & KLEIN, C. D. 2011. Arylcyanoacrylamides as inhibitors of the Dengue and West Nile virus proteases. *Bioorg Med Chem,* 19**,** 7318-37.
NOBLE, C. G., SEH, C. C., CHAO, A. T. & SHI, P. Y. 2012. Ligand-bound structures of the dengue virus protease reveal the active conformation. *J Virol,* 86**,** 438-46.
OGATA, N., ALTER, H. J., MILLER, R. H. & PURCELL, R. H. 1991. Nucleotide sequence and mutation rate of the H strain of hepatitis C virus. *Proc Natl Acad Sci U S A,* 88**,** 3392-6.
OTWINOWSKI, Z. & MINOR, W. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol,* 276**,** 307-326.
OZEN, A., HALILOGLU, T. & SCHIFFER, C. A. 2011. Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *J Mol Biol,* 410**,** 726-44.
OZEN, A., HALILOGLU, T. & SCHIFFER, C. A. 2012. HIV-1 Protease and Substrate Coevolution Validates the Substrate Envelope As the Substrate Recognition Pattern. *J Chem Theory Comput,* 8.

OZEN, A., LIN, K. H., KURT YILMAZ, N. & SCHIFFER, C. A. 2014. Structural basis and distal effects of Gag substrate coevolution in drug resistance to HIV-1 protease. *Proc Natl Acad Sci U S A,* 111**,** 15993-8.

OZEN, A., SHERMAN, W. & SCHIFFER, C. A. 2013. Improving the Resistance Profile of Hepatitis C NS3/4A Inhibitors: Dynamic Substrate Envelope Guided Design. *J Chem Theory Comput,* 9**,** 5693-5705.

PAN, A. C., BORHANI, D. W., DROR, R. O. & SHAW, D. E. 2013. Molecular determinants of drug-receptor binding kinetics. *Drug Discov Today,* 18**,** 667-73.

PARADA, C. A. & ROEDER, R. G. 1996. Enhanced processivity of RNA polymerase II triggered by Tat-induced phosphorylation of its carboxy-terminal domain. *Nature,* 384**,** 375-8.

PARTALEDIS, J. A., YAMAGUCHI, K., TISDALE, M., BLAIR, E. E., FALCIONE, C., MASCHERA, B., MYERS, R. E., PAZHANISAMY, S., FUTER, O., CULLINAN, A. B. & ET AL. 1995. In vitro selection and characterization of human immunodeficiency virus type 1 (HIV-1) isolates with reduced sensitivity to hydroxyethylamino sulfonamide inhibitors of HIV-1 aspartyl protease. *J Virol,* 69**,** 5228-35.

PATICK, A. K., DURAN, M., CAO, Y., SHUGARTS, D., KELLER, M. R., MAZABEL, E., KNOWLES, M., CHAPMAN, S., KURITZKES, D. R. & MARKOWITZ, M. 1998. Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir. *Antimicrob Agents Chemother,* 42**,** 2637-44.

PATICK, A. K., MO, H., MARKOWITZ, M., APPELT, K., WU, B., MUSICK, L., KALISH, V., KALDOR, S., REICH, S., HO, D. & WEBBER, S. 1996. Antiviral and resistance studies of AG1343, an orally bioavailable inhibitor of human immunodeficiency virus protease. *Antimicrob Agents Chemother,* 40**,** 292-7.

PERELSON, A. S., NEUMANN, A. U., MARKOWITZ, M., LEONARD, J. M. & HO, D. D. 1996. HIV-1 dynamics in vivo: virion clearance rate, infected cell life-span, and viral generation time. *Science,* 271**,** 1582-6.

PEREZ-VALERO, I. & ARRIBAS, J. R. 2011. Protease inhibitor monotherapy. *Curr Opin Infect Dis,* 24**,** 7-11.

PERNI, R. B., ALMQUIST, S. J., BYRN, R. A., CHANDORKAR, G., CHATURVEDI, P. R., COURTNEY, L. F., DECKER, C. J., DINEHART, K., GATES, C. A., HARBESON, S. L., HEISER, A., KALKERI, G., KOLACZKOWSKI, E., LIN, K., LUONG, Y. P., RAO, B. G., TAYLOR, W. P., THOMSON, J. A., TUNG, R. D., WEI, Y., KWONG, A. D. & LIN, C. 2006. Preclinical profile of VX-950, a potent, selective, and orally bioavailable inhibitor of hepatitis C virus NS3-4A serine protease. *Antimicrob Agents Chemother,* 50**,** 899-909.

PERRYMAN, A. L., LIN, J. H. & MCCAMMON, J. A. 2004. HIV-1 protease molecular dynamics of a wild-type and of the V82F/I84V mutant: possible contributions to drug resistance and a potential new target site for drugs. *Protein Sci,* 13**,** 1108-23.

PETTIT, S. C., LINDQUIST, J. N., KAPLAN, A. H. & SWANSTROM, R. 2005. Processing sites in the human immunodeficiency virus type 1 (HIV-1) Gag-Pro-

Pol precursor are cleaved by the viral protease at different rates. *Retrovirology,* 2**,** 66.

PETTIT, S. C., SHENG, N., TRITCH, R., ERICKSON-VIITANEN, S. & SWANSTROM, R. 1998. The regulation of sequential processing of HIV-1 Gag by the viral protease. *Adv Exp Med Biol,* 436**,** 15-25.

POORDAD, F., MCCONE, J., JR., BACON, B. R., BRUNO, S., MANNS, M. P., SULKOWSKI, M. S., JACOBSON, I. M., REDDY, K. R., GOODMAN, Z. D., BOPARAI, N., DINUBILE, M. J., SNIUKIENE, V., BRASS, C. A., ALBRECHT, J. K., BRONOWICKI, J. P. & INVESTIGATORS, S.-. 2011. Boceprevir for untreated chronic HCV genotype 1 infection. *N Engl J Med,* 364**,** 1195-206.

PRABU-JEYABALAN, M., KING, N. M., NALIVAIKA, E. A., HEILEK-SNYDER, G., CAMMACK, N. & 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**,** 1518-21.

PRABU-JEYABALAN, M., NALIVAIKA, E. & SCHIFFER, C. A. 2002. Substrate shape determines specificity of recognition for HIV-1 protease: analysis of crystal structures of six substrate complexes. *Structure,* 10**,** 369-81.

RABI, S. A., LAIRD, G. M., DURAND, C. M., LASKEY, S., SHAN, L., BAILEY, J. R., CHIOMA, S., MOORE, R. D. & SILICIANO, R. F. 2013. Multi-step inhibition explains HIV-1 protease inhibitor pharmacodynamics and resistance. *J Clin Invest,* 123**,** 3848-60.

RHEE, S. Y., GONZALES, M. J., KANTOR, R., BETTS, B. J., RAVELA, J. & SHAFER, R. W. 2003. Human immunodeficiency virus reverse transcriptase and protease sequence database. *Nucleic Acids Res,* 31**,** 298-303.

RICO-HESSE, R. 1990. Molecular evolution and distribution of dengue viruses type 1 and 2 in nature. *Virology,* 174**,** 479-93.

RIEMANN, J. F. & KOHLER, B. 1988. [AIDS and gastroenterologic endoscopy. Survey of current practice]. *Z Gastroenterol,* 26**,** 293-6.

ROBINSON, B. S., RICCARDI, K. A., GONG, Y. F., GUO, Q., STOCK, D. A., BLAIR, W. S., TERRY, B. J., DEMINIE, C. A., DJANG, F., COLONNO, R. J. & LIN, P. F. 2000. BMS-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents. *Antimicrob Agents Chemother,* 44**,** 2093-9.

RODENHUIS-ZYBERT, I. A., WILSCHUT, J. & SMIT, J. M. 2010. Dengue virus life cycle: viral and host factors modulating infectivity. *Cell Mol Life Sci,* 67**,** 2773-86.

ROMANO, K. P., ALI, A., AYDIN, C., SOUMANA, D., OZEN, A., DEVEAU, L. M., SILVER, C., CAO, H., NEWTON, A., PETROPOULOS, C. J., HUANG, W. & SCHIFFER, C. A. 2012. The molecular basis of drug resistance against hepatitis C virus NS3/4A protease inhibitors. *PLoS Pathog,* 8**,** e1002832.

ROMANO, K. P., ALI, A., ROYER, W. E. & 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,* 107**,** 20986-91.

ROMANO, K. P., LAINE, J. M., DEVEAU, L. M., CAO, H., MASSI, F. & SCHIFFER, C. A. 2011. Molecular mechanisms of viral and host cell substrate recognition by hepatitis C virus NS3/4A protease. *J Virol,* 85**,** 6106-16.

ROSE, J. R., SALTO, R. & CRAIK, C. S. 1993. Regulation of autoproteolysis of the HIV-1 and HIV-2 proteases with engineered amino acid substitutions. *J Biol Chem,* 268**,** 11939-45.

ROSENQUIST, A., SAMUELSSON, B., JOHANSSON, P. O., CUMMINGS, M. D., LENZ, O., RABOISSON, P., SIMMEN, K., VENDEVILLE, S., DE KOCK, H., NILSSON, M., HORVATH, A., KALMEIJER, R., DE LA ROSA, G. & BEUMONT-MAUVIEL, M. 2014. Discovery and development of simeprevir (TMC435), a HCV NS3/4A protease inhibitor. *J Med Chem,* 57**,** 1673-93.

ROSS, T. M. 2010. Dengue virus. *Clin Lab Med,* 30**,** 149-60.

SADLER, B. M. & STEIN, D. S. 2002. Clinical pharmacology and pharmacokinetics of amprenavir. *Ann Pharmacother,* 36**,** 102-18.

SALYKIN, A., KUZMIC, P., KYRYLENKO, O., MUSILOVA, J., GLATZ, Z., DVORAK, P. & KYRYLENKO, S. 2013. Nonlinear regression models for determination of nicotinamide adenine dinucleotide content in human embryonic stem cells. *Stem Cell Rev,* 9**,** 786-93.

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. & KWONG, A. D. 2007a. Dynamic hepatitis C virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir. *Gastroenterology,* 132**,** 1767-77.

SARRAZIN, C., ROUZIER, R., WAGNER, F., FORESTIER, N., LARREY, D., GUPTA, S. K., HUSSAIN, M., SHAH, A., CUTLER, D., ZHANG, J. & ZEUZEM, S. 2007b. SCH 503034, a novel hepatitis C virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders. *Gastroenterology,* 132**,** 1270-8.

SASTRY, G. M., ADZHIGIREY, M., DAY, T., ANNABHIMOJU, R. & SHERMAN, W. 2013. Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments. *J Comput Aided Mol Des,* 27**,** 221-34.

SAYER, J. M., LIU, F., ISHIMA, R., WEBER, I. T. & LOUIS, J. M. 2008. Effect of the active site D25N mutation on the structure, stability, and ligand binding of the mature HIV-1 protease. *J Biol Chem,* 283**,** 13459-70.

SCHRÖDINGER 2015. Release 2015-3: Maestro, version 10.3, version 10.3 ed., Schrödinger, LLC, New York, NY.

SEELMEIER, S., SCHMIDT, H., TURK, V. & VON DER HELM, K. 1988. Human immunodeficiency virus has an aspartic-type protease that can be inhibited by pepstatin A. *Proc Natl Acad Sci U S A,* 85**,** 6612-6.

SETH, R. B., SUN, L., EA, C. K. & CHEN, Z. J. 2005. Identification and characterization of MAVS, a mitochondrial antiviral signaling protein that activates NF-kappaB and IRF 3. *Cell,* 122**,** 669-82.

SHAFER, R. W. 2006. Rationale and uses of a public HIV drug-resistance database. *J Infect Dis,* 194 Suppl 1**,** S51-8.

SHAM, H. L., KEMPF, D. J., MOLLA, A., MARSH, K. C., KUMAR, G. N., CHEN, C. M., KATI, W., STEWART, K., LAL, R., HSU, A., BETEBENNER, D., KORNEYEVA, M., VASAVANONDA, S., MCDONALD, E., SALDIVAR, A., WIDEBURG, N., CHEN, X., NIU, P., PARK, C., JAYANTI, V., GRABOWSKI, B., GRANNEMAN, G. R., SUN, E., JAPOUR, A. J., LEONARD, J. M., PLATTNER, J. J. & NORBECK, D. W. 1998. ABT-378, a highly potent inhibitor of the human immunodeficiency virus protease. *Antimicrob Agents Chemother,* 42**,** 3218-24.

SHEN, L., PETERSON, S., SEDAGHAT, A. R., MCMAHON, M. A., CALLENDER, M., ZHANG, H., ZHOU, Y., PITT, E., ANDERSON, K. S., ACOSTA, E. P. & SILICIANO, R. F. 2008. Dose-response curve slope sets class-specific limits on inhibitory potential of anti-HIV drugs. *Nat Med,* 14**,** 762-6.

SHIRYAEV, S. A., KOZLOV, I. A., RATNIKOV, B. I., SMITH, J. W., LEBL, M. & STRONGIN, A. Y. 2007. Cleavage preference distinguishes the two-component NS2B-NS3 serine proteinases of Dengue and West Nile viruses. *Biochem J,* 401**,** 743-52.

SHIVAKUMAR, D., WILLIAMS, J., WU, Y., DAMM, W., SHELLEY, J. & SHERMAN, W. 2010. Prediction of Absolute Solvation Free Energies using Molecular Dynamics Free Energy Perturbation and the OPLS Force Field. *J Chem Theory Comput,* 6**,** 1509-19.

SIMMONDS, P., BUKH, J., COMBET, C., DELEAGE, G., ENOMOTO, N., FEINSTONE, S., HALFON, P., INCHAUSPE, G., KUIKEN, C., MAERTENS, G., MIZOKAMI, M., MURPHY, D. G., OKAMOTO, H., PAWLOTSKY, J. M., PENIN, F., SABLON, E., SHIN, I. T., STUYVER, L. J., THIEL, H. J., VIAZOV, S., WEINER, A. J. & WIDELL, A. 2005. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. *Hepatology,* 42**,** 962-73.

SIMMONS, C. P. 2015. A Candidate Dengue Vaccine Walks a Tightrope. *N Engl J Med,* 373**,** 1263-4.

SUN, Z., LU, W., JIANG, A., CHEN, J., TANG, F. & LIU, J. N. 2009. Expression, purification and characterization of aprotinin and a human analogue of aprotinin. *Protein Expr Purif,* 65**,** 238-43.

SUSSER, S., WELSCH, C., WANG, Y., ZETTLER, M., DOMINGUES, F. S., KAREY, U., HUGHES, E., RALSTON, R., TONG, X., HERRMANN, E., ZEUZEM, S. & SARRAZIN, C. 2009. Characterization of resistance to the protease inhibitor boceprevir in hepatitis C virus-infected patients. *Hepatology,* 50**,** 1709-18.

SVAROVSKAIA, E. S., MARTIN, R., MCHUTCHISON, J. G., MILLER, M. D. & MO, H. 2012. Abundant drug-resistant NS3 mutants detected by deep sequencing in hepatitis C virus-infected patients undergoing NS3 protease inhibitor monotherapy. *J Clin Microbiol,* 50**,** 3267-74.

THEURETZBACHER, U. & MOUTON, J. W. 2011. Update on antibacterial and antifungal drugs - can we master the resistance crisis? *Curr Opin Pharmacol,* 11**,** 429-32.

TONG, X., BOGEN, S., CHASE, R., GIRIJAVALLABHAN, V., GUO, Z., NJOROGE, F. G., PRONGAY, A., SAKSENA, A., SKELTON, A., XIA, E. & RALSTON, R.

2008. Characterization of resistance mutations against HCV ketoamide protease inhibitors. *Antiviral Res,* 77**,** 177-85.

TOZSER, J., BLAHA, I., COPELAND, T. D., WONDRAK, E. M. & OROSZLAN, S. 1991. Comparison of the HIV-1 and HIV-2 proteinases using oligopeptide substrates representing cleavage sites in Gag and Gag-Pol polyproteins. *FEBS Lett,* 281**,** 77-80.

TURNER, S. R., STROHBACH, J. W., TOMMASI, R. A., ARISTOFF, P. A., JOHNSON, P. D., SKULNICK, H. I., DOLAK, L. A., SEEST, E. P., TOMICH, P. K., BOHANON, M. J., HORNG, M. M., LYNN, J. C., CHONG, K. T., HINSHAW, R. R., WATENPAUGH, K. D., JANAKIRAMAN, M. N. & THAISRIVONGS, S. 1998. Tipranavir (PNU-140690): a potent, orally bioavailable nonpeptidic HIV protease inhibitor of the 5,6-dihydro-4-hydroxy-2-pyrone sulfonamide class. *J Med Chem,* 41**,** 3467-76.

UNAIDS. 2016. Global AIDS Update.

VACCA, J. P., DORSEY, B. D., SCHLEIF, W. A., LEVIN, R. B., MCDANIEL, S. L., DARKE, P. L., ZUGAY, J., QUINTERO, J. C., BLAHY, O. M., ROTH, E. & ET AL. 1994. L-735,524: an orally bioavailable human immunodeficiency virus type 1 protease inhibitor. *Proc Natl Acad Sci U S A,* 91**,** 4096-100.

VAN MARCK H, E. A. 2007. Unravelling the complex resistance pathways of darunavir using bioinformatics resistance determination (BIRD). *Antivir Ther* 12**,** S141-S141.

VERMEHREN, J. & SARRAZIN, C. 2012. The role of resistance in HCV treatment. *Best Pract Res Clin Gastroenterol,* 26**,** 487-503.

VERMEIREN, H., VAN CRAENENBROECK, E., ALEN, P., BACHELER, L., PICCHIO, G., LECOCQ, P. & VIRCO CLINICAL RESPONSE COLLABORATIVE, T. 2007. Prediction of HIV-1 drug susceptibility phenotype from the viral genotype using linear regression modeling. *J Virol Methods,* 145**,** 47-55.

WEBER, I. T., MILLER, M., JASKOLSKI, M., LEIS, J., SKALKA, A. M. & WLODAWER, A. 1989. Molecular modeling of the HIV-1 protease and its substrate binding site. *Science,* 243**,** 928-31.

WELSCH, C., DOMINGUES, F. S., SUSSER, S., ANTES, I., HARTMANN, C., MAYR, G., SCHLICKER, A., SARRAZIN, C., ALBRECHT, M., ZEUZEM, S. & LENGAUER, T. 2008. Molecular basis of telaprevir resistance due to V36 and T54 mutations in the NS3-4A protease of the hepatitis C virus. *Genome Biol,* 9**,** R16.

WELSCH, C. & ZEUZEM, S. 2012. Clinical relevance of HCV antiviral drug resistance. *Curr Opin Virol,* 2**,** 651-5.

WENSING, A. M., CALVEZ, V., GUNTHARD, H. F., JOHNSON, V. A., PAREDES, R., PILLAY, D., SHAFER, R. W. & RICHMAN, D. D. 2014. 2014 Update of the drug resistance mutations in HIV-1. *Top Antivir Med,* 22**,** 642-50.

WHO 2009. *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition.* Geneva.

WHO. 2014. *Guidelines for the Screening, Care and Treatment of Persons with Hepatitis C Infection* [Online]. Available: http://www.ncbi.nlm.nih.gov/pubmed/25535634.

WITTEKIND, M. S., WEINHEIMER, S., ZHANG, Y. & GOLDFARB, V. 2002. *Modified forms of hepatitis C NS3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes*.

XU, T., SAMPATH, A., CHAO, A., WEN, D., NANAO, M., CHENE, P., VASUDEVAN, S. G. & LESCAR, J. 2005. Structure of the Dengue virus helicase/nucleoside triphosphatase catalytic domain at a resolution of 2.4 A. *J Virol,* 79**,** 10278-88.

YANG, C. C., HU, H. S., WU, R. H., WU, S. H., LEE, S. J., JIAANG, W. T., CHERN, J. H., HUANG, Z. S., WU, H. N., CHANG, C. M. & YUEH, A. 2014. A novel dengue virus inhibitor, BP13944, discovered by high-throughput screening with dengue virus replicon cells selects for resistance in the viral NS2B/NS3 protease. *Antimicrob Agents Chemother,* 58**,** 110-9.

YAO, N., REICHERT, P., TAREMI, S. S., PROSISE, W. W. & 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**,** 1353-63.

YAP, T. L., XU, T., CHEN, Y. L., MALET, H., EGLOFF, M. P., CANARD, B., VASUDEVAN, S. G. & LESCAR, J. 2007. Crystal structure of the dengue virus RNA-dependent RNA polymerase catalytic domain at 1.85-angstrom resolution. *J Virol,* 81**,** 4753-65.

YENI, P. G., HAMMER, S. M., CARPENTER, C. C., COOPER, D. A., FISCHL, M. A., GATELL, J. M., GAZZARD, B. G., HIRSCH, M. S., JACOBSEN, D. M., KATZENSTEIN, D. A., MONTANER, J. S., RICHMAN, D. D., SAAG, M. S., SCHECHTER, M., SCHOOLEY, R. T., THOMPSON, M. A., VELLA, S. & VOLBERDING, P. A. 2002. Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society-USA Panel. *JAMA,* 288**,** 222-35.

YI, M., MA, Y., YATES, J. & 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.

YILDIZ, M., GHOSH, S., BELL, J. A., SHERMAN, W. & HARDY, J. A. 2013. Allosteric inhibition of the NS2B-NS3 protease from dengue virus. *ACS Chem Biol,* 8**,** 2744-52.

YIN, Z., PATEL, S. J., WANG, W. L., CHAN, W. L., RANGA RAO, K. R., WANG, G., NGEW, X., PATEL, V., BEER, D., KNOX, J. E., MA, N. L., EHRHARDT, C., LIM, S. P., VASUDEVAN, S. G. & KELLER, T. H. 2006a. Peptide inhibitors of dengue virus NS3 protease. Part 2: SAR study of tetrapeptide aldehyde inhibitors. *Bioorg Med Chem Lett,* 16**,** 40-3.

YIN, Z., PATEL, S. J., WANG, W. L., WANG, G., CHAN, W. L., RAO, K. R., ALAM, J., JEYARAJ, D. A., NGEW, X., PATEL, V., BEER, D., LIM, S. P., VASUDEVAN, S. G. & KELLER, T. H. 2006b. Peptide inhibitors of Dengue virus NS3 protease. Part 1: Warhead. *Bioorg Med Chem Lett,* 16**,** 36-9.

YUN, C. H., MENGWASSER, K. E., TOMS, A. V., WOO, M. S., GREULICH, H., WONG, K. K., MEYERSON, M. & ECK, M. J. 2008. The T790M mutation in EGFR kinase causes drug resistance by increasing the affinity for ATP. *Proc Natl Acad Sci U S A,* 105**,** 2070-5.

YUSOF, R., CLUM, S., WETZEL, M., MURTHY, H. M. & PADMANABHAN, R. 2000. Purified NS2B/NS3 serine protease of dengue virus type 2 exhibits cofactor NS2B dependence for cleavage of substrates with dibasic amino acids in vitro. *J Biol Chem,* 275**,** 9963-9.

ZBYSZEK OTWINOWSKI, W. M. 1997. Processing of X-ray diffraction data collected in oscillation mode. *Method Enzymol* 276**,** 307-326.

ZHANG, L. & PADMANABHAN, R. 1993. Role of protein conformation in the processing of dengue virus type 2 nonstructural polyprotein precursor. *Gene,* 129**,** 197-205.

ZHANG, X., GE, P., YU, X., BRANNAN, J. M., BI, G., ZHANG, Q., SCHEIN, S. & ZHOU, Z. H. 2013. Cryo-EM structure of the mature dengue virus at 3.5-A resolution. *Nat Struct Mol Biol,* 20**,** 105-10.

ZHANG, Y., ZHANG, W., OGATA, S., CLEMENTS, D., STRAUSS, J. H., BAKER, T. S., KUHN, R. J. & ROSSMANN, M. G. 2004. Conformational changes of the flavivirus E glycoprotein. *Structure,* 12**,** 1607-18.

ZHANG, Y. M., IMAMICHI, H., IMAMICHI, T., LANE, H. C., FALLOON, J., VASUDEVACHARI, M. B. & SALZMAN, N. P. 1997. Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its Gag substrate cleavage sites. *J Virol,* 71**,** 6662-70.