

eScholarship@UMassChan

Structure and Dynamics of Viral Substrate Recognition and Drug Resistance: A Dissertation

Item Type	Doctoral Dissertation
Authors	Ozen, Aysegul
DOI	10.13028/M2JK5V
Publisher	University of Massachusetts Medical School
Rights	Copyright is held by the author, with all rights reserved.
Download date	2026-05-19 04:33:09
Link to Item	https://hdl.handle.net/20.500.14038/32031

STRUCTURE AND DYNAMICS OF VIRAL SUBSTRATE
RECOGNITION AND DRUG RESISTANCE

A Dissertation Presented

By

AYŞEGÜL ÖZEN

Submitted to the Faculty of the
University of Massachusetts Graduate School of Biomedical Sciences, Worcester
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

MAY 29, 2013

BIOCHEMISTRY & MOLECULAR PHARMACOLOGY

STRUCTURE AND DYNAMICS OF VIRAL SUBSTRATE
RECOGNITION AND DRUG RESISTANCE

A Dissertation Presented

By

AYŞEGÜL ÖZEN

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

CELIA A. SCHIFFER, PH.D., Thesis Advisor

C. ROBERT MATTHEWS, PH.D., Member of Committee

FRANCESCA MASSI, PH.D., Member of Committee

KONSTANTIN ZELDOVICH, PH.D., Member of Committee

WOODY SHERMAN, PH.D., Member of Committee

AMY ANDERSON, PH.D., Member of Committee

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

MELISSA J. MOORE, PH.D., Chair of Committee

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

ANTHONY CARRUTHERS, PH.D.,
Dean of the Graduate School of Biomedical Sciences
BIOCHEMISTRY & MOLECULAR PHARMACOLOGY

MAY 29, 2013

Table of Contents

List of Tables	vii
List of Figures	viii
Nomenclature	xiii
List of Third Party Copyrighted Material	xv
Dedication	xv
Acknowledgements	xvi
Abstract	xx
Preface	xxii
I Introduction	1
1.1 Human Immunodeficiency Virus	1
1.2 Structure and Function of HIV-1 Protease	2
1.3 HIV-1 Protease Inhibitors as Antivirals	10
1.4 Viral Resistance to HIV-1 Protease Inhibitors	13
1.5 Substrate Envelope in Drug Design	14
1.6 Generality of the Substrate Envelope Hypothesis	16
1.7 Another Quickly Evolving Virus: Hepatitis C	17
1.8 NS3/4A Protease as an Antiviral Target	18
1.9 Protein Dynamics	25
1.9.1 Protein Dynamics is Key to Molecular Recognition	25
1.9.2 Change in Dynamics as a Drug Resistance Mechanism	28
1.9.3 Protein Dynamics is Often Neglected in Drug Design	29
1.10 Scope of Thesis	31

II Dynamics of Preferential Substrate Recognition in HIV-1 Protease: Redefining the Substrate Envelope	33
2.1 Abstract	34
2.2 Introduction	35
2.3 Results	38
2.3.1 Molecular Interactions Between Protease and Substrates: Analysis of vdW Contacts	38
2.3.2 Molecular Interactions Between the Protease and Substrates: Analysis of Hydrogen Bonds	51
2.3.3 Analysis of Atomic Fluctuations	56
2.3.4 Dynamics of the Substrate Envelope	62
2.4 Discussion	83
2.5 Methods	88
2.5.1 Protease-Substrate Complex Structures	88
2.5.2 MD Simulations	89
2.5.3 Modeling of the Dynamic Substrate Envelope	90
2.5.4 Estimation of the vdW Potential	94
2.5.5 Evaluation of Hydrogen Bonding	95
2.5.6 Fluctuation Dynamics	95
2.6 Supporting Information	96
III HIV-1 Protease and Substrate Co-evolution Validates the Substrate Envelope as the Substrate Recognition Pattern	100
3.1 Abstract	101
3.2 Introduction	102
3.3 Results and Discussion	105
3.3.1 Fit within the Dynamic Substrate Envelope	105
3.3.2 vdW Interactions Between the Protease and Substrates	113
3.3.3 Hydrogen Bonds between the Protease and Substrates	121
3.3.4 Analysis of Atomic Fluctuations	129
3.4 Conclusions	136
3.5 Methods	140
3.5.1 Nomenclature	140
3.5.2 Protease-Substrate Complex Structures	141
3.5.3 MD simulations	143
3.5.4 Fit within the Substrate Envelope	144
3.5.5 Estimation of vdW Potential	145
3.5.6 Evaluation of Hydrogen Bonding	146

3.5.7	Fluctuation Dynamics	147
-------	--------------------------------	-----

**IV Dynamic Substrate Envelope in Hepatitis C Virus NS3/4A Protease:
Predicting Resistance and Guiding the Design of Novel Inhibitors 148**

4.1	Abstract	149
4.2	Introduction	150
4.3	Results and Discussion	154
4.3.1	Substrate Shape and the Dynamic Substrate Envelope . . .	158
4.3.2	Inhibitor Shape	167
4.3.3	Molecular Flexibility and Atomic Fluctuations - Substrates	174
4.3.4	Molecular Flexibility and Atomic Fluctuations - Inhibitors	182
4.3.5	vdW Contacts at the Binding Surface	187
4.3.6	Electrostatic Network	194
4.4	Conclusion	208
4.5	Methods	211
4.5.1	Protease-Substrate/Inhibitor Complex Structures	211
4.5.2	Structure Preparation	212
4.5.3	MD simulations	214
4.5.4	Quantitative Definition of the Dynamic Substrate Envelope	215
4.5.5	Visualizing the Dynamic Substrate Envelope	216
4.5.6	Estimating the vdW Contact Potential	216
4.5.7	Hydrogen Bonds and Salt bridges	217
4.5.8	Root-Mean-Squared Fluctuations and Time-Delayed Cor- relations	218

**V Structural Insights into the Drug Resistance Conferred by R155K/V36M
in HCV NS3/4A Protease 219**

5.1	Abstract	220
5.2	Introduction	221
5.3	Results and Discussion	226
5.3.1	Drug Susceptibility and Enzyme Inhibition Assays	226
5.3.2	Crystallographic Binding Mode of Inhibitors	229
5.3.3	Intermolecular Hydrogen Bonds	237
5.3.4	Structural Changes in the Protease Backbone	241
5.3.5	Intramolecular Salt Bridges	251
5.3.6	vdW Interactions at the Binding Surface	254
5.3.7	Inter-residue Distances in the MD Simulations	263

5.3.8	Correlation of Equilibrium Fluctuations from Gaussian Network Model	277
5.4	Conclusion	280
5.5	Materials and Methods	282
5.5.1	Mutagenesis and Gene Information	282
5.5.2	Expression and Purification	282
5.5.3	Crystallization	283
5.5.4	Data Collection and Structure Solution	284
5.5.5	Drug Susceptibility and Enzyme Inhibition	285
5.5.6	Double-Difference Maps	285
5.5.7	Hydrogen Bonds	286
5.5.8	Salt Bridges	286
5.5.9	van der Waals Interactions	287
5.5.10	MD simulations	287
5.5.11	Molecular Modeling	287
5.5.12	Gaussian Network Model	288
VI	Discussion	289
6.1	Robust Drug Design: Hitting Multiple Targets at a Time	290
6.2	Drug Target's Function and Protein Dynamics	291
6.3	Dynamic Substrate Envelope: Integrating Dynamics into Drug Discovery while Avoiding Resistance	297
6.4	Predicting Substrate Co-evolution using the Dynamic Substrate Envelope	302
6.5	Concluding Remarks	303
	Appendices	306
A	Assessing van der Waals Contacts	307
1.1	A Simplified Lennard-Jones Potential	307
1.2	Forcefield	312
1.3	Limitations	313
	References	314

List of Tables

1.1	Cleavage site sequences for HIV-1 and HCV NS3/4A proteases . . .	7
1.2	Reported sites of drug resistance mutation in HCV NS3/4A	23
2.1	Hydrogen bonds within the substrates of HIV-1 protease	56
3.1	Effect of drug resistance on intermolecular hydrogen bonding . . .	126
3.2	Effect of drug resistance on intramolecular substrate hydrogen bonding	127
3.3	HIV-1 protease-substrate complexes investigated for the molecu- lar basis of substrate co-evolution	142
4.1	The cleavage site products and macrocyclic small-molecule in- hibitors of HCV NS3/4A Protease	155
4.2	The intermolecular hydrogen bonds in MD simulations	196
5.1	Drug susceptibilities against wild-type and resistant HCV clones and inhibitory activities against NS3/4A proteases	228
5.2	X-ray data collection and crystallographic refinement statistics . .	230
5.3	Protease-inhibitor hydrogen bonds	240
5.4	Salt bridges along the binding site of resistant NS3/4A variants . .	253

List of Figures

1.1	HIV-1 and HCV NS3/4A protease structures	6
1.2	The substrate envelope hypothesis	9
1.3	FDA-approved HIV-1 protease inhibitors	12
1.4	HCV NS3/4A protease inhibitors	22
2.1	Contact potentials between HIV-1 protease and seven substrates are well represented by crystal structures	41
2.2	HIV-1 protease has a similar profile of interaction with non-homologous substrates	44
2.3	Substrates maintain a consensus vdw interaction potential with the protease independent of sequence	48
2.4	Fluctuations in the vdW interaction potential within/among sub- strate residues are negatively correlated, revealing their interde- pendency	50
2.5	Backbone hydrogen bonds are more conserved across varied sub- strates and more stable than the side-chain hydrogen bonds	54
2.6	Interdependency within substrate residues revealed by cross-correlations of mean-square fluctuations between protease and substrate residues	59

2.7	Unique intrinsic dynamic coupling revealed by cross-correlations of the substrate mean square fluctuations	61
2.8	The static and dynamic substrate envelopes share the same overall trend in V_{out} , i.e., the volume lying outside the envelope	65
2.9	Distributions of V_{out} , V_{in} , and V_{tot} values throughout the MD simulations are unimodal for each substrate	67
2.10	Substrate size appears to determine how well the substrate fits within the substrate envelope, except for CA-p2, NC-p1, and p1-p6	70
2.11	The overall dynamics of the substrate in the active site is dominated by side-chain fluctuations	72
2.12	Intrinsic flexibility appears to play an important role in substrate fit within the substrate envelope for substrates MA-CA, CA-p2, NC-p1, and p1-p6	74
2.13	The dynamic substrate envelope is a probability distribution	76
2.14	Static and dynamic shapes of HIV-1 protease substrates	78
2.15	Residue-based substrate V_{out} values	82
2.16	Effect of structural alignment on the substrate envelope	99
3.1	Protease-substrate co-evolution preserves the dynamic substrate envelope	108
3.2	The compensatory mutation LP1'F in the p1-p6 cleavage site preserves the substrate envelope in drug-resistant D30N/N88D protease variant	112

3.3	Residue-based view of the effect of drug resistance on protease-substrate vdW interactions	117
3.4	The compensatory mutation in the p1-p6 cleavage site, LP1'F, causes major structural rearrangements reducing vdW interaction potential of the invariant Asp P2	120
3.5	Substrate backbone hydrogen bonds are robust against the local conformational rearrangements due to drug resistance mutations .	124
3.6	Effect of drug resistance on intrinsic dynamic cooperativity of NC-p1 substrate variants	132
3.7	Effect of drug resistance on intrinsic dynamic cooperativity of p1-p6 substrate variants	135
4.1	HCV protease substrate and inhibitor structures	157
4.2	Substrate flexibility reduces the consensus volume	160
4.3	P4-P1 substrate region is less mobile compared to P6 and P5 residues, contributing to the highly conserved consensus volume	163
4.4	The HCV NS3/4A dynamic substrate envelope is more conserved in the P4 to P1 region	166
4.5	Macrocyclic inhibitors are more rigid than the substrates	170
4.6	Macrocyclic inhibitors do not fully occupy the dynamic substrate envelope	173
4.7	Compound flexibility is dictated by a combination of structural features	177

4.8	Correlation of atomic fluctuations within P4-P1 region survives longer compared to those across P6-P5 and P4-P1	181
4.9	Inhibitors are highly cooperative in atomic fluctuations.	186
4.10	Differences between the binding surfaces of inhibitors versus substrates provide opportunity for resistance mutations	189
4.11	Differences between the binding surfaces of inhibitors versus substrates provide opportunity for resistance mutations	191
4.12	P6 Asp/Glu of substrates participates in a dynamic salt bridge network with the protease residues R119, R123, R161 and K165 . . .	200
4.13	R155 is shared by D81 and D168 in a salt bridge network along the binding surface	204
4.14	The R155-D168 salt bridge is stabilized by D81-H57 salt bridge in inhibitor complexes	207
5.1	Interdependency within substrate residues revealed by cross-correlations of mean-square fluctuations between protease and substrate residues	224
5.2	Binding modes of inhibitors in crystal structures	234
5.3	Conformational flexibility of danoprevir's P2 moiety	236
5.4	Hydrogen bonds between HCV NS3/4A and inhibitors; telaprevir, boceprevir, and danoprevir	239
5.5	Double-difference maps	243
5.6	Changes in the atomic fluctuations upon inhibitor binding	246
5.7	Changes in the atomic fluctuations upon resistance mutations . . .	250
5.8	vdW interactions at the NS3/4A binding surface	256

5.9	vdW interactions of the inhibitor moieties	258
5.10	Crystallographic vdW interactions at the NS3/4A binding surface .	262
5.11	C_{α} - C_{α} distance between V/M36 and F43	266
5.12	C_{α} - C_{α} distance between F43 and K136, R/K155, A156, A157 . .	268
5.13	C_{α} - C_{α} distance between A39 and R161	270
5.14	C_{α} - C_{α} distance within the catalytic triad	272
5.15	C_{α} - C_{α} distance between H57 and K136, A156, and A157	274
5.16	C_{α} - C_{α} distance between NS3 and the cofactor NS4A	276
5.17	The correlation of equilibrium fluctuations in the slowest modes from GNM	279
6.1	Guiding drug design using the dynamic substrate envelope	294
6.2	Incorporating drug resistance and protein dynamics into drug design	300

List of Third Party Copyrighted Material

- Chapter I: Adapted from *Viruses*, 2 (11), A. Ali., R. M. Bandaranayake, Y. Cai, N. M. King, M. Kolli, S. Mittal, J. F. Murzycki, M. N. L. Nalam, E. A. Nalivaika, A. Özen, M. M. Prabu-Jeyabalan, K. Thayer, C. A. Schiffer. Molecular Basis for Drug Resistance in HIV-1 Protease. 2509, 2010, permission is not required.
- Figure 1.2A: Adapted from *Chemistry & Biology*, 11, N. M. King, M. Prabu-Jeyabalan, E. A. Nalivaika, C. A. Schiffer, Combating susceptibility to drug resistance: Lessons from HIV-1 protease, 1333, ©(2004), permission requested from Elsevier.
- Figure 1.2B: Adapted from *Proc Natl Acad Sci USA*, 107 (49), K. P. Romano, A. Ali, W. E. Royer, C. A. Schiffer. Drug resistance against HCV NS3/4A inhibitors is defined by the balance of substrate recognition versus inhibitor binding, 20986, (2010), permission is not required for noncom-

mercial and educational purposes.

- Chapter II: Adapted from *Journal of Molecular Biology*, 410, A. Özen, T. Haliloğlu, C. A. Schiffer, Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope, 726, ©(2011), with permission from Elsevier (License number: 3147820834265).
- Chapter III: Adapted with permission from *Journal of Chemical Theory and Computation*, 8 (2), A. Özen, T. Haliloğlu, C. A. Schiffer, HIV-1 Protease and Substrate Coevolution Validates the Substrate Envelope As the Substrate Recognition Pattern, 703. ©(2012) American Chemical Society.

*To the my parents, Ayşe and Tuncel, for their unconditional love and
encouragement and strong support throughout my life...*

Canım anneciğim ve babacığım...

*Sizin bana inancınız ve sonsuz desteğiniz olmadan, bu dereceyi elde etmem asla
mümkün olamazdı. Bana verdiğiniz, beni ben yapan herşey için
çok teşekkür ederim.*

Acknowledgements

“I myself knew very little biology, except in a rather general way, till I was over thirty, since my first degree was in physics. It took me a little time to adjust to the rather different way of thinking necessary in biology. It was almost as if one had to be born again. Yet such a transition is not as difficult as all that and is certainly well worth the effort.”

Francis Crick, 1988

Biology in the classic sense indeed requires a different way of thinking. However, that physics, chemistry, and biology is a continuum, not distinct fields was the first and the most important lesson I learned from my greatest mentor Dr. Celia Schiffer. Science is inherently interdisciplinary. Without contribution from fields such as chemistry, physics, mathematics, and many more, we are unlikely to understand how life works, why it stops working properly in the disease state, and what we can do to cure diseases. I owe great thanks to all the faculty members, students and postdocs at the department of biochemistry and molecular pharmacology. Being exposed to various research fields and countless experimental ap-

proaches in the scientifically diverse, intellectually rich, and highly collaborative environment at UMass Medical School has equipped me with a deeper understanding of biological systems and events.

I am deeply grateful to my advisor, Dr. Celia Schiffer. I've been extremely lucky to be a student of Celia's. She has always given me the independence to explore my interests while making sure to guide me at critical times with her useful advices. Growing as a young female scientist, she's been an excellent role model for me. Carefully observing her, I have learned so much about being a successful scientist, which is not limited to conducting exceptional science in laboratory but includes being proactive and persistent, forming smart hypotheses based on scientific questions that matter, paying attention to details while not getting lost in them, nicely blending multiple disciplines to address relevant questions, initiating and maintaining fruitful collaborations, and maintaining the delicate balance between personal life and work. Celia has always treated her students with care and respect, which has only encouraged me to care more, think deeper, work harder and respect her more. Thank you, Celia.

As very nicely put by Dr. Francis Crick, the initial adjustment to the rather different way of thinking necessary in biology took a little time for me as well since I had a formal training in engineering. Special thanks to Drs. Kendall Knight and Anthony Carruthers for their vision and for providing the opportunities for me to learn thinking in biology. I would not have been able to continue my graduate studies without the one-on-one mentoring sessions from Drs. Craig Peterson and Nick Rhind. In graduate school, I have come to realize that the most important

thing you can ask from a PI is their time and both Craig and Nick were willing to invest time on a student who was not even going to join their laboratories. They are not just excellent in conducting research but also dedicated to training young scientists, which I am grateful for. And I'm thankful to Celia again for her solid support throughout this process.

I sincerely thank all the members of my thesis committee, Drs. Melissa Moore, Robert Matthews, Francesca Massi, Konstantin Zeldovich, Akbar Ali, and Woody Sherman for tracking the progress of my research over the years and providing valuable suggestions. I am especially thankful to Woody for his availability at all times when I needed guidance with my calculations and many fruitful discussions. I thank Dr. Amy Anderson for being an external examiner in my defense committee; her insights into drug resistance added value to my dissertation.

I also thank Dr. Nese Kurt Yilmaz for critically reading my manuscripts and for several insightful conversations. I can only wish she had joined the lab much earlier. Many thanks to Dr. William Royer for being a great teacher and always receiving my crystallography questions with infinite patience. I could have easily said that I taught molecular dynamics myself, had I not learned from Drs. Kelly Thayer, Yufeng Cai, and Brittany Morgan, great thanks for sharing their experience and giving me ideas especially at the beginning of my graduate school years. I must also thank Alan Ritacco for patient technical assistance with software installation, Christine Pruitte not just for helping with conference registrations, room scheduling, travel arrangements, paperwork preparation, but also for the moral support at several occasions when much needed, and Luca Leone for

always receiving me with a smiling face and solving my computer problems in the blink of an eye.

Science is certainly not an individual activity. Solid prior work and effective collaborations are essential for productivity. I would like to thank all the former and current members of the Schiffer lab for the great science and collaborative environment.

Nothing is entirely fun without friends. Many thanks to Arda, Aslı, Brian, Celia, Chris, Ebru, Efsun, Hülya, Kenny, Laura, Lori, Madhavi, Markus, Michelle, Naomi, Özlem, Sezin, Sagar for sharing all the good and the bad moments in life.

Noah Cohen deserves special thanks for making my life better and happier.

I carefully chose my friends but not family; I was simply just lucky. My cousin Nurhan and cousin-in-law Uğur traveled all the way from Turkey to be with me on my graduation day. Being thousands of miles away from family has been tough at times but my sister and brother-in-law, Gülden and Mert, and my lovely niece, Ada, gave me all the joy and comfort from the other side of the ocean. My parents, Ayşe and Tuncel, could not have been more supportive and encouraging to give me the strength, dedication and perseverance I needed to complete my graduate program. My greatest thanks go to my family...

Abstract

Drug resistance is a major problem in quickly evolving diseases, including the human immunodeficiency (HIV) and hepatitis C viral (HCV) infections. The viral proteases (HIV protease and HCV NS3/4A protease) are primary drug targets. At the molecular level, drug resistance reflects a subtle change in the balance of molecular recognition; the drug resistant protease variants are no longer effectively inhibited by the competitive drug molecules but can process the natural substrates with enough efficiency for viral survival. Therefore, the inhibitors that better mimic the natural substrate binding features should result in more robust inhibitors with flat drug resistance profiles. The native substrates adopt a consensus volume when bound to the enzyme, the substrate envelope. The most severe resistance mutations occur at protease residues that are contacted by the inhibitors outside the substrate envelope. To guide the design of robust inhibitors, we investigate the shared and varied properties of substrates with the protein dynamics taken into account to define the dynamic substrate envelope of both viral proteases. The NS3/4A dynamic substrate envelope is compared with inhibitors to detect the structural and dynamic basis of resistance mutation patterns. Com-

parative analyses of substrates and inhibitors result in a solid list of structural and dynamic features of substrates that are not shared by inhibitors. This study can help guiding the development of novel inhibitors by paying attention to the subtle differences between the binding properties of substrates versus inhibitors.

Preface

Sections of Chapter I has previously been published as:

A. Ali, Akbar, R. M. Bandaranayake, Y. Cai, N. M. King, M. Kolli, S. Mittal, J. F. Murzycki, M. N. L. Nalam, E. A. Nalivaika, A. Özen, M. M. Prabu-Jeyabalan, K. Thayer, C. A. Schiffer (2010). Molecular Basis for Drug Resistance in HIV-1 Protease. *Viruses* 2, 11, 2509-2535.

Author contributions: All authors contributed equally to this review article. The author names were listed alphabetically. The text has been modified for this thesis.

Chapter II has previously been published as:

A. Özen, T. Haliloğlu, C. A. Schiffer (2011). Dynamics of preferential substrate recognition in HIV-1 protease: redefining the substrate envelope. *Journal of Molecular Biology* 410, 726-744.

Chapter III has previously been published as

A. Özen, T. Haliloğlu, C. A. Schiffer (2012). HIV-1 protease and substrate coevolution validates the substrate envelope as the substrate recognition pattern. *Journal of Chemical Theory & Computation* 8, 703-714.

Author contributions: For both Chapters II and III, AÖ, TH, and CAS conceived and designed the experiments. AÖ performed the experiments. AÖ, TH, and CAS analyzed the data and wrote the published manuscript.

Chapter IV has been submitted for publication as:

A. Özen, W. Sherman, C. A. Schiffer (2013). Guiding the Design of Novel Inhibitors of NS3/4A Protease by using the Dynamic Substrate Envelope. [*submitted*].

Author contributions: AÖ, WS, and CAS conceived and designed the experiments. AÖ performed the experiments. AÖ, WS, and CAS analyzed the data and wrote the manuscript.

Chapter V has been a collaborative study that is currently under preparation for

publication as:

A. Özen, K. Romano, K. H., Lin, C. Aydın, D. Tavella, C. A. Schiffer (2013).
Structural and dynamic basis for drug resistance conferred by R155K/V36M mutation in HCV NS3/4A protease. [*in preparation*].

Author contributions: AÖ, RKH, and CAS conceived and designed the experiments. AÖ, RKH, KHL, CA, and DT performed the experiments. AÖ, DT, and CAS analyzed the data and AÖ and CAS is currently writing the manuscript.

Chapter I

Introduction

1.1 Human Immunodeficiency Virus

Human Immunodeficiency Virus (HIV) is a lentivirus of the *Retroviridae* family that infects vital cells in the human immune system including CD4+ T cells, macrophages, and dendritic cells. HIV infection causes the acquired immunodeficiency syndrome (AIDS). Without effective treatment, the infection can result in the collapse of the immune system. UNAIDS reports, globally, 34 million people were living with HIV at the end of 2011 with 2.5 million new infections and 1.7 million people died from AIDS-related causes. [1] Two types of HIV and several groups within each type exist in the genus of lentivirus with distinct patterns of spread and progression to AIDS. [2] HIV Type 1 (HIV-1) is responsible for the AIDS pandemic. As the worldwide epidemic finishes its third decade, a cure for HIV-1 still eludes the biomedical community. In the absence of a cure for

HIV-1 pathogenesis, suppressing viral replication and maintaining it at low to undetectable levels have become critical goals in the field of HIV-1 research. To this end, highly active antiretroviral therapy (HAART) has become a successful strategy in providing long, quality life for infected individuals [3,4] and is the global standard of care for AIDS treatment. There are many steps in the viral life cycle for potential pharmacological intervention. The US Food and Drug Administration (FDA) has approved more than 30 drugs that target various stages of viral replication cycle including fusion and entry, reverse transcription, integration and proteolytic processing of viral polyproteins. Maturation requires a viral protease to process the virally encoded polyproteins to yield infectious virions. Due to this essential function in viral maturation, HIV-1 protease is an antiviral drug target.

1.2 Structure and Function of HIV-1 Protease

HIV-1 protease is an aspartyl protease that processes the viral polyproteins, allowing the virus to mature and become infectious. The enzyme is a symmetrical homodimer of 99 amino acids each (Figure 1.1). Substrate binding disrupts this symmetry, allowing the recognition of asymmetric substrates by this otherwise symmetric protease. Each monomer contains a flap comprising two anti-parallel β -strands connected by a β -turn and situated on top of the catalytic site. The conformation of the highly flexible flap region differs significantly in the apo and substrate-bound forms. The dimer is stabilized by four antiparallel β -strands, two from each subunit, which form an interdigitated β -sheet. Each subunit also con-

tributes a catalytic aspartate residue, located at the dimer interface. The catalytic triad in the active site (Asp25-Thr26-Gly27) is responsible for the substrate hydrolysis.

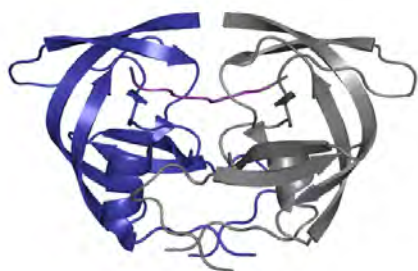
In total, twelve proteolytic reactions are required to process Gag, Gag-Pro-Pol, and Nef. These reactions occur late in the viral life cycle, during virion assembly and maturation at the cell surface. The cleavage is a highly regulated and ordered stepwise process. The scissile bond between Nucleocapsid (NC) and p2 (MA-CA-p2↓NC-p1-p6) is cleaved first, followed by the hydrolysis of Matrix (MA) and Capsid-p2 (CA-p2)(MA↓CA-p2). Subsequently, p6 is separated from NC-p1 (NC-p1↓p6). Finally, the spacer peptides p1 and p2 are removed from NC (NC↓p1) and Ca-p2 (CA↓p2), respectively. NC-p1 is the rate-limiting cleavage reaction. [5]

Temporal regulation is achieved with a unique intrinsic proteolytic rate at each site and is critical for precise processing underlying the importance of substrate specificity. Although hydrophobic residues are favored at P1/P1' positions, the cleavage sites are non-homologous in sequence and asymmetric in size and charge (Table 1.1). This asymmetry prompts the question as to how a symmetric protease could recognize and cleave an asymmetric substrate. The amino acid sequence by itself is not the specificity determinant for substrate recognition. On the other hand, the symmetry observed in the apo enzyme juxtaposed with the asymmetric nature of the substrates [6] and all these asymmetric substrates have similar binding modes in an extended conformation. [7] Structural studies have shown that the various cleavage site peptides adopt a conserved shape when bound to HIV-1 pro-

tease. [7] This conserved shape was initially characterized by visualizing the van der Waals volume shared by four of the six substrates, which was later on referred to the substrate envelope (Figure 1.2). The substrate envelope was hypothesized as the recognition motif for HIV-1 protease.

Figure 1.1: HIV-1 and HCV NS3/4A protease structures. (A) HIV-1 protease has two monomers (blue and gray) and displayed here bound to a decameric peptide (magenta) corresponding to the matrix-capsid cleavage site on Gag bound to the active site. (PDB ID: 1KJ4) (B) The full-length HCV NS3/4A has two domains: helicase (light pink) and protease (teal). NS4A, the cofactor for the protease, is shown in blue. C-terminal 6 residues of the helicase domain, the self-cleavage product, is observed at the binding site of the protease domain. (PDB ID: 1CU1)

(A) HIV-1 Protease



(B) HCV NS3/4A Protease/Helicase

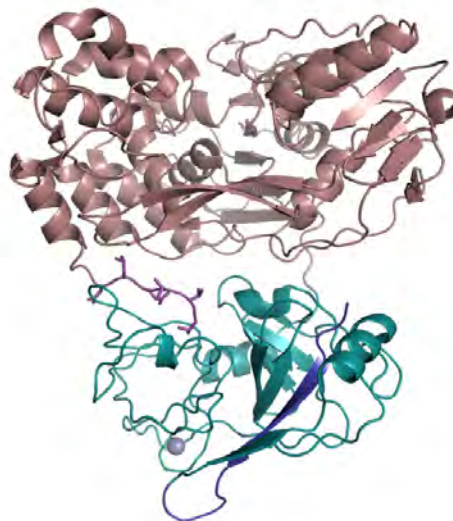


Figure 1.1

Table 1.1: Cleavage site sequences for HIV-1 and HCV NS3/4A proteases

HIV-1 Protease Substrates

Substrate ^a	P4	P3	P2	P1	*	P1'	P2'	P3'	P4'
Matrix-Capsid	S	Q	N	F	*	P	I	V	Q
Capsid-p2	A	R	V	L	*	A	E	A	M
p2-Nucleocapsid	A	T	I	M	*	M	Q	R	G
Nucleocapsid-p1	R	Q	A	N	*	F	L	G	K
p1-p6	P	G	N	F	*	L	Q	S	K
Nucleocapsid-Transframe	R	Q	A	N	*	F	L	R	E
Transframe-p6 ^{pol}	D	L	A	F	*	L	Q	G	K
p6 ^{pol} -Protease	S	F	N	F	*	P	Q	I	T
Autoproteolysis	Q	I	T	L	*	W	Q	R	P
PR-Rev. Transcriptase	T	L	N	F	*	P	I	S	P
Rev. Transcriptase-RNase H	A	E	T	F	*	Y	V	D	G
RNase H-Integrase	R	R	I	L	*	F	L	D	G

HCV NS3/4A Protease Substrates

Substrate ^a	P6	P5	P4	P3	P2	P1	*	P1'	P2'	P3'	P4'
3-4A	D	L	E	V	V	T	*	S	T	W	V
4A-4B	D	E	M	E	E	C	*	S	Q	W	L
4B-5A	E	C	T	T	P	C	*	S	G	S	W
5A-5B	E	D	V	V	C	C	*	S	M	S	Y

Amino acids are colored according to their physicochemical properties.

RK (purple), DE (red), ILVA (yellow), H (light blue), CM (green)

FYW (dark blue), NQST (pink), PG (gray).

^a Protease cleaves the scissile bond P1/P1'.

^b Host cellular proteins.

Figure 1.2: The substrate envelope hypothesis. (A) HIV-1 protease substrate and inhibitor envelopes are colored blue and red, respectively. The two envelopes were superimposed to highlight the regions where inhibitors protrude beyond the substrate to make more extensive contacts with some protease residues. These protease residues corresponded to the previously known sites of drug resistance. Figure modified from King et. al., 2004. [8] (B) HCV NS3/4A protease substrate envelope (blue) and a small-molecule inhibitor of NS3/4A protease, Danoprevir, are shown in comparison along with the binding site residues. Figure modified from Romano et. al., 2010. [9]

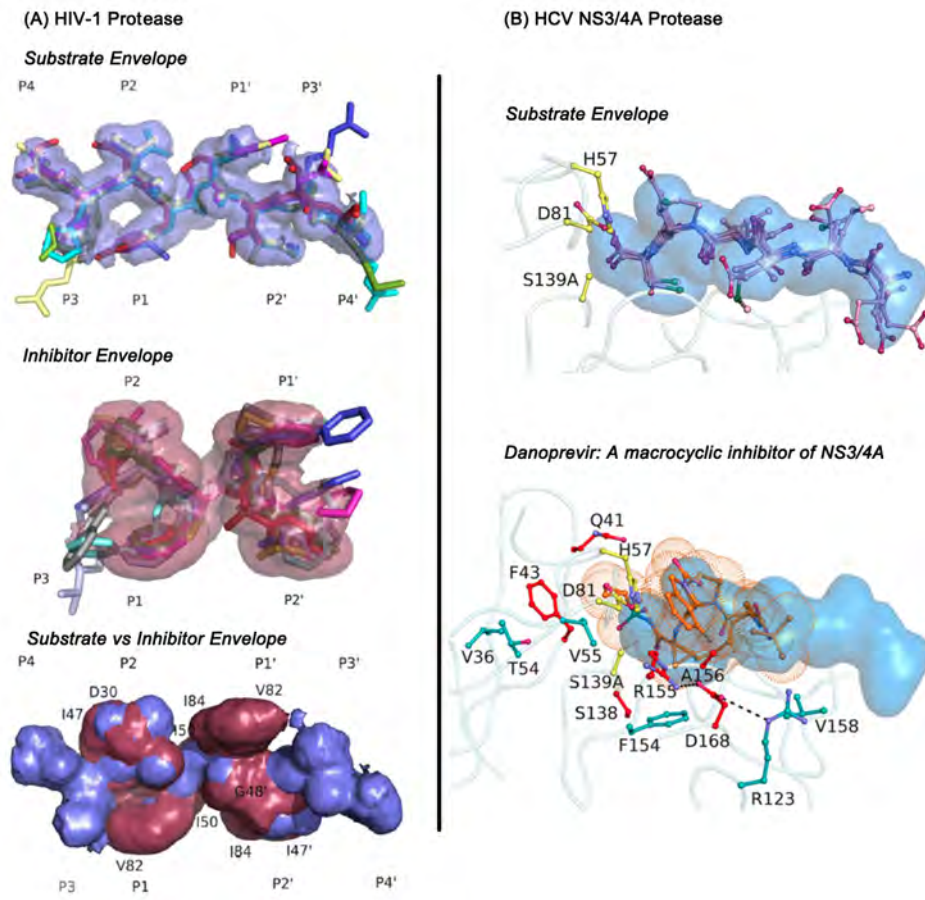


Figure 1.2

1.3 HIV-1 Protease Inhibitors as Antivirals

HIV-1 protease is a major drug target because of its key function in viral assembly and maturation. [10] FDA has so far approved nine protease inhibitors (PIs) for clinical use: Saquinavir (SQV) [11], Indinavir (IDV) [12], Ritonavir (RTV) [13], Nelfinavir (NFV) [14], Amprenavir (APV) [15], Lopinavir (LPV) [16], Atazanavir (ATV) [17], Tipranavir (TPV) [18], and Darunavir (DRV) [19–21] (Figure 1.3). All these PIs are competitive active site inhibitors and, except for TPV, they are all peptidomimetics. They bind to the protease with the flaps of the enzyme tightly closed over the active site, thus mimicking the transition state between substrate binding and cleavage reaction and thereby effectively inactivating the enzyme. PIs generally have large, hydrophobic moieties that interact with the mainly hydrophobic S2-S2' pockets in the active site. Although they are chemically different relatively low molecular weight compounds, their three dimensional shape and electrostatic character are fairly similar. Majority of these PIs occupy a similar volume, the inhibitor envelope, mostly contacting the same residues at the active site (Figure 1.2). [8]

Figure 1.3: FDA-approved HIV-1 Protease Inhibitors. They are all chemically different; yet share a fairly similar three-dimensional shape and electrostatic character.

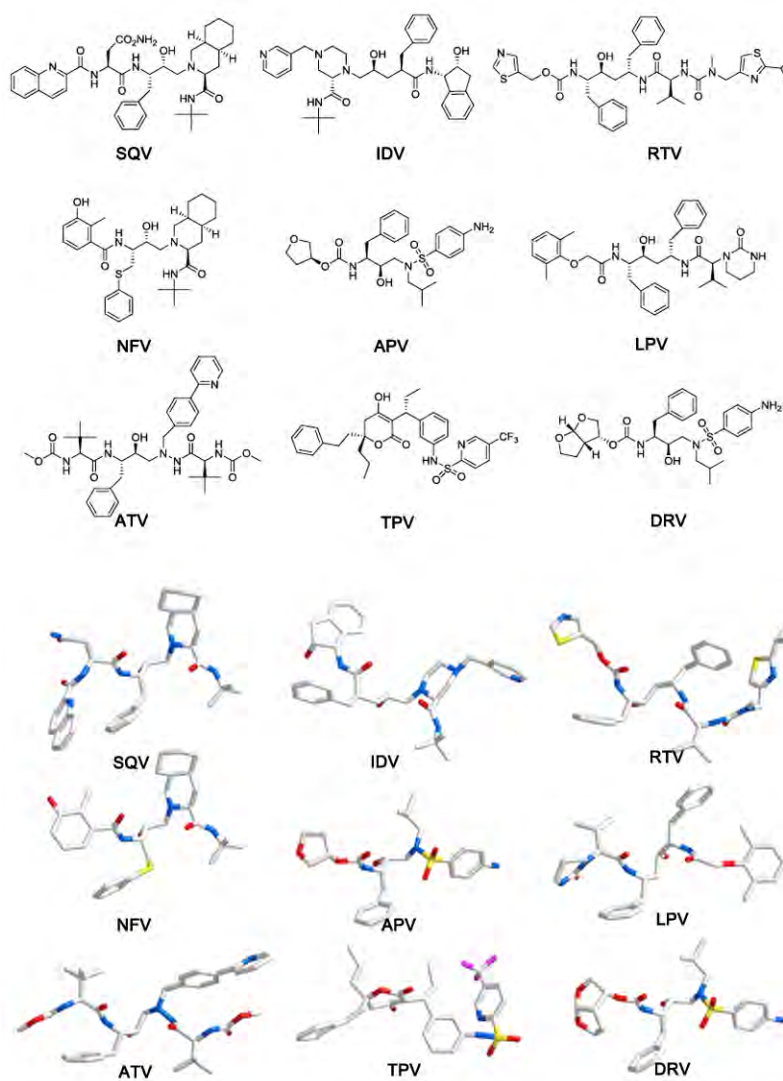


Figure 1.3

1.4 Viral Resistance to HIV-1 Protease Inhibitors

The quality and life expectancy of HIV infected patients has improved since the introduction of HAART. However, high frequency of random nucleotide misincorporation by the error-prone reverse transcriptase (about three mutations per virion per round of replication) and a huge reservoir of replicating virus (10^{10} infected cells in an average patient) diversify the viral population. [22] The selective pressure of therapy, especially combined with low drug adherence due to toxicity and high pill burden, facilitates the emergence of drug resistance variants. [23] PIs are essential components of HAART. [24, 25] Consequently, drug resistance to PIs has become a major issue in the failure of HAART. [23]

Structural studies showed that drug resistance occurs where the inhibitor atoms protrude beyond the substrate envelope and contact protease residues. [8] Thus, mutations at these sites would specifically impact inhibitor binding while substrate recognition and cleavage remains relatively unaffected. The fact that most drug resistance mutation sites do not contact the substrates led to the substrate envelope hypothesis: Inhibitors that fit well within the substrate envelope would be less susceptible to drug resistance, because a mutation that affects inhibitor binding would simultaneously impact the recognition and processing of the majority of the substrates. [8]

More than half of the residues within the protease mutate in different combinations and lead to drug resistance. [26,27] Following the accumulation of resistance mutations within the protease, mutations also develop within the substrate cleav-

age sites, including NC-p1 and p1-p6, altering the susceptibility to various PIs. [28–32] Primary PI mutations, especially in the active site, reduce both protease catalytic efficiency and viral replicative capacity. [33–36] Nevertheless, evolution of mutations within the cleavage sites leads to improved viral fitness compared to the viral variants carrying protease resistance mutations only. [28, 30, 37, 38] Substrates exhibit subtle differences in their binding modes. Some substrates protrude beyond the substrate envelope at regions where they contact the sites of drug resistance in the protease. The primary mutation sites, contacted by a substrate outside the substrate envelope, can lead to impaired recognition and cleavage of that substrate. [32, 39, 40] Therefore, that particular substrate would be more susceptible to the drug resistance mutations in the protease. [40] In conclusion, co-evolved mutations within the cleavage sites play an important role in the development of resistance and affect virological response during therapy. The substrate envelope hypothesis not only explains specificity of the substrates but also the development of resistance to various PIs and substrate co-evolution. [8, 39, 40]

1.5 Substrate Envelope in Drug Design

The substrate envelope hypothesis provides a framework to designing robust HIV-1 PIs that avoid resistance. A survey of five approved drugs indicates that the volume of the inhibitors outside the substrate envelope correlates with the loss of affinity to mutant proteases. [41] Meanwhile, DRV, the most potent of the currently prescribed inhibitors, fits well within the substrate envelope although not

designed using the substrate envelope as a constraint. [42, 43] The ability of the substrate envelope to correlate with resistance mutations prompted the use of substrate envelope constraints in the design of new inhibitors. [44–48] According to the substrate envelope hypothesis, the optimum strategy to minimize resistance is to design inhibitors that fit within the substrate envelope. To test the substrate envelope hypothesis, various groups took different approaches and designed new HIV-1 PIs on the hydroxyethylamine scaffold. Two computational methods incorporated structural constraints of the substrate envelope as an *a priori* consideration during the design stage of the inhibitors while the third method employed a structure activity relationship (SAR) that does not include the substrate envelope constraint explicitly. The first computational design, based on optimized docking, resulted in two good candidates exhibiting flat affinity profiles against multi-drug resistant mutants, although the binding affinity of these candidates were in the nM range. [48] The second computational design systematically explored the combinatorial space for three constituent R groups on the same scaffold in two rounds of design, synthesis, testing, and retrospective structural analysis. The second round resulted in low nanomolar-picomolar range compounds, the majority of which have flatter resistance profiles against resistant variants. [45] As a negative control, the inhibitors designed with the SAR approach resulted in picomolar inhibitors, however they were significantly less potent against the resistant variants. These studies validated the use of the substrate envelope as a constraint during the development of HIV-1 PIs with low susceptibility to resistance and yielded several leads for potential new drugs. [46]

1.6 Generality of the Substrate Envelope Hypothesis

Application of the substrate envelope hypothesis to other quickly evolving targets is beginning to emerge. In a recent study, applicability of the substrate envelope hypothesis has been tested for five prospective drug targets from a diverse set of diseases: Abl kinase, chitinase, thymidylate synthase, dihydrofolate reductase, and neuraminidase. [49] The volume of inhibitors protruding beyond the native substrate envelope correlated with average mutation sensitivity. This suggests that inhibitor design would benefit from a similar reverse engineering strategy for these enzymes. The substrate envelope has also been applied in the development of tenofovir, a reverse transcriptase inhibitor. [50] Similar to the case of HIV-1 protease, the drugs AZT and 3TC protrude beyond the consensus volume, creating an opportunity for the reverse transcriptase to develop resistant mutations. The newly designed inhibitor, lacking such protrusions, is expected to evade resistance mutations as an improvement over its predecessors. Thus, the substrate envelope hypothesis appears to be a valid general strategy for avoiding drug resistance. Recently, high resolution crystal structures showed that the substrate cleavage products of Hepatitis C viral protease NS3/4A adopt a consensus volume at the binding site and the most severe resistance mutations correspond to the sites that are contacted by inhibitors outside the substrate envelope. [9] NS3/4A emerges as a good candidate to target with the substrate envelope approach.

1.7 Another Quickly Evolving Virus: Hepatitis C

Hepatitis C is an infectious liver disease with significant global impact, which is caused by an RNA virus of *Flaviviridae* family. The Hepatitis C virus (HCV) infection can lead to liver cirrhosis and hepatocellular sarcoma and is the most common reason for liver transplants in the United States. The World Health Organization estimates 150 million people worldwide are infected with HCV and 3-4 million new infections coming up every year with more than 350,000 cases of death from HCV-related liver diseases. [51] HCV is genetically highly diverse. So far, six major HCV genotypes and several subtypes within each genotype have been identified. [52] The high viral replication rate combined with the error-prone RNA-dependent RNA polymerase causes large inter-patient genetic diversity as well as viral diversity within a single infected individual. [53, 54]

The genetic heterogeneity of the virus across and within patients has greatly challenged the development of robust direct-acting antiviral agents (DAAs) that retain efficacy against multiple genotypes and drug-resistant variants of these genotypes since the discovery of HCV in 1989 as the cause of the Hepatitis C. [55, 56] Until 2011, the standard of care for HCV was weekly injections of pegylated interferon α combined with ribavirin (Peg-IFN/RBV), which can result in undetectable levels of HCV in 70-80% of people with genotypes 2 and 3 but only 40-50% of people with genotype 1. [51] Genotype 1, the most difficult genotype to treat, is also the most common form of HCV in the US accounting for about 75-80% of the cases. [57, 58] In addition, the severe side effect pro-

file of this combination therapy amplifies the need to develop more effective and better-tolerated DAAs.

1.8 NS3/4A Protease as an Antiviral Target

Among the drug targets in HCV is the non-structural protein 3 (NS3), which is a 631 amino acid bifunctional protein, with a serine protease domain located in the N-terminal one-third and an NTPase/RNA helicase domain in the C-terminal two-third. The reason for the protease and helicase domains to be physically linked is not fully known. Although their interplay has been reported [59–62], both domains fold independently and are active in the absence of the other [59, 61–64]. The most efficient proteolytic activity of NS3 requires a cofactor NS4A, a 54-amino acid peptide that is tightly associated with the protease. [51] The HCV genome encodes a single polyprotein of ≈ 3000 acids, which is processed by a series of host and viral proteases into at least 10 structural and nonstructural proteins. The viral NS3/4A hydrolyzes the polyprotein precursor at four known cleavage sites (3-4A, 4A-4B, 4B-5A, 5A-5B) yielding non-structural proteins essential for viral maturation. The first proteolytic event occurs at 3-4A junction in *cis* as a unimolecular reaction, while processing of the remaining junctions 4A-4B, 4B-5A, and 5A-5B occurs bimolecularly in *trans*. [65] These cleavage sites are non-homologous except for an Asp/Glu at P6, Cys/Thr at P1 and Ser/Ala at P1' (Table 1.1). NS3/4A also confounds the innate immune response to viral infection by cleaving the human cellular targets TRIF and MAVS and to block toll-like re-

ceptor 3 signaling and RIG-I signaling, respectively. [66–69] Cleavage of another cellular target, TC-PTP, at two separate sites enhances EGF signaling and basal Akt activity. [70] Very recently, DDB1, a core subunit of the Cul4-based ubiquitin ligase complex, was reported to play a critical role in HCV replication and get cleaved by NS3/4A. [71] Thus, in addition to blocking the viral maturation, effective inhibition of the proteolytic activity of the NS3/4A may also exert indirect antiviral effects, further interfering with viral replication.

NS3/4A adopts a chymotrypsin-like fold with two β -barrel domains (Figure 1.1). The catalytic triad is formed by His-57, Asp-81 and Ser-139 and is located in a cleft separating the two domains. The structure is stabilized by a Zn^{+2} ion that is coordinated by Cys-97, Cys-99, Cys-145, and His-149. NS4A aids in the proper folding of NS3; the central 11 amino acids of NS4A inserts as a β -strand to the N-terminal β -barrel of NS3. The binding site of the protease is very shallow presenting a challenge to develop high affinity and low molecular weight inhibitors.

Product inhibition by the N-termini of the trans-cleavage sites [72,73] formed the basis for the development and optimization of peptidomimetic inhibitors of the NS3/4A protease. [74] The proof-of-concept for antiviral efficacy was first demonstrated in 2002 with the macrocyclic inhibitor BILN-2061 (ciluprevir), which was later discontinued due to concerns about its cardiotoxicity. [75–77] In May 2011, the FDA approved two NS3/4A PIs, telaprevir [78,79] and boceprevir [80], for treatment of chronic HCV genotype 1 infection to be used in combination with Peg-IFN/RBV. Both telaprevir and boceprevir are acyclic ketoamide inhibitors

that associate with the protease through a reversible, covalent bond with the catalytic serine (S139) as well as short-range molecular interactions with the binding site. The limitation of these drugs to a single genotype pushes the research for developing inhibitors with broader activity.

Several non-covalent NS3/4A inhibitors, including macrocyclic compounds, are at various stages of clinical development (Figure 1.4). The non-covalent acylsulfonamide inhibitors contain a macrocycle connecting either P1 and P3 groups (ITMN-191 or danoprevir [81]) or alternatively P2 and P4 groups (MK-5172 [82], MK-7009 or vaniprevir [83]). These compounds have higher efficacies than telaprevir and boceprevir against the wild-type enzyme. However, in addition to the reported resistance in replicon studies, HCV quickly evolves to confer resistance to these protease inhibitors even at early stages of clinical trials compromising their high efficacy (Table 1.2). [84–89]

Figure 1.4: HCV NS3/4A protease inhibitors. The canonical nomenclature for drug moiety positioning is indicated using telaprevir. Telaprevir and boceprevir, recently approved for clinical use, are linear, reversible-covalent ketoamide inhibitors. Danoprevir, vaniprevir, and MK-5172 are macrocyclic acylsulfonamide inhibitors. Danoprevir has a P1-P3 macrocycle, while vaniprevir and MK-5172 have a P2-P4 macrocycle. Vaniprevir and MK-5172 differ in their P2 moieties.

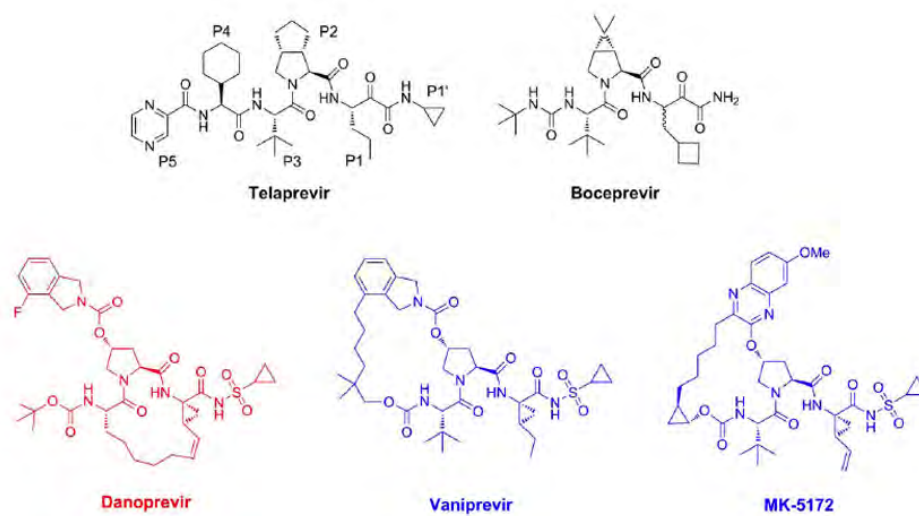


Figure 1.4

Table 1.2: Reported sites of drug resistance mutation in HCV NS3/4A .

Drug	Site	Mutation	Drug	Site	Mutation
<i>Telaprevir</i>	V36	M, A, L, G, I	<i>Vaniprevir</i>	F43	S
	T54	A, S		R155	K, T, G
	I132	V		D168	A, G, V, Y
	R155	K, T, M, G			
	A156	F, N, S, T, V			
	D168	N			
<i>Boceprevir</i>	V36	M, A	<i>MK-5172</i>	R155	G, Q
	T54	A, S, G, C		A156	T, V
	V55	A		D168	Y, E, T, A
	R155	K, T			
	A156	S, T, V			
	V158	I			
	D168	N			
	I170	A, T			
	M175	L			
<i>Danoprevir</i>	Q41	R			
	F43	S, C			
	S138	T			
	R155	K, T, I, M, G, L, S			
	A156	V, T, S, I			
	D168	A, V, E			

An important lesson learned from studying HIV-1 is that the key to designing robust inhibitors is to pay attention to the details of the balance between the natural substrate recognition and inhibitor binding. Mimicking the characteristics of the substrate recognition motifs in the designed compounds results in inhibitors that retain potency against drug resistant viral variants. The native function of the viral protease imposes an evolutionary constraint on the protease under the selective pressure of drug therapy. According to the substrate envelope hypothesis, resistance reflects a subtle change in the balance of molecular recognition events, in favor of substrate processing versus inhibitor binding. The drug-resistant protease variants are no longer effectively inhibited by the competitive drug molecules but can process the viral Gag and Gag-Pro-Pol polyproteins efficient enough for viral survival. Therefore, to guide the design of robust inhibitors, the specificity determinants of substrate recognition and how these determinants are maintained by the resistant variants should be elucidated for HCV as previously done for HIV-1.

High resolution co-crystal structures have been determined for the NS3/4A protease domain with the cleavage products as well as inhibitors, including danoprevir, MK-5172, and vaniprevir. [9, 90] In these structures, the products, despite the low sequence homology, adopted a consensus volume at P6 to P1 residues; *the substrate envelope*. The most severe resistance mutations occur at protease residues that are contacted by the inhibitors outside the substrate envelope.

1.9 Protein Dynamics

1.9.1 Protein Dynamics is Key to Molecular Recognition

Proteins are inherently flexible, undergoing functionally relevant conformational transitions under native state conditions on a wide range of scales, both in time and space. Atomic fluctuations and internal motions are key components of intra- and intermolecular communication, thus protein dynamics are essential for biological function and activity.

Thinking about proteins as dynamic entities that are in constant motion is fundamental to our understanding of molecular recognition, including substrate binding-catalysis-turnover and small-molecule inhibitor binding. The initial *lock-and-key* theory of molecular recognition conceptualized a rigid receptor accommodating a small molecule without undergoing any conformational rearrangements. [91] More recent binding models account for conformational changes caused by ligand binding, *induced fit*, [92] and random conformers of receptor that are selectively bound by ligand, conformational selection. [93–96] Both induced fit and conformational selection models are plausible and likely exist in Nature depending on the extent of inherent flexibility of a receptor protein.

The interplay between the internal motions of an enzyme and its activity has been reported repeatedly for various systems. Independently determined crystal structures of ligand-bound and free states of some enzymes have substantially different conformations. [97, 98] Catalytic activities of many enzymes are closely

associated with loop motions over the active site that orient key residues in the binding site relative to the substrate, including HIV-1 protease. [99–106] Equilibrium fluctuations in enzymes are also indispensable for substrate binding and product release. [107, 108] More indirect indications of protein dynamics include non-additive effects of double mutations at distal regions within proteins and substituted functional groups of inhibitors, suggesting a dynamic cooperativity of molecular interactions. [109–113]

The research cited here is only a small fraction of ever-growing evidence for the essential role of conformational dynamics in molecular recognition and function. Protein dynamics is a complex problem to access experimentally and researchers have used several techniques to address this problem. While techniques such as fluorescence spectroscopy, [114] spin label electron paramagnetic resonance (EPR), [115] and small angle X-ray scattering [116] have applications in elucidating some aspects of protein flexibility, x-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are the key experimental methods for characterizing molecular structures. X-ray crystallography has traditionally provided snapshots of one or only some of the conformations accessible to proteins in cryogenic conditions. However, technological advances in synchrotron X-ray sources have recently allowed time-resolved measurements on single crystals with the possibility of obtaining probability density for atoms. [117] Also recently developed computational tools for electron-density sampling, model refinement, and molecular packing analysis allow monitoring room-temperature conformational ensembles with x-ray crystallography that reveal motions in crystals crucial for

catalysis, ligand binding, and allosteric regulation. [118] In contrast to crystallography, NMR experiments are performed in solution and can allow direct observation of flexibility of proteins in structure and molecular interactions with ligands. The general consensus is that solution mimics the biological environment better than crystal, although molecular crowding has been reported to exert changes in structure. [119, 120] Protein size has been a limitation of this technique, but increasing field strengths in spectrometers will not only expand the range of protein size tractable with NMR but also enhance the resolution of spectra leading to the identification of more conformers accessible to the protein. [117]

Experimentally obtained structural data, whether by means of crystallography or NMR, provide a basis for understanding protein dynamics but currently cannot fully elucidate the details of conformational flexibility in full-atomistic scale. Molecular dynamics (MD) simulations are being more and more combined with experiments because MD is a powerful computation tool that tracks rapid processes across a spatiotemporal domain at atomic resolution for many biologically relevant systems at reasonable computational cost. MD simulations generate successive conformations of a system, so-called *trajectory*, by iterating Newton's second law of motion. MD has been successfully applied to many systems to address numerous biologically relevant questions, including protein folding, [121] enzyme catalysis, [122] and binding kinetics. [123] While the force field describing the molecular interactions can still benefit from improvements, the advancements in parallel computing architectures and algorithms have tremendously revolutionized the MD field and increased the accessibility of wider timescales and

larger systems. The earlier simulations of a ≈ 900 atom protein lasted 9 ps [124] while now protein folding simulations as long as 1 ms [125] or 50 ns simulation of an intact virion of 1 million atoms [126] can be performed.

Each of the techniques mentioned here has its own advantages and limitations. Being aware of the drawbacks and accordingly taking advantage of multiple techniques in concert would serve best for the purpose of obtaining a complete “movie” of protein dynamics at atomic resolution.

1.9.2 Change in Dynamics as a Drug Resistance Mechanism

Characteristics of the structural fluctuations contribute to the activity of an enzyme. Both nuclear magnetic resonance experiments and molecular dynamics simulations show that the flap region of HIV-1 protease is highly mobile and can adopt a large number of conformations on the nanosecond to microsecond timescale. [101, 102, 105] The exchange between these conformational states is correlated with structural reorganization of the active site, and the active site pre-organization is a rate-limiting factor in native function of the protease. [127] Binding of a small-molecule inhibitor and recognition of a natural substrate may have differential protein dynamics, even for a competitive active site inhibitor. Additionally, single or multiple amino acid substitutions in the protease may alter dynamics of protein-ligand interactions. In fact, NMR relaxation experiments combined with extensive molecular dynamics simulations revealed differential dynamics in the flaps between the wild-type and a drug resistant variant of HIV-1 protease. [128] Consistently, mutations that stabilize a closed-like conformation

correlate with enzymes of lowered activity and with higher affinity for inhibitors, which is corroborated by a further increase in the fractional occupancy of the closed state upon addition of inhibitor or substrate-mimic. [129] In addition to the flaps, a change in the flexibility of the hydrophobic core of HIV-1 protease, caused by various mutations selected under drug therapy, also modulates the enzymatic activity. [127, 130] All these results suggest that modulation of protein dynamics is a molecular mechanism of drug resistance. To avoid resistance in drug discovery, crystallographic studies should be coupled with methods probing for the altered dynamics in resistant variants relative to the wild-type to fully characterize the molecular mechanism of drug resistance.

1.9.3 Protein Dynamics is Often Neglected in Drug Design

Identifying potential drug leads by brute force, making as many compounds as possible with the hope that some would eventually have desirable activity, is no longer an effective method for drug discovery but structure-based drug design is now an integral part of most industrial drug discovery programs. [131] Nevertheless, structure-based drug design typically relies on static protein structures. Although incorporating ligand flexibility into design is routine, target flexibility is often ignored despite significant evidence for the need to consider the role of dynamics in recognition. [132] Instead, a common strategy has been to generate hydrophobic and conformationally constrained ligands with an entropically-driven binding affinity. This strategy raises the issue of drug resistance since conformationally constrained ligands cannot easily adapt to changes in the geometry of

the binding site but the enthalpically-driven inhibitors can be less susceptible to resistance. [133]

A single protein structure is only useful to identify inhibitors for that particular narrow state of the conformational ensemble. To obtain promising leads and accurately predict affinity of existing compounds, targeting multiple structures is a better strategy. Although assessing protein dynamics is not yet an established step of structure-based drug design in pharmaceutical industry, researchers have long been working on this problem. [134]

In earlier efforts, keeping the protein fixed but using a soft scoring function implicitly accounted for flexibility. [135] This method, *soft docking*, allows for overlap between the ligand and the protein, giving a small estimate of the plasticity of the receptor. Soft docking is computationally cheap due to the fixed scoring function but is not very effective in identifying binders with varied size and conformation. Another approach is the *relaxed complex method*, in which ligands are docked to an ensemble of conformations of the target, and the best complexes are re-scored to provide predictions of optimal binding geometries. [136, 137] Receptor-based pharmacophore models that use a collection of crystal structures have been proposed to account for the inherent protein flexibility in structure-based drug design. [138, 139] Keeping the protein backbone rigid but sampling the side chain conformations of the receptor is another technique to account for flexibility in the binding site. This technique uses *rotamer libraries*, which are collections of low energy amino acid side chain conformers and their frequency of observation, compiled from high-resolution x-ray structures. [140] Computa-

tionally more rigorous simulated annealing was used for conformational search with a rapid grid-based method of energy evaluation in *flexible docking* algorithms. [141, 142] More recently, use of multiple crystal structures has been combined with regular and replica exchange molecular dynamics simulations to sample the conformational space accessible to the binding surface. [143]

Although flexibility has not been a major concern in earlier stages of drug design, with the recent innovations in computational and experimental methods to characterize protein dynamics, new approaches based on ensemble structures are expected to emerge. Ultimately, accurate incorporation of conformational flexibility will improve the productivity of drug discovery efforts. [117, 144]

1.10 Scope of Thesis

The substrate envelope hypothesis provides a fundamental explanation to the very complicated problems of substrate specificity, drug resistance, and co-evolution of the natural cleavage sites with the drug resistant variants. Although the substrate envelope, defined by the crystal structures, has greatly enhanced our understanding of the structural basis for these three phenomena, the picture it draws cannot be complete when protein dynamics is left out. All biological macromolecules, including enzymes and their substrates, sample ensembles of conformations. Even the native state cannot be thought of as one static structure, but a collection of similar conformations sampled around a common minimum in the energy landscape. Therefore, it is crucial to take into account these local fluc-

tuations around the native state, a static image of which is captured in a crystal structure. In my thesis, I attempt to fill this gap and evaluate the generality of the substrate envelope hypothesis from a protein dynamics point of view. I first introduce the dynamic substrate envelope by sampling the local conformational space accessible to the substrates bound to HIV-1 protease and characterize the dynamics of preferential substrate recognition by the wild-type HIV-1 protease. Then I validate the HIV-1 dynamic substrate envelope by investigating the dynamic properties of three drug-resistant protease-substrate complexes that have co-evolved under therapy. Next, I assess the generality of the dynamic substrate envelope by applying it to the NS3/4A protease of another quickly evolving virus, Hepatitis C. This part of my thesis also highlights considerable deviations in inhibitors' dynamic interactions from those of substrates and suggests inhibitor modifications to improve resistance profiles. Finally, I highlight the importance of complementing the structural studies with protein dynamics by showing a distal mutation's effect on the NS3/4A binding site structural organization and the overall protease dynamics.

Chapter II

Dynamics of Preferential Substrate Recognition in HIV-1 Protease: Redefining the Substrate Envelope

2.1 Abstract

HIV-1 protease (PR) permits viral maturation by processing the Gag and Gag-Pro-Pol polyproteins. Though HIV-1 PR inhibitors (PIs) are used in antiviral therapy, the emergence of drug resistance has limited their efficacy. The rapid evolution of HIV-1 necessitates the consideration of drug resistance in novel drug-design strategies. Drug-resistant HIV-1 PR variants, while no longer efficiently inhibited, continue to efficiently hydrolyze the natural viral substrates. Though highly diverse in sequence, the HIV-1 PR substrates bind in a conserved shape we defined as the “substrate envelope”. Resistance mutations arise where PIs protrude beyond the substrate envelope, as these regions are crucial for drug binding but not for substrate recognition. Here, we extend this model by considering the role of protein dynamics in the interaction of HIV-1 PR with its substrates. Seven molecular dynamics simulations of PR-substrate complexes were performed to estimate the conformational flexibility of substrates in their complexes. Interdependency of the substrate-protease interactions may compensate for the variations in cleavage-site sequences, and explain how a diverse set of sequences can be recognized as substrates by the same enzyme. This diversity may be essential for regulating sequential processing of substrates. We also introduce a *dynamic substrate envelope* as a more accurate representation of PR-substrate interactions. The dynamic substrate envelope, described by a probability distribution function, is a powerful tool for design efforts targeting ensembles of resistant HIV-1 PR variants to develop drugs that are less susceptible to resistance.

2.2 Introduction

Drug resistance reflects a subtle change in the balance of molecular recognition: the resistant viral enzyme is no longer effectively inhibited by drugs, but can still recognize and process its natural substrates. Targeting resistance in drug design, therefore, requires a detailed understanding of substrate recognition. The principles underlying substrate recognition are especially important in the case of HIV-1 protease, a 99-residue, homodimeric aspartyl protease, since all currently available FDA-approved protease inhibitors are competitive inhibitors. The natural substrates of HIV-1 PR are cleavage sites on the Gag and Gag-Pro-Pol polyproteins, whose processing is crucial for the maturation of infectious virions. The cleavage sites, however, share little sequence homology, making the determinants of substrate specificity difficult to identify.

To address the substrate recognition problem, Prabu *et. al.*, has solved crystal structures of an inactive variant (D25N) HIV-1 PR bound to decameric peptides corresponding to the cleavage sites within the Gag and Gag-Pro-Pol polyproteins. [6, 7] These crystallographic studies revealed that the common structural feature of the natural substrates is the conserved consensus volume that they adopt within the binding site of the enzyme. This consensus volume, the substrate envelope, is hypothesized to be the key element for substrate recognition rather than a specific amino acid sequence. [7]

The structural basis for drug resistance came from further crystallographic studies on protease-inhibitor complexes. [145] The inhibitor molecules in these

crystal structures also adopted a conserved consensus volume, the *inhibitor envelope*. The superposition of the inhibitor and substrate envelopes showed that the inhibitors protrude beyond the substrate envelope and make contact with some protease residues. These residues corresponded to the previously known drug resistance mutation sites in the protease. This observation led to the hypothesis, that the inhibitors, when designed to fit within the substrate envelope, are less likely to elicit drug resistance mutations. Following this hypothesis, the substrate envelope has been used successfully as a constraint to guide the design of new drug molecules that are effective inhibitors of protease variants, including highly resistant variants. [44–46, 48]

Some cleavage sites co-evolve with the drug resistant protease variants, presumably complementing the emergence of drug resistance mutations in the protease. Previous crystallographic and structural modeling studies on resistant protease-substrate complexes suggested that these co-evolving substrates make extensive contacts with drug resistance mutation sites on the protease and hence co-mutate with them. [39, 40] However, the compensatory mutations on these co-evolving cleavage sites are not always at the positions that interact with the drug resistance mutation site on the protease, but sometimes they occur at another residue on the cleavage site. Modeling studies on Nelfinavir-resistant PR_{D30N/N88D} in complex with p1-p6_{WT} and p1-p6_{LP1'F} suggested that the protease-substrate interactions lost in PR_{D30N/N88D} are recovered by the LP1'F substitution in p1-p6. [39] Similarly, the multi-drug resistant PR_{V82A} does not interact with the wild-type NC-p1 cleavage site as favorably as it does with NC-p1_{AP2V}. [40] According to the sub-

strate envelope hypothesis, substrates that protrude beyond the envelope and make contact with the drug resistance mutation sites on the protease co-evolve with resistant protease variants.

In this study we elucidate the balance between the variation and commonality of the natural substrates of HIV-1 PR in terms of differential substrate recognition by investigating the conformational ensembles of protease-substrate complexes. Clues to substrate recognition by HIV-1 PR can be found in two characteristics of substrate cleavage. The sites in gag and gag-pro-pol polyproteins are cleaved each with specific unique enzyme kinetics and in a tightly regulated required order. [146] This variation in substrate processing, complements the commonality between the substrates found in the conserved shape. Although initially we analyzed various crystal structures of HIV-1 PR-substrate complexes, these crystal structures only represent an average of the trapped low energy states of the enzyme. However, the activity of HIV-1 PR is intimately related to its dynamics, from local to global motions. Consideration of this flexibility is essential to understanding how the specificity of substrate recognition is determined. To capture aspects of this dynamics, the substrate complexes were simulated using molecular dynamics (MD). From these simulations, the conformations and dynamic cooperativity of the substrates with the protease were analyzed as a dynamic ensemble. Most of the protease-substrate interactions observed in the crystal structures were conserved during the MD simulations, validating our understanding of substrate recognition. The dynamic ensemble was further analyzed in terms of a dynamic substrate envelope, to ascertain how molecular dynamics affects the consensus

volume of the bound substrates. The dynamic substrate envelope reproduces the main characteristics of the static substrate envelope, providing a more realistic and accurate representation of the consensus volume. Essentially the dynamic substrate envelope is a probability distribution function in contrast with the static substrate envelope, which is a step function. Thus, the dynamic substrate envelope is potentially a more powerful tool in computer-aided drug design, easily incorporated as an added constraint in structure-based design algorithms.

2.3 Results

2.3.1 Molecular Interactions Between Protease and Substrates:

Analysis of vdW Contacts

As the HIV-1 protease active-site is mainly hydrophobic, understanding the preferential substrate recognition requires quantitatively analyzing van der Waals (vdW) contact potentials between the protease and its substrates. Such an elucidation is essential to reveal and compare how well a particular substrate fits into the binding site.

Comparison of crystal structures and molecular dynamics trajectories. Crystal structures of substrate complexes provide a starting conformation for examining vdW contacts between the enzyme and its substrate but do not assess the stability of these contacts. Therefore, the crystal structures were examined quantitatively for the PR-substrate vdW interactions, and then compared with those

from the MD trajectories. An average contact potential is shown in Figure 2.1 for each protease residue interacting with at least one substrate. Each average was calculated for each protease residue over the set of seven PR-substrate crystal structures (Figure 2.1, black bars) and their respective MD trajectories (light gray bars), enabling a quantitative comparison of static and dynamic contacts. Although the standard deviations of individual residue vdW contact potentials from the dynamic ensembles are larger than those from the various crystal structures, the general pattern of protease residue potentials is very similar. Thus, the crystal structures provide a good representation of the contact potentials between the protease and substrates.

Figure 2.1: Contact potentials between the protease and seven substrates are well represented by crystal structures. For each protease residue contacting at least one substrate, its potential energy of vdW contact is averaged over seven PR-substrate crystal structures (black bars), and their respective MD trajectories (light gray bars). Each panel shows one monomer of the protease. Stars represent PR drug-resistance mutation sites according to the Stanford Database of HIV Drug Resistance. [27] Most of these sites do not make extensive contacts with the natural substrates, suggesting that these residues are more important for inhibitor binding than substrate recognition. A few of these sites contact natural substrates and have been reported to co-evolve with natural substrates, i.e., D30N/N88D with p1-p6, I50V with both NC-p1 and p1-p6, V82A with NC-p1.

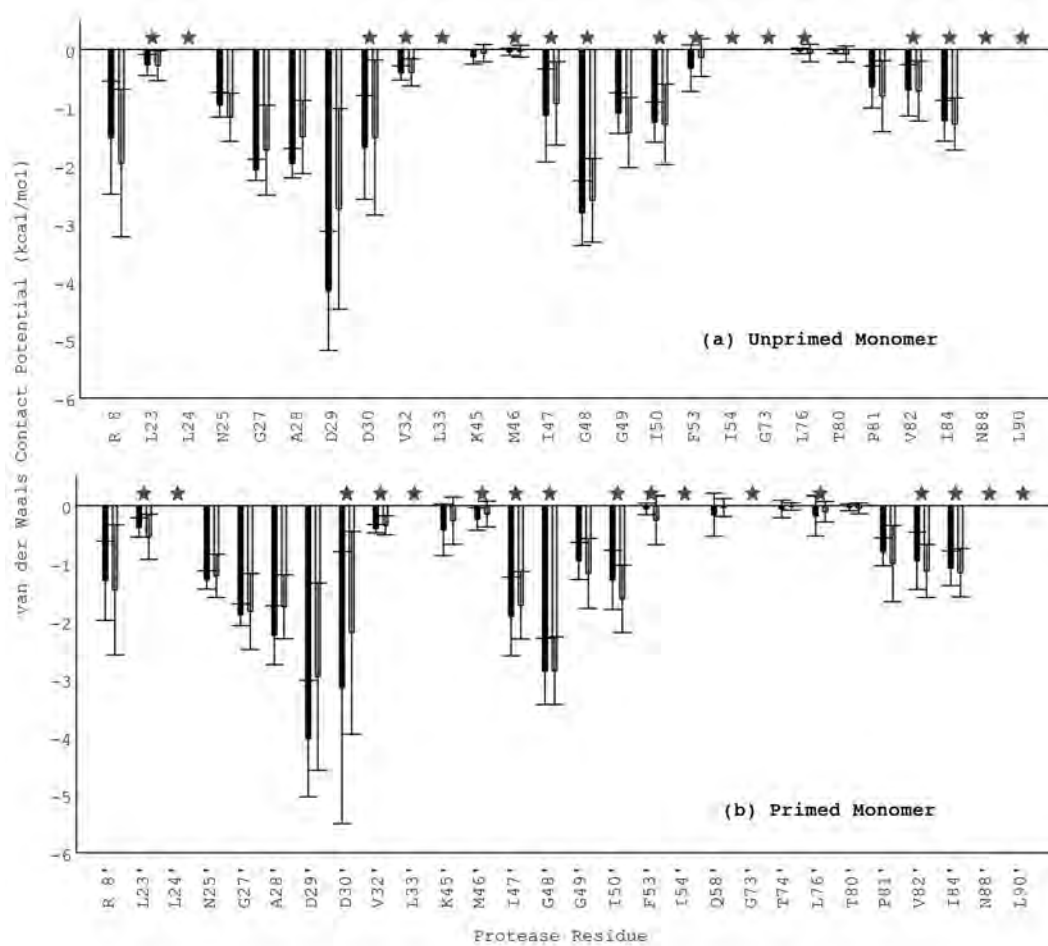


Figure 2.1

Interaction of drug-resistance mutation sites with substrates. Drug-resistance mutation sites (shown as stars in Figure 2.1) tend to have less contact potential with substrates than other sites. This occurs presumably as a result of these sites are more important for inhibitor binding than substrate recognition. To investigate the change in role of these PR residues for recognizing particular substrates, both the crystal structure and the dynamic ensemble of each substrate were examined. The overall trend of vdW contacts averaged over various crystal structures matches the trend of time-averaged contact potentials. However, for a few protease residues, including: I47/I47', F53/F53' and L76', the contact decreases with certain substrates and no contact is formed with other substrates. These residues are among the list associated with drug-resistance in the Stanford Database. [27] Since these residues do not preserve their crystal contacts with any substrates, they are indeed less important for substrate recognition than inhibitor binding.

At some protease residues, the interactions with the substrates are more complex. One such special case occurs with D30 and D30', which make fewer contacts with p2-NC and CA-p2, respectively, during the course of MD simulations than in the crystal structures. Nonetheless, D30 preserves high interaction potential with the majority of substrates. In the case of p1-p6, D30 forms a dynamically stable interaction with the substrate. The alteration of the interaction by the D30N mutation may provide a structural explanation for the observed co-evolution of this cleavage site. Generalization of this analysis requires further attention to co-evolved residues in the protease and substrates.

Figure 2.2: Protease structures have similar interaction profiles although they are in complex with non-homologous substrates. (A) Unprimed monomer (B) Primed monomer.

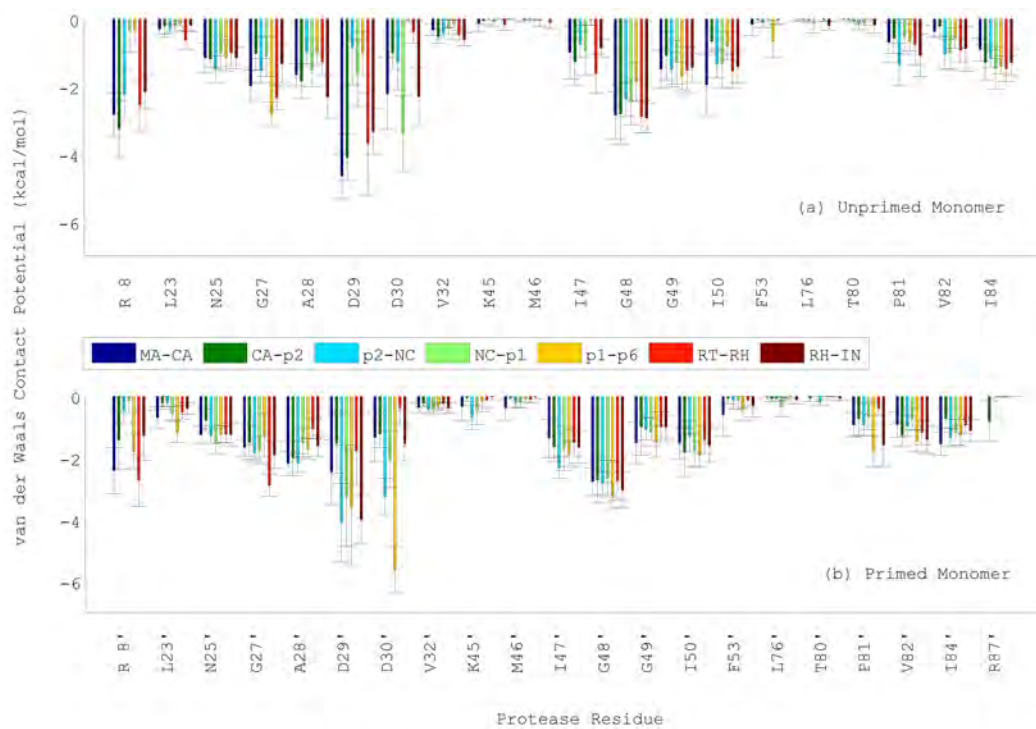


Figure 2.2

Protease-substrate vdW interactions in the context of preferential substrate recognition. Each non-homologous substrate cleavage site forms unique interactions with the mainly hydrophobic protease active site, although the protease residues have fairly similar vdW interaction profiles (Figure 2.2). The overall vdW potential energy of substrate-protease binding is conserved at approximately -45 kcal/mol (Figure 2.3), suggesting an optimal interaction for the cleavage sites despite their low sequence homology. In all substrates, however, individual contributions to the overall interaction energy by the eight residues, P4 through P4' positions, are not the same (Figure 2.3). P4 and P4' positions, in general, have lower interaction potential probably because they are more flexible and solvent exposed. The majority of substrates interact extensively with the protease through the P3-P3' region, though each interaction is unique. For example, P3 and P2' have the highest vdW contact potential in CA-p2 due to relatively bulkier amino acids at these positions (R-P3 and E-P2') while in p2-NC, Q-P2' and R-P3' are the main interacting residues. MA-CA and RT-RH have a more symmetrical interaction profile along their sequence presumably due to their pseudo-symmetry around the cleavage site. Symmetry of vdW contact potential around the cleavage site appears to be associated with symmetry in the side-chain size of substrate residues. Thus, overall vdW interaction potential is an important indicator for a sequence to be recognized by HIV-1 protease as a substrate. However, these substrates have varied contacts due to their diverse sequences, suggesting that the cleavage-sites have been evolutionarily optimized to maintain a particular degree of vdW interaction within the protease active site.

The hypothesis of evolutionary optimization of cleavage-site sequences to maintain their overall interaction potential with the protease can hold true only if there is dynamic interdependency among the cleavage-site residues. A PR interaction lost at one substrate residue should be compensated by an increased interaction at another residue. To examine this possibility, the vdW contact potential of each substrate residue was calculated over time in MD simulations, and a correlation coefficient was computed for each possible residue pair (Figure 2.4). All substrates showed a similar pattern of coupled residues, which compensate each other for vdW interaction (shown in blue in Figure 2.4). This consensus interdependency between the substrate amino acid positions appears to be critical for recognition by HIV-1 protease, supporting the hypothesis above.

Figure 2.3: Substrates maintain a consensus vdW interaction potential with the protease independent of sequence. (A) Despite low sequence homology, substrates have an optimal total vdW interaction with the PR. (B) The interaction potential between the PR and substrate differs by the individual contribution of substrate residues P4 to P4'.

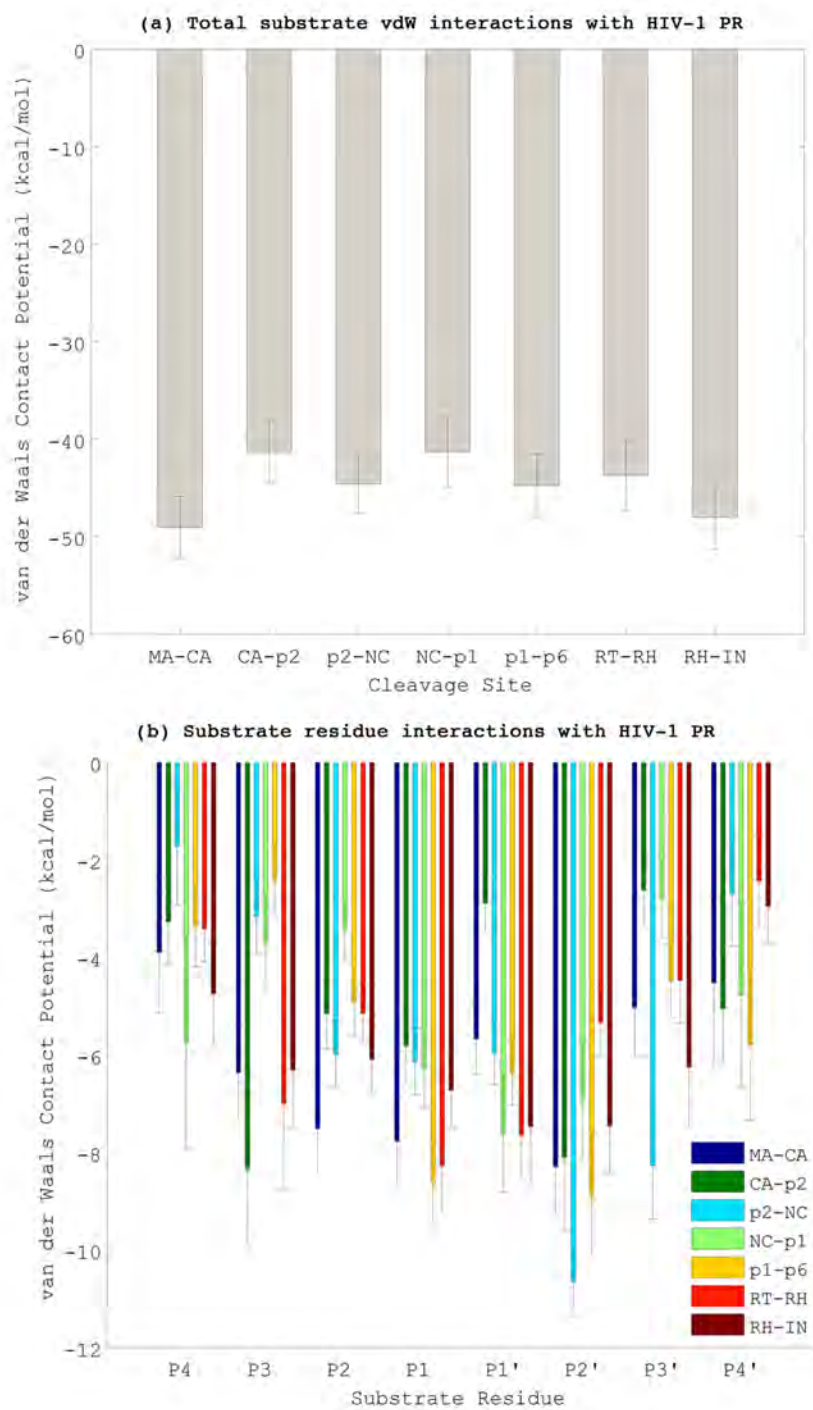


Figure 2.3

Figure 2.4: Fluctuations in the vdW interaction potential within/among substrate residues are negatively correlated, revealing their interdependency.

The vdW contact potential of each substrate residue with the PR was calculated over time. Fluctuations in this potential with respect to every other residue in the same substrate were estimated by computing Pearson correlation coefficients. The correlation coefficients are contoured for each substrate and color-coded from red (fully correlated) to blue (fully anti-correlated). (A) MA-CA, (B) CA-p2, (C) p2-NC, (D) NC-p1, (E) p1-p6, (F) RT-RH, (G) RH-IN.

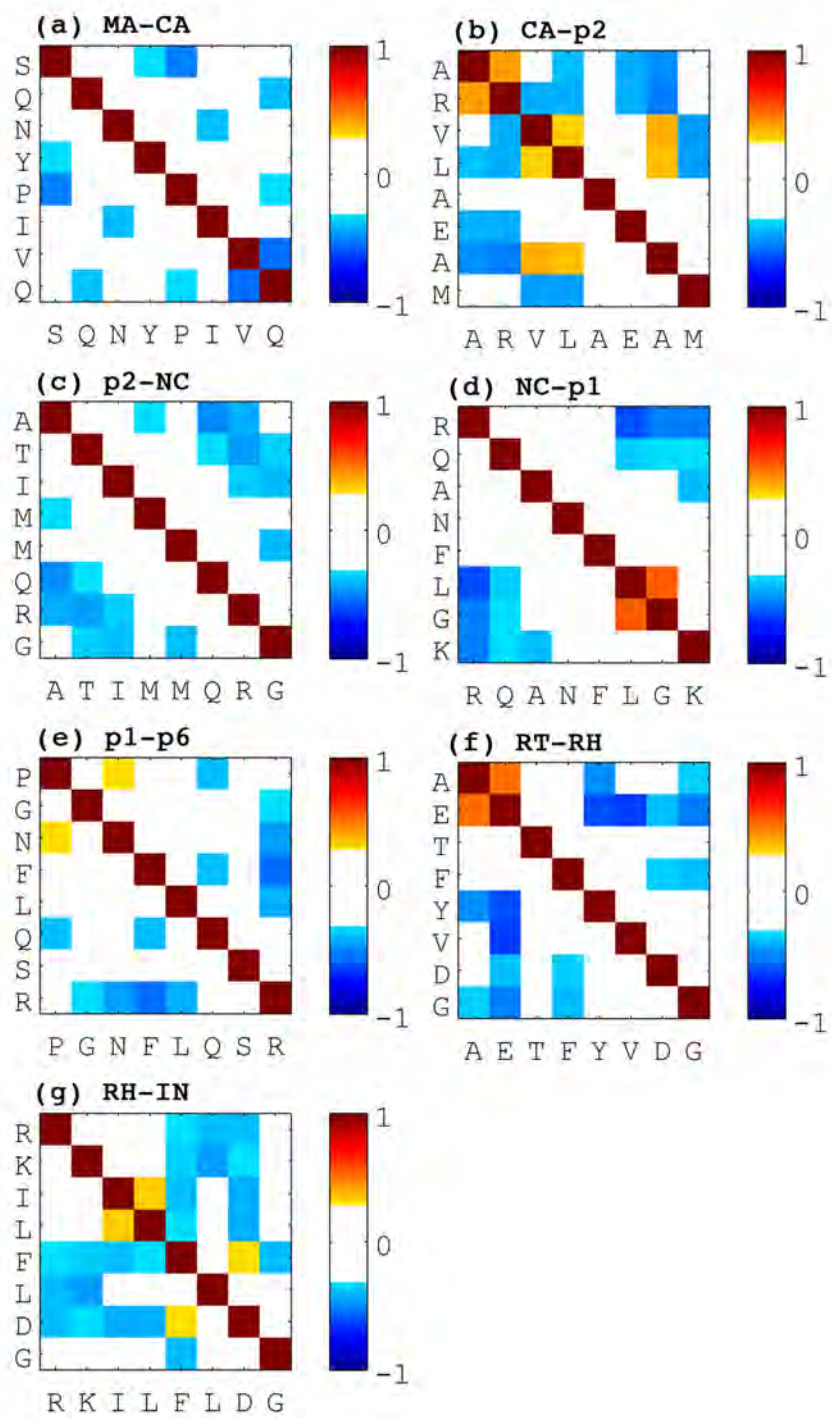


Figure 2.4

2.3.2 Molecular Interactions Between the Protease and Substrates:

Analysis of Hydrogen Bonds

Though vdW interactions are fairly important for stabilizing the substrates in the mainly hydrophobic binding groove, analysis of hydrogen bonds in variant substrates is necessary for better understanding the specificity of substrate recognition. To this end, both the intermolecular hydrogen bonding networks between the protease and substrates and the intramolecular hydrogen bonds within the substrates were investigated.

The hydrogen bonds between the protease and substrates were analyzed in two groups: the substrate backbone and substrate side-chain hydrogen bonds regardless of whether the acceptor/donor is in the protease backbone or side-chains. This analysis showed that the backbone hydrogen bonds are more stable throughout the trajectories, also more conserved across the substrates than the hydrogen bonds formed by substrate side-chains (Figure 2.5). Majority of the backbone hydrogen bonds appear to be consistently observed within the course of the simulations with varied frequencies (Figure 2.5). There are a few exceptions; the bonds formed by P4 and P3 backbones do not appear to be stable in the simulations. In addition, the protease forms hydrogen bonds with substrate backbones via D29/D29' in both monomers, yet these hydrogen bonds are not conserved across the substrates in the simulations. No additional hydrogen bonds are formed in the MD simulations except for those formed with N25. In all the crystal structures, a hydrogen bond was observed between N25' and the substrate backbone at P1 position. In

the MD simulations, however, this hydrogen bond keeps alternating between N25 and N25', more than 50% of the time in some substrates. The active site is formed by the two D25/D25' in the catalytically active species. Thus, this pattern, alternating between the two monomers, of hydrogen bonding is in agreement with the necessity of the two Asp residues for catalytic activity. [147] The substrates interact with both the ASP residues in a balanced manner, hence a single mutation at this position in tethered dimers results catalytic inactivity.

Figure 2.5: Backbone hydrogen bonds are more conserved across varied substrates and more stable than the side-chain hydrogen bonds. Frequency of occurrence of hydrogen bonds during the simulations between any protease atom and substrate backbone (A) and substrate side-chains (C), and the hydrogen bonds observed in the crystal structures between any protease atom and substrate backbone (B) and substrate side-chains (D). The side-chain hydrogen bonds are shown by residue (B, D) for the crystal structures and simulations, respectively. In panels A and C, the hydrogen bonds are color-coded with respect to the time percentage of occurrence throughout the simulations with red being 100%. Only the hydrogen bonds that existed more than 50% of the time were shown, below 50% occurring bonds are shown in white. In panels B and D, gray indicates a hydrogen bond observed in the crystal structure and white shows that the bond does not exist in that structure.

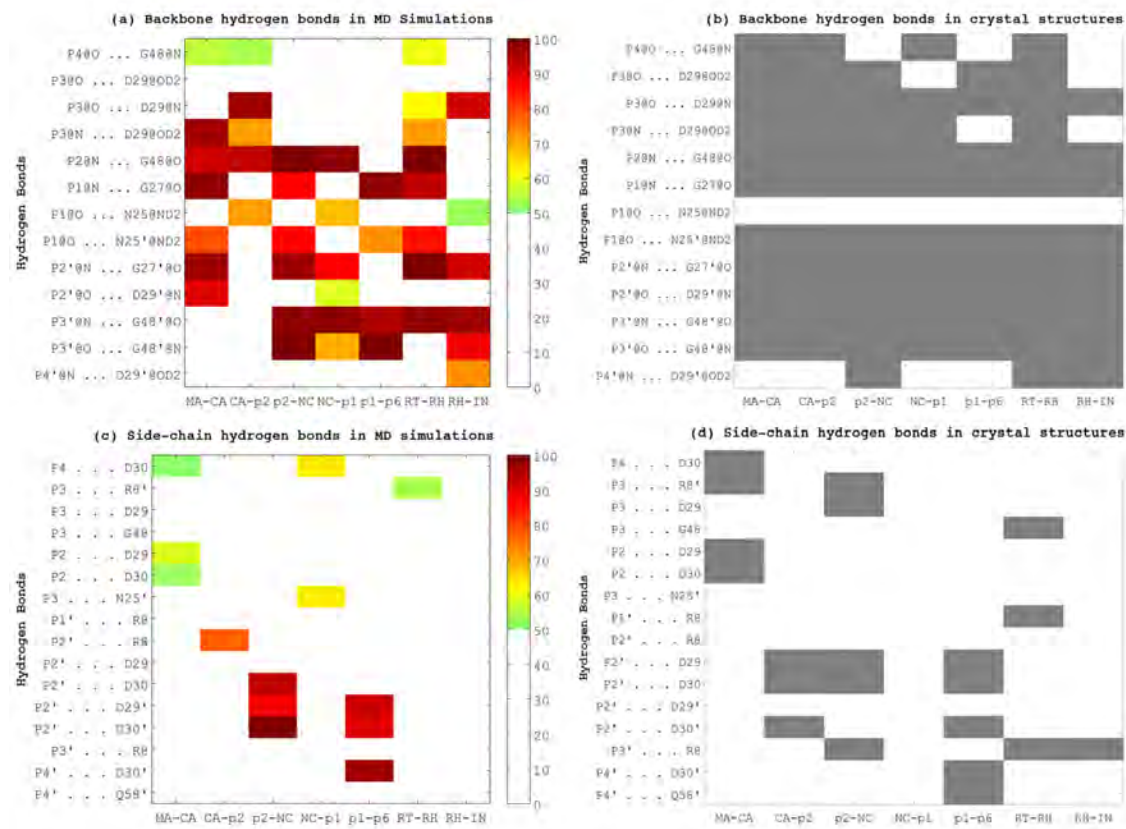


Figure 2.5

The side-chain hydrogen bonds have variations between the crystal structures and MD simulations when analyzed at the atomistic level. In contrast, when analyzed in a residue-based manner, hydrogen bonding potential of any protease-substrate residue pair does not deviate from the crystal structures significantly whether or not the acceptor-donor atoms change in the simulations (Figure 2.5).

The protease residues cluster in the specificity of their interactions with the substrates. Some residues form hydrogen bonds only with the backbones of almost all substrates, while others make hydrogen bonds with substrate side-chains in a more specific way. D29 and D30 interact with the substrate side-chains in either monomer, while N25, G27, and G48 typically make hydrogen bonds with the substrate backbones. p2-NC and p1-p6 appear to have very stable side-chain hydrogen bonding patterns mainly at their P2' position. RT-RH and RH-IN side-chains lose their hydrogen bonds with the protease immediately in the simulations, while they maintain most backbone hydrogen bonds. Despite the obvious similarity in the backbone hydrogen bond patterns of varied substrates, the side-chain hydrogen bonds appear to be unique for individual substrates.

Almost every substrate has an intramolecular hydrogen bond within its sequence (Table 2.1). Most of these intramolecular hydrogen bonds are formed within the substrate backbone and all between the i^{th} and $i+2^{th}$ residues within the P3-P3' region. Although the hydrogen bonding patterns may have been affected by the D25N substitutions, these hydrogen bonds likely stabilize the backbone conformation within the active site, contributing to the conserved consensus volume occupied by the substrates.

Table 2.1: Hydrogen bonds within the substrates of HIV-1 protease that exist more than 50% of the time .

Hydrogen Bond	Substrate (%)
P3 O ... P1 N	NC-p1 (61.8%) RH-IN (64.4%)
P2 O ...P1' N	p2-NC (91.7%) RT-RH (83.6%)
P2 OD1 ... P1' N	p1-p6 (87.3%)
P1' O ... P3' N	CA-p2 (56.6%)

2.3.3 Analysis of Atomic Fluctuations

Dynamic coupling between the protease and its substrate. Proteins are dynamic; even a relatively stable, highly populated enzyme-substrate complex state can visit multiple local conformations. During sampling of these conformations, the substrate complexes of HIV protease fluctuate around the native states in a coupled manner. Correlation coefficients of mean-square fluctuations between HIV-1 PR residues and substrate residues are shown for all seven PR-substrate complexes (Figure 2.6). Protease residues that fluctuate in correlation with any substrate residue are the same in all complexes, corresponding to mainly three regions around positions 30, 50, and 80 in both monomers. The substrate residues coupled with these protease regions are not, however, the same for all substrates. In all substrates but CA-p2, the unprimed side moves in correlation with the region around position 50 in the unprimed monomer. In CA-p2, the primed side is coupled to the same region of the protease. The 50' region in the other monomer is highly correlated to P3' and P4' of CA-p2, while the correlation of the 50' region seems more dispersed with the other substrates. p1-p6 appears to be moderately correlated with the 50 region compared to the other substrates. RH-IN has a dual character; the prime side correlates with the primed PR monomer, and the unprimed side with the unprimed monomer. The primed and unprimed sides of RH-IN fluctuate independently of each other, almost like two distinct domains (Figure 2.7).

Figure 2.6: Interdependency within substrate residues revealed by cross-correlations of mean-square fluctuations between protease and substrate residues.

Atomic positional fluctuations are calculated for all residues in the PR-substrate complex. How one PR residue fluctuates with respect to a substrate residue was estimated by computing cross-correlations (see "Methods"). The correlation coefficients are contoured and color-coded from red (fully correlated) to blue (fully anti-correlated). (A) MA-CA (B) CA-p2 (C) p2-NC (D) NC-p1 (E) p1-p6 (F) RT-RH (G) RH-IN.

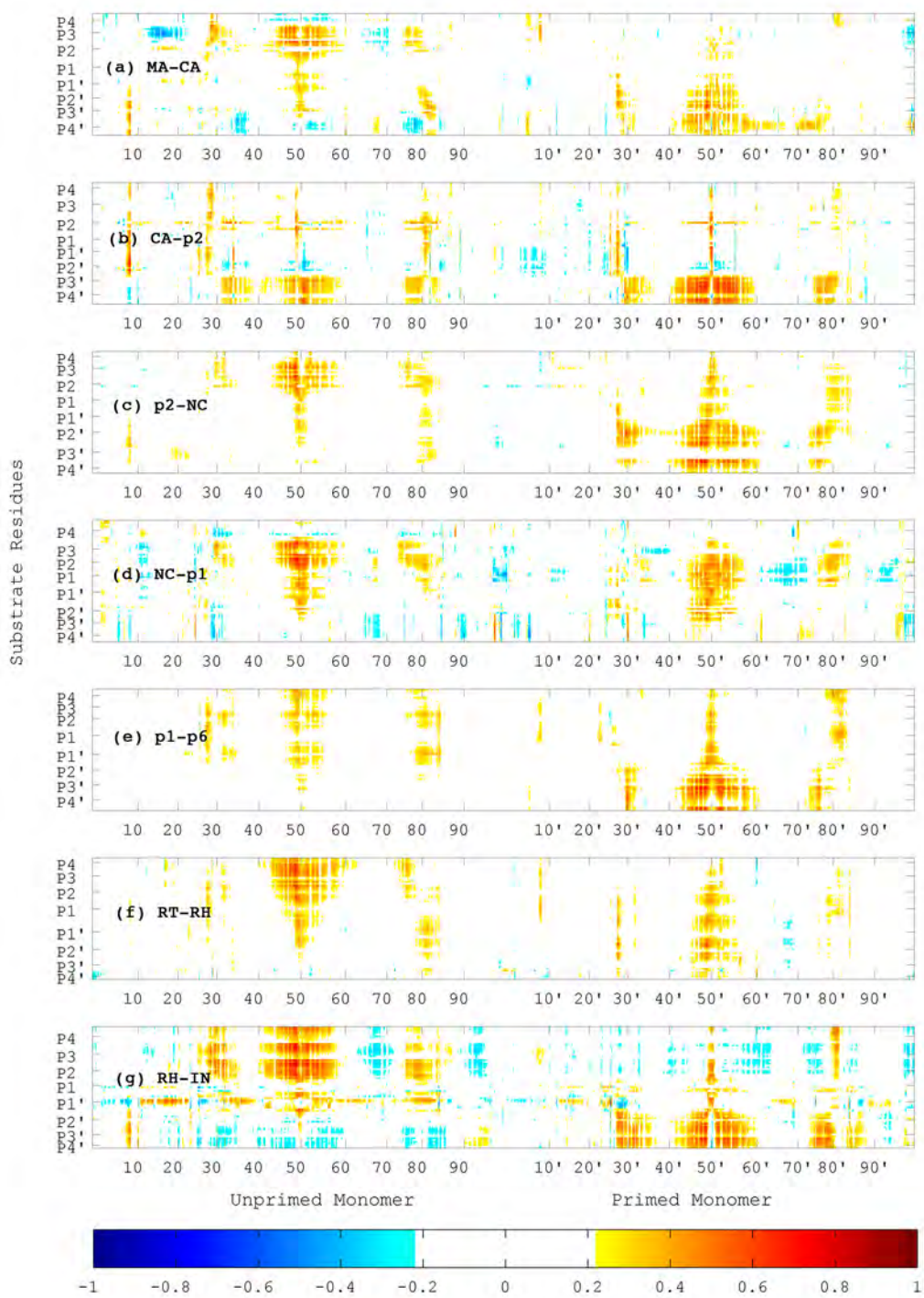


Figure 2.6

Figure 2.7: Unique intrinsic dynamic coupling revealed by cross-correlations of the substrate mean square fluctuations. The same cross-correlation analysis in Figure 2.6 was repeated for the substrate residues. The correlation coefficients are contoured and color-coded from red (fully correlated) to blue (fully anti-correlated). (A) MA-CA, (B) CA-p2, (C) p2-NC, (D) NC-p1, (E) p1-p6, (F) RT-RH, (G) RH-IN.

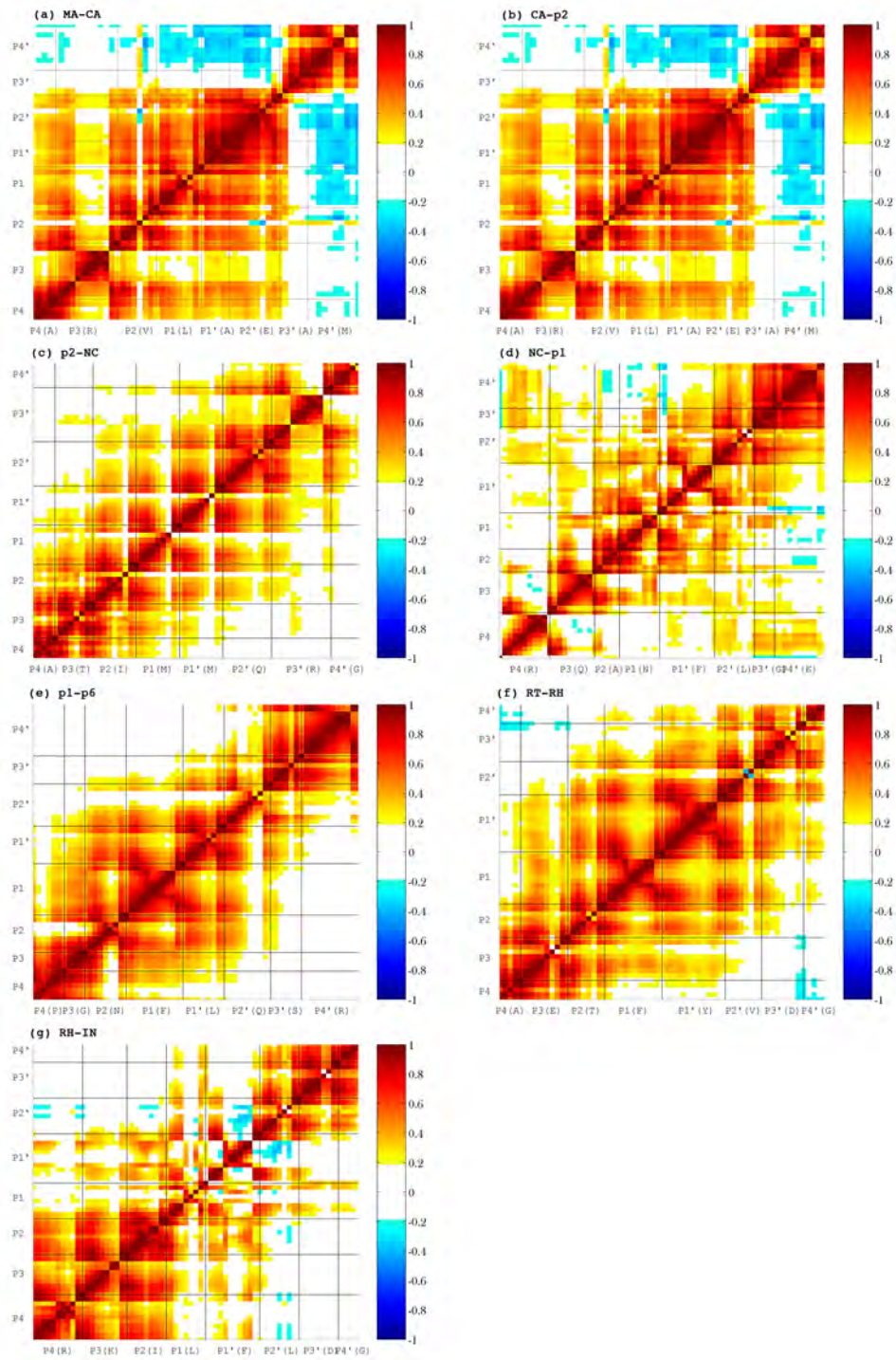


Figure 2.7

This variation in PR-substrate dynamic coupling is unexpected, especially because the substrates are only eight amino acids long. CA-p2 and p1-p6 also have two dynamically independent domains, whereas p2-NC, NC-p1, and RT-RH exhibit autocorrelation among their primed and unprimed residues. The correlations of these substrates with the PR monomers are also distributed throughout the whole substrate sequence as opposed to clustering in certain residues as for P3'-P4' in CA-p2 and P2'-P4' in RH-IN. Thus, the protease shows a conserved pattern in maintaining cooperativity with its substrates, but substrates can differ in their patterns of reciprocating signals from the protease. In other words, multiple correlation patterns satisfy the dynamic requirements of cooperativity of the protease. This conclusion is only possible if the residues within a substrate are interdependent, as shown also by vdW contact analysis.

2.3.4 Dynamics of the Substrate Envelope

Static versus dynamic substrate envelopes. Incorporating protein flexibility into modeling the substrate envelope may make the envelope a more realistic representation of the substrates. Both the static and dynamic substrate envelopes were computed and the fit of each substrate within these envelopes was evaluated by using a three-dimensional grid-based approach. The static substrate envelope was defined using crystal structures of seven substrates in complex with the inactive D25N protease variant. These crystal structures were used as initial structures for MD simulations, and the dynamic substrate envelope was defined using snapshots throughout these trajectories. The substrate volumes lying within the

dynamic substrate envelope (V_{in}^{dyn}), within the static substrate envelope (V_{in}^{stat}), outside the dynamic substrate envelope (V_{out}^{dyn}), and outside the static substrate envelope (V_{out}^{stat}) are shown in Figure 2.8, along with the total individual substrate volumes (V_{tot}^{stat} and V_{tot}^{dyn}). The total volumes of the static and dynamic substrate envelopes are $1552 \pm 100 \text{ \AA}^3$ and $1413 \pm 66 \text{ \AA}^3$, respectively. These numbers show that the fluctuations around the native state of substrates when bound to the protease decreased their consensus volume. According to the substrate envelope hypothesis, the volume of a substrate protruding beyond the substrate envelope, i.e. V_{out} determines whether or not that substrate will be susceptible to mutations in the protease. Because the substrates have different total volumes, although V_{in} is similar for each, they differ in their V_{out} . This is reflected in altered interactions with protease residues (Figure 2.8). Although the dynamic envelope is slightly smaller than the static envelope, the relative fit of a particular substrate within the envelope is the same. For a particular substrate the V_{out} values do not vary dramatically throughout the simulations (Figure 2.8). The histograms of V_{in} , V_{out} , and V_{tot} values for all substrates are shown in Figure 2.9. All substrates exhibit a unimodal distribution for their V_{out} values, suggesting that they do not have multiple preferences in terms of their fit into the substrate envelope. Thus, although the absolute values of static and dynamic V_{out} vary, their distribution is similar.

Figure 2.8: The static and dynamic substrate envelopes share the same overall trend in V_{out} , i.e., the volume lying outside the envelope. The absolute values, however, vary for static and dynamic cases, most likely due to the dynamic substrate envelope being more heterogeneous than the static envelope. The dynamic envelope is denser in the interior region, which consists of substrate backbones. Towards its surface, the dynamic envelope is less well-defined than its interior. This contrast is not as pronounced in the static substrate envelope. Hence, V_{out}^{stat} is smaller than V_{out}^{dyn} for all substrates.

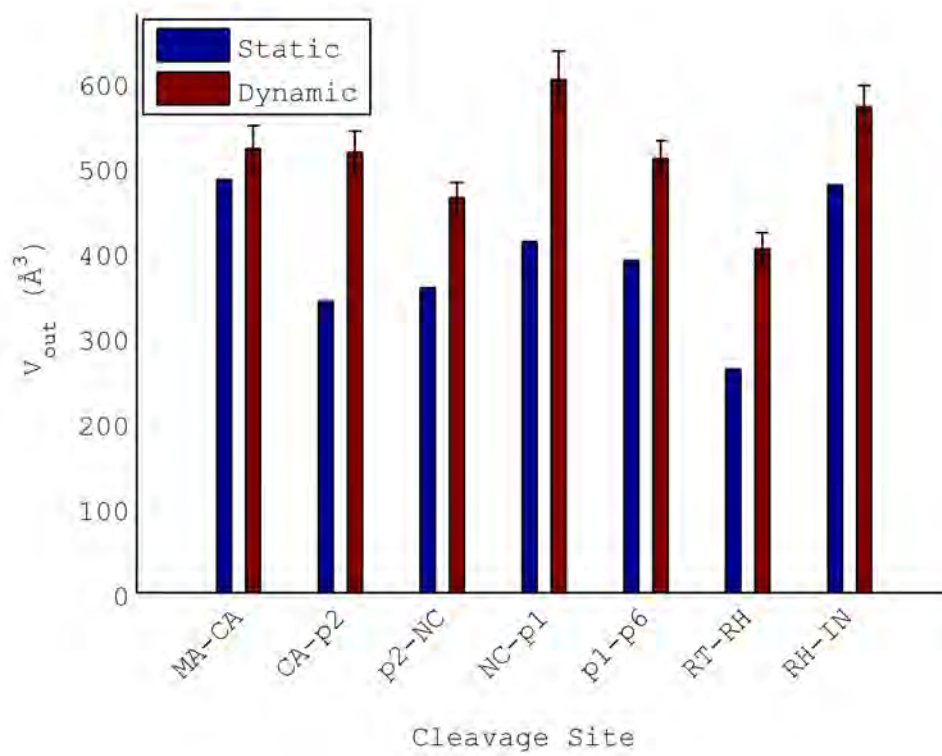
**Figure 2.8**

Figure 2.9: Distributions of V_{out} , V_{in} , and V_{tot} values throughout the MD simulations are unimodal for each substrate. Mean of data was shown as a red line in each histogram.

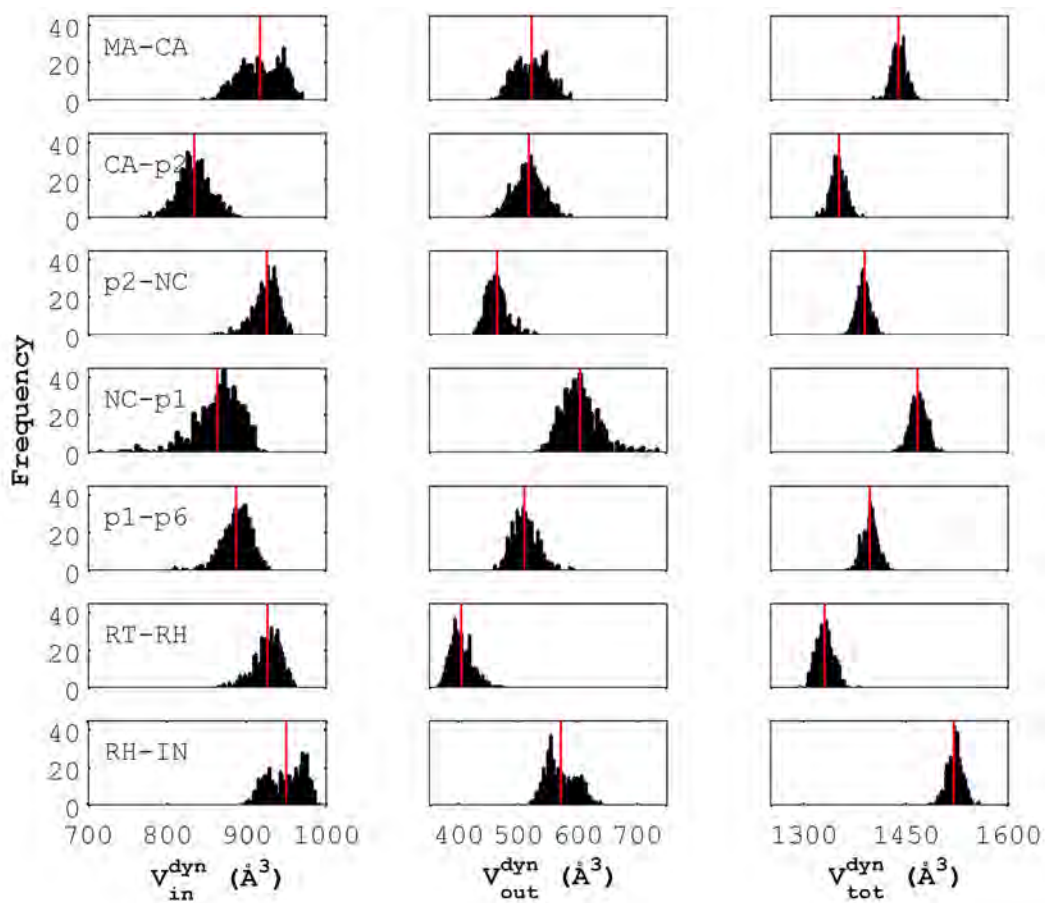


Figure 2.9

How a particular substrate fits within the substrate envelope is influenced by both substrate dynamics and size. In general, substrates with bulky side-chains protrude more extensively beyond the substrate envelope than smaller ones. To test this expectation, the change in V_{out} was plotted as a function of V_{tot} (Figure 2.10). The overall volume of a particular substrate correlates with how much that substrate protrudes beyond the substrate envelope, fitting a straight line with $R^2=0.88$. However, CA-p2, NC-p1, and p1-p6 are the exceptions; these substrates protrude beyond the dynamic substrate envelope more than one would predict based on their size. This behavior is explained by these substrates being more dynamic. Of these three substrates, p1-p6 is the least dynamic and comes closest to fitting the regression line in Figure 2.10. More dynamic substrates sample a wider conformational space, resulting in a higher deviation from the crystal structure and worse fit within the substrate envelope. Mean-square fluctuations of substrate residues were calculated for non-hydrogen atoms: (1) for the backbone, (2) for side-chain, and (3) for the entire residue (Figure 2.11). Correlations were then calculated between these fluctuations and the extent to which each residue protruded beyond the dynamic substrate envelope (Figure 2.12). Among the seven substrates, the greatest variation is in the center of mass fluctuations of their respective side-chains. CA-p2, NC-p1, and p1-p6 are all more dynamic and protrude more from the envelope. For MA-CA, on the other hand, both the substrate size and flexibility/mobility appear to determine how much it protrudes beyond the substrate envelope.

Figure 2.10: Substrate size appears to determine how well the substrate fits within the substrate envelope, except for CA-p2, NC-p1, and p1-p6.

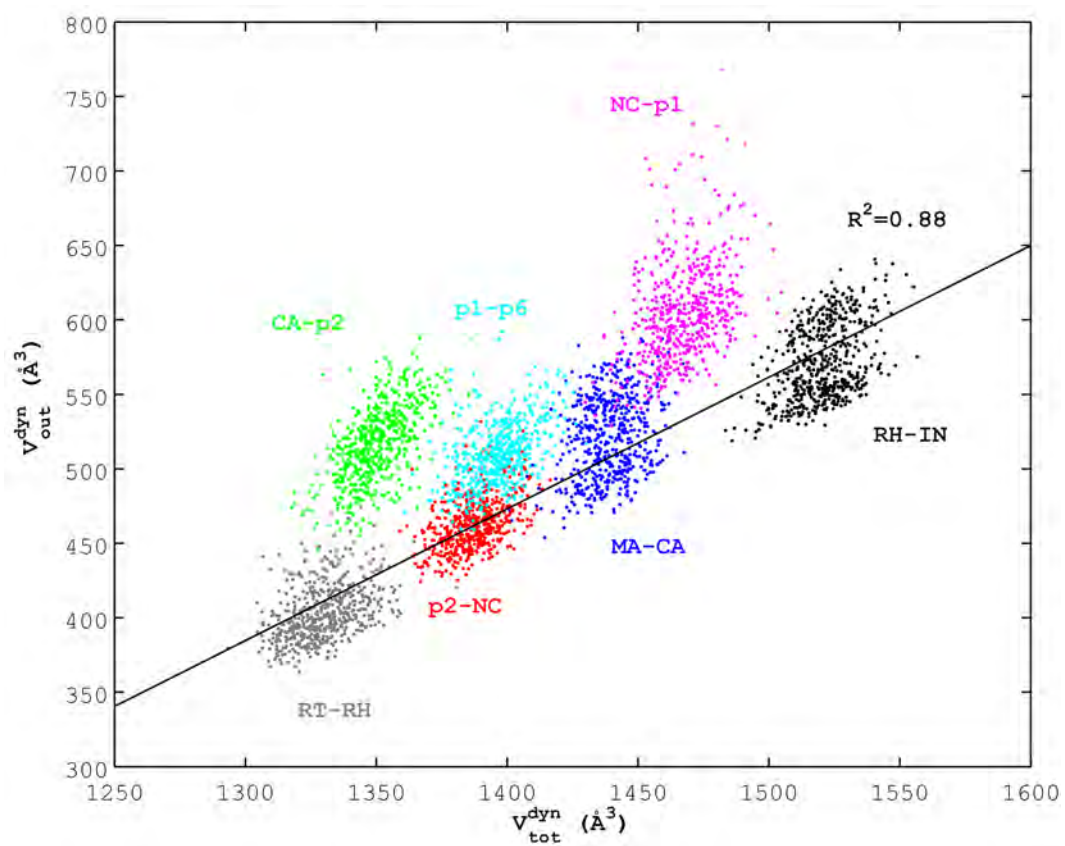


Figure 2.10

Figure 2.11: The overall dynamics of the substrate in the active site is dominated by side-chain fluctuations. Mean-square fluctuations of the center of mass for each substrate are plotted for the backbone, side-chain, and the entire substrate residue. Two substrates, NC-p1 and CA-p2, which protrude beyond the substrate envelope more than their total volume, appear to have highly dynamic centers of mass.

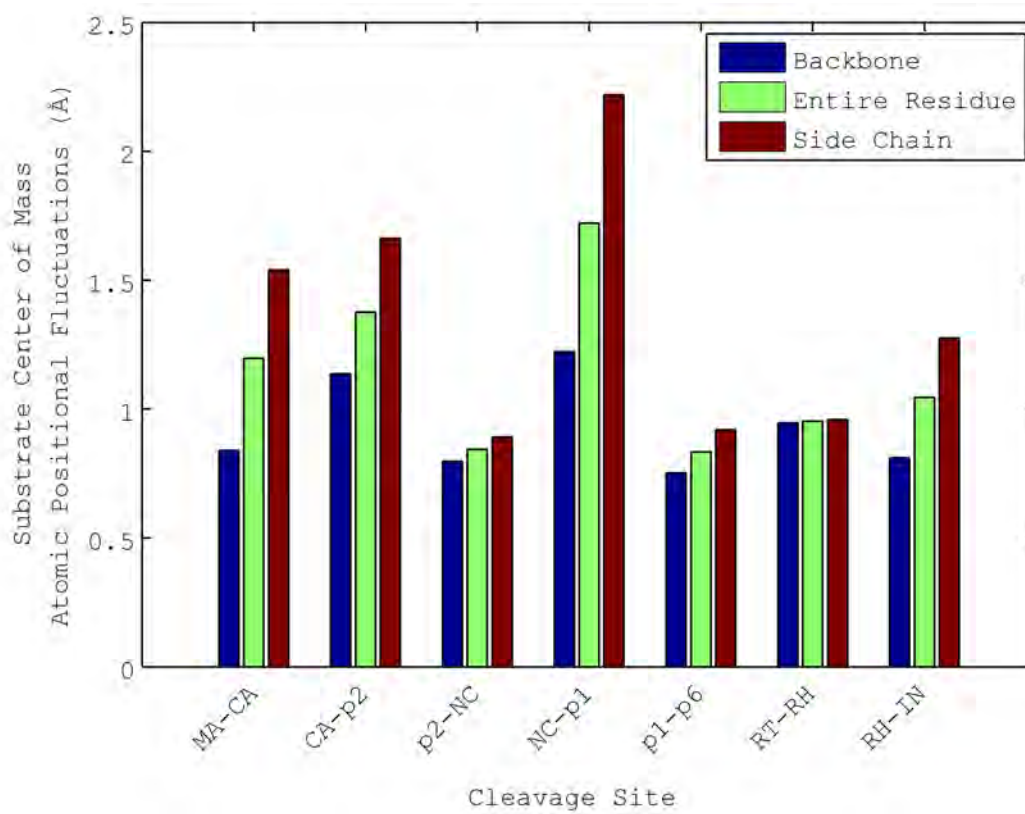


Figure 2.11

Figure 2.12: Intrinsic flexibility appears to play an important role in substrate fit within the substrate envelope for substrates MA-CA, CA-p2, NC-p1, and p1-p6. Correlation coefficients between center of mass fluctuations and V_{out} for each substrate are plotted using the backbone, side-chains, and entire substrate residues.

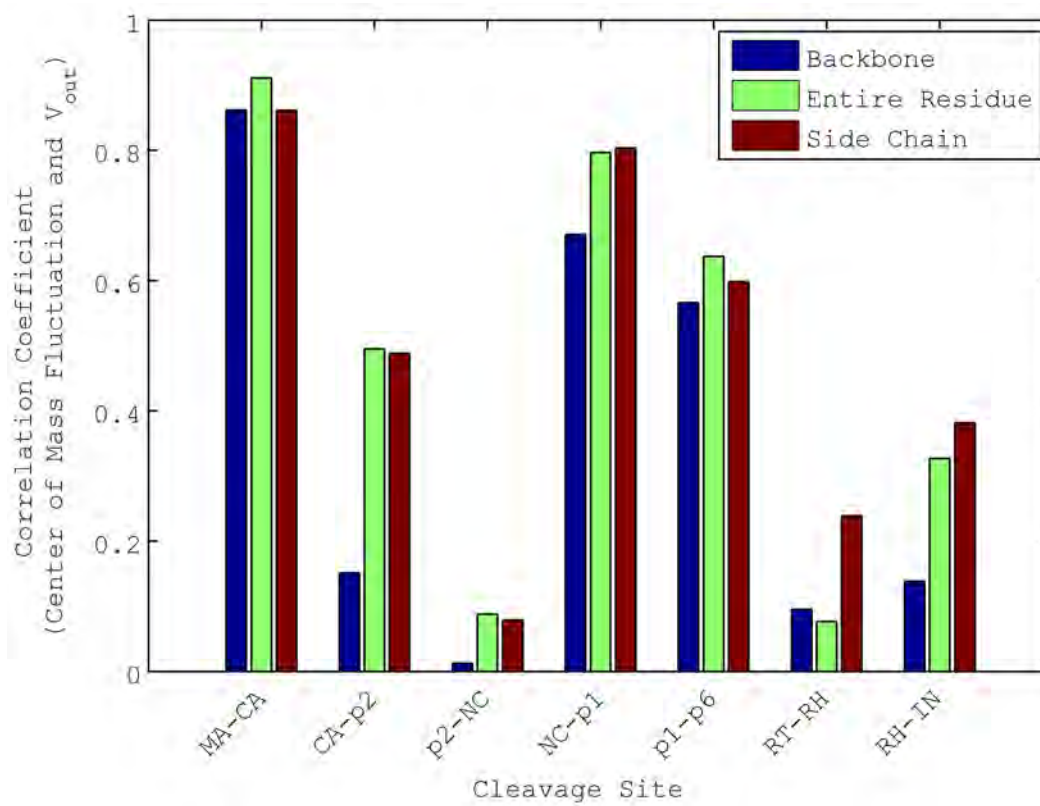
**Figure 2.12**

Figure 2.13: The dynamic substrate envelope gives a probabilistic consensus volume, which is easier and more accurate to incorporate into structure-based drug design protocols than the static envelope, which is essentially a step function. (A) Side and (B) top views of static substrate envelope, (C-D) respective views of dynamic substrate envelope. The grid cell centers are color-coded from high occupancy (red) to low occupancy (blue). Both envelopes are visualized as superposed onto the wild-type PR-CA-p2 structure. The dynamic substrate envelope is denser in the interior where the backbone is more rigid than the side-chains. The static substrate envelope can be considered homogeneous compared to the dynamic envelope. Individual substrate volumes for both the crystal structures and conformational ensembles are also shown in Figure 2.14.

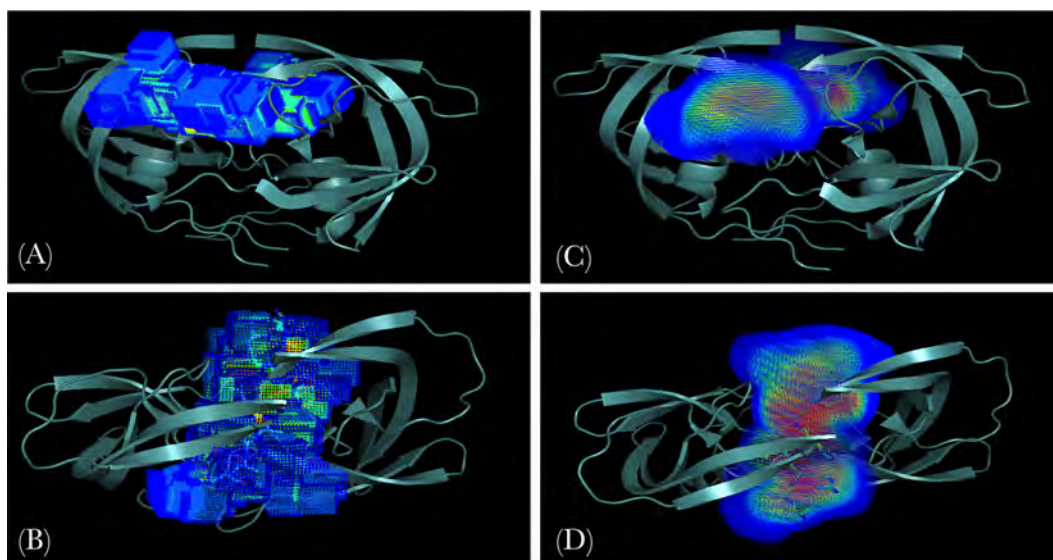


Figure 2.13

Figure 2.14: Static and dynamic shapes of HIV-1 protease substrates. The vdW volume of each substrate was generated using the *map_set* function of PyMOL from the crystal structure and compared to the dynamic vdW volume from the MD simulation.







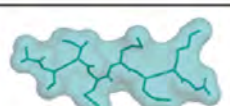
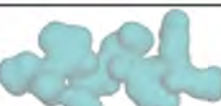

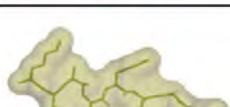





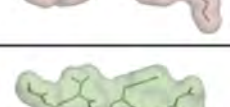


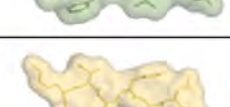

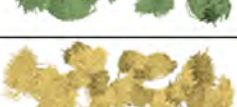
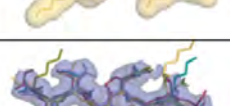

	Crystal Structures	Conformational Ensemble	
		Consensus volume	Conformers from MD
MA-CA			
CA-p2			
p2-NC			
NC-p1			
p1-p6			
RT-RH			
RH-IN			
Substrate Envelope			

Figure 2.14

Mapping numerous substrate conformations onto a three-dimensional grid defines a probabilistic substrate envelope rather than a deterministic envelope defined by discrete boundaries. The static substrate envelope [7, 8] and the dynamic substrate envelope are visualized in (Figure 2.13). The probabilistic consensus volume of the dynamic substrate volume was color-coded by occupancy of the grid cells, red being highly occupied and blue being less occupied. As in the static envelope the dynamic substrate envelope is better defined in the positions close to the cleavage site, but the envelope becomes less well defined at the substrate end residues (P4 and P4'), which are exposed to solvent and highly flexible. Unlike in the static representation, this ensemble envelope transitions smoothly with the relative degree of occupancy.

The dynamic substrate envelope reproduces the essential characteristics of the static substrate envelope. This more realistic dynamic envelope, however, appears to be slightly smaller than the static one. The reason is that, in MD simulations, unlike crystal lattices, atoms are allowed to fluctuate in solvated periodic boundary conditions, reducing the residence time in the native state. In a crystal structure, the probability of an atom being in an exact set of coordinates is 0 or 1 unless the electron density is ambiguous. However, in an ensemble of conformations generated in MD simulations, this probability ranges from 0 to 1. As a result, the dynamic substrate envelope is more heterogeneous than the static substrate envelope, being denser in the interior where the backbone is relatively rigid and less dense towards the solvent-exposed regions where the side-chains are relatively more flexible. Consequently, the consensus volume, which distinct substrate pep-

tides adopt, becomes smaller than the crystal structures impose. Therefore, V_{out}^{dyn} values are greater than those of V_{out}^{stat} . This finding needs to be considered in designing protease inhibitors, which fit better within the substrate envelope, in order to minimize the likelihood of the emergence of drug resistance.

Individual analysis of amino acids for their fit into the substrate envelope is also likely to help understand the nature of and to more realistically characterize this probabilistic substrate envelope. Amino acids with larger and/or more mobile/flexible side-chains generally contribute more to the overall V_{out} than other amino acids in the corresponding residue of other substrates. This observation was quantified by decomposing the overall protrusion of each substrate into single-residue protrusions (Figure 2.15). In general, when an amino acid residue of a particular substrate is larger than the corresponding amino acids in the same position of the other substrates, that particular substrate tends to protrude more than others in that position. Another general observation is that asymmetry around the cleavage site is preserved in terms of residue-based V_{out} values just as for vdW contacts and mean-square fluctuations (except for RT-RH which is, to a certain extent, symmetric in its V_{out} values, Figure 2.15).

Figure 2.15: Residue-based V_{out} values for (A) MA-CA, (B) CA-p2, (C) p2-NC, (D) NC-p1, (E) p1-p6, (F) RT-RH, and (G) RH-IN. The amino acids with both flexible and bulky side-chains, i.e., Arg and Lys, tend to protrude beyond the substrate envelope more than other amino acids with fewer degrees of freedom and/or smaller side-chains, i.e., Met, and/or Val, Ala. The P4 and P4' positions usually have high V_{out} values, likely due to the less well-defined characteristic of the substrate envelope in the more solvent-exposed regions of these substrate end residues. Some substrates, however, have higher V_{out} values even in their more buried positions, i.e., P1' in NC-p1, P1 in p1-p6. According to our hypothesis, these substrates co-evolve with drug-resistant PR because their greater protrusion beyond the substrate envelope makes them more sensitive to changes in the PR.

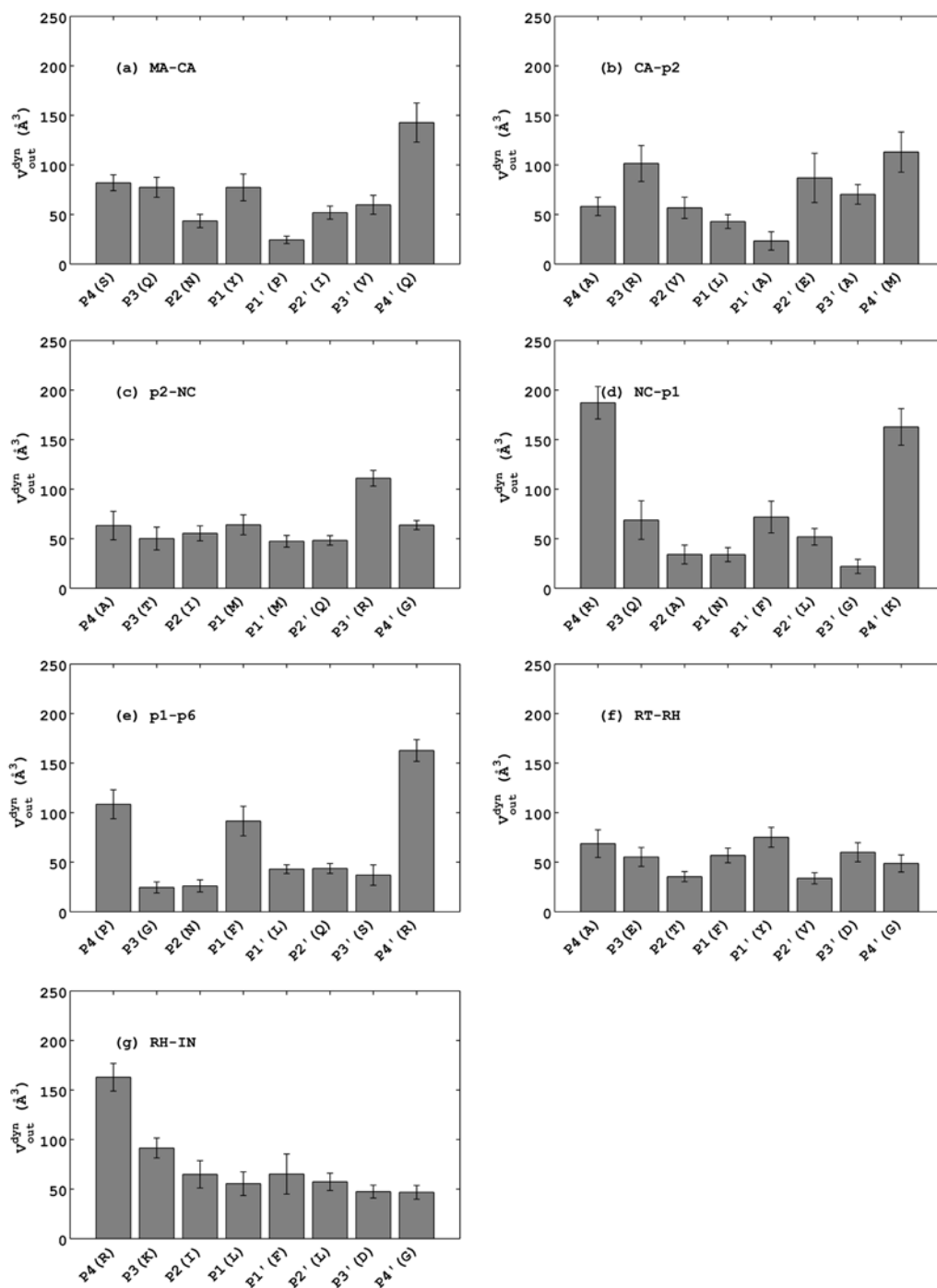


Figure 2.15

A detailed look at the individual substrate residues in terms of their fit within the substrate envelope explains the high-level p1-p6 protrusion despite its smaller absolute volume compared with the other peptides. The results of this analysis show that p1-p6 protrudes extensively beyond the dynamic substrate envelope mainly at positions P4, P1, and P4'. (Figure 2.15). The residues at these positions are Pro, Phe, and Arg, respectively. Arg and Phe are both large amino acids. Pro at P4 and Arg at P4' fluctuate the most among the eight residues of the whole p1-p6 sequence. Although the residues at P4 and P4' protrude beyond the substrate envelope, the co-evolutionary mutations in p1-p6 in response to drug-resistant protease mutations occur mainly at P1' and P3' positions, which fit better within the substrate envelope than P4, P1, or P4'. One possible explanation for this observation is the interdependency among substrate residues discussed above regarding vdW and dynamic cross-correlation analysis. Because of this interdependency, a mutation at one position may result in another position fitting better within the substrate envelope. Future investigation of mutant variants will elucidate this phenomenon.

2.4 Discussion

HIV-1 protease recognizes a conserved shape in non-homologous cleavage sites, yet the cleavage sites are processed differentially. In this paper, we shed light on this balance in substrate processing by investigating the effect of substrate dynamics on the protease-substrate interactions, cooperative motions of the complex systems, and the individual and consensus substrate volumes.

The protease-substrate interactions in the complex structure are fairly stable during the time course of MD simulations. Most of the interactions oscillate around their measured value in the crystal structure. The substrates share a conserved pattern in their backbone hydrogen bonds. This highly conserved pattern of backbone hydrogen bonds may be a prerequisite for recognition by the protease. However, no single side-chain hydrogen bond is conserved across the substrates. This variation in the side-chain hydrogen-bonding pattern likely causes the substrates to be recognized and/or processed differentially. The balance between the conserved backbone and varied side-chain hydrogen bonding characteristics makes the substrates all recognizable to the protease, yet uniquely processed, contributing to the accurate and precise cleavage. One caveat and an area for future investigations is that all our simulations were performed with the inactive D25N HIV protease. The active form of the enzyme may alter the overall stability of the hydrogen bonds, but only would exist transiently before cleavage. In the inactive form simulated here the highly stable and well-conserved backbone hydrogen bonds contributes to the conserved consensus substrate volume in the active site, by holding the substrates in place and stabilizing their shared conformational preference, we expect this conformation is relevant for the active enzyme as well.

The crystal structures provide a good representation of the vdW contact potentials between the protease and the substrates. The extent of the fluctuation around the crystal structures varies for each protease/substrate residue pair due to the variations in flexibility across the complex structure. However, the high dynamic stability of these interactions provides insights to why the rigid structural studies

have been so far successful in structure-based design of protease inhibitors, especially considering the fact that protease inhibitors have less degree of freedom than the natural substrates of HIV-1 protease and therefore less flexible.

The same overall vdW interaction potential with the protease is maintained by substrates, although each one has a unique residual profile. HIV-1 PR differentiates among its natural substrates both in cellular and molecular contexts. [146, 148] This variation in substrate processing may be influenced by the variation in the distribution of the interaction potential along the cleavage site. That is, the diversity in vdW contact potentials of substrate residues may contribute to preferential substrate recognition. While the overall vdW interaction potential may be a prerequisite for recognition by the protease, the substrate specific sequences may regulate their order of processing as well as the reaction kinetics.

The protease residues that interact with the substrates in the active site are usually evolutionarily conserved since they are functionally important. Drug resistance mutation sites 30, 48, 50, and 82 are among the exceptions. The protease interacts with one or more of the natural substrates in these positions. As a result, mutations at positions 30, 50, and 82 are associated with p1-p6 and NC-p1 cleavage site mutations. Thus, predicting co-evolution is possible through a combined investigation of the protease-substrate interaction profiles and evaluating the fit of the wild-type and co-evolved substrate variants within the substrate envelope. For example, G48 interacts extensively with the substrates, suggesting the natural substrates will likely co-evolve when G48 mutates to confer resistance. This prediction awaits statistical confirmation by analyzing viral sequences from patients

failing therapy for associations between mutations in position 48 and in cleavage sites. Such predictions may validate our understanding of the substrate recognition and co-evolution.

The interplay between the conserved and varied properties of non-homologous cleavage sites enables the substrate processing to be regulated. A sequence is required to possess conserved properties to be recognized by the HIV-1 PR, yet the unique properties of the sequences allow for sequential substrate recognition. This interplay plays a key role in the tight regulation of gag and gag-pro-pol processing.

Non-homologous natural substrates of HIV-1 protease maintain conserved properties through interdependency within their residues. Among these properties the most striking one is their ability to adopt a conserved consensus volume in the active site despite of the lack of an obvious pattern in their sequence. Moreover, these separate substrates have a conserved vdW interaction potential with the protease, although the interaction profiles along their sequences do not follow a common trend. In addition, these substrates are dynamically coupled to the same protease residues, although the intrinsic conformational dynamics of the bound substrates are highly varied.

This adaptation ability, which is common for all non-homologous substrates investigated in this study, should originate from the interdependent characteristics of the cleavage site residues. This interdependency enables the protease to recognize a series of non-homologous sequences as specific substrates. Interdependency may not be sufficient, but appears to be necessary for substrate recognition. The combination of the low homology in cleavage site sequences and interdepen-

gency among their residues appear to contribute to preferential substrate processing by both providing a common basis for substrate specificity and making each substrate complex unique. This uniqueness results in highly specific sequential substrate processing.

Dynamic Substrate Envelope: A more realistic representation of the consensus volume. The substrate envelope hypothesis, which is based on crystallographic data, explains the structural basis for the significantly represented mutation patterns in HIV-1 PR and its substrates. The crystal structure represents a static image of the native state of a protein. Protein recognition, however, is inherently a dynamic event as proteins adopt an ensemble of different conformations around their native states. Consideration of this flexibility around the native state is essential to arrive at a better understanding of the specificity determinants of substrate recognition. Introducing molecular dynamics into the substrate envelope allows a more realistic evaluation of the substrate envelope hypothesis.

Dynamic Substrate Envelope: A potential tool for structure-based drug design. Combining the molecular dynamics simulations with the three-dimensional grid-based approach to model the substrate envelope allows the construction of a probabilistic consensus volume. The edges of this probabilistic consensus volume are not defined by a sharp cutoff. The dynamic substrate envelope is essentially a smooth probability distribution while the static substrate envelope is a step function. This probability distribution function can easily be incorporated as a constraint into the potential functions in structure-based computer-aided drug-design algorithms, especially grid-based docking softwares like AutoDock. [149]

Optimizing this energetically smoother potential function to minimize V_{out} will be more feasible than fitting candidate inhibitor molecules into a homogeneous substrate envelope with discrete boundaries. Ultimately, optimizing the dynamic substrate envelope as a part of a potential function should be an easier and more accurate way to incorporate the substrate envelope into structure-based drug design.

2.5 Methods

2.5.1 Protease-Substrate Complex Structures

Seven PR substrates in complex with an inactive, isosteric PR variant were analyzed. Five of these peptides correspond to substrate sequences within the gag polyprotein (matrix-capsid [MA-CA], capsid-p2 [CA-p2], p2-nucleocapsid [p2-NC], nucleocapsid-p1 [NC-p1] and p1-p6), and two correspond to substrate sequences within the gag-pro-pol polyprotein (reverse transcriptase-RNaseH [RT-RH] and RNaseH-integrase [RH-IN]). The crystal structures of protease bound to CA-p2 [6], MA-CA, p2-NC, p1-p6, RT-RH, and RH-IN [7], and NC-p1 [40] were used as the starting structures in the MD simulations (PDB ID: 1F7A, 1KJ4, 1KJ7, 1KJF, 1KJG, 1KJH, 1TSU).

2.5.2 MD Simulations

The AMBER [150, 151] simulation package (version 8) with the *ff03* force field was used in all simulations. All structures were solvated explicitly in a truncated octahedron box using the TIP3P water mode. [152] Each system was neutralized by adding Cl^- counterions using a Coulombic potential on a 1 Å grid with the preparatory program *tleap* of AMBER. The initial structures were first minimized at constant volume with convergence criteria of either maximum 90,000 cycles or an RMSD value of 0.01 Å by steepest descent integration algorithm for 50 steps and then switched to conjugate gradient algorithm. Bonds involving hydrogens were constrained by the SHAKE algorithm [153] with a relative geometrical tolerance of $10\text{E-}5$ Å. Initial atom velocities corresponding to a temperature of 10 K were generated from a Maxwellian distribution and the temperature was gradually raised to 300 K. The temperature was maintained at 300 K, and the pressure was maintained at 1 bar by the Berendsen weak-coupling approach. [154] Constant pressure-periodic boundary conditions were used with isotropic position scaling. The particle mesh Ewald (PME) method [155] was used to calculate the full electrostatic energy of a periodic box, bypassing pair list creation and non-bonded force and energy evaluation by calling special PME functions to calculate the Lennard-Jones and electrostatic interactions with a cutoff distance of 9 Å. A time step of 2 fs was employed in the Leapfrog integrator. Coordinates and energies were written every 0.4 ps. All the complexes were simulated for 11 ns, during which root-mean-square-deviation of backbone trace were monitored for

convergence. From these simulations, snapshots, taken at regular intervals of 20 ps, were used to model the dynamic substrate envelope, calculate V_{in} and V_{out} and to investigate the molecular interactions. Time evolution of V_{out} was also investigated for convergence.

2.5.3 Modeling of the Dynamic Substrate Envelope

The substrate envelope was previously modeled by a quantitative approach to evaluate the fit of protease inhibitors within the envelope. [41] This approach was based on a three-dimensional grid with side length 10 Å and grid spacing 0.2 Å centered on the active site. The occupancy of each grid cell, g_{ijk} , was assigned an initial value of 0. Then the number of times that a grid cell was within the vdW volume of any peptide was counted. This value eventually became the overall occupancy of that grid cell. The total volume of the substrate envelope was essentially the summation over all grid cells with occupancy greater than 0 normalized by the total number of structures used.

In this study, this quantitative model was modified in two essential ways. The most important modification was to incorporate the molecular dynamics into the static substrate envelope, which was defined by the PR-substrate complex crystal structures. To include protein dynamics in this model, multiple conformations for each PR-substrate complex were extracted from MD trajectories at different time frames. These conformations were mapped onto the grid centered in the active site, instead of the crystal structures. This approach defined the dynamic substrate envelope.

The second modification was in the method for superimposing these multiple conformations, which is extremely important as it directly determines the consensus volume. The best way to see conformational differences in a series of structures is to overlay them using the most rigid nodes in the structure; otherwise the motion could be propagated to regions that are relatively immobile. Therefore, instead of overlaying the PR-substrate structures based on their backbone atoms, we determined the common hinge residues in protease bound to various substrates from MD trajectories. When the trajectories of all seven complex structures were analyzed, we found that the most immobile backbone carbons in both monomers were residues 24-26 and 85-90. Using these residues, all frames were superimposed on the wild-type protease-CA-p2 complex structure by the RMSD trajectory tool of the molecular visualization and trajectory analysis software, VMD. [156]

With this modified model for the substrate envelope, the substrates were then compared to each other in terms of how much they protruded beyond the static versus dynamic substrate envelopes. Modeling the substrate envelope using this grid-based approach allows construction of a probabilistic consensus volume without sharp edges, which is likely more realistic than the static substrate envelope. The static substrate envelope was previously modeled using the mathematical formulation below. [41] The volume of an individual substrate lying outside the substrate envelope, V_{out} , was calculated as follows, where d is the edge length of a single grid cell and N is the total number of structures:

$$V_{out}^{stat} = \frac{d^3}{N} \sum_{i,j,k}^{inside} (N - g_{ijk}) \quad (2.1)$$

Similarly, the volume of a substrate that fits within the substrate envelope was computed as

$$V_{in}^{stat} = \frac{d^3}{N} \sum_{i,j,k}^{inside} (g_{ijk}) \quad (2.2)$$

Based on this model, equations were redefined in this study for the dynamic substrate envelope. For that, the vdW volume of an individual substrate was computed from N_1 time points in the MD trajectory of that substrate. An initial value of 0 was again assigned to each grid cell. Then a value of $g_{ijk,1}$ was incremented by 1 for every snapshot where a grid cell (i, j, k) was considered to be contained by a substrate frame if it lay within the AMBER vdW radius of any non-hydrogen atom of the substrate. The resulting grid occupancy values range between 0 (outside all conformations of the individual substrate) and N_1 (inside all conformations of the individual substrate), which is the total number of time points from a single MD simulation. In the meantime, another variable, $g_{ijk,2}$, was updated to map the dynamic substrate envelope on the grid. $g_{ijk,2}$ values vary between 0 (outside all conformations of all substrates) and N_2 (inside all conformations of all substrates), which is the total number of frames from all seven simulation trajectories ($= 7 \times N_1$).

The effective volume of an individual substrate outside the dynamic substrate envelope (V_{out}^{dyn}) was computed by summing the values of the grid points $g_{ijk,1}$ multiplied by $N_2 - g_{ijk,2}$, which is a measure of how dense the dynamic substrate

envelope is at that grid point, and normalizing this sum by $N_1 \times N_2$:

$$V_{out}^{dyn} = d^3 \sum_{i,j,k}^{\text{all grid points}} \frac{(N_2 - g_{ijk,2})}{N_2} \times \frac{g_{ijk,1}}{N_1} \quad (2.3)$$

The final result for V_{out}^{dyn} was obtained after conversion to volume by multiplying by the volume of a single grid cell d^3 . The volume of the substrate within the dynamic substrate envelope was computed as follows:

$$V_{in}^{dyn} = d^3 \sum_{i,j,k}^{\text{all grid points}} \frac{g_{ijk,2}}{N_2} \times \frac{g_{ijk,1}}{N_1} \quad (2.4)$$

The total volume is computed by summing all occupancy values, multiplying by d^3 , and normalizing by the number of structures used. This total volume for each substrate was compared to the sum of V_{out} and V_{in} as a control for numerical errors.

These new equations allow comparison of two probabilistic vdW volumes quantifying the fit of an individual substrates probabilistic vdW volume within the probabilistic dynamic substrate envelope.

Finally, the defined structures of the peptides in the crystal structures span slightly different lengths of the cleavage sites including or not including P5 or P5'; however, these differences likely have minimal or no impact on the bound conformation of the substrate. [40] These positions are not buried in the active site, but are highly solvent exposed, and in some crystal structures they were not modeled, as these residues showed no obvious electron density. If P5 or P5' residues were

ordered in the crystal structure, they were included in the simulations. However, they were highly mobile and their coordinates were not included in calculating the substrate envelope. Only the region P4-P4' was used for the substrate envelope calculation.

2.5.4 Estimation of the vdW Potential

How a particular ligand interacts with a particular protease variant can be assessed in terms of their detailed molecular interactions, i.e., hydrogen bonds and vdW contact potentials. While hydrogen bonds can be easily counted based on distance and geometry, assessing vdW contacts accurately is a more subtle analysis. A simple distance criterion of 4.2 - 5.0 Å often does not accurately distribute the energetic contributions. In this study, substrate-protease vdW contacts were estimated by a simplified Lennard-Jones potential $V(r)$ using the relation $4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, where r is the PR-substrate interatomic distance, and ϵ and σ are the well depth and hard sphere diameter, respectively, for each PR-substrate atom pair. $V(r)$ for all possible PR-substrate atom pairs was computed within 5 Å, and when the distance between non-bonded pairs was less than ϵ , $V(r)$ was considered equal to ϵ . Using this simplified potential for each non-bonded PR-substrate pair, $\sum V(r)$ was then computed for each protease and each substrate residue.

2.5.5 Evaluation of Hydrogen Bonding

A hydrogen bond was defined by a distance between the donor and the acceptor of less than 3.5 Å and a donor-hydrogen-acceptor angle of greater than 120°. The percentage of time these hydrogen bonds existed was calculated using the trajectory analysis utility *ptraj*.

2.5.6 Fluctuation Dynamics

To see the similarities and differences in flexible regions of PR-substrate complexes, atomic positional fluctuations were computed based on MD simulation trajectories. The cross-correlations of these fluctuations were also calculated to assess how the protease and substrate dynamics are coupled and if the protease dynamics is affected in a different manner by binding of different substrates. The normalized cross-correlations of residue pairs were defined as

$$CO_{i,j} = \frac{\langle \Delta R_i \Delta R_j \rangle}{\langle \Delta R_i^2 \rangle^{1/2} \langle \Delta R_j^2 \rangle^{1/2}} \quad (2.5)$$

where ΔR_i is the fluctuation in the position vector R of site i and ΔR_j is the fluctuation in the position vector R of site j . The brackets represent time averages over recorded snapshots. The cross-correlations vary in the range $[-1, 1]$, with the lower and upper limits indicating fully anti-correlated and correlated atomic fluctuations, respectively. $CO_{i,j} = 0$ gives uncorrelated atomic fluctuations.

2.6 Supporting Information

Assessing the impact of structural alignment on the substrate envelope

Structural alignment may affect the shape of the dynamic substrate envelope. Aligning the structures on their mobile residues may misrepresent the rigid regions as being highly dynamic and reduce the consensus volume defined as the volume occupied by the majority of the conformers. Hence, the alignment should be based on relatively less mobile residues. However, even using different subsets of these relatively less mobile regions as reference for structural alignment could have affected the results. To address this issue, we redefined the wild-type dynamic substrate envelope based on five separate structural alignments using the C_{α} atoms of (1) all protease residues (1-99), (2) residues on the dimerization interface excluding the flaps as flaps are highly mobile (1-9, 86-99), (3) the least mobile residues in the MD simulations (24-26, 86-90), (4) catalytic triad (25-27), and finally (5) only the highly flexible GLY-rich region of the flaps (48-51). The purpose of the first four alignments was to probe for the effect of aligning different sets of rigid residues on the dynamic substrate envelope. The last one was performed to see the effect of using more mobile residues during superposition on the shape of the dynamic substrate envelope.

All the trajectories were aligned onto the wild-type CA-p2 crystal structure using the VMD software. The substrate conformers from the aligned trajectories were loaded to PyMOL and the van der Waals volume maps were generated for each substrate and the dynamic substrate envelope. The consensus vdW volumes

defined by each set of conformers are illustrated in Figure 2.16.

These results show that the shape of the dynamic substrate envelope is not sensitive to the selection of the reference residues for structural alignment as long as these reference residues are not in a highly mobile region of the protease. However, aligning the structures on the very flexible GLY-rich flap residues resulted in a slightly smaller dynamic substrate envelope as expected.

Figure 2.16: Effect of structural alignment on the substrate envelope. The effect of residue selection on the shape of the dynamic substrate envelope was assessed by aligning the trajectories using the $C\alpha$ atoms of all protease residues, dimerization interface excluding the flaps, catalytic triad, and finally only the flaps. The results showed that the shape of the dynamic substrate envelope is not sensitive to the selection of the reference residues for structural alignment as long as these reference residues are not on a highly mobile region of the protease.









































Consensus vdW Volume					
Substrate	Alignment #1 Protease Backbone 1-99	Alignment #2 Dimerization Interface 1-9, 86-99	Alignment #3 Least Mobile Residues 24-26, 86-90	Alignment #4 Catalytic Triad 25-27	Alignment #5 GLY-Rich Flap Region 48-51
MA-CA					
CA-p2					
p2-NC					
NC-p1					
p1-p6					
RT-RH					
RH-IN					
DSE					

Figure 2.16

Chapter III

HIV-1 Protease and Substrate

Co-evolution Validates the Substrate

Envelope as the Substrate

Recognition Pattern

3.1 Abstract

Drug resistance of HIV-1 protease alters the balance in the molecular recognition events in favor of substrate processing versus inhibitor binding. To develop robust inhibitors targeting ensembles of drug resistant variants, the code of this balance needs to be cracked. For this purpose, the principles governing the substrate recognition are required to be revealed. Previous crystallographic studies on the wild-type protease-substrate complexes showed that the substrates have a conserved consensus volume in the protease active site despite their low sequence homology. This consensus volume is termed as the substrate envelope. The substrate envelope was recently reevaluated by taking the substrate dynamics into account and the dynamic substrate envelope was reported to better define the substrate specificity for HIV-1 protease. Drug resistance occurs mostly through mutations in the protease, occasionally accompanied by cleavage site mutations. In this study, three co-evolved protease-substrate complexes (AP2V NC-p1 $_{V82A}$, $^{LP1'F}$ p1-p6 $_{D30N/N88D}$ and $^{SP3'N}$ p1-p6 $_{D30N/N88D}$) were investigated for structural and dynamic properties by molecular modeling and dynamics simulations. The results show the substrate envelope is preserved by these cleavage site mutations in the presence of drug-resistance mutations in the protease, if not enhanced. This study on the conformational and mutational ensembles of protease-substrate complexes validates the substrate envelope as the substrate recognition motif for HIV-1 protease. The substrate envelope hypothesis allows for the elucidation of possible drug resistance mutation patterns in the polyprotein cleavage sites.

3.2 Introduction

Human immunodeficiency virus type-1 (HIV-1) protease (PR) is a key enzyme in the viral life cycle that processes the Gag and Gag-Pro-Pol viral polyproteins at 12 cleavage sites, yielding mature, infectious virions. [10] HIV-1 PR, a 99 residue, homodimeric aspartyl protease, [157, 158] is essential for viral maturation, [10] hence is a main drug target. The US Food and Drug Administration (FDA) has approved nine protease inhibitors (PIs) for clinical use. These PIs, used as part of highly active antiretroviral therapy, have significantly improved the management of disease condition by lengthening the life-span and increasing the quality of life of HIV-infected patients. [159] However, the high rate of viral replication [22] combined with the lack of proofreading mechanism in viral reverse transcriptase [160] generates massive amounts of genetically distinct viral variants. Within these viral quasispecies, the selective pressure of drug therapy populates the viral variants that are not completely inhibited by antiviral drugs targeting viral enzymes.

At a molecular level, drug resistance reflects a subtle change in the balance of enzyme recognition events, in favor of natural substrate processing versus inhibitor binding. Drug-resistant protease variants have mutations that significantly alter inhibitor binding but do not drastically change substrate processing. In addition, emergence of resistance to protease inhibitors does not always depend solely on protease mutational plasticity. The natural substrates can also mutate in association with drug therapy. [29, 30, 37, 161] Two examples of this protease-substrate

co-evolution are NC-p1 and p1-p6, two cleavage sites on the Gag polyprotein that co-evolve with the viral protease to confer resistance to protease inhibitors. In the NC-p1 cleavage site, Ala in the P2 position mutates to a Val in response to the V82A multi-drug-resistance protease mutation (AP2V NC-p1 $_{V82A}$). [28,29,37,162] The p1-p6 cleavage site mutates predominantly at the P1' or P3' positions [39] associated with the protease double mutation D30N/N88D ($^{LP1'F}$ p1-p6 $_{D30N/N88D}$ and $^{SP3'N}$ p1-p6 $_{D30N/N88D}$), which is a signature of nelfinavir treatment. [163,164] Replicative capacity assays showed that the D30N/N88D viruses with the compensatory mutations in p1-p6 do not have improved fitness relative to viruses with mutations in the protease alone. [32] V82A virus has an even lower replicative capacity in combination with mutations at Gag 431, which corresponds to Ala-P2 of NC-p1 cleavage site, compared to those without this mutation. [32] However, these co-occurring protease-substrate mutations were shown to significantly affect the protease inhibitor susceptibilities. [32]

The structural basis for protease-substrate co-evolution in AP2V NC-p1 $_{V82A}$ variant came from analyzing the crystal structures of inactive D25N wild-type (WT) and V82A HIV-1 protease in complex with their respective WT and AP2V mutant NC-p1 substrates. [7] The crystal structures revealed that WT NC-p1 binds to HIV-1 protease less optimally than AP2V NC-p1 with fewer hydrogen bonds and fewer van der Waals (vdW) contacts. In addition, Ala-P2 was observed to incompletely fill the P2 pocket. For protease-substrate co-evolution in $^{LP1'F}$ p1-p6 $_{D30N/N88D}$ or $^{SP3'N}$ p1-p6 $_{D30N/N88D}$ variants, however, no experimentally determined structures have been reported so far.

Completely understanding the molecular basis of substrate recognition is crucial to develop robust inhibitors that better compete with natural substrates of highly resistant protease variants. Co-evolution of protease and the natural substrates allows for the study of the interdependency between HIV-1 protease and its natural substrates, which facilitates the substrate recognition. The principles underlying substrate recognition by HIV-1 protease are not sequence specific because the amino acid sequences of the cleavage sites on the Gag and Gag-Pro-Pol polyproteins do not have an obvious sequence homology. These non-homologous substrates, however, occupy a conserved consensus volume in the binding site of the protease in crystal structures. [7, 8] This conserved three-dimensional shape, describing the “substrate envelope”, was recently redefined by incorporating the substrate dynamics within the protease binding site. [165] In that previous study, the dynamic substrate envelope was shown to better define the substrate specificity for HIV-1 protease, compared to the static substrate envelope.

In the present study the dynamics of three separate examples of HIV-1 protease-substrate co-evolution were investigated, to elucidate the interdependence of substrate recognition with mutations in the protease. The molecular interactions of each set were analyzed, including with respect to the dynamic substrate envelope. [165] The compensatory mutations in the cleavage sites were shown to fit within the substrate envelope better than WT substrates. This was achieved by a variety of changes in interactions, but not in one conserved manner for all three substrates. Thus, the substrate envelope is preserved by the cleavage site mutations in the presence of drug-resistance mutations in the protease, validating the

substrate envelope as the substrate recognition pattern for HIV-1 protease, whether or not the substrate evolves. The dynamic substrate envelope is potentially a powerful tool to predict the co-evolution of the cleavage sites of HIV-1 protease.

3.3 Results and Discussion

The three separate cases of HIV-1 protease-substrate co-evolution investigated in this study are listed in Table 3.3. The variants of the protease and the cleavage site are denoted by a subscript and a superscript, respectively. Details of the nomenclature used throughout the paper are described in the section 3.5 section.

3.3.1 Fit within the Dynamic Substrate Envelope

Substrate fit within the binding groove was compared over the course of molecular dynamics (MD) simulations with and without the effect of drug resistance by first calculating the substrate volume within and outside the dynamic substrate envelope and the overall vdW interactions of the substrate with the protease (Figure 3.1). Substrate volume in the binding groove can be analyzed in two components: the substrate volume protruding beyond the substrate envelope (V_{out}), and the substrate volume lying within the substrate envelope (V_{in}). High degree of protrusion beyond the substrate envelope (quantified as high V_{out} values) was shown in our earlier work to indicate substrates susceptible to drug-resistance mutations in the protease, e.g., both p1-p6 and NC-p1 substrates have higher V_{out} values than expected based on their molecular volume. [165]

The current study shows that compensatory mutations in the cleavage sites optimize the portion of the substrate volume that stays within the substrate envelope. This observation is quantified in Figure 3.1A, B, and C where V_{in} is plotted for three cases of co-evolution (^{AP2V}NC -p1 $_{V82A}$, $^{LP1'F}$ p1-p6 $_{D30N/N88D}$ and $^{SP3'N}$ p1-p6 $_{D30N/N88D}$) with reference to their respective wild-type complex structure ($\Delta V_{in} = V_{in}^{MUT} - V_{in}^{WT}$). In each case, the substrate mutation compensates for the decrease in V_{in} as a result of primary protease mutation.

Figure 3.1: Protease-substrate co-evolution preserves the dynamic substrate envelope. ΔV_{in} , ΔV_{out} and overall protease-substrate $\Delta v d W$ interactions are shown in panels A-C, D-F, and G-I, respectively. The difference in these three properties for each protease-substrate variant compared to the wild-type protease-substrate complex is plotted as bars. Labels of the x-axes correspond to the mutations in either the protease (V82A or D30N/N88D) or the cleavage site (AP2V, LP1'F or SP3'N) or both. ($\Delta V_{in} = V_{in}^{MUT} - V_{in}^{WT}$, $\Delta V_{out} = V_{out}^{MUT} - V_{out}^{WT}$, $\Delta v d W = v d W^{MUT} - v d W^{WT}$)

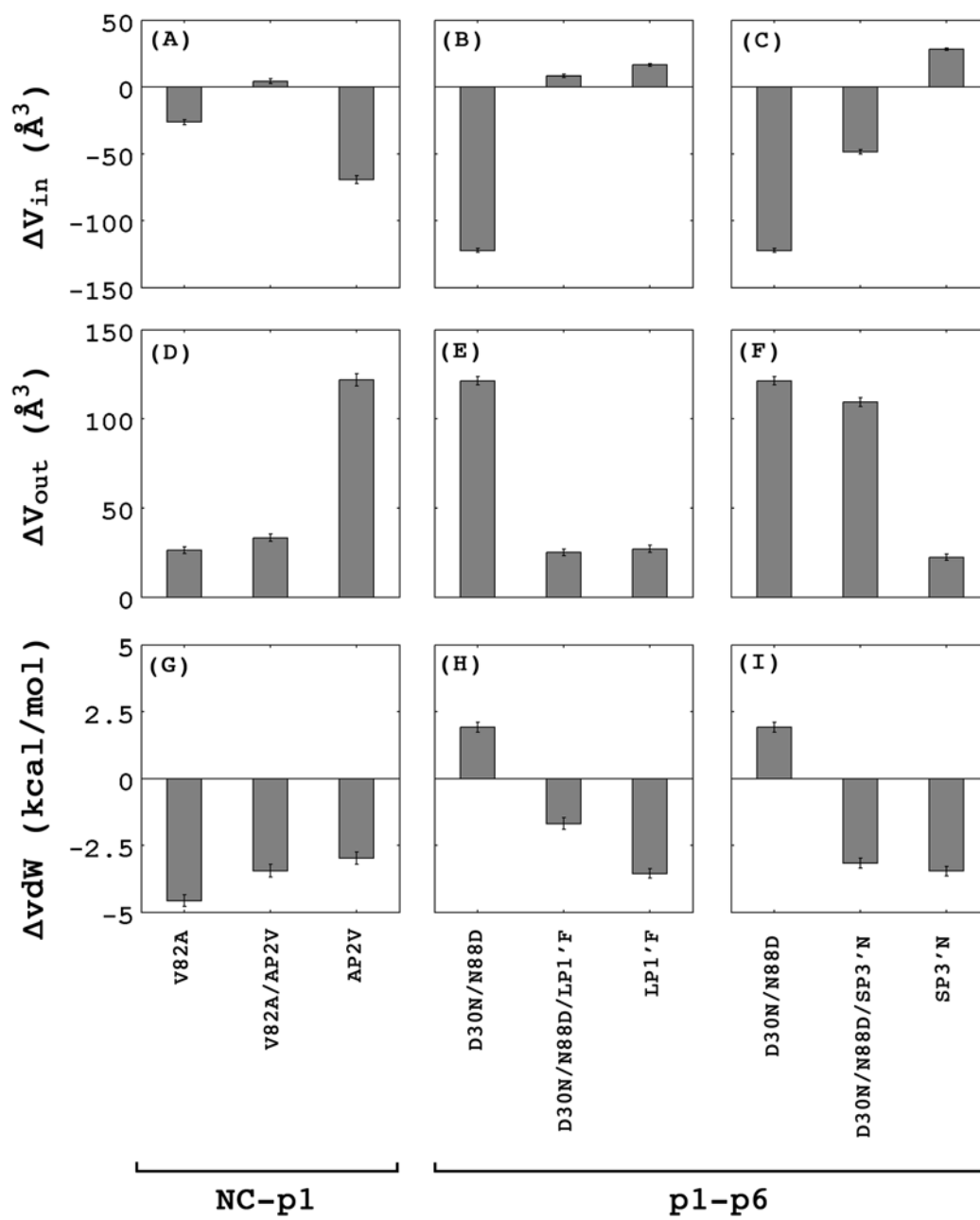


Figure 3.1

In co-evolution of NC-p1 cleavage site, ΔV_{in} increases from $-26.3 \pm 1.9 \text{ \AA}^3$ in $^{WT}\text{NC-p1}_{V82A}$ complex to $4.2 \pm 1.8 \text{ \AA}^3$ in $^{AP2V}\text{NC-p1}_{V82A}$ (Figure 3.1A) while in co-evolution of p1-p6 cleavage site, ΔV_{in} increases from $-122.2 \pm 1.5 \text{ \AA}^3$ in $^{WT}\text{p1-p6}_{D30N/N88D}$ to $8.2 \pm 1.3 \text{ \AA}^3$ in $^{LP1'F}\text{p1-p6}_{D30N/N88D}$ (Figure 3.1B) and to $-48.5 \pm 1.8 \text{ \AA}^3$ in $^{SP3'N}\text{p1-p6}_{D30N/N88D}$ (Figure 3.1C).

$^{AP2V}\text{NC-p1}$ is slightly larger than $^{WT}\text{NC-p1}$ because of two additional methyl groups on the side chain. Despite 2.5% larger volume, $^{AP2V}\text{NC-p1}$ fills the dynamic substrate envelope better than $^{WT}\text{NC-p1}$, i.e., the improvement in V_{in} ($30.5 \pm 2.6 \text{ \AA}^3$, Figure 3.1A) is more pronounced than the increase in V_{out} ($6.9 \pm 2.8 \text{ \AA}^3$, Figure 3.1D) in $^{AP2V}\text{NC-p1}_{V82A}$ compared to $^{WT}\text{NC-p1}_{V82A}$. Similarly, in the presence of D30N/N88D mutations in the protease, LP1'F and SP3'N substitutions in p1-p6 improve V_{in} by $130.4 \pm 2.0 \text{ \AA}^3$ (Figure 3.1B) and $73.7 \pm 2.3 \text{ \AA}^3$ (Figure 3.1C), respectively, making the substrate better fill the substrate envelope. On the other hand, the reduction in V_{out} is $96.2 \pm 2.8 \text{ \AA}^3$ (Figure 3.1E) and $11.8 \pm 3.3 \text{ \AA}^3$ (Figure 3.1F) upon LP1'F and SP3'N, respectively, occurs in the context of D30N/N88D protease variant. These compensatory mutations in p1-p6 optimize the substrate fit within the envelope by both increasing V_{in} and decreasing V_{out} .

Preservation of substrate envelope upon compensatory cleavage site mutations in drug-resistant protease variants is visualized in Figure 3.2 in the case of $^{LP1'F}\text{p1-p6}_{D30N/N88D}$. The altered fit of p1-p6 within the dynamic substrate envelope in $^{WT}\text{p1-p6}_{D30N/N88D}$ (Figure 3.2) and $^{LP1'F}\text{p1-p6}_{D30N/N88D}$ (Figure 3.2B) complexes is illustrated by mapping the ΔV_{in} grid matrix onto the co-crystal struc-

ture. The negative values of the difference matrix ΔV_{in} , shown in cyan, correspond to the regions where the substrate in mutant complex fits poorly within the dynamic substrate envelope compared to $^{WT}p1-p6_{WT}$. The positive values of the difference matrix ΔV_{in} , shown in orange, correspond to the regions where the substrate in mutant complex fits better within the dynamic substrate envelope compared to $^{WT}p1-p6_{WT}$. The improvement in ΔV_{in} upon the compensatory mutation in the p1-p6 cleavage site shown in Figure 3.1B is reflected by the complete loss of negative regions in Figure 3.2B.

Figure 3.2: The compensatory mutation LP1'F in the p1-p6 cleavage site preserves the substrate envelope in drug-resistant D30N/N88D protease variant.

ΔV_{in} is visualized for (A) $^{WT}p1-p6_{D30N/N88D}$ and (B) $^{LP1'F}p1-p6_{D30N/N88D}$ by mapping the difference matrix ($V^{MUT}-V^{WT}$) onto the co-crystal structure. The negative values of the difference matrix ΔV_{in} , shown in cyan, correspond to the regions where the substrate in mutant complex fits poorly within the dynamic substrate envelope compared to $^{WT}p1-p6_{WT}$. The positive values of the difference matrix ΔV_{in} , shown in orange, correspond to the regions where the substrate in mutant complex fits better within the dynamic substrate envelope compared to $^{WT}p1-p6_{WT}$. To get clear images, only the grid cells with an occupancy factor of 0.5 or higher are visualized. The protease dimer is represented as light gray ribbon. p1-p6 side-chains are shown as sticks. In both panels, the substrate is oriented from N- to C-terminus (left to right). The figure was prepared with the molecular visualization software, PyMOL. [166]

(A) ^{WT} p1-p6_{D30N/N88D}

(B) ^{LP1'F} p1-p6_{D30N/N88D}

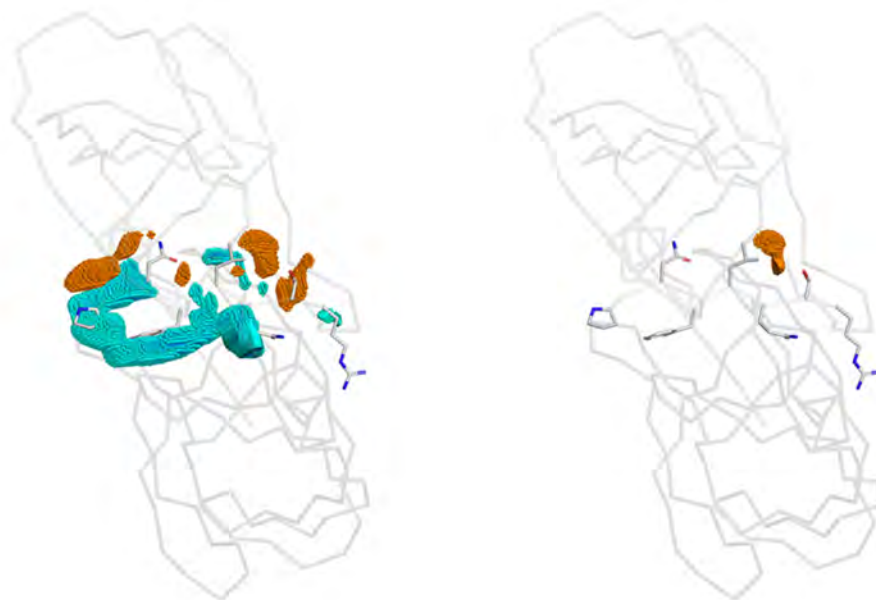


Figure 3.2

3.3.2 vdW Interactions Between the Protease and Substrates

Protease and/or cleavage-site mutations conferring drug resistance do not drastically alter the overall vdW contact potential between the protease and substrates (Figure 3.1G-I). This result is consistent with our earlier observation²⁰ that a conserved consensus overall vdW interaction potential might be a prerequisite for a sequence to be recognized by the protease. This observation, combined with the ability of drug-resistant protease variants to still process cleavage sites, leads to the expectation that cleavage sites in the resistant viruses will naturally maintain this optimal overall interaction potential. In fact, this optimal vdW potential is preserved by all the drug-resistant protease-substrate complex variants investigated in this study, with overall Δ vdW on the order of ≈ 2.5 kcal/mol within the range of WT contact potential (Figure 3.1G-I, middle bars). For example, the protease-substrate vdW interaction potential of $^{WT}p1-p6_{D30N/N88D}$ is 1.9 ± 0.2 kcal/mol (Figure 3.1H) less favorable than that of $^{WT}p1-p6_{WT}$. The loss of this interaction due to D30N/N88D protease mutations is restored with LP1'F and SP3'N substitutions in the cleavage site to a level of -1.7 ± 0.2 (Figure 3.1H) and -3.2 ± 0.2 (Figure 3.1I) kcal/mol, respectively. In contrast, $^{WT}NC-p1_{V82A}$ has -4.6 ± 0.2 kcal/mol more favorable vdW interactions compared to $^{WT}NC-p1_{WT}$ (Figure 3.1G). NC-p1 is inherently a very flexible substrate due to the P3'-Gly and P4'-Lys residues. Substitution of Val-82 to an Ala results in a larger local volume for this flexible peptide to sample. These conformations may not be accessible to the peptide in $^{WT}NC-p1_{WT}$ complex because of the additional methyl groups in Val-82. Provid-

ing extra volume in the variant allows the peptide to sample conformations that have more favorable interactions with the protease, as is evident by the change in vdW interaction potential.

AP2V mutation in NC-p1 substrate brings down vdW interactions by 1.1 ± 0.3 kcal/mol, but still keeps it more favorable than the WT level. This increase in the overall vdW contact potential might contribute to AP2V NC-p1 being better recognized than WT NC-p1 by the wild-type protease. [161] However, the processing efficiency of each cleavage site should be optimal for viral fitness. A significantly lower efficiency leaves unprocessed viral polyproteins packaged in the virion, while a significantly higher efficiency might interfere with the accurate and precise regulation of Gag and Gag-Pro-Pol processing. NC-p1 is the rate-limiting cleavage site in HIV-1 Gag during processing by the viral protease. [167, 168] Better-than-optimum cleavage efficiency for this site might introduce temporal imprecision in the regulation of its hydrolysis, leading to immature Gag processing. This may be the reason why (1) on average AP2V NC-p1_{V82A} variants have lower replicative capacity than WT NC-p1_{V82A}16, and (2) the more efficiently cleaved AP2V NC-p1 is not highly populated in the absence of drug resistance mutations in the protease. [40]

Although drug resistance does not severely impact the overall protease-substrate contact potential, the vdW contact potential profiles of substrate residues with any protease atom significantly deviate from the WT substrate behavior (Figure 3.3). The nature of this deviation, however, is not shared by all substrate variants. The loss of interactions on the primed side of the NC-p1 cleavage site is compensated

by the gain of interactions on the unprimed side (Figure 3.3A). This compensation supports that the substrate amino acid residues are interdependent despite the very short substrate sequence. Consequently, this interdependence among the amino acid residues helps maintain the overall vdW interaction potential at an optimal level within ≈ 3 kcal/mol of the WT range. [165]

Figure 3.3: Residue-based view of the effect of drug resistance on protease-substrate vdW interactions. Substrate residual Δ vdW interactions with the protease for (A) NC-p1 variants ($^{WT}NC-p1_{V82A}$, $^{AP2V}NC-p1_{V82A}$, $^{AP2V}NC-p1_{WT}$), (B) the first set of p1-p6 variants ($^{WT}p1-p6_{D30N/N88D}$, $^{LP1'F}p1-p6_{D30N/N88D}$, $^{LP1'F}p1-p6_{WT}$) and (C) the second set of p1-p6 variants ($^{WT}p1-p6_{D30N/N88D}$, $^{SP3'N}p1-p6_{D30N/N88D}$, $^{SP3'N}p1-p6_{WT}$) with reference to their respective WT complexes (Δ vdw= $vdW^{MUT}-vdW^{WT}$).

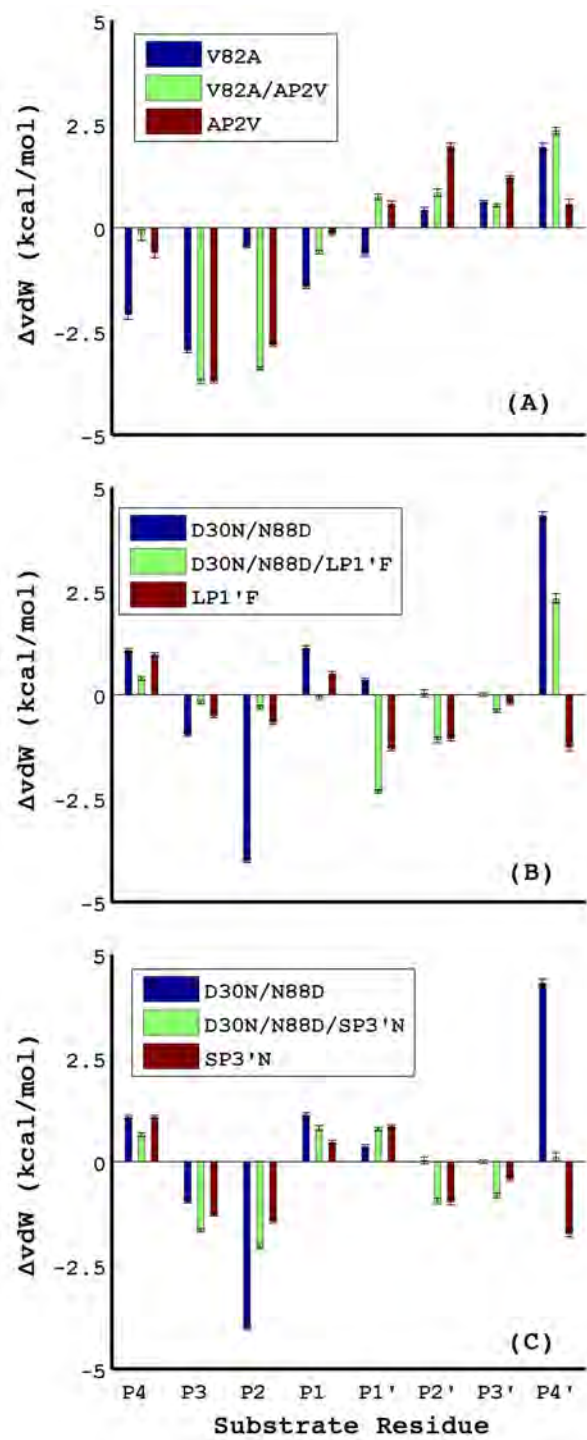


Figure 3.3

In contrast, the compensatory mechanism is more symmetric in the p1-p6 cleavage site than in NC-p1 (Figure 3.3B, and C). WT p1-p6_{D30N/N88D} loses 4.3 ± 2.0 kcal/mol vdW interactions at the P4' position and gains comparable interactions at the P2 position (-4.0 ± 0.9 kcal/mol, Figure 3.3B, blue bars). This severe deviation from the WT profile at these two positions is moderated with either LP1'F or SP3'N substitutions in p1-p6. $^{LP1'F}$ p1-p6_{D30N/N88D} has more favorable P4' interactions than WT p1-p6_{D30N/N88D}, although the loss of interaction potential at this position is considerably higher than WT p1-p6_{WT}. Loss of interactions at P4' position in $^{LP1'F}$ p1-p6_{D30N/N88D} with respect to WT p1-p6_{WT} (2.3 ± 0.1 kcal/mol, Figure 3.3B, green bars) is compensated by the gain of interactions at two other substrate residues, P1' (-2.3 ± 0.1 kcal/mol) and P2' (-1.1 ± 0.1 kcal/mol).

In addition, P2 in $^{LP1'F}$ p1-p6_{D30N/N88D} does not interact with the protease as favorably as P2 in WT p1-p6_{D30N/N88D} (Δ vdW = -4.0 ± 0.1 kcal/mol), even though P2 is invariant. The dramatic change in the vdW contacts of this invariant position is caused by the side chain rearrangements within the substrate and the binding site upon LP1'F. The conformational flexibility of the substrate combined with the interdependency within the substrate sequence allows for the altered vdW interactions in an invariant position (Asp P2) as a result of a mutation in another position, LP1'F (Figure 3.4), within the cleavage site. Thus, a single mutation manages to alter the whole vdW interaction profile along the sequence.

Figure 3.4: The compensatory mutation in the p1-p6 cleavage site, LP1'F, causes major structural rearrangements reducing vdW interaction potential of the invariant Asp P2. The side chains of the protease residues that interact with Asp P2 at 10 ns are displayed as dotted spheres for (A) ^{WT}p1-p6_{D30N/N88D} and (B) ^{LP1'F}p1-p6_{D30N/N88D}. The substrate residues were colored based on the vdW interaction potential with any protease atom and the protease residues were colored based on the vdW interaction potential with any substrate atom. The substrate residues are shown in italic and the mutated residues in bold.

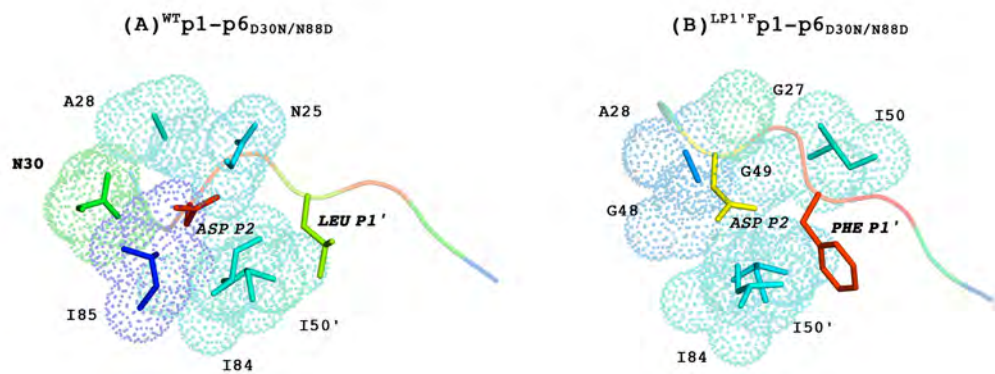


Figure 3.4

In addition, SP3'N has a similar impact on the per residue vdW interaction profile of p1-p6. Compared to the WT p1-p6 $_{WT}$ complex, P1 and P1' in $^{SP3'N}$ p1-p6 $_{D30N/N88D}$ have slightly less favorable interactions with the protease, on the order of total ≈ 1 kcal/mol (Figure 3.3C). This decreased interaction, however, is compensated by the improved contact potentials of the surrounding residues, P3, P2, P2' and P3'. Overall, the vdW contacts are preserved, however through a different interaction profile, likely facilitated by the interdependence within the substrate sequences.

3.3.3 Hydrogen Bonds between the Protease and Substrates

The hydrogen bonds formed between the protease and substrates were analyzed in two groups: hydrogen bonds formed by the (1) substrate backbone, and (2) substrate side chains. The backbone hydrogen bonds are conserved among various natural substrates, while no single side-chain hydrogen bond is shared by all or most of the substrates. [7, 165] Therefore, hydrogen bonds between the side chains and any protease atoms are more likely than the backbone hydrogen bonds to contribute to substrate specificity. [165] The less specific backbone hydrogen bonds are not altered by drug-resistance mutations in the protease and/or substrate despite remarkable local conformational rearrangements (Figure 3.5A-D). The three complexes of WT p1-p6 $_{WT}$, WT p1-p6 $_{D30N/N88D}$ and $^{LP1'F}$ p1-p6 $_{D30N/N88D}$ are superimposed to highlight these rearrangements (Figure 3.5A and E.). The percentage of time the hydrogen bonds existed throughout the simulations is listed in Table 3.1. The most stable NC-p1 backbone hydrogen bonds are formed be-

tween P2 and G48, P3' and G48' and P2' and G27' existing more than 90% of the time in most variants. These hydrogen bonds should contribute to stabilizing the substrate when bound to the protease active site. The hydrogen bonds formed by NC-p1 side-chains, on the other hand, are more variable (second main column, Table 3.1). The bonds formed in $^{WT}NC-p1_{WT}$ between the side chain of P1 and I50, N25' and G27 were not stable throughout the MD trajectories. However, both $^{AP2V}NC-p1_{WT}$ and $^{AP2V}NC-p1_{V82A}$ has more consistent hydrogen bonds formed by P1 and I50 and G27. These presumably stronger hydrogen bonds might be important for the increased substrate affinity in both $^{AP2V}NC-p1_{WT}$ and $^{AP2V}NC-p1_{V82A}$ variants, consistent with $^{AP2V}NC-p1$ being more efficiently cleaved than $^{WT}NC-p1$ by the wild-type HIV-1 protease. [161]

Figure 3.5: Substrate backbone hydrogen bonds are robust against the local conformational rearrangements due to drug resistance mutations. The side chain hydrogen bonding pattern, however, is altered by protease mutations but restored by cleavage site mutation. Coordinates of WT p1-p6 $_{WT}$, WT p1-p6 $_{D30N/N88D}$, and $^{LP1'F}$ p1-p6 $_{D30N/N88D}$ after 10 ns were superimposed to reveal the local conformational rearrangements due to the drug resistance mutations (A, E). Substrate backbone (A) and side chain (E) atoms are displayed as sticks. The backbone and side chain hydrogen bonds are shown as dashed lines for (B and F) WT p1-p6 $_{WT}$, (C and G) WT p1-p6 $_{D30N/N88D}$, and (D and H) $^{LP1'F}$ p1-p6 $_{D30N/N88D}$. The substrate residues are labeled in italic, mutated residues in bold.

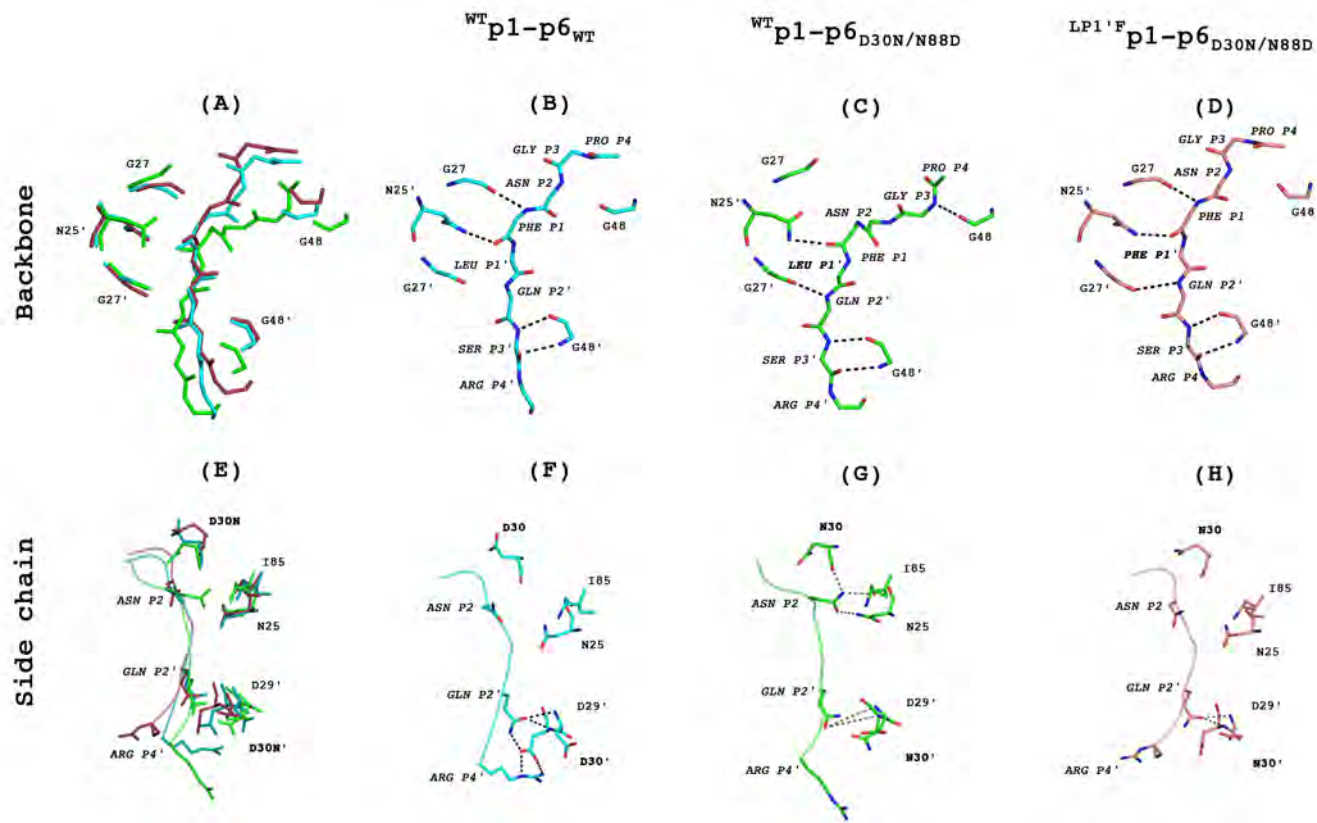


Figure 3.5

WT p1-p6 $_{WT}$ forms substrate backbone hydrogen bonds at P1 and P3' positions (Figure 3.5B). The majority of these bonds are preserved in drug-resistant WT p1-p6 $_{D30N/N88D}$ although the bond between backbone nitrogen of P1 and backbone oxygen of G27 no longer exists in WT p1-p6 $_{D30N/N88D}$ (Figure 3.5C). This loss of a hydrogen bond, however, is compensated by the hydrogen bonds newly formed by substrate P3 and P2' positions. $^{LP1'F}$ p1-p6 $_{D30N/N88D}$ has almost the same pattern of hydrogen bonding network as WT p1-p6 $_{WT}$ (Figure 3.5D). In addition, the P1, P2' and P3' positions make backbone hydrogen bonds with the protease in $^{SP3'N}$ p1-p6 $_{D30N/N88D}$ (Table 3.1). In complex with the WT protease, these mutant substrates have more backbone hydrogen bonds with decreased percent time of existence. Drug resistance, conferred by mutations in the protease and/or the substrate, does not alter the backbone hydrogen bonding pattern drastically. Even though the side chains vary by drug resistance, the substrates can still form these less specific hydrogen bonds with the protease, presumably stabilizing the substrates in the binding groove.

Table 3.1: Effect of drug resistance on intermolecular hydrogen bonding pattern

Backbone					Side-chain				
	WT	V82A	AP2V/V82A	AP2V		WT	V82A	AP2V/V82A	AP2V
P3...D29	<50	92.3	99.8	99.1	P4...D30	62.7	73.8	52.4	79.8
P2...G48	97.6	80.1	98.7	100.0	P1...I50	<50	<50	98.4	95.5
P1...N25	67.5	83.7	59.6	54.4	P1...N25'	63.1	<50	<50	<50
P1...N25'	<50	79.1	<50	<50	P1...G27	<50	<50	77.883.6	
P2'...G27'	86.0	91.5	53.8	<50					
P2'...D29'	58.0	<50	<50	<50					
P3'...G48'	96.9	92.1	78.2	66.2					
P4'...G48'	<50	<50	62.0	<50					
Backbone					Side-chain				
	WT	D30N/N88D	D30N/N88D/LP1'F	LP1'F		WT	D30N/N88D	D30N/N88D/LP1'F	LP1'F
P4...R8'	<50	<50	<50	<50	P2...N25	<50	84.7	<50	<50
P3...G48	<50	90.7	<50	55.5	P2...D30	<50	83.1	<50	<50
P3...D29	<50	<50	<50	<50	P2...I85	<50	52.7	<50	<50
P2...G48	<50	<50	<50	72.4	P2'...D29'	91.6	81.3	74.7	86.8
P1...G27	97.3	<50	98.8	97.1	P2'...D30'	89.1	87.6	86.8	81.8
P1...N25'	72.9	95.9	81.2	64.3	P4'...D30'	96.0	<50	<50	91.1
P2'...G27'	<50	91.5	76.5	94.9					
P3'...G48'	99.8	98.3	99.7	99.3					
P4'...D29'	<50	<50	<50	<50					
Backbone					Side-chain				
	WT	D30N/N88D	D30N/N88D/SP3'N	SP3'N		WT	D30N/N88D	D30N/N88D/SP3'N	SP3'N
P4...R8'	<50	<50	<50	60.7	P2...N25	<50	84.7	57.8	<50
P3...G48	<50	90.7	<50	70.0	P2...D30	<50	83.1	60.7	<50
P3...D29	<50	<50	87.3	59.1	P2...I85	<50	52.7	<50	<50
P2...G48	<50	<50	<50	83.3	P2'...D29'	91.6	81.3	87.1	58.0
P1...G27	97.3	<50	<50	98.2	P2'...D30'	89.1	87.6	96.2	61.1
P1...N25'	72.9	95.9	92.4	89.1	P4'...D30'	96.0	<50	<50	98.2
P2'...G27'	<50	91.5	98.4	97.5					
P3'...G48'	99.8	98.3	93.1	88.7					
P4'...D29'	<50	<50	63.3	50.7					

Table 3.2: Effect of drug resistance on intramolecular substrate hydrogen bonding pattern. Duration of hydrogen bonds within substrates that exist greater than 50% of the time are shown for substrate variants. The table is color-coded based on the frequency of hydrogen bonds as a spectrum; red being highest frequency and green being lowest frequency hydrogen bonds.

	<i>WT</i> NC-p1 _{WT}	<i>WT</i> NC-p1 _{V82A}	<i>AP2V</i> NC-p1 _{V82A}	<i>AP2V</i> NC-p1 _{WT}
P1(N)...P3(O)	61.8	<50	<50	<50
P1(ND2)...P3(OE1)	<50	<50	78.0	81.6
	<i>WT</i> p1-p6 _{WT}	<i>WT</i> p1-p6 _{D30N/N88D}	<i>LP1'F</i> p1-p6 _{D30N/N88D}	<i>LP1'F</i> p1-p6 _{WT}
P2(O)...P1'(N)	<50	79.9	<50	<50
P2(OD1)...P1'(N)	87.3	<50	83.7	69.0
P3'(O)...P2'(NE2)	<50	61.6	73.7	<50
	<i>WT</i> p1-p6 _{WT}	<i>WT</i> p1-p6 _{D30N/N88D}	<i>SP3'N</i> p1-p6 _{D30N/N88D}	<i>SP3'N</i> p1-p6 _{WT}
P2(O)...P1'(N)	<50	79.9	<50	<50
P2(OD1)...P1'(N)	87.3	<50	<50	71.6
P3'(O)...P2'(NE2)	<50	61.6	<50	<50

WT p1-p6_{D30N/N88D} has consistent protease-substrate side-chain hydrogen bonds formed by P2 and P2' residues (Figure 3.5G). P2' side-chain hydrogen bond is not affected by drug resistance (Figure 3.5F, H). However, both SP3'N and LP1'F substitutions weaken P2 hydrogen bonds from $\approx 85\%$ to 60% and less than 50%, respectively. These three P2 bonds (P2...N25, P2...D30, and P2...I85) do not occur in the WT complex, either. The viral advantage of weakening highly consistent hydrogen bonds formed in $^{LP1'F}$ p1-p6_{D30N/N88D} could be to optimize substrate binding for optimal viral fitness. Stronger hydrogen bonding between the protease and substrate could negatively impact substrate turnover, possibly resulting in product inhibition.

In WT p1-p6_{WT}, the substrate side chains have very consistent hydrogen bonds between P2'-D29', P2'-D30', and P4'-D30' existing more than 90% of the time (Figure 3.5F). These hydrogen bonds are preserved in all mutant complexes from 75 to 96% of the time; hence they should be crucial for the recognition of p1-p6 substrate (Figure 3.5G and H).

The natural substrates, in their bound conformation, have intramolecular hydrogen bonds that stabilize the preferred conserved consensus volume. [165] The drug-resistant protease-substrate complexes were evaluated for these intramolecular hydrogen bonds (Table 3.2). In WT NC-p1_{WT}, backbone nitrogen of the P1 position makes a hydrogen bond with the backbone oxygen of P3 position. This bond is lost upon V82A mutation in the protease, but compensated by the secondary mutation AP2V in AP2V NC-p1_{V82A}. Similarly, in $^{LP1'F}$ p1-p6_{D30N/N88D}, LP1'F mutation restores the hydrogen bond between P2 and P1' positions of the

substrate that is lost upon D30N/N88D protease mutations in WT p1-p6_{D30N/N88D}. In SP3N p1-p6_{D30N/N88D} complex, on the other hand, no intramolecular hydrogen bond exists more than 50% of the time, suggesting that these hydrogen bonds are not necessary for this variant to adopt the required three-dimensional shape that fits the binding groove. These results point out that maintaining a stable intramolecular hydrogen bonding network is only one of the various mechanisms through which the recognition motif is preserved by compensatory mutations.

3.3.4 Analysis of Atomic Fluctuations

The interdependency within the substrate residues was earlier shown [165] to play a key role in maintaining the balance between the conserved (i.e. their consensus volume and overall vdW contact potential with the protease) and varied (i.e. distribution of vdW contact potential across the substrate sequence and variations in the mean square fluctuations of the substrate residue side-chains) properties of the substrates. The dynamic cooperativity within the substrate sequence as a way to maintain this balance was evaluated for drug-resistant variants by computing the cross-correlation of the atomic positional fluctuations within the substrate. The heat maps for the NC-p1 and p1-p6 complexes are shown in Figure 3.6 and Figure 3.7, respectively, where the correlation coefficients are color coded as a red to blue spectrum for full positive correlation to full negative correlation in the fluctuations. This analysis showed that WT NC-p1_{V82A} is less cooperative compared to WT NC-p1_{WT} (Figure 3.6A versus B). The cleavage site mutation AP2V appear to improve cooperativity (increased red regions in Figure 3.6C). The

P4-P4' residues of $^{AP2V}NC-p1_{WT}$ are even more interdependent. This enhanced intrinsic cooperativity in $^{AP2V}NC-p1$ should cause better communication within the substrate residues. The fact that $^{AP2V}NC-p1$ is a better-processed substrate for even the wild-type protease supports that this interdependency within the substrate residues is a critical aspect of efficient substrate processing.

Figure 3.6: Effect of drug resistance on intrinsic dynamic cooperativity of NC-p1 substrate variants. The cross-correlations of the atomic positional fluctuations are color-coded (red: positively correlated, blue: negatively correlated).

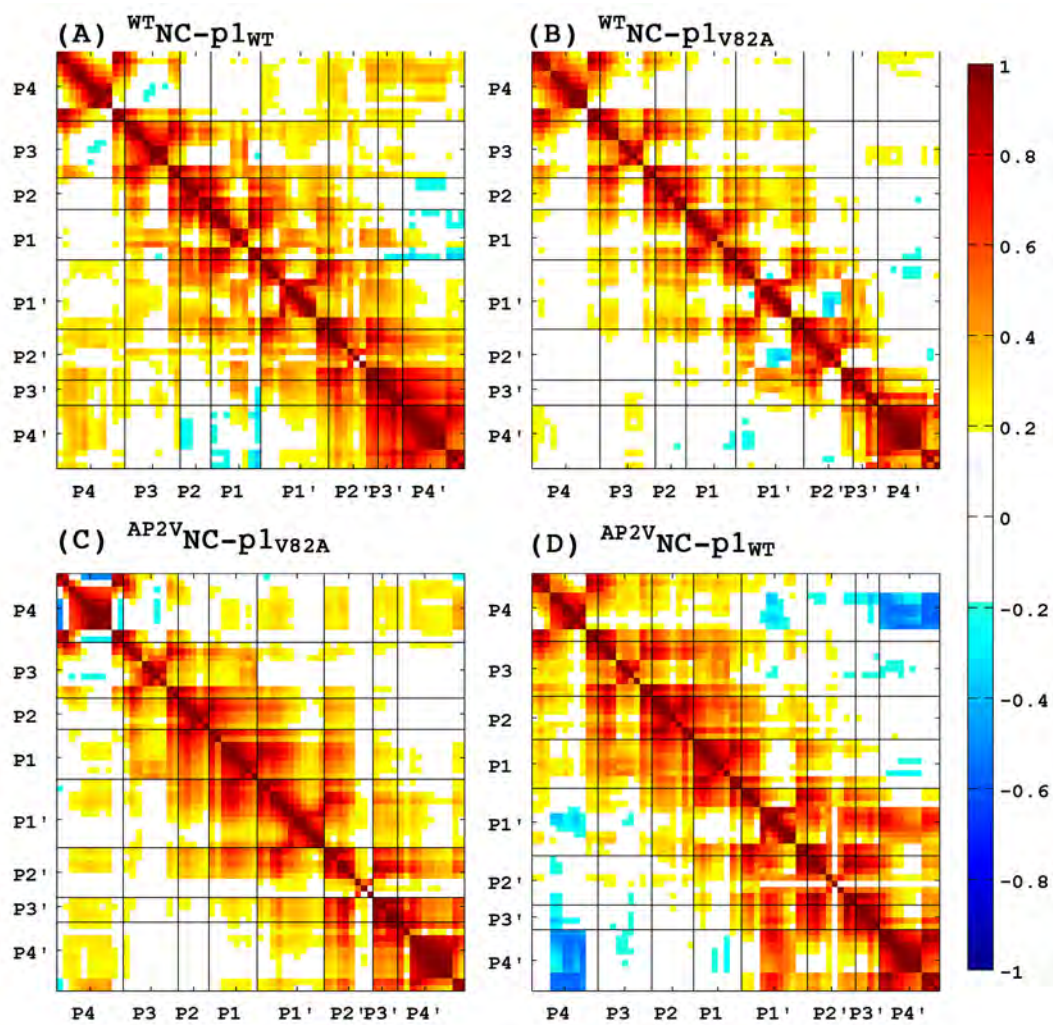


Figure 3.6

p1-p6 substrate appears to be highly cooperative in both $^{LP1'F}$ p1-p6_{D30N/N88D} and $^{SP3'N}$ p1-p6_{D30N/N88D} complexes compared to WT p1-p6_{WT} (Figure 3.7). In WT p1-p6_{WT} complex, the regions P4-P2' and P3'-P4' fluctuate almost like two independent domains while the residues in each region are highly correlated with other residues in the same region (Figure 3.7A). The same substrate sequence behaves slightly differently in WT p1-p6_{D30N/N88D} (Figure 3.7B), where the P4-P1 and P1'-P4' residues define the two regions that fluctuate independently. Upon cleavage site mutations, these two distinct regions in the heat maps (Figure 3.7C and E) become less obvious. The cooperativity is evenly distributed across the substrate sequence through neighboring residues. These results show that correlated motion of the peptide sequence may be important to maintain the interdependent nature of the cleavage sites. In principle, both the substrate structures and protease-substrate interactions should have an impact on this interdependence. The alterations in the substrate conformations and protease-substrate interactions are not independent but rather highly interdependent.

Figure 3.7: Effect of drug resistance on intrinsic dynamic cooperativity of p1-p6 substrate variants. The cross-correlations of the atomic positional fluctuations are color-coded (red: positively correlated, blue: negatively correlated).

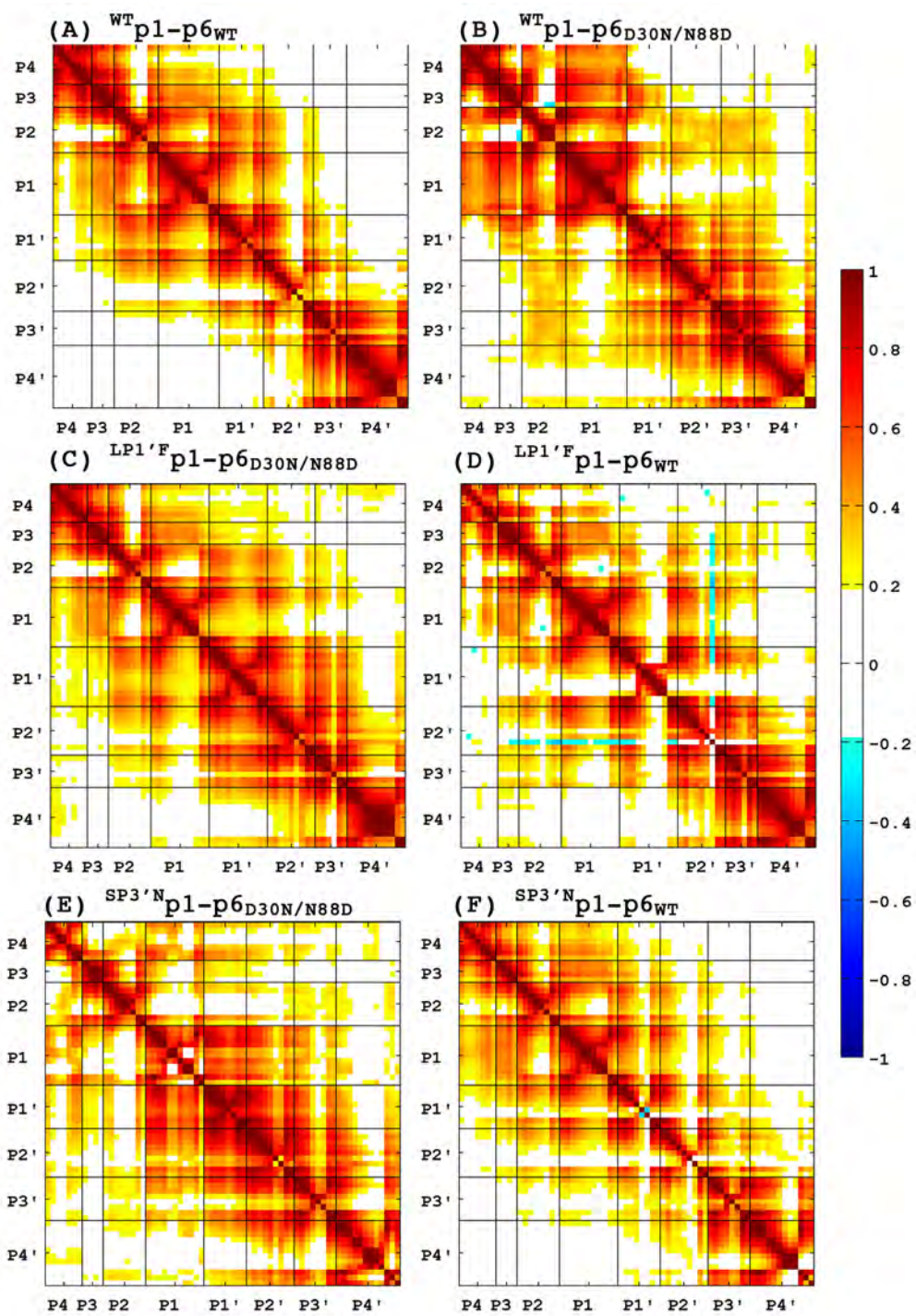


Figure 3.7

3.4 Conclusions

Drug-resistant HIV-1 protease variants maintain substrate specificity, although they do not bind as tightly to protease inhibitors as the wild-type enzyme, which the inhibitors were designed to target. Resistance against these inhibitors is usually conferred by mutations in the protease gene. However, cleavage site mutations are also associated with resistance to certain protease inhibitors. Some of these substrate mutations co-occur with the primary drug resistance mutations in the protease. In this study, structural properties of two drug resistant protease variants (V82A and D30N/N88D) and the substrates that co-evolved with these two variants (AP2V NC-p1 and $^{LP1'F}$ p1-p6/ $^{SP3'N}$ p1-p6, respectively) were investigated in conformational ensembles obtained from molecular dynamics simulations. We found that the substrate envelope is preserved by the cleavage site mutations in the presence of primary drug resistance mutations in the protease. These compensatory mutations make the substrate fit better within the envelope in the context of drug-resistant protease by increasing V_{in} . Consistency of these results ensures our understanding of the specificity determinants of substrate recognition. Protease-substrate co-evolution validates the substrate envelope as the substrate recognition motif for HIV-1 protease.

The results of substrate co-evolution and flexibility can potentially predict what other cleavage sites may be susceptible to co-evolution. The increase in V_{in} is achieved by a variety of mechanisms; recovery of protease-substrate hydrogen bonding, van der Waals interactions or the dynamic cooperativity within the

substrate sequence. For each substrate we studied this increase in V_{in} is achieved by a different mechanism, but in every case upon mutation the amino acid side chain becomes larger (Ala to Val in NC-p1 and Leu to Phe or Ser to Asp in p1-p6) resulting in substrate variants that better fill the substrate envelope. Thus we would expect that other compensatory mutations in the cleavage site would likely involve mutations to larger residues. Our results also highlight the importance of the substrate conformational plasticity in the binding site. The mutant substrates that are intrinsically more flexible undergo local conformational rearrangements that assist a better fit within the substrate envelope. However this flexibility also can cause a larger V_{out} making the sequence more susceptible to co-evolution. The exact effect of the substrate flexibility on the fit within the substrate envelope cannot be predicted based solely on the cleavage site sequence. Our approach of combining the dynamic substrate envelope with the molecular dynamics simulations of model structures of the mutant complexes can assist in this prediction. Theoretically, ΔV_{in} upon a single amino acid substitution could be estimated by combining *in silico* mutagenesis followed by energy minimization, molecular dynamics simulations and three-dimensional grid based volume calculations. The substrate mutations resulting in higher ΔV_{in} values should be beneficial for the drug resistant protease variants. Ultimately, the approach we present here could be useful in predicting the enzyme-substrate mutations that will more likely be tolerated in drug resistant viral variants.

Being able to accurately predict the substrate co-evolution is critical to avoid drug resistance in drug design process. Our work highlights the importance of

paying attention to the details of natural substrate recognition in designing inhibitors targeting resistant viral variants. We demonstrate, in this study, the impact of drug resistance mutations on substrate recognition in addition to inhibitor binding. Protease inhibitors, when designed fit within the substrate envelope, are less likely to elicit protease mutations that can ultimately be compensated by mutations in the cleavage sites. In designing new inhibitors, staying within the substrate envelope the potency of the inhibitors must be maintained. Some inhibitors may protrude beyond the envelope to make favorable contacts with a protease residue to preserve affinity. In those cases, accurately predicting how likely for one or more substrates to co-evolve with a potential resistance mutation in that particular protease residue is crucial. The availability of this probabilistic information can be used to decide at which location an inhibitors protrusion beyond the envelope is more affordable in terms of avoiding the emergence of drug resistance via cleavage site mutations.

Researchers have so far used evolutionary information to predict the structural and functional features of proteins and their complexes. Sequence conservation allows for modeling the structure of a protein based on the experimentally determined structure of a homologous protein, while sequence variation, especially in the form of correlated mutations, is useful to infer functional features. This study, on the other hand, proposes the substrate envelope based on structural data, as a tool to predict the evolution due to drug pressure for a protease-substrate system. Using the dynamic substrate envelope, the substrates can be assigned a probability of mutating in the presence of a particular protease mutation. To the best of our

knowledge, this is the first structure-based approach to predict evolution under the selective pressure of drug therapy.

Using the substrate envelope in prediction of co-evolution in other systems

In general, viral replication is a highly interdependent process that involves very complex interactions between separate viral macromolecules (proteins and polynucleotides) as well as host-virus interactions. This highly interdependent character of viral life cycle should make the co-evolution of separate viral genes inevitable. In particular, HIV is a well-studied model system, with a massive amount of structural, biochemical, and sequence data on HIV are available in the published literature and online databases, [27, 169–171] allowing for indepth analysis to occur. In other viral systems far less data is available, but we expect the substrate envelope hypothesis to be applicable. For example, co-evolution of individual sites within the Hepatitis B Viral genome has been reported to contribute to the rate of drug resistance. [172] In addition, correlated mutations in four of the influenza proteins have been reported to be critical for host adaptation and pathogenicity [173], demonstrating the interdependence of the viral genomic sites in evolution.

Hepatitis C Virus (HCV) is also among these systems. The viral RNA-dependent RNA-polymerase, just like the HIV-1 reverse transcriptase, is inherently inaccurate accounting for a very high mutation rate. [174] A very recent Bayesian network analysis of viral polyprotein sequences has shown that the intrahost evolution of HCV is a complex process encoded in the interrelationships among many sites along the entire viral polyprotein. [175] Crystallographic studies from our

group on the serine protease of this virus, NS3/4A, complexed with its natural substrates have shown that the substrate envelope hypothesis is valid also for HCV in explaining (1) the specificity determinants of the natural substrate recognition and (2) emergence of primary drug resistance mutations. [9] The most severe resistance-conferring mutations occur where the inhibitors protrude from the NS3/4A substrate envelope, as these changes selectively weaken inhibitor binding without compromising the binding of substrates. The high mutation rate of HCV can also impact the sequence variability of the cleavage sites of NS3/4A, as a lesson learned from studying HIV-1 protease-substrate co-evolution. No data on substrate co-evolution has been reported so far. Nevertheless, the characteristics of the protease-substrate co-evolution in HCV and other viral systems should be investigated. Our combined approach, involving molecular modeling, molecular dynamics, and grid-based volume calculations, will better elucidate the interdependency of viral evolution by predicting the general sequence variability of each cleavage site and the particular cleavage site mutations that are likely to be selected in the context of a primary drug resistance protease mutation.

3.5 Methods

3.5.1 Nomenclature

HIV-1 protease variant (WT, V82A or D30N/N88D) and the cleavage site variant (AP2V, LP1'F or SP3'N) in a protease-substrate complex are designated by a

subscript and a superscript to the name of the cleavage site. For example, ^{LP1'F}p1-p6_{D30N/N88D} denotes a complex of D30N/N88D protease variant with the LP1'F mutant of the p1-p6 cleavage site where LP1'F refers to a Leu-to-Phe mutation at P1' position of the cleavage site.

3.5.2 Protease-Substrate Complex Structures

Two PR substrates and their mutant variants in complex with inactive, isosteric PR variants were investigated for structural and dynamic properties. All these inactive protease variants have D25N mutation that has been shown to minimally alter the overall structure of the protease complexes [176]. This D25N mutation is not associated with drug-resistance; therefore we refer to D25N as WT throughout the paper for simplicity in nomenclature. The crystal structures of ^{WT}p1-p6_{WT} [7] and ^{WT}NC-p1_{WT} [40] and ^{AP2V}NC-p1_{V82A} variants were used as the starting structures in the MD simulations (PDB ID: 1KJF, 1TSU, 1TSQ). The other structures were modeled by side-chain mutations performed in PyMOL [166] by selecting the most probable rotamer from the rotamer library followed by energy minimization as described below. The protease-substrates complex variants are listed in Table 3.3.

Table 3.3: HIV-1 protease-substrate complexes

Protease-substrate complex variant	Drug-resistance mutations		Starting structure (PDB Code)
	Protease (D25N) ^a	Substrate	
^{WT} p1-p6 _{WT}	—	—	^{WT} p1-p6 _{WT} (1KJF)
^{LP1'F} p1-p6 _{WT}	—	LP1'F	^{WT} p1-p6 _{WT} (1KJF)
^{SP3'N} p1-p6 _{WT}	—	SP3'N	^{WT} p1-p6 _{WT} (1KJF)
^{WT} p1-p6 _{D30N/N88D}	D30N/N88D	—	^{WT} p1-p6 _{WT} (1KJF)
^{LP1'F} p1-p6 _{D30N/N88D}	D30N/N88D	LP1'F	^{WT} p1-p6 _{WT} (1KJF)
^{SP3'N} p1-p6 _{D30N/N88D}	D30N/N88D	SP3'N	^{WT} p1-p6 _{WT} (1KJF)
^{WT} NC-p1 _{WT}	—	—	^{WT} NC-p1 _{WT} (1TSU)
^{AP2V} NC-p1 _{WT}	—	AP2V	^{WT} NC-p1 _{WT} (1TSU)
^{WT} NC-p1 _{V82A}	V82A	—	^{WT} NC-p1 _{WT} (1TSU)
^{AP2V} NC-p1 _{V82A}	V82A	AP2V	^{AP2V} NC-p1 _{V82A} (1TSQ)

^aAll 10 PR variants have D25N mutation. This mutation inactivates the PR allowing for natural substrate co-crystal structure determination and is known to hardly alter the structure. [176] D25N is not a resistance mutation; hence, the PR with D25N is referred WT to simplify the notation.

3.5.3 MD simulations

The AMBER [150, 151] simulation package (version 8) with the *ff03* force field was used in all simulations. All structures were solvated explicitly in a truncated octahedron box using the TIP3P water model. [152] Overall charge of each system was neutralized by adding the appropriate number of Cl^- counterions using a Coulombic potential on a 1 Å grid with the preparatory program *tleap* of AMBER. The initial structures were first minimized at constant volume with convergence criteria of either maximum 90,000 cycles or an RMSD value of 0.01 Å by steepest descent integration algorithm for 50 steps and then switched to conjugate gradient algorithm. Initial atom velocities corresponding to a temperature of 10 K were generated from a Maxwellian distribution and the temperature was gradually raised to 300 K. The temperature was maintained at 300 K, and the pressure was maintained at 1 bar by the Berendsen weak-coupling approach. [154] Constant pressure-periodic boundary conditions were used with isotropic position scaling. The particle-mesh-Ewald (PME) method [155] was used to calculate the full electrostatic energy of a periodic box, bypassing pair list creation and nonbonded force and energy evaluation by calling special PME functions to calculate the Lennard-Jones and electrostatic interactions with a cutoff distance of 9 Å. Covalent bonds involving hydrogen atoms were constrained by the SHAKE algorithm [153] with a relative geometrical tolerance of $10\text{E-}5$ Å. A time step of 2 fs was employed in the Leapfrog integrator. Various system properties (Total, kinetic and potential energies, temperature, pressure, density, root-mean-square-deviation of the backbone

atoms and $V_{in/out}$) were monitored during equilibration to ensure the stability of the simulations. Coordinates and energies were written every 0.4 ps during the 11 ns production phase.

3.5.4 Fit within the Substrate Envelope

The dynamic substrate envelope was previously modeled by a quantitative approach using the molecular dynamics of seven protease-substrate wild-type complexes. [165] This approach was based on a three-dimensional grid with side length 10 Å and grid spacing 0.2 Å located in the binding site. The occupancy of each grid cell, g_{ijk} , was assigned an initial value of 0. Then the number of times that a grid cell was within the vdW volume of any peptide was counted. This value eventually became the overall occupancy of that grid cell. The total volume of the substrate envelope was essentially the summation over all grid cells with occupancy greater than 0 normalized by the total number of structures used. This dynamic substrate envelope was used in this study as a basis for the conserved consensus volume that is occupied by the substrates in the binding groove.

In this study, the substrate vdW volumes were mapped on the same grid to define probability distributions of the volume occupied by each substrate variant. Before this mapping, all frames from each trajectory were superimposed using the least flexible/mobile residues 24-26 and 85-90 on the $^{WT}CA-p2_{WT}$ complex structure using the RMSD trajectory tool of the molecular visualization and trajectory analysis software, VMD. [156] The effect of residue selection on the shape of the dynamic substrate envelope was assessed in chapter II by aligning the trajectories

using the $C\alpha$ atoms of all protease residues, dimerization interface excluding the flaps, catalytic triad, and finally only the flaps. The results showed that the shape of the dynamic substrate envelope is not sensitive to the selection of the reference residues for structural alignment as long as these reference residues are not on a highly mobile region of the protease (Figure 2.16).

The substrates in various complexes were then compared to each other in terms of how well they fit within (V_{in}) and how much they protruded beyond the dynamic substrate envelope (V_{out}). For this comparison, the vdW volume of an individual substrate was computed from N_1 time points in the MD trajectory of that substrate as described earlier. [165]

3.5.5 Estimation of vdW Potential

Protease-substrate vdW contacts were estimated by a simplified Lennard-Jones potential $V(r)$ using the relation $4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, where r is the PR-substrate interatomic distance, and ϵ and σ are the well depth and hard sphere diameter, respectively, for each PR-substrate atom pair. $V(r)$ for all possible PR-substrate atom pairs was computed within 5 Å, and when the distance between non-bonded 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. [165] Using this simplified potential for each non-bonded protease-substrate pair, $\sum V(r)$ was then computed for each protease and each substrate residue.

For comparison with the wild-type values, the property of interest (V_{in} , V_{out} , or vdW) was calculated for each frame taken from an MD trajectory as described

above. These time-series data were first plotted as a histogram to ensure Gaussian distribution and the sample mean and standard error were calculated. The difference between two distributions (for example $\Delta V_{in} = V_{in}^{MUT} - V_{in}^{WT}$) was computed by subtracting the sample mean of one from the other. The corresponding error was estimated by combining the individual standard errors in quadrature assuming V_{in}^{WT} and V_{in}^{MUT} are independent variables.

$$SE_{\Delta\bar{x}} = \sqrt{SE_{\bar{x}^{MUT}}^2 + SE_{\bar{x}^{WT}}^2} \quad (3.1)$$

Finally, the calculated parameters were reported as $\Delta x \pm SE_{\Delta\bar{x}}$ in Figures 1 and 3, where x is V_{in} , V_{out} , or vdW. The details of the mathematical proof can be found elsewhere. [177]

3.5.6 Evaluation of Hydrogen Bonding

The trajectory analysis program *ptraj* was used to calculate the percentage of time a hydrogen bond existed during the simulated trajectories. 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°. Hydrogen bonds that existed more than 50% of the time were analyzed.

3.5.7 Fluctuation Dynamics

The normalized cross-correlations of residue pairs were defined as

$$CO_{i,j} = \frac{\langle \Delta R_i \Delta R_j \rangle}{\langle \Delta R_i^2 \rangle^{1/2} \langle \Delta R_j^2 \rangle^{1/2}} \quad (3.2)$$

where ΔR_i is the fluctuation in the position vector R of site i and ΔR_j is the fluctuation in the position vector R of site j . The brackets represent time averages over recorded snapshots. The cross-correlations vary in the range $[-1, 1]$, with the lower and upper limits indicating fully anticorrelated and correlated atomic fluctuations, respectively. $CO_{i,j} = 0$ gives uncorrelated atomic fluctuations.

The computer programs for calculating V_{in} , V_{out} , vdW contact potential, and the cross correlations of the atomic fluctuations were written in FORTRAN programming language. The vdW radii necessary for the volume and vdW contact potential calculations were taken from the *amber03* forcefield. The error analysis was performed using MATLAB[®]. [178]

Chapter IV

Dynamic Substrate Envelope in Hepatitis C Virus NS3/4A Protease: Predicting Resistance and Guiding the Design of Novel Inhibitors

4.1 Abstract

Drug resistance is a major problem in quickly evolving diseases, including infection by Hepatitis C Virus (HCV). The viral protease NS3/4A is a primary drug target for HCV. At the molecular level, drug resistance reflects a subtle change in the balance of molecular recognition by NS3/4A; the drug resistant protease variants are no longer effectively inhibited by the competitive active site inhibitors but can still process the natural substrates with enough efficiency for viral survival. Therefore, the inhibitors that better mimic the natural substrate molecular recognition features, including the dynamic characteristics of the complex, should result in more robust inhibitors that are less susceptible to drug resistance. When bound to the enzyme, the native substrates of NS3/4A adopt a consensus volume, termed the substrate envelope. The most severe drug resistance mutations occur at protease residues that are contacted by the inhibitors but are outside the substrate envelope. To guide the design of robust inhibitors, we investigate the enzyme complexes of four natural substrates using molecular dynamics and define the dynamic substrate envelope of NS3/4A. Comparative analyses of substrates and inhibitors distinguish unique structural and dynamic features of substrates that are not shared by inhibitors and suggest changes to inhibitors to improve resistance profiles. This study provides a template for guiding the development of novel inhibitors that will be more robust against resistance by mimicking the static and dynamic binding characteristics of natural substrates.

4.2 Introduction

Drug resistance presents a great challenge in treating quickly evolving diseases, including the hepatitis C viral (HCV) infection. Direct-acting anti-viral agents specifically target the viral enzymes, inhibiting viral replication and eventually the disease progress. Two inhibitors targeting HCV NS3/4A protease, telaprevir and boceprevir, are approved by the US Food and Drug Administration (FDA) for treatment of chronic HCV genotype 1 infection in combination with pegylated interferon and ribavirin. Both telaprevir and boceprevir are peptidomimetic small-molecule inhibitors that associate with NS3/4A through a reversible covalent linkage to the catalytic serine (S139) as well as short-range molecular interactions with the binding site. The narrow selectivity of the two FDA-approved drugs to a single genotype renders them susceptible to drug resistance and creates a demand for methods to assist in developing inhibitors with broader selectivity profiles.

Several non-covalent NS3/4A inhibitors, including macrocyclic compounds, are at various stages of clinical development. These inhibitors contain a macrocycle connecting either the P1 and P3 groups (ITMN-191 or danoprevir [81]) or alternatively the P2 and P4 groups (MK-5172 [82], MK-7009 or vaniprevir [179]). These compounds are more potent than telaprevir and boceprevir against wild-type virus. Nevertheless, HCV quickly evolves due to the high replication rate combined with the lack of proofreading in the viral RNA-dependent RNA polymerase. Under the selective pressure of drug therapy, resistant variants are rapidly

populated even at early stages of clinical trials, compromising the high efficacy of protease inhibitors and eventually restricting their usage to treatment-naïve patients for a limited period of time. [174, 180, 181]

Ideally, the molecular recognition aspects of drug resistance should be elucidated and incorporated a priori into the structure-based design process to develop more robust inhibitors. To achieve this goal, structural requirements for the protease to fulfill its enzymatic function in the viral life cycle should be thoroughly studied, since the biological function imposes evolutionary constraints on the protease under the selective pressure of drug therapy. An important lesson learned from studying human immunodeficiency virus type 1 (HIV-1), also a quickly evolving enzyme, is that designing robust protease inhibitors can be facilitated by an understanding of the key molecular recognition elements of the natural substrates. For example, inhibitors should stay within the substrate recognition regions (the “substrate envelope”) in order to minimize the likelihood that the enzyme can mutate to reduce the inhibitor efficacy while still maintaining sufficient enzymatic activity on the substrates. [45, 46, 48, 49, 182] Resistance to protease inhibitors reflects a subtle change in the balance of molecular recognition events, in favor of substrate processing versus inhibitor binding.

Crystallographic studies from our group have shown that the substrate recognition motif for HIV-1 protease is not a consensus sequence of the substrates but a consensus volume adopted by the non-homologous cleavage site sequences within the binding site, which is termed the substrate envelope. [7] In addition, protein dynamics have been incorporated into the substrate envelope to assess the effects

of local conformational fluctuations of the bound substrate. [165] The dynamic substrate envelope better captures the specificity determinants of substrate recognition [165] and is preserved even in drug resistant protease variants in the presence of compensatory mutations in the cleavage sites. [183] The primary drug resistance mutations in HIV occurs at protease residues that are contacted by the inhibitors outside the substrate envelope. [8] These residues are expected to be more important for inhibitor binding than substrate recognition. Therefore, similar to HIV-1 protease, the specificity determinants of substrate recognition and how these determinants are maintained by the resistant variants should be elucidated to guide the design of robust HCV NS3/4A protease inhibitors.

HCV NS3/4A protease is the N-terminal domain of the viral bifunctional protein NS3, which also has a C-terminal helicase domain. Efficient proteolytic activity of NS3 requires a cofactor NS4A, a 54-amino acid peptide that tightly associates with the protease. [184] The viral genome is translated as a single polyprotein along the endoplasmic reticulum by host cell machinery. NS3/4A hydrolyzes the polyprotein precursor at four known cleavage sites (3-4A in *cis*, 4A-4B, 4B-5A, 5A-5B) yielding structural and non-structural proteins essential for viral maturation. [65, 185] The full-length apo NS3/4A crystal structure has the C-terminal six residues of the helicase sitting at the binding site of the protease domain. [186] The binding mode of these six residues, which correspond to the unprimed side of the 3-4A junction, provided the basis for structure-based drug design efforts targeting the NS3/4A protease. The four known substrates of HCV share little sequence homology except for an Asp/Glu at P6, Cys/Thr at P1, and

Ser/Ala at P1' positions. Our group reported co-crystal structures of the wild-type protease domain bound to substrate cleavage products and inhibitors that included the macrocyclic compounds ITMN-191, MK-5172, and MK-7009, which are used in the current study. [90] In the co-crystal structures, the unprimed side substrate products adopt a consensus volume despite the low sequence homology. Similar to HIV-1 protease, the most severe resistance mutations in NS3/4A correspond to the residues that are contacted by the inhibitors outside the substrate envelope defined by the cleavage products.

Protein dynamics is especially critical to consider when characterizing the substrate envelope of HCV NS3/4A, as the very shallow binding site may allow more flexibility for the bound ligands. In this study, the cleavage products of the natural substrates and three inhibitors in complex with HCV NS3/4A protease are compared using the dynamic substrate envelope. The structural and dynamic analyses reveal several features of the intermolecular interactions between the protease and ligands providing insights into drug resistance. This analysis in particular (1) allows predictions of potential drug resistance mutation sites, at which the enzyme can afford amino acid substitutions with minimal effect on substrate recognition and (2) reveals the untapped regions of the HCV NS3/4A protease binding site that satisfy the substrate envelope and can be explored by new inhibitors to potentially gain affinity without compromising flat resistance profiles.

4.3 Results and Discussion

Comparative analysis of the binding modes and molecular interactions of substrates versus inhibitors is critical to guide the design of new small-molecule inhibitors that better overlap the natural substrate binding features, minimizing the likelihood of resistance. Toward this goal, the regions of the substrate binding surface that are untapped by the current inhibitors as well as the regions the inhibitors contact on the protein surface beyond the substrate envelope were revealed using molecular dynamics simulations.

Three macrocyclic NS3/4A protease inhibitors, ITMN-191, MK-5172, and MK-7009, were compared to the natural substrates, 3-4A, 4A-4B, 4B-5A, and 5A-5B, in terms of dynamics in the bound state and interactions with the protease (Table 4.1). For substrates, each of the six amino acid residues defines a site (P6 to P1, Figure 4.1A); the inhibitor functional groups were assigned positions by superimposing the inhibitor complexes onto the substrates and grouping the inhibitor atoms. Because none of the inhibitors span the P6-P5 region, only four chemical groups (P4 to P1, Figure 4.1B) and one group outside the envelope (P1') were defined for the inhibitors.

Table 4.1: The cleavage site products and macrocyclic small-molecule inhibitors of HCV NS3/4A Protease

<i>Substrates</i>											
	P6	P5	P4	P3	P2	P1	PDB ID	# of atoms	V_{out} (\AA^3)	V_{tot} (\AA^3)	vdW (kcal/mol)
3-4A	D	L	E	V	V	T	1CU1	47	341 ± 31	1042 ± 14	-26.9 ± 2.8
4A-4B	D	E	M	E	E	C	3M5M	50	431 ± 40	1097 ± 15	-31.9 ± 2.5
4B-5A	E	C	T	T	P	C	3M5N	43	300 ± 26	950 ± 17	-27.7 ± 2.7
5A-5B	E	D	V	V	C	C	3M5O	44	318 ± 24	993 ± 14	-27.9 ± 2.2
<i>Inhibitors</i>											
							PDB ID	# of atoms	V_{out} (\AA^3)	V_{tot} (\AA^3)	vdW (kcal/mol)
ITMNN-191 (Danoprevir)							3M5L	51	567 ± 18	1082 ± 15	-36.5 ± 2.0
MK-5172							3SUD	54	633 ± 16	1165 ± 16	-35.0 ± 2.8
MK-7009 (Vaniprevir)							3SU3	53	606 ± 14	1152 ± 14	-36.6 ± 1.8

Figure 4.1: HCV protease substrate and inhibitor structures. (A) The overlaid substrate cleavage products (3-4A, 4A-4B, 4B-5A, and 5A-5B) are displayed in sticks with the substrate sites indicated by distinct colors. (B) The inhibitors (ITMN-191, MK-5172, and MK-7009) superimposed onto the substrates, inhibitor chemical moieties colored in the same scheme as the substrates (additionally P1': white). The substrates do not have a P1' position in the co-crystal structures due to the residual activity of the S139A protease construct used in crystallization.

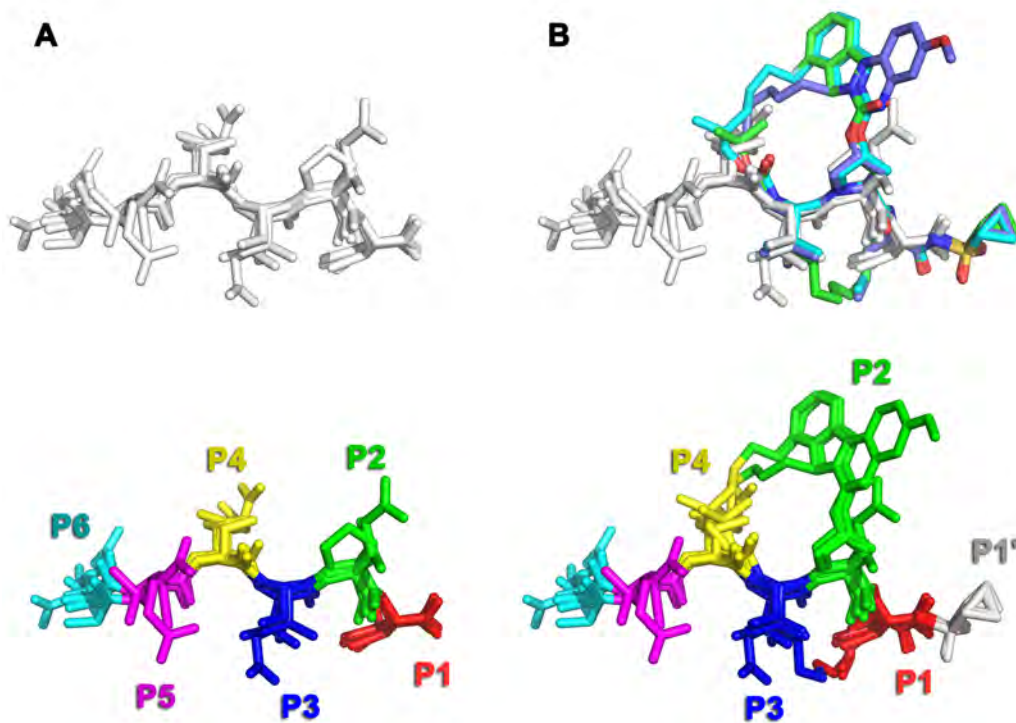


Figure 4.1

4.3.1 Substrate Shape and the Dynamic Substrate Envelope

The conformations of four substrates in complex with NS3/4A protease adopt a similar overall shape in each of their crystal structures (Figure 4.2A), and the overlapping consensus volume defines the static substrate envelope (Figure 4.2B). To incorporate the dynamic nature of the protease-substrate complex, the most probable volume occupied by at least 75% of each substrates conformers during MD simulations was calculated (labeled as dynamic in Figure 4.2A). However, the somewhat arbitrary cutoff of 75% and the high mobility of substrates resulted in a smaller and more restrictive dynamic consensus volume. Therefore, a continuous representation of the dynamic substrate envelope was defined by mapping vdW volumes of P6 to P1 residues of substrate conformers onto a three-dimensional grid, similar to the case for HIV-1 protease. [165, 183] The dynamic substrate envelope is a probability distribution of vdW occupancies at each grid point (Figure 4.2B). This probabilistic dynamic substrate envelope allows all the information from MD conformers to be properly incorporated for a more accurate comparison between substrate and inhibitor complexes.

Figure 4.2: (A) Substrate flexibility reduces the consensus volume as reflected in dynamic volumes smaller than static substrate volumes. The vdW volume occupied by 75% of substrate conformers during MD simulations (*Dynamic*) is compared to the vdW volume of the crystal structure conformation (*Static*) for substrates 3-4A, 4A-4B, 4B-5A, 5A-5B. **(B) Dynamic substrate envelope is a more realistic representation of the conserved substrate volume.** The dynamic consensus volume is the volume occupied by at least 75% of substrate conformers in the MD simulation trajectories, whereas the static consensus volume is the volume occupied by at least 3 of 4 substrates in the crystal structures. The dynamic substrate envelope is the probability distribution of vdW volume occupied by all substrate conformers in the MD simulations. Substrate co-crystal structures in stick representation are superposed (P6 to P1 orientation left to right) onto the grid matrix of the dynamic substrate envelope.

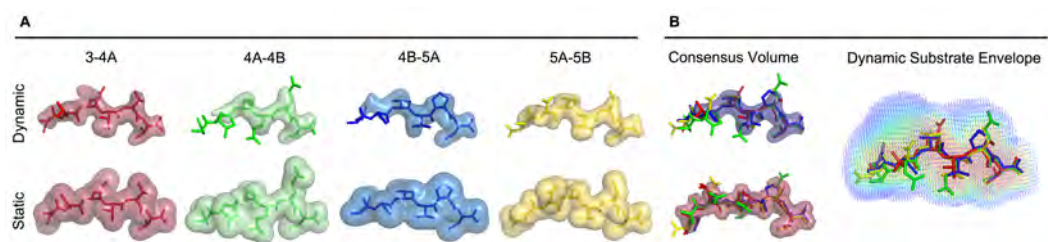


Figure 4.2

In the dynamic substrate envelope, the most conserved and dynamically restricted region is defined by the P4 to P1 residues, as indicated by the low V_{out} values (V_{out} range $26 \pm 4 \text{ \AA}^3$ to $56 \pm 22 \text{ \AA}^3$) (Figure 4.4A). The 4A-4B cleavage site is an exception, with larger V_{out} values, especially at P2 position ($80 \pm 19 \text{ \AA}^3$). Two factors contribute to this: first, 4A-4B has a bulky glutamic acid at P2 with a higher degree of side-chain conformational freedom; and second, 4A-4B appears to be inherently more flexible than the other substrates as suggested by alpha carbon root-mean-squared fluctuations (RMSF of 1.6 \AA , average for other substrates is $1.0\text{-}1.2 \text{ \AA}$; Figure 4.3) resulting in a smaller dynamic volume for this substrate than other substrates (Figure 4.2A). With 4A-4B being the exception, the P4-P1 region is the key conserved volume within which the inhibitors should be designed to fit with minimal protrusion to avoid resistance mutations.

Figure 4.3: P4-P1 substrate region is less mobile compared to P6 and P5 residues, contributing to the highly conserved consensus volume. The root-mean-squared-fluctuations (RMSF) of C_{α} atoms of each substrate were calculated from MD simulations. The relatively high mobility of 4A-4B likely causes the substrate to sample a wider conformational space, resulting in a smaller dynamic volume.

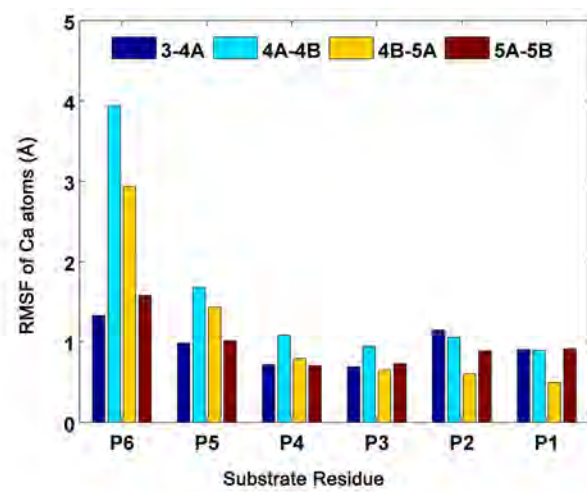


Figure 4.3

P5 and P6 have higher mobility during the MD simulations (Figure 4.3). The P6 positions of all four substrates have relatively high V_{out} values compared to the other residues in the range of $82 \pm 15 \text{ \AA}^3$ for 3-4A to $132 \pm 18 \text{ \AA}^3$ for 4B-5A, which corresponds to 23% to 40% of the total V_{out} of the respective substrate. Similarly at the P5 position, the substrates have a lower consensus volume. In general, the lower density at P6 and P5 is consistent with the higher temperature factors of P6 and P5 in the crystal structures. [9] The high dynamics and relatively lower occupancies at P6 and P5 positions is also reflected in the inhibitors, as none of the more potent macrocyclic inhibitors have P6 and P5 groups. Indeed, telaprevir, an FDA-approved non-macrocyclic ketoamide inhibitor, has a P5 moiety that does not improve its binding compared to the macrocyclic compounds without a P5 moiety. The K_i for telaprevir is $34.4 \pm 3.0 \text{ nM}$ against wild type enzyme, while macrocyclic compounds without a P5 moiety are more potent (ITMN-191= $1.0 \pm 0.1 \text{ nM}$, MK-5172= $0.74 \pm 0.07 \text{ nM}$, MK-7009= $0.14 \pm 0.02 \text{ nM}$). [90] Thus extending inhibitors into the P5-P6 subpockets observed in the crystal structures may not result in more potent inhibitors unless moieties with highly specific, possibly electrostatic interactions with the protease are introduced.

Figure 4.4: The HCV NS3/4A dynamic substrate envelope is more conserved in the P4 to P1 region. The volume outside the dynamic substrate envelope for each (A) substrate and (B) inhibitor moiety, V_{out} , is colored in a spectrum of blue to red. Warmer and cooler colors indicate a higher and lower level of protrusion beyond the dynamic substrate envelope, respectively. The macrocyclic inhibitors protrude from the substrate envelope mainly at P2 and P1' positions.

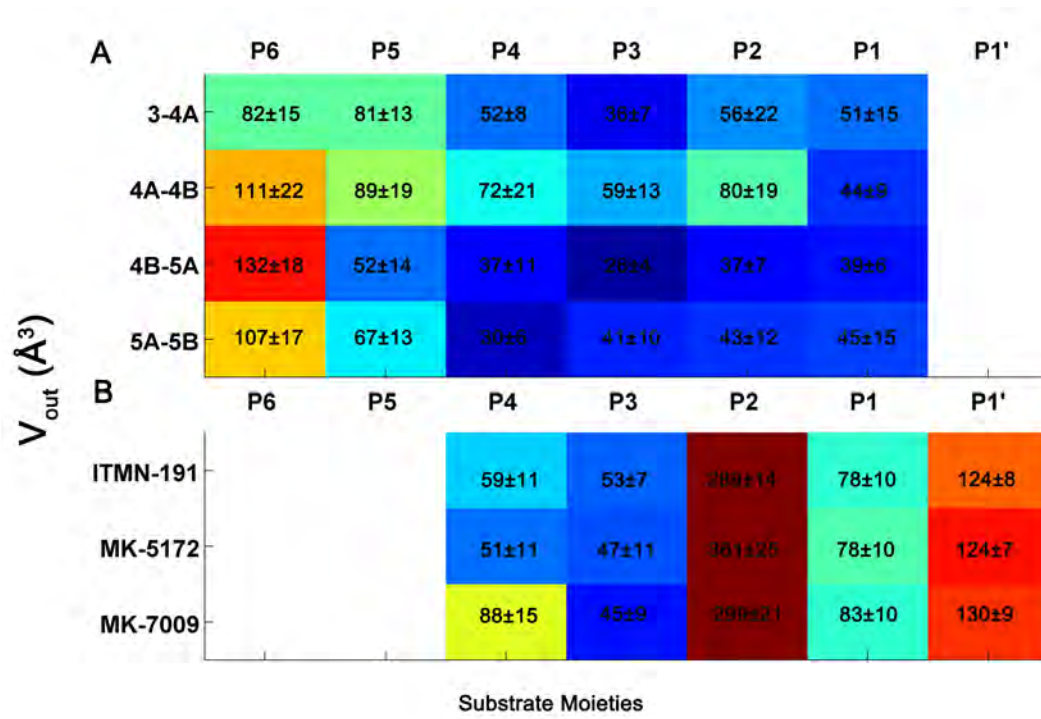


Figure 4.4

Overall, the shared features of the P4-P1 region appear to be the common motif recognized by the protease. However, the fact that the flanking residues do not adopt a consensus volume does not exclude their role in substrate recognition. These residues can change the context of the cleavage site as has been shown in substrate cleavage efficiency in the viral protease of HIV. [187] The fact that the P5' residue in HIV-1 protease has co-evolved with drug resistant HIV-1 protease variants suggests that this position plays a role in substrate binding and/or processing. [32] In HCV, the P6 substrate residue makes extensive electrostatic interactions with the protease binding site residues R119, R123, R161 or K165, as discussed below. Flanking residues may also have key interactions with the protease in the cellular context that aid substrate recognition. Although these possible additional interactions of protease with viral polyprotein precursor in the cell are not captured by peptide complex structures, the details of molecular interactions within the binding site may be essential to develop robust small-molecule inhibitors.

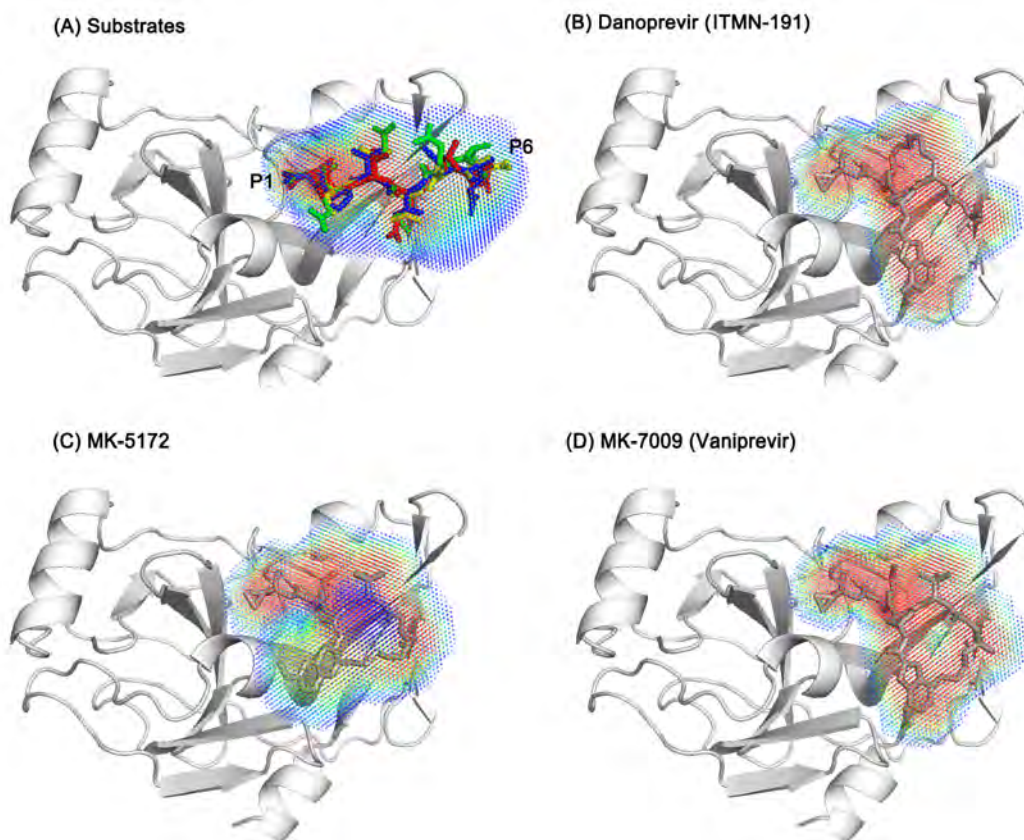
4.3.2 Inhibitor Shape

To assess the three macrocyclic compounds in terms of their mimicry of the substrate envelope, the probabilistic volume occupied by each inhibitor during MD simulations was compared with the dynamic substrate envelope. The inhibitors are not as conformationally flexible as the substrates and therefore they sample a relatively restricted local conformational space than the substrates (more red regions in Figure 4.5). This conformational restriction results in the more

homogeneous volume distribution for inhibitors. The inhibitors, being relatively rigid, are expected to be more severely affected by mutations at the binding site compared to the flexible substrates. In fact, the recently published co-crystal structures of these inhibitors with the wild type and three of the most severe drug resistant protease variants (R155K, A156T, D168A) revealed a complete shift in the binding mode of the inhibitor. [90] This shift in the binding mode is likely due to the rigidity of the inhibitors, which prevent them to adapt to amino acid changes in their target enzyme.

For each inhibitor, the portion of the volume that fails to fit within the dynamic substrate envelope was quantified as V_{out} (Table 4.1). The contribution of each moiety (P4 to P1') to V_{out} was also calculated (Figure 4.4B). The P2 group in each inhibitor is completely outside the substrate envelope. This group makes up, on average, 33% of an inhibitor's total volume. The P1 moiety, although not as severely as P2, fails to fit exclusively within the substrate envelope, with a V_{out} value of $78 \pm 10 \text{ \AA}^3$ for ITMN-191 and MK-5172 and $83 \pm 10 \text{ \AA}^3$ for MK-7009. The inhibitors also have a P1' moiety that protrudes beyond the substrate envelope. A corresponding residue for P1' does not exist in the product crystal structures; in fact no crystallographic data is currently available on the binding mode of the primed side of the cleavage sites. The inhibitor moieties that do not overlap with the substrate volume, especially within the P2 groups, are responsible for the loss of binding affinity upon resistant mutations.

Figure 4.5: Macrocyclic inhibitors are more rigid than the substrates, resulting in a more homogeneous dynamic volume. The four superposed substrates, 3-4A, 4A-4B, 4B-5A, 5A-5B, are shown as sticks in red, green, blue, yellow, respectively. The dynamic substrate envelope and the volume distribution of individual inhibitors are represented with a spectrum from red to blue, reflecting respectively more and less likely occupied regions.

**Figure 4.5**

In addition to staying within the substrate envelope, a robust inhibitor ideally fills the substrate envelope efficiently to gain maximum potency. The three inhibitors were assessed in terms of their efficient use of the substrate envelope by calculating the volume of the dynamic substrate envelope that is not being occupied by inhibitor atoms over the course of the MD simulations, $V_{remaining}$ (Figure 4.6). The volume of inhibitors that are outside the substrate envelope (V_{out} values are $567 \pm 18 \text{ \AA}^3$, $633 \pm 20 \text{ \AA}^3$, $606 \pm 19 \text{ \AA}^3$, Table 4.1) is comparable to the part of the dynamic substrate envelope that is not occupied by any inhibitor atom ($V_{remaining}$ values are $506 \pm 15 \text{ \AA}^3$, $489 \pm 16 \text{ \AA}^3$, and $474 \pm 16 \text{ \AA}^3$, Figure 4.6). Therefore, in principle, a compound with the same total volume as the existing inhibitors can be designed to fit fully within the envelope by removing the parts outside the substrate envelope and adding atoms to the regions of the substrate envelope currently not filled. The challenging goal of developing robust inhibitors that avoid drug resistance mutations may be achieved by optimizing the compounds to minimize V_{out} and $V_{remaining}$ metrics while retaining potency.

Figure 4.6: Macrocyclic inhibitors do not fully occupy the dynamic substrate envelope. The portion of the dynamic substrate envelope that is not used by the inhibitor molecules, $V_{remaining}$, is shown on the structures of (A) ITMN-191, (B) MK-5172, and (C) MK-7009. Stereo images were prepared with PyMOL. [166]

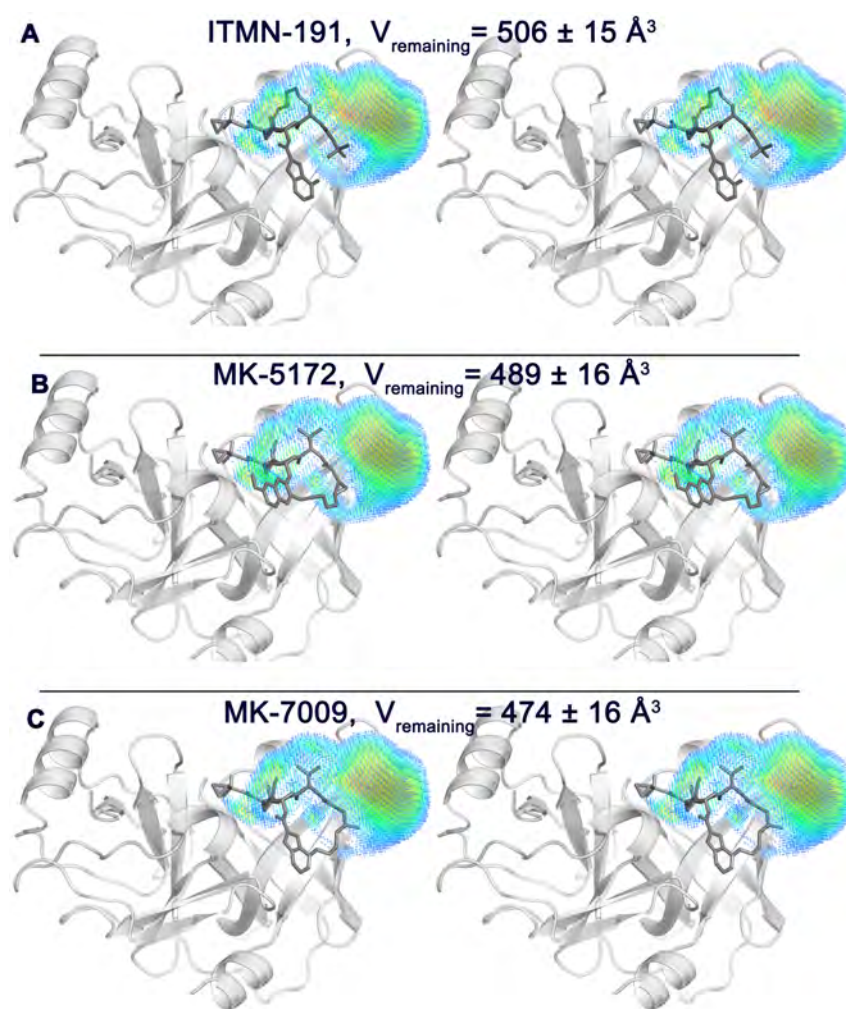


Figure 4.6

4.3.3 Molecular Flexibility and Atomic Fluctuations - Substrates

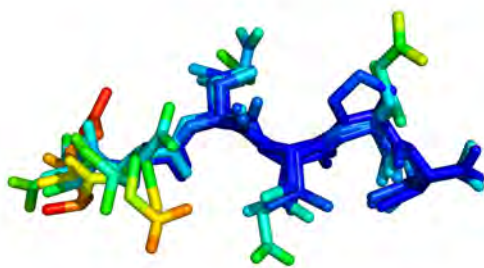
Protease-bound substrates undergo conformational fluctuations around the native state of the complex, reflected in the probabilistic profile of the dynamic substrate envelope. [165] Molecular flexibility is also critical for the substrates to adapt to mutations in the protease and to be recognized as substrates. [183] An ideal inhibitor should mimic aspects of the dynamic characteristics of the natural substrates to have the flexibility to achieve a flat resistance profile. For this reason, the dynamic features of substrates were characterized to compare with those of the inhibitors, by calculating root-mean-squared-fluctuations (RMSF) of alpha carbon atoms in MD simulations, and the intramolecular cooperativity of these fluctuations in the bound state.

Fit of a substrate within the substrate envelope is not solely dependent on the size of the side chains but also the inherent flexibility of the substrate. [165] The two substrates 3-4A and 4A-4B are comparable in size ($1042 \pm 14 \text{ \AA}^3$ and $1097 \pm 15 \text{ \AA}^3$, respectively) and larger than 4B-5A and 5A-5B ($950 \pm 17 \text{ \AA}^3$ and $993 \pm 14 \text{ \AA}^3$, respectively). However, 3-4A is inherently less flexible than the other substrates (Figure 4.3) and has a more conserved volume (Figure 4.2A). In addition, the P4-P1 regions of all substrates are more rigid than the P6-P5 (Figure 4.7A), resulting in a more homogeneous volume distribution of the dynamic substrate envelope at P4-P1 and a relatively heterogeneous volume at P6-P5 sites (Figure 4.2B). Hence, the most likely volume that a substrate adopts in the enzyme binding site should be evaluated based not only on sequence and the binding con-

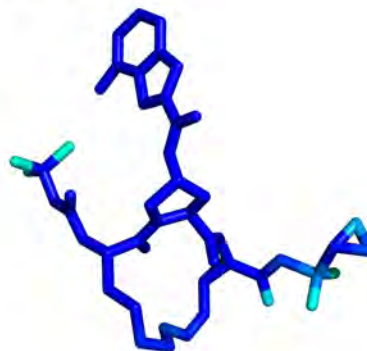
formation observed in the crystal structure, but also on the local conformational plasticity of the substrate in complex with the protease.

Figure 4.7: Compound flexibility is dictated by a combination of features (status/location of the macrocycle, connection points of the linker, and identity of the P2 moiety connected by the linker). The RMSF of substrate and inhibitor atoms are mapped onto the structures of (A) four substrates, (B) ITMN-191, (C) MK-5172, and (D) MK-7009. Warm colors indicate more flexible regions while cool colors correspond to rigid regions.

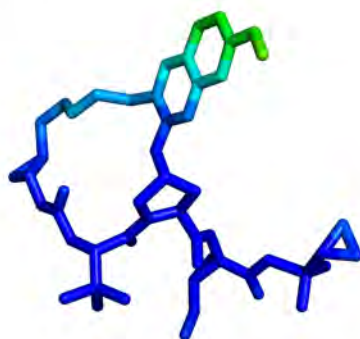
(A) Substrates



(B) ITMN-191



(C) MK-5172



(D) MK-7009

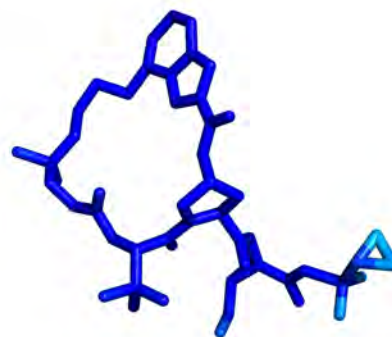


Figure 4.7

Interdependence of the dynamics of substrate residues plays a role in maintaining the molecular shape necessary for substrate recognition. [165, 183] This can be assessed as a time-dependent correlation of atomic fluctuations among the residues of each substrate (Figure 4.8). The residues that fluctuate in a cooperative manner are dynamically coupled. Strikingly, positive correlations within the P4-P1 region are more pronounced than the correlations across the P6-P5 and P4-P1 regions for all four substrates. These correlations within the P4-P1 region last longer than others and still exist even after 1 ns. The positive correlation within the P4-P1 region is likely to aid in maintaining the substrate consensus volume conserved across the four substrates, thereby contributing to substrate recognition.

Substrates exhibit subtle variations in the positional fluctuations and their relative time-evolved interdependence (Figure 4.8). For 3-4A and 4A-4B, atomic fluctuations of the P6 residue are negatively correlated to those of the P3-P1 region. The cooperativity within the 4A-4B substrate is considerably higher than the other substrates even after 1 ns. This higher cooperativity may correlate with the number of charged residues and the higher net charge in 4A-4B. Since electrostatic interactions are long range, movements in one charged residue are more likely to propagate and influence a distal residue if the distal residue is also charged. In fact, movement of the negatively charged Asp-P6 is coupled to the negatively charged Glu-P3 and Glu-P2 since their movement in opposite directions would maximize their separation. This kind of cooperativity may result in interdependence within the substrate sequence. Because 4A-4B protrudes beyond the substrate envelope,

especially at P2 site, more than the other substrates (Figure 4.4), the 4A-4B cleavage site is expected to be more susceptible to co-evolution with drug resistant protease variants. The cooperativity within the 4A-4B sequence suggests that a compensatory mutation in the cleavage site may occur at a position other than P2 that would allow the substrate to adapt to drug resistance mutations in the protease.

Figure 4.8: Correlation of substrate atomic fluctuations within P4-P1 region survives longer compared to those across P6-P5 and P4-P1. Residues are separated by lines, the P6-P5 and P4-P1 regions are separated by a thicker black line.

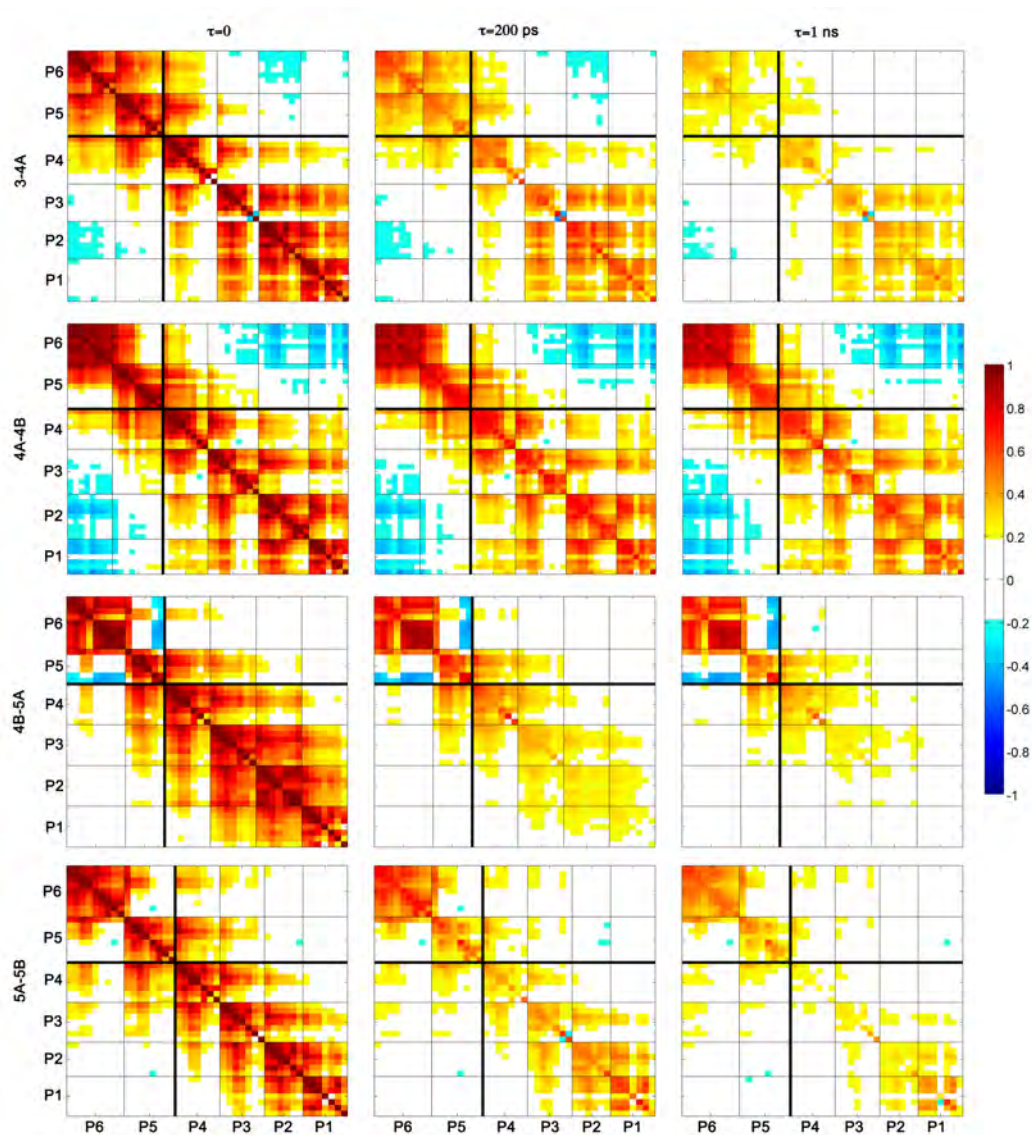


Figure 4.8

4.3.4 Molecular Flexibility and Atomic Fluctuations - Inhibitors

The conformational flexibility of an inhibitor is dictated by a combination of its structural features. Macrocyclic inhibitors may be more susceptible to resistance mutations than more flexible acyclic analogues. Location of a macrocycle linkage (whether the linkage connects the P1 group to P3 or P2 to P4) and the particular atom involved with the linking may also impact the inhibitor's flexibility. The P2-P4 moieties in NS3/4A inhibitors are highly solvent-exposed; hence they are relatively free to move while the P1-P3 moieties are fully buried within the binding site and therefore more conformationally restricted. As a result, a macrocycle connecting the more flexible P2 and P4 moieties may be more effective in constraining the small molecule inhibitor than a linker connecting P1 and P3 moieties that are already relatively rigid. However, the packing of the flexible P2 group may also affect the flexibility of an inhibitor in the binding site. Very tight interactions may lock the inhibitor in a single conformation.

For example, both MK-5172 and MK-7009 have a P2-P4 linker; however their dynamic envelopes suggest that more local conformational states are accessible to MK-5172 than MK-7009, especially for the P2 group and the linker. MK-5172 has an ether-linked quinoxaline P2 group and a linker to C2 of quinoxaline, while MK-7009 has a carbamate-linked isoindoline P2 group and a linker to the C4 of the isoindoline (Table 4.1). In MK-5172, the quinoxaline group is pointing away from the flexible C2-linker and the local fluctuations of this group creates enough momentum to affect the local fluctuations of the linker. In contrast, the isoindoline

in MK-7009 is linked at C4 position, which is located at the tip of the P2 group and renders the constraint more effective. Thus the C4 connection in MK-7009 is more restrictive than the C2 connection.

Interestingly, ITMN-191 entirely lacks this linker, which makes the isoindoline more flexible than MK-7009. Yet the enhanced flexibility of C2-linked quinoxaline in MK-5172 is comparable to the free P2 group of ITMN-191. Thus a P2-P4 macrocycle does not guarantee a severe dynamic constraint in a compound, but the connection point of the P2-P4 linker also has a role in the extent of flexibility. All three effects combined together (location of the macrocycle, the connection points of the linker, and the identity of the P2 moiety connected by the linker) are responsible for the differential dynamics observed in the bound state (Figure 4.7).

In contrast, the backbone of the inhibitors is always stable, consistent with the substrate P4-P1 backbone stability (Figure 4.7). P1-P3 moieties of all three inhibitors have very homogeneous volume distributions due to their rigidity in the bound state (Figure 4.5). The interplay between the rigidity of the P1-P3 groups and the stability of the favorable interactions with the corresponding binding subsite contributes to the tight binding characteristics to these inhibitors. The structural rigidity of these very tight binding compounds, relative to natural substrates, gives rise to high orientational correlations of motion within a compound (Figure 4.9). The effect of a slight movement in one moiety is propagated rapidly to the rest of the molecule because of this rigidity. This propagation of movement is facilitated by the linkers (Figure 4.9), which couple the moieties that they con-

nect. The cross-correlations of atomic fluctuations within the P2-P4 macrocyclic compounds, MK-5172 and MK-7009, are much higher and last a longer timescale than those within the P1-P3 linked compound, ITMN-191, (Figure 4.9). The high cooperativity may prevent the inhibitors from conformational readjusting to adapt to resistance mutations in the protease. For example, the P2-P4 linked inhibitors are more susceptible to A156T mutation than the P1-P3 linked compound (The fold-change in K_i relative to the wild-type is 4429 and 1295 for MK-5172 and MK-7009, respectively but 45 for ITMN-191 [90]), because the relatively more rigid and highly cooperative macrocyclic compounds cannot adjust a single moiety to avoid resistance mutations. To overcome this opportunity for resistance, robust inhibitors to NS3/4A protease should be designed to have an intrinsic flexibility consistent with natural substrates.

Figure 4.9: Inhibitor atomic fluctuations are highly correlated and correlations survive longer in compounds with P2-P4 macrocycle (MK-5172 and MK-7009). Residues are separated by lines, and the macrocycle linker atoms are indicated by darker lines.

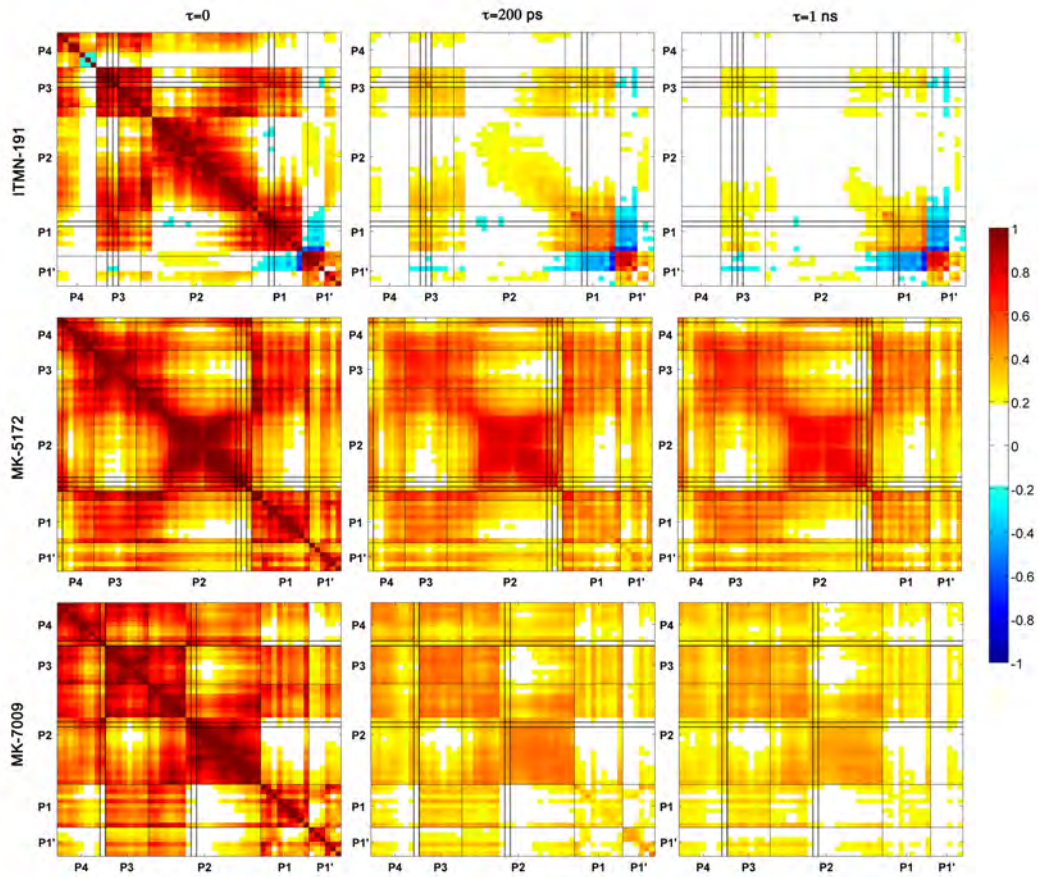


Figure 4.9

4.3.5 vdW Contacts at the Binding Surface

Van der Waals (vdW) interaction energies at a protein interaction surface represent the local packing. A comparative view of protease vdW energies with substrates versus inhibitors during MD simulations reveals the regions on the substrate binding surface that are untapped by the inhibitors, as well as the protease residues that should be avoided in designing new inhibitors to minimize the likelihood of drug resistance (Figure 4.10 and 4.11). The protease residues that are contacted by substrates and the extent of these interactions are conserved across the four substrates (Figure 4.10). The vdW interaction energy of each protease residue was averaged over four substrates and mapped onto the binding surface (Figure 4.10). The interactions with the catalytic H57 appear to be critical for all substrates as well as K136, A156, A157, V158, and S159. A comparison of the natural polymorphisms and drug resistance mutation sites and the contact potential of these sites with the natural substrates reveal that the drug resistance mutations have been selected at sites that are not major contact points for substrates, namely Q41, T42, F53, K80, S138, and D168 (Figure 4.10E). However, R155, A156, and V158 mutate to confer drug resistance although they have extensive, stable interactions with the natural substrates. Based on this structural observation, the viral variants carrying the mutations at these three sites are expected to pay a fitness cost because of a likely interference with substrate recognition, yet the selective pressure of the drug therapy populates the viral variants carrying these mutations.

Figure 4.10: Differences between the binding surfaces of inhibitors versus substrates provide opportunity for resistance mutations. The protease residues that contact (A) the substrates, (B) ITMN-191, (C) MK-5172, and (D) MK-7009 are colored with respect to their vdW contact potential energies. The catalytic triad residues, *H57*, *D81* and *S139*, are indicated by underlined labels in italics. The substrate residues P6 and P1 are labeled to show the orientation of the cleavage products while all five moieties (P4 to P1') are labeled for inhibitors. (E) vdW contact potentials for protease residues. The black and red stars correspond to the natural polymorphic and drug resistance mutation sites, respectively.

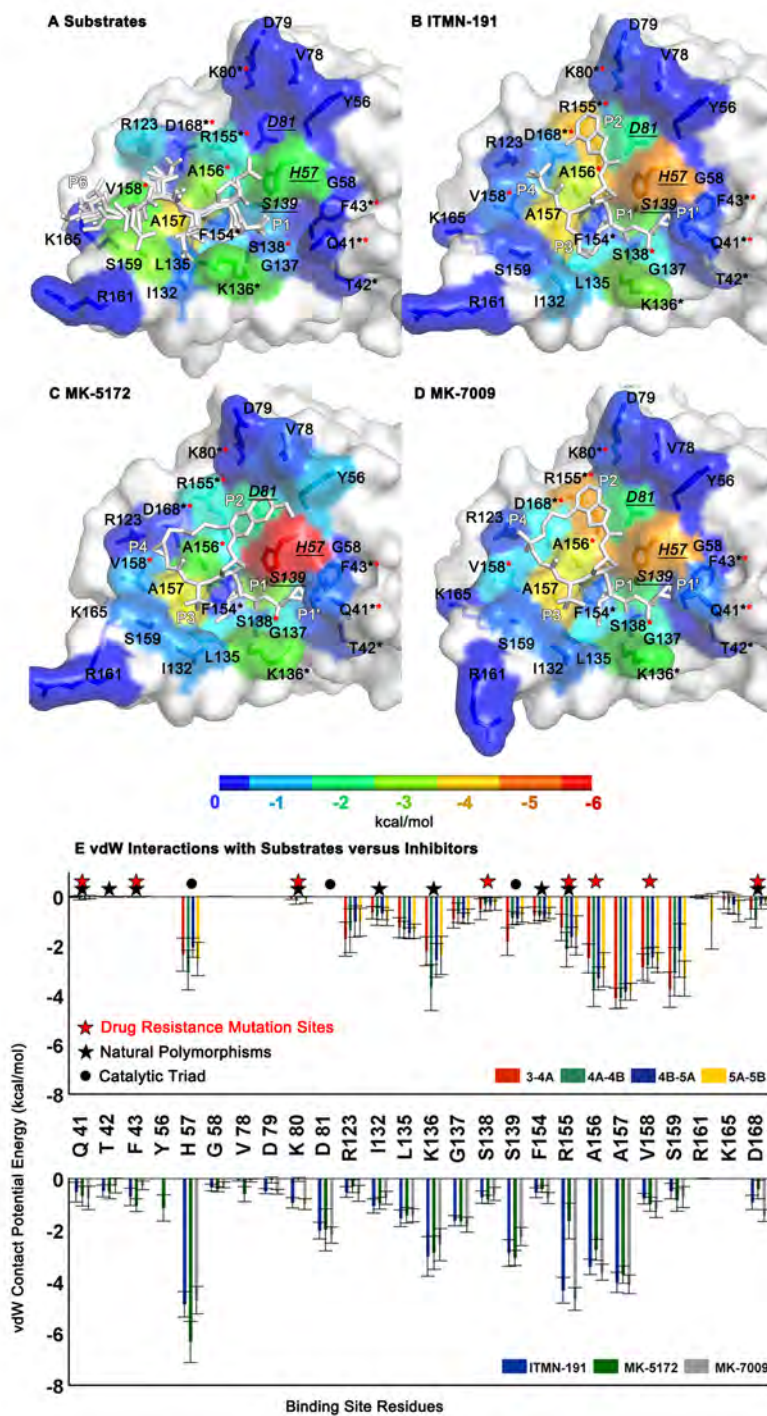


Figure 4.10

Figure 4.11: Differential interaction profiles of inhibitors versus substrates suggest potential improvement at P1 moiety of inhibitors. The vdW contact potential of (A) substrate residues P6 to P1, and (B) inhibitor moieties P4 to P1'.

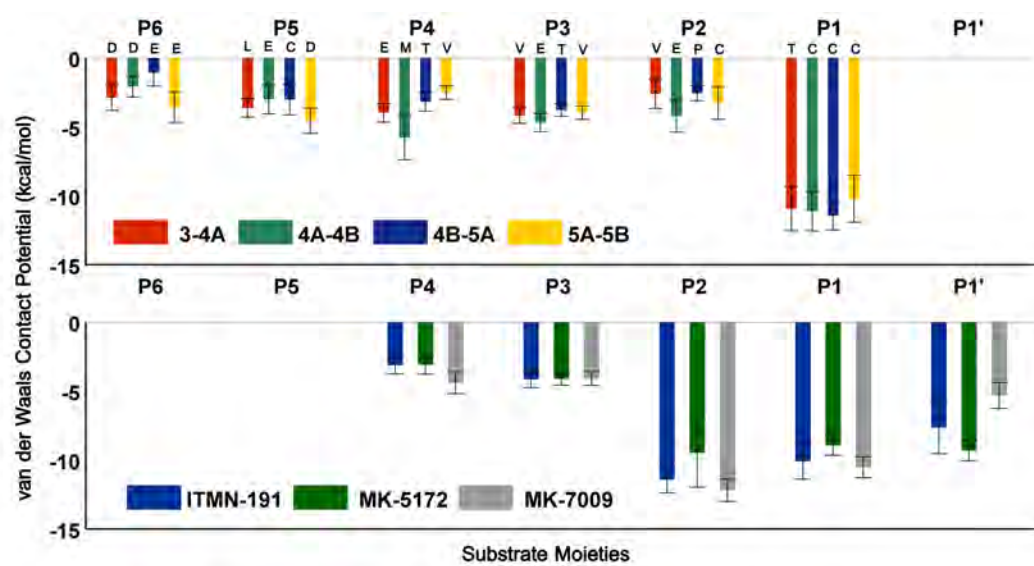


Figure 4.11

The subtleties between the overlapping binding surfaces of the substrates versus inhibitors are reflected in their differential vdW interaction energies with the protease (Figure 4.10). The substrate complexes have larger fluctuations than the inhibitor complexes (higher standard deviations in Figure 4.10E), likely as a result of their greater structural flexibility. The lower standard deviation of inhibitor interactions indicates that a particular interaction has remained stable throughout the simulated time interval. The total vdW interaction energy of the three inhibitors is comparable, -36.5 ± 2.0 , -35.0 ± 2.8 , and -36.6 ± 1.8 kcal/mol for ITMN-191, MK-5172, and MK-7009, respectively. The catalytic histidine is the main contact for all inhibitors. The extensive interactions with H57 may contribute to the broad activity of MK-5172 across different genotypes and resistant variants. [188] The residues H57, S139, R155, and D168 have more favorable interactions with the inhibitors than substrates, while the residues Q41, T42, F43, Y56, G58, V78, D79, K80, and D81 do not contact the substrates at all but interact with the inhibitors. These residues have low contact potential (in the range of 0 to -1 kcal/mol) but may have a collective impact on the binding affinity. Taken together, targeting the protease residues that are major contact points for natural substrates, including the catalytic triad (H57, D81, S139), and minimizing interactions with the variable sites is a promising strategy for avoiding resistance without compromising selectivity.

While the maximum possible interaction potential is desired for potent inhibitors targeted against an ensemble of resistant variants, the substrates also need to be specifically recognized to permit enzyme cleavage. Suboptimal interaction

potential between the protease and a substrate may negatively impact the recognition at the binding interface, decreasing the association rate and/or increasing the dissociation rate, while higher-than-optimal interaction potential may interfere with product release, causing product inhibition. 4A-4B is a larger substrate with 50 non-hydrogen atoms in the P6-P1 sequence (Table 4.1), making slightly more favorable vdW contacts with the protease (-31.9 ± 2.5 kcal/mol) while 3-4A has a lower contact potential (-26.9 ± 2.8 kcal/mol). The more favorable contacts of 4A-4B likely compensate for the high degree of inherent conformational flexibility of this substrate, ensuring sustained affinity. In contrast for 3-4A, making less favorable contacts than average may be tolerated as 3-4A is processed in a cis cleavage reaction, increasing the effective concentration through the intramolecular nature of the processing of this particular site.

A large portion of substrate vdW energy with the protease ($\approx 40\%$) involves interactions with P1 position (Figure 4.11). Indeed, the Cys in 3-4A and Thr in the other substrates at P1 position are nicely packed in this small cavity compared to the other substrate residues, which contact the relatively shallower parts of the binding surface. The macrocyclic inhibitors make more favorable vdW contacts than the substrates at the P2 position due to the bulky isoindoline (ITMN-191/MK-7009) or quinoxaline (MK-5172) groups. While the P2 interactions contribute to the high affinity to the wild-type protease, most of them are outside the substrate envelope (Figure 4.4) and therefore render the inhibitors susceptible to resistance mutations.

4.3.6 Electrostatic Network

Electrostatic interactions are typically more specific and longer range than vdW interactions and contribute significantly to molecular recognition. All of the protease-substrate hydrogen bonds, except for the one formed by R123, are conserved across the substrates in the crystal structures. [9] A majority of these hydrogen bonds also persists during the MD simulations (Table 4.2), especially the ones involving A157 and C/S159. The protease-substrate hydrogen bonds formed by the substrate backbone are more stable, with all except one (between the $\epsilon 2$ oxygen of the Glu-P4 of 3-4A substrate) existing for at least 50% of the simulated time. Backbone hydrogen bonds are independent of sequence as all amino acids share the same backbone, hence the same hydrogen bond donors/acceptors. This relative non-specificity of shared backbone hydrogen bonds suggests that binding specificity is due to other molecular recognition motifs. Although the hydrogen bonds are conserved across the substrates, the percent time that these interactions exist throughout the simulations vary between the complexes. Substrate 4B-5A preserves the complete hydrogen-bonding network with the protease in the simulation while 4A-4B preserves half of the hydrogen bonds, consistent with higher flexibility. All P1 terminal carboxyl groups sit in the oxyanion hole, hydrogen bonding with the $\epsilon 2$ nitrogen of the catalytic histidine (H57), which exists 97% of the time in 4B-5A complex while in other substrates this bond exists $\approx 70\%$ of the time. Similarly, the backbone hydrogen bond between P1 and G137 is present 89% of the time in the 4B-5A complex whereas the same bond exists 59%

of the time or less for other substrates. These variations likely arise from other differences in molecular interactions with the enzyme and inherent differences in substrate flexibilities and contribute to the substrate specificity.

Table 4.2: The frequency of the protease-substrate/inhibitor hydrogen bonds in MD simulations

<i>Substrates</i>		Cleavage site			
Protease		3-4A	4A-4B	4B-5A	5A-5B
H57 – NE2	P1 – O	72	71	97	69
G137 – N	P1 – O	55	<50	89	59
S139 – OG	P1 – O	53	<50	58	<50
R155 – O	P1 – N	<50	<50	56	<50
A157 – N	P3 – O	69	85	76	80
A157 – O	P3 – N	87	82	72	84
R123 – NH2	P4 – OE2	51	<50	<50	<50
C/S159 – N	P5 – O	78	63	57	81
S159 – O	P5 – N	<50	<50	<50	59
<i>Inhibitors</i>		Inhibitor			
Protease		ITMN-191	MK-5172	MK-7009	
G137 – N	P1'	78	77	74	
S139 – N	P1'	71	67	69	
G137 – N	P1	<50	<50	<50	
S139 – OG	P1	95	95	90	
R155 – O	P1	89	53	93	
A157 – N	P3	89	87	89	
A157 – O	P3	76	90	88	

Interestingly, despite not being an essential part of the dynamic substrate envelope, the P6 residue is conserved at the amino acid level (Asp or Glu), suggesting functional importance. Structural evidence suggests that this arises from the potential electrostatic interactions of P6 acid with the protease and possibly in long-range substrate recognition, and thus is included in the definition of the dynamic substrate envelope. In all crystal structures except for the product complex 4B-5A, K165 forms salt bridges with the P6 acids. [189] In the MD trajectories, the P6 acid-K165 salt bridge criterion was satisfied for 12%, 26%, 35%, and 57% of 100 ns simulated time for 3-4A, 4A-4B, 4B-5A, 5A-5B, respectively. The percentages are relatively low for the first three complexes, however the disrupted salt bridge reforms in a recurring manner. In addition, the percentage of time P6 acid forms a salt bridge with any one of the four binding site basic residues (R113, R119, R161, K165) is 53%, 68%, 50%, and 83% for 3-4A, 4A-4B, 4B-5A, 5A-5B, respectively (Figure 4.12). These numbers suggest that the P6 acid, in solution, interacts with at least three more basic residues on the binding surface in a dynamic manner in addition to K165 observed in the crystal structures. The P6 position, while forming salt bridges with a series of binding site residues, does not adopt a conserved consensus volume. The highly dynamic interaction pattern of the P6 acid with the protease binding surface rationalizes the strong conservation of Asp or Glu at this position as well as the dramatic reduction in cleavage efficiencies of substrates with positively charged P6 substitutions. To mimic the salt bridges between the protease and the P6 acid, inhibitors would require a negatively charged P6 and/or P5 moiety that is flexible enough to form interchanging salt bridges with a series

of binding site residues. The probabilistic character of the dynamic substrate envelope provides a basis for building the P5 and/or P6 moieties with the relevant charge and flexibility requirements, preventing potential pitfalls and misleading results in case the static P6 crystallographic volume were to be satisfied.

Figure 4.12: P6 Asp/Glu of substrates participates in a dynamic salt bridge network with the protease residues R119, R123, R161 and K165. The salt bridge distance over simulation time (left) and the corresponding histogram (right) is displayed for (A) 3-4A, (B) 4A-4B, (C) 4B-5A, and (D) 5A-5B. Salt bridges were defined as an interaction between a side-chain oxygen atom of Asp or Glu within 4.0 Å of a nitrogen atom of Arg or Lys (indicated by the red horizontal and vertical lines).

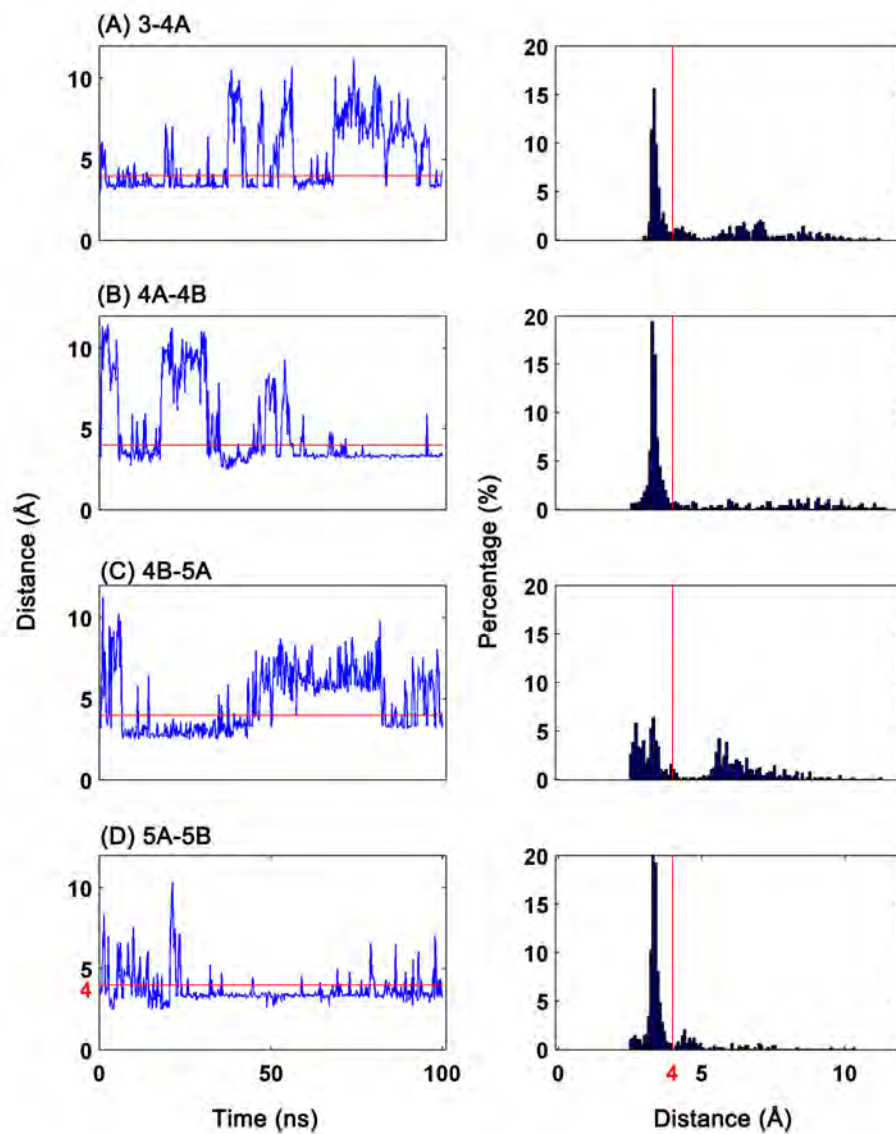


Figure 4.12

The protease-inhibitor hydrogen bonds are also formed by the backbone of protease binding site residues, except for the γ oxygen of the catalytic serine S139 with the P1 group of each inhibitor. The backbone hydrogen bonds should in theory be preferred while designing new drugs to avoid losing a specific hydrogen bond donor/acceptor in a side chain upon an amino acid substitution, however a drug resistance mutation may also cause slight backbone structural changes, which may or may not affect the distance/angle between the donor and the acceptor of a hydrogen bond. For example, R155, which is a site of drug resistance mutations, makes a backbone hydrogen bond with the P1 position of all three inhibitors, with MK-5172 being less stable. The stability of this backbone hydrogen bond in the simulations of mutant complexes deserves further attention in a follow-up study. A157 makes two backbone hydrogen bonds with P3 position of all four substrates, which is successfully mimicked by the inhibitors. The backbone hydrogen bond formed by G137 and P1 of at least three substrates is also formed by the P1' moiety of inhibitors. In addition, the P1' moiety makes a hydrogen bond with the backbone nitrogen of the catalytic serine $\approx 70\%$ of the simulated time. The inhibitors fail to mimic the hydrogen bond with the catalytic histidine that appears to be conserved across the substrates.

The electrostatic network along one surface of the bound cleavage products between protease residues R123, D168, R155, and the catalytic D81 has been proposed to increase the stability of the bound state and therefore contribute to substrate recognition. [189] The salt bridges D81-R155 and R155-D168 in the substrate complexes appear to be mutually exclusive, suggesting that R155 is shared

between D81 and D168 (Figure 4.13). This exchange of salt bridges along the binding surface does not occur in inhibitor complexes as the rigid tight binding inhibitors lock the conformation of the side chains at the binding site stabilizing individual salt bridges. In the inhibitor-bound complexes, R155-D168 salt bridge is extremely stable (Figure 4.14). The D81-H57 salt bridge exists throughout the inhibitor simulations, as R155 is not available for salt bridging to D81.

Figure 4.13: R155 is shared by D81 and D168 in a salt bridge network along the binding surface. The salt bridges formed along the binding surface by three pairs of protease residues (D81-R155, R155-D168, and D168-R123) were monitored throughout the 100 ns MD simulation trajectory for four substrate complexes (3-4A, 4A-4B, 4B-5A, 5A-5B).

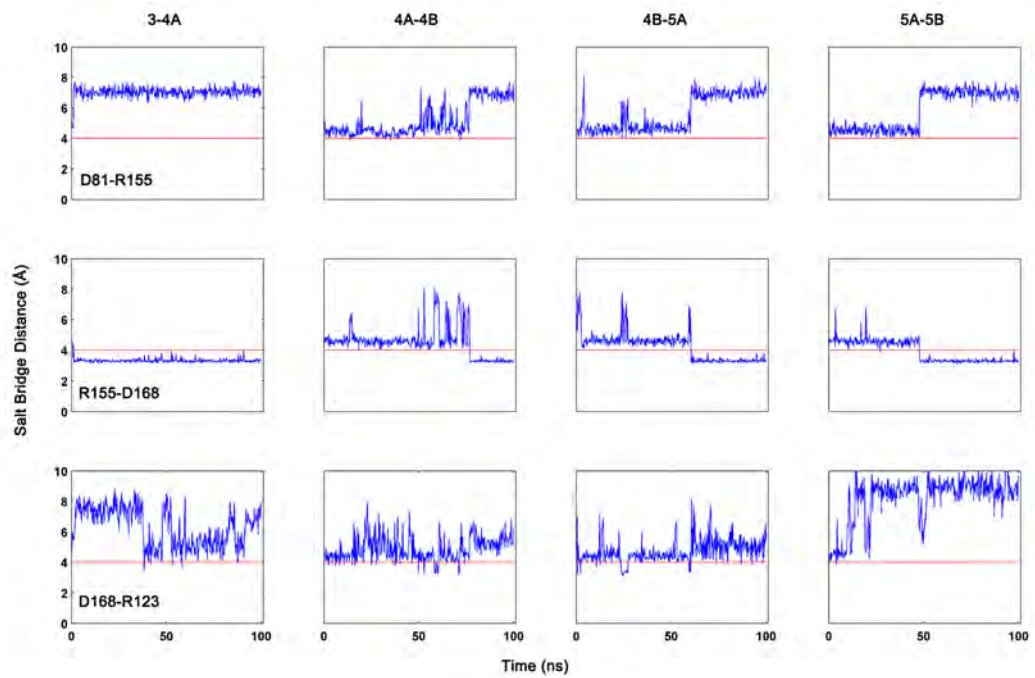


Figure 4.13

The D168-R123 salt bridge is an exception, which appears to be transient in both substrate and inhibitor MD simulations (Figure 4.13 and 4.14). The enhanced flexibility towards the P6 site in the substrates may have given a relative conformational freedom to the R123 side chain while the inhibitors have minimal interactions with this residue due to the lack of P6-P5 moieties. Two of the most severe drug resistance mutations, R155K or D168A, could disrupt this salt bridge causing an energetically unfavorable perturbation in the inhibitor complexes. The maintenance of this electrostatic network within the protease residues should also be assessed in inhibitor design.

In addition to the electrostatic interactions within the protease substrate or inhibitor complexes, recent work also suggests that water molecules can play an important role in molecular recognition. [190–192] While the substrates do not have conserved water-mediated hydrogen bonds with the protease in the crystal structures, the inhibitors ITMN-191 and MK-7009 have eight crystallographic water molecules that are within 4 Å of both protease and inhibitor. The solvent exposure of the binding site combined with the effects of crystal contacts may present challenges in accurately assessing the relative importance of these water molecules.

Figure 4.14: The R155-D168 salt bridge is stabilized by D81-H57 salt bridge in inhibitor complexes. The salt bridges formed along the binding surface by three pairs of protease residues (D81-H57, R155-D168, and D168-R123) were monitored throughout the 100 ns MD simulation for three inhibitor complexes (ITMN-191, MK-5172, and MK-7009).

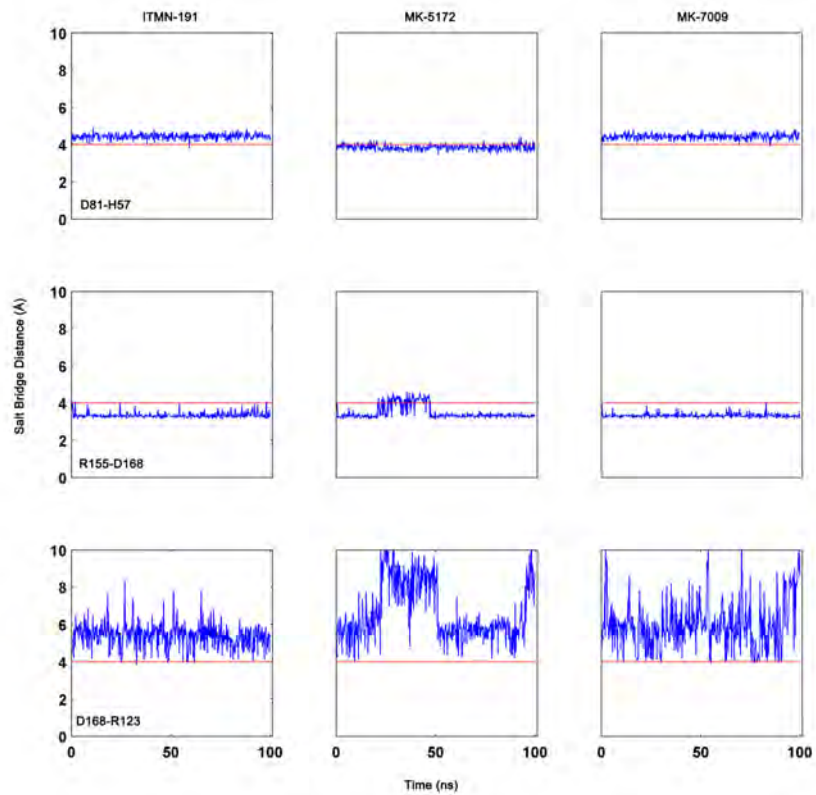


Figure 4.14

4.4 Conclusion

Understanding differential interaction patterns between substrates and inhibitors can help guide the design of new inhibitors that better target resistant protease variants. Comparison of a series of structural and dynamic features of inhibitors and substrates reveals key insights for structure-based design of inhibitors that are robust against resistance mutations. One aspect of substrate specificity that can be used to optimize inhibitors is the dynamic substrate envelope. This is a two-sided optimization problem: (1) the volume of the compound that stays outside the envelope (V_{out}) should be minimized to avoid contacts with protease residues that are not important for substrate recognition and at the same time (2) the portion of the substrate envelope that is not being filled by the inhibitor ($V_{remaining}$) should be minimized to efficiently use the substrate envelope when growing the compounds. These need to be achieved while maintaining the potency of the inhibitors.

Considering the effect of protein dynamics in substrate binding modes and defining the substrate envelope as a probability distribution is especially critical for the HCV NS3/4A protease. The high solvent exposure of NS3/4A binding site allows the substrates to be highly flexible. Accessibility of more conformational states to the substrates means they are less constrained within the consensus volume defined by the crystal structure conformations; thus, the increased space explored by the substrates results in a smaller consensus volume. The dynamic substrate envelope, defined as a probability distribution of the substrate volume in the binding site, better captures the substrate binding features and allows for a

more accurate assessment of the inhibitors during drug design. This probabilistic dynamic substrate envelope allows for a detailed investigation of the individual moieties on an inhibitor and an accurate evaluation of the similarities/differences between substrate and inhibitor complexes.

The most conserved substrate volume, when protein dynamics is taken into account, is defined by the P1 to P4 residues while P5 and P6 have more conformational freedom. P6 is able to serve as a hub in an intermolecular electrostatic network involving the protease residues R119, R123, R161 or K165, dynamically flipping between these basic residues. In theory, an inhibitor can be designed to have a charged P6 moiety with molecular flexibility consistent with the natural substrates to facilitate these electrostatic interactions with the protease. Including the P6 residues of the substrates in the dynamic substrate envelope will provide a useful basis to guide the design of these charged compounds. Based on this profile, inhibitors should be designed such that their regions spanning P1-P4 portion of the substrate envelope are more restrained while the moieties spanning P5-P6 sites are more flexible. Ideal inhibitors of NS3/4A protease should be designed to have the intrinsic flexibility consistent with the natural substrates.

To minimize the susceptibility of an inhibitor to drug resistance mutations, interactions with functional residues, which include the catalytic triad and also the residues that play a key role in substrate recognition in HCV protease, should be enhanced. In contrast, the interactions with known drug resistance mutation sites or polymorphic sites should be minimized. R155, which is among the most severe drug resistance mutation sites, is in extensive contact with the P2 groups

of all three inhibitors. During the optimization of new compounds, favoring the P2 conformation over D81 and H57 away from R155, as in MK-5172, is expected to be useful for better performance against the R155K variant. In fact, the fold-change in the inhibition constant K_i against the R155K variant relative to the wild type is 162, 6, and 749 for ITMN-191, MK-5172, and MK-7009, respectively. [90] In addition, the P2-P4 macrocycle enhances the protease interactions through D168 and V158. Breaking this macrocycle could reduce the resistance severity of the D168A mutation. In fact, fold changes in K_i values for ITMN-191 and MK-7009 are 208 and 3561, respectively, against the D168A variant. [90] Finally, P1-P3 macrocycle appears to hinder P1 to fill the S1 pocket completely. P1 may be modified to explore the possibility of enhancing the interactions with F154 in the S1 pocket, consistent with the substrates very favorable interactions at P1 site.

Another binding interaction to optimize is the hydrogen bonding. When placing the hydrogen bond donors/acceptors in the compounds, the protease backbone donors/acceptors could be targeted at polymorphic sites. While a conformational readjustment is possible with an amino acid substitution at these sites, a hydrogen bond with the side chain is more likely to be lost upon a mutation. In addition, the substrate P6 residues salt bridges with a series of protease binding site residues. Maintenance of this electrostatic network within the protease could also be taken into account during inhibitors design.

In conclusion, the binding volume of a robust inhibitor should ideally not fall outside of the dynamic substrate envelope to achieve a better resistance profile. When searching the chemical space in the design process, ranking the candidates

based on their V_{out} and $V_{remaining}$ in combination with vdW contacts and the electrostatic interactions may predict which compounds will be more susceptible to resistance. The exact location where the inhibitor protrudes beyond the substrate envelope is also critical. Some regions of the dynamic substrate envelope is more heterogeneous, therefore the compound could have more freedom at those regions. Relative importance of subpockets should be explored for each compound in development. Protruding beyond the substrate envelope might be tolerable to pick up interactions with the catalytic triad since the catalytic residues are evolutionarily conserved and highly unlikely to mutate to confer drug resistance. In addition, modifications to ligands may be needed to alter properties other than target affinity and selectivity (i.e. solubility, metabolism, permeability, or efflux). [193] Extending beyond the substrate envelope in regions that do not contact the enzyme could be used to tune such properties and modifications in these regions are likely to be better tolerated than those that contact the enzyme. Finally, designing inhibitors to have flexibility consistent with substrates should improve the chances of avoiding resistance mutations.

4.5 Methods

4.5.1 Protease-Substrate/Inhibitor Complex Structures

Structural and dynamic properties of four substrates and three small-molecule inhibitors in complex with the NS3/4A serine protease were investigated (Ta-

ble 4.1). Crystal structures of the three protease-substrate complexes, which correspond to the cleavage sites 4A-4B, 4B-5A, and 5A-5B on the viral polyprotein, were available (PDB ID: 3M5M, 3M5N, 3M5N). [9] The structure of the fourth cleavage site, 3-4A, was modeled based on the full-length helicase/protease structure [186] (PDB ID: 1CU1) by deleting the helicase domain but keeping the C-terminal eight residues in the enzyme binding site. This C-terminal region corresponds to the product of the cleavage site 3-4A. In addition, the co-crystal structures of three macrocyclic inhibitors ITMN-191, MK-5172, and MK-7009 (PDB ID: 3SU0, 3SUD, 3SU3) were used. [9]

4.5.2 Structure Preparation

For the co-crystal structures in which more than one molecule exists in the asymmetric unit, all of the molecules were inspected for major structural changes in the backbone, average temperature factors, the number of missing side chains, and the ambiguity of the electron density for the substrate or inhibitor in the binding site. A choice of molecule was made when molecules were equivalent in terms of structural quality. The complex molecules with protease chain IDs A, B, D, and A were chosen for the simulations of 3-4A, 4A-4B, 4B-5A, and 5A-5B, respectively. The crystallographic waters within 4.0 Å of any protein or ligand atom were kept, however all the buffer salts were removed from the coordinate files.

The protease in the substrate co-crystal structures, except for the 3-4A complex, has the S139A substitution in the active site, which was originally thought to inactivate the enzyme, prevent hydrolysis, and allow capturing the intact sub-

strates in the active site. To get a more realistic model of the active enzyme interactions with the substrates, the A139S back-mutation was introduced *in silico* to the crystal structures by deleting the Ala side chain and predicting the conformation of the Ser side chain using the software Prime. The N-terminal residues in the protease construct, GSHMASMKKK, were extremely flexible in the first set of trial simulations and did not form stable secondary structure. In addition, these residues did not interact with either protease or substrate. Therefore, this region was deleted in subsequent simulations to reduce the computational cost of the simulations.

As the x-ray structures are not suitable for immediate use in MD simulations, [194] the above described structures were prepared using the Protein Preparation Wizard from Schrödinger. The process adds hydrogen atoms, builds side chains with missing atoms, and determines the optimal protonation states for ionizable side chains and ligand groups. In addition, the hydrogen bonding network is optimized by flipping the terminal chi angle of Asn, Gln, and His residues and sampling hydroxyl/thiol polar hydrogens. The exhaustive sampling option with the inclusion of water orientational sampling was used. Following this step, the structure was minimized in vacuum with restraints on heavy atoms using the Impact Refinement module with the OPLS2005 force field and terminated when the root-mean square deviation (RMSD) reached a maximum cutoff of 0.3 Å. This step allows hydrogen atoms to be freely minimized, while allowing heavy-atom movement to relax strained bonds, angles, and clashes.

4.5.3 MD simulations

Desmond [195] with OPLS2005 force field [152, 196] was used in all simulations. The prepared systems were solvated in an orthorhombic solvent box with the SPC water model extending 10 Å beyond the protein in all directions using the *System Builder* utility. The overall charge of the system was neutralized by adding the appropriate number of counterions (Na^+ or Cl^-).

Each system was relaxed using a protocol consisting of an initial minimization restraining the solute heavy atoms with a force constant of $1000 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ for 10 steps with steepest descent and with LBFGS method up to 2000 total steps with a convergence criterion of $50.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$. 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 $1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ with 5000 steps (250 steepest descent plus 4750 LBFGS) for each value of force constant (1000, 500, 250, 100, 50, 10, $1.0 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$) 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 $50 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ 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 short-range 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. [155, 197] 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 50 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 100 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) [155] 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.

4.5.4 Quantitative Definition of the Dynamic Substrate Envelope

All MD trajectories were aligned onto the 4A-4B co-crystal structure using the alpha carbon of binding site residues 137-139 and 154-160, which are less flexible compared to the rest of the molecule. The effect of structural alignment on the resulting substrate envelope was previously assessed to show that the selection of residues for alignment does not have a significant impact on the final substrate envelope as long as the selected residues are located in a relatively less mobile region of the enzyme. [183] The structural alignments were performed using the Visual Molecular Dynamics (VMD) software. [156]

Substrate conformers were extracted from the aligned MD trajectories. The

vdW volumes of all conformers were mapped onto the three-dimensional grid placed on the binding site of the enzyme and normalized by the total number of conformers to obtain a probability distribution. The vdW radii were taken from OPLS2005 force field. The mathematical details of the grid-based volume calculations can be found elsewhere. [165]

4.5.5 Visualizing the Dynamic Substrate Envelope

The substrate conformations from pre-aligned snapshots of MD simulations were loaded into PyMOL. A vdW map was generated for each conformer using the *map_new* functionality. The maps were then summed using the *map_set* function. The final map was shown in surface representation with a contour level that corresponds to 25% of the total number of conformations used.

4.5.6 Estimating the vdW Contact Potential

The vdW contact energy between the protease and a substrate/inhibitor was estimated by a simplified Lennard-Jones potential, $V(r_{ij})$, using the equation

$$V(r_{ij}) = 4\epsilon \left[\left(\frac{\sigma}{r_{ij}} \right)^{12} - \left(\frac{\sigma}{r_{ij}} \right)^6 \right] \quad (4.1)$$

where r_{ij} is the interatomic distance between the protease atom i and the substrate/inhibitor atom j . The ϵ and σ are the well depth and hard sphere diameter, respectively, for the ij protease-substrate/inhibitor atom pair. $V(r_{ij})$ for all possible protease-substrate/inhibitor atom pairs was computed within 6 Å, and when

the distance between nonbonded pairs are less than that corresponds to ϵ , $V(r_{ij})$ was equated to ϵ . The rationale for this modification to the original 6-12 Lennard-Jones potential was previously described in detail. [165] Using this simplified potential for each nonbonded protease-substrate/inhibitor pair, $\sum V(r)_{ij}$ was then computed for each protease and substrate/inhibitor residue. The vdW parameters, ϵ and σ , were taken from OPLS2005. For the pairs involving two separate atom types, the vdW parameters were geometrically averaged.

4.5.7 Hydrogen Bonds and Salt bridges

The percentage of time a hydrogen bond existed between the protease and a substrate/inhibitor was calculated using VMD. [156] A hydrogen bond was defined by a distance between the donor and acceptor of less than 3.5 Å and a hydrogen-donor-acceptor angle of less than 30°. Only the hydrogen bonds that existed more than 50% of the time were considered in the analyses. Salt bridges were defined as an interaction between a side-chain oxygen atom of Asp or Glu within 4.0 Å of a nitrogen atom of Arg or Lys.

4.5.8 Root-Mean-Squared Fluctuations and Time-Delayed Correlations

The normalized time-delayed orientational cross-correlations between residue pairs are defined as

$$CO_{i,j}(\tau) = \frac{\langle \Delta R_i(t) \Delta R_j(t + \tau) \rangle}{\langle \Delta R_i(t)^2 \rangle^{1/2} \langle \Delta R_j(t + \tau)^2 \rangle^{1/2}} \quad (4.2)$$

where $\Delta R_i(t)$ is the fluctuation in the position vector R_i of site i at time t around its average position throughout the trajectory and $\Delta R_j(t + \tau)$ is the fluctuation in the position vector R_j of site j at time $t + \tau$. The brackets represent time averages. The cross-correlations vary in the range $[-1, 1]$ with the upper and lower limits indicating fully correlated (moving in the same direction) and fully anti-correlated (moving in the opposite direction) atomic fluctuations, respectively. $\tau = 0$ gives the equal-time cross-correlations of atomic fluctuations.

Chapter V

**Structural Insights into the Drug
Resistance Conferred by
R155K/V36M in HCV NS3/4A
Protease**

5.1 Abstract

NS3/4A protease inhibitors telaprevir and boceprevir were recently approved by FDA for treatment of chronic HCV genotype 1 infection to be used in combination with Peg-IFN/RBV. Both telaprevir and boceprevir are acyclic ketoamide covalent inhibitors. Several non-covalent NS3/4A inhibitors are at various stages of clinical development, including danoprevir (ITMN-191), which is a macrocyclic acylsulfonamide inhibitor. The co-occurring protease mutations R155K/V36M have been reported to confer drug resistance both in replicon assays and in clinic. R155K is an active site mutation that directly alters the electrostatic characteristics of the binding surface on the enzyme. In contrast, V36M is a distal mutation that is not in direct contact with either bound inhibitors or active site residues but in direct contact with the cofactor NS4A and the B1 β -strand of NS3. The molecular mechanism by which V36M in the presence of R155K confers higher levels of resistance than R155K alone is not known. In this chapter, the molecular basis of drug resistance conferred by V36M in the R155K background has been investigated using a combination of drug susceptibility assays, x-ray crystallography, and molecular dynamics simulations. R155K and R155K/V36M co-crystal structures of each inhibitor differ mainly in residues on α 2 helix and E2 strand near the active site. Telaprevir-bound R155K/V36M had alternate conformations at the site of V36M mutation and the two binding site residues, D168 and R123, suggesting a coupling between the structural destabilization at the binding surface and distal V36M mutation site. The R155K mutation disrupts the

electrostatic network along the binding surface. Danoprevirs P2 moiety is highly flexible in the MD simulations since its P2 moiety packs against R155 and upon R155K mutation; danoprevir loses significant interactions with both R/K155 and catalytic H57 and D81. Finally the significant changes in the internal distance distributions of residues 36 and 43 as well as 43 and a series of binding surface residues 136/155/156/157 are considerably different in the wild-type, R155K, and R155K/V36M complexes. The unbound form also samples conformations where the internal distances vary within the three variants. Based on these observations, one possible mechanism is that the impact of V36M mutation may be propagated to the binding site through a conformational shift in B1 β -strand, which is a spatial neighbor of both V36M and the binding site. V36M mutation may also interfere with the binding of the cofactor NS4A, which aids in the proper folding of NS3.

5.2 Introduction

The new standard of care for Hepatitis C patients emerged in 2011 with the regulatory approval of the first generation NS3/4A protease inhibitors, telaprevir (Incivek(®)) and boceprevir (Victrelis(®)). These protease inhibitors, combined with Peg-IFN/RBV, offer genotype-1 infected patients significantly improved sustained virologic response rates (68-75%) and the potential for shorter durations of therapy. [198–200] However, efficacy of the approved inhibitors is limited to a single genotype. In addition, severe side effects such as anemia, rash or depression [201] combined with pill burden, food requirements, and emerging drug

resistance [202–206] are serious drawbacks to these PIs. Thus, HCV therapeutics continues to evolve and there are several drugs currently in development targeting NS3/4A. Among these drugs, danoprevir (RG7227/ITMN-191) is a macrocyclic inhibitor that is currently being administered in Phase II clinical trials to genotype-1 patients.

Response to therapy is varied in HCV patients, but cross-resistance to PIs is a common issue. In patients with genotype 1a, the R155K mutation causes resistance against nearly all inhibitors, [207] though it rarely occurs in genotype 1b patients due to high genetic barrier in 1b. [85, 208–214] In genotype 1b patients, resistance mutations arise depending on the class of inhibitors use; A156 mutates in response to treatment with linear ketoamide protease inhibitors [85, 213, 214] while macrocyclic PIs mainly select for D168A and R155K variants. [207, 209, 210] Mutations at V36, T54, V36/R155 were initially reported to be associated with resistance to ketoamide inhibitors. [85, 213, 214] However, recent data released at the Liver Meeting 2012 [215] shows that an interferon-free mericitabine/danoprevir combination therapy resulted in a confirmed viral breakthrough in 21% of patients. These patients had danoprevir-resistant variants, including R155K/V36M.

Figure 5.1: Sequence and topology of HCV NS3/4A protease

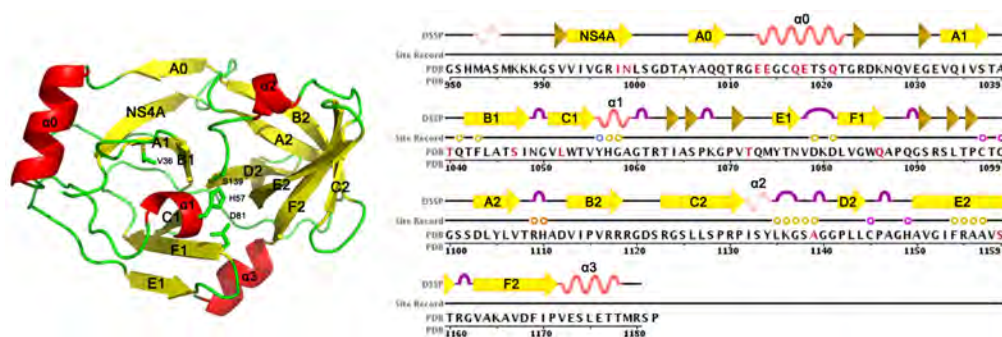


Figure 5.1

NS3/4A, a serine protease with a chymotrypsin-like fold, has two globular domains surrounding a cleft containing the active site residues, H57-D81-S139, on the N-terminal β -barrel. The same nomenclature introduced by Barbato *et al.* was used for the secondary structural organization of NS3/4A (Figure 5.1). [216] R155 is located at the binding site of NS3/4A and is in direct contact with the inhibitors while V36 is a distal mutation site, which is located on A1 β -strand about 13 Å away from the catalytic histidine and 15 Å from R155 (Figure 5.1). Crystallographic studies showed that R155K disrupts the electrostatic network among R123-D168-R155-D81 altering the charge distribution along the binding surface and affecting the P2 cyclopentylproline and P4 cyclohexylalanine moieties of telaprevir. [90] Meanwhile, danoprevir loses a favorable cation- π interaction at the P2 isoindoline upon R155K, shifting the P2 moiety. Pan *et al.* performed molecular dynamics simulations and per-residue binding free energy composition on the wild-type and four variants of NS3/4A protease carrying the active site drug resistance mutations (A156V, D168A, D168E, and R155K) and reported that the destruction of the salt bridge between 168 and 155 causes large conformational changes in the binding pocket and A156V mutation changes the binding pocket with an extra bulky methyl group possibly resulting in resistance. [217] While these studies provided a structural mechanism of resistance for R155K, the molecular basis of the distal mutation V36M is unknown. Welsch *et al.*, based on molecular modeling studies, hypothesized that the distal mutations at V36 and T54 result in impaired interaction of the protease residues with telaprevir's cyclopropyl group, however this hypothesis has not been experimentally proven. [218]

In this study, we take a multidisciplinary approach to elucidate the molecular basis of the susceptibility of telaprevir, boceprevir, and danoprevir to R155K/V36M variant of NS3/4A. Through a combination of drug susceptibility assays, crystallography, and molecular dynamics simulations, we reveal possible molecular mechanisms underlying the drug resistance conferred by the distal V36M mutation.

5.3 Results and Discussion

5.3.1 Drug Susceptibility and Enzyme Inhibition Assays

Resistance to inhibitors was tested by measuring the enzymatic inhibition constant, K_i , with both isolated full-length and single domain NS3/4A protease and determining the cellular half-maximal inhibition constant, IC_{50} . Susceptibilities of telaprevir and danoprevir against the R155K variant were recently reported as a fold-change relative to the wild-type. [90] For this study, these two PIs were tested against R155K/V36M variant in addition to boceprevir. Telaprevir loses about 20-fold inhibitory activity against R155K variant of both the full-length and the single domain isolated proteases (Table 5.1). More than 200-fold reduction in K_i was measured for telaprevir against both R155K/V36M constructs. Boceprevir, despite its binding mode being very similar to telaprevir, loses 7- and 31-fold inhibitory activity against the isolated R155K and R155K/V36M protease and the loss is 11- and 29-fold against the full-length NS3/4A. Danoprevir, despite being

much more potent against the wild-type, loses more than 100- and almost 200-fold inhibitory activity against the full-length and single domain proteases. On the other hand, the reason for danoprevir having a lower IC_{50} than K_i is not clear because typically higher drug concentrations are needed to inhibit the protease in a replicon system than an isolated protease. Nevertheless, the general trend of the data from the replicon and enzymatic assays is consistent. Antiviral activity (IC_{50}) loss is more dramatic against the HCV clones carrying R155K/V36M than those carrying R155K, which trends with the enzymatic inhibition assays. Telaprevir and danoprevir appear to be more susceptible to R155K/V36M than boceprevir reflected in the fold-change relative to the respective IC_{50} against the wild-type. In conclusion, the susceptibility of all three inhibitors tested here increased with V36M in the presence of R155K, which is striking considering the distal location of V36M from the binding site.

Table 5.1: 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 IC₅₀ (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)

^aNumbers in paranthesis reflect fold-change relative to wild-type; > indicates IC₅₀ and K_i values higher than the maximum drug concentration tested in the assay.

5.3.2 Crystallographic Binding Mode of Inhibitors

To shed light on the molecular basis of resistance conferred by R155K/V36M, high-resolution crystal structures of the R155K/V36M variant were determined in complex with telaprevir, boceprevir, and danoprevir (Table 5.2). Our group recently published the crystal structures of the telaprevir- and danoprevir-bound R155K variant along with a series of other drug resistance mutants. [90] Bennett *et al.* reported a crystal structure of the wild-type protease in complex with boceprevir, however the resolution was only 2.65 Å with relatively higher temperature factors around the binding site and R_{free} was considerably higher than the R_{factor} (27% versus 19%). [219] Therefore, to determine a higher resolution structure, boceprevir was also crystallized with the wild-type in addition to the R155K and R155K/V36M variants. The structure of the wild-type apo NS3/4A protease domain was also determined. No crystallographic data were collected for the apo R155K and R155K/V36M proteases; but these variants were modeled using the crystal structures of the danoprevir-bound R155K and R155K/V36M variants and the wild-type apo protease as templates by removing danoprevir from the mutant protease complexes and introducing the mutations to the apo structure, respectively.

Table 5.2: X-ray data collection and crystallographic refinement statistics

Drug Protease variant	Telaprevir		Boceprevir		Danoprevir	
	Apo WT	R155K/V36M	WT	R155K R155K/V36M	R155K/V36M	
Resolution (Å)	1.3	1.6	1.7	1.8	1.7	1.1
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
a (Å)	54.8	54.9	55.1	55.4	55.4	55.3
b (Å)	58.6	58.7	58.8	59.0	58.9	58.5
c (Å)	60.9	60.9	60.1	59.8	60.0	59.8
Molecules in AU ^a	1	1	1	1	1	1
R _{merge} (%) ^b	4.8	7.2	4.5	3.2	3.4	5.2
Completeness(%)	93.1	99.8	95.8	90.8	90.8	97.2
I/σ	14.8	20.6	22.8	15.4	11.6	12.6
Measured reflections	190364	209787	96791	48641	41927	334751
Unique reflections	45595	26953	21118	15757	19992	69496
Redundancy	4.2	7.8	4.1	3.1	2.1	4.8
RMSD ^c Bonds (Å)	0.009	0.010	0.012	0.006	0.007	0.009
RMSD ^c Angles (°)	1.294	1.337	1.492	1.117	1.128	1.250
R _{free} (%)	19.1	19.3	19.3	18.8	20.0	16.0
R _{factor} (%)	16.1	17.5	15.6	15.2	16.3	14.1
No of waters	166	156	247	186	212	186

^a AU, asymmetric unit.

^b $R_{merge} = \sum_h \sum_i |I_i - \langle I \rangle| / \sum_h \sum_i I_i$.

^c RMSD, root mean squared deviation.

Overall, the inhibitors superimpose well within the binding site except for danoprevir's flexible isoindoline at P2 moiety (Figure 5.2). The direct contact of this moiety with the site of mutation causes a shift in the binding mode upon R155K and loss of favorable interactions (Figure 5.3). The P2 isoindoline is destabilized with R155K mutation, consistent with the simulation results of Pan *et al.*. This verifies that the extreme destabilization of P2 is not forcefield dependent and is reproducible with alternate forcefields. This high flexibility causes loss of favorable interactions with the catalytic residues D81 and H57 while in the crystal structure they are stable. On the other hand, structural perturbations are observed in key protease residues that are in contact with telaprevir upon R166K/V36M mutation (Figure 5.2). The R155K/V36M protease, when bound to telaprevir, has alternate conformations at the M36 position. Substitution of a valine at a relatively buried position to a larger methionine side chain likely alters the local packing causing the M36 side chain to assume alternate conformations. This alteration must somehow be transmitted to the binding site because the residues D168 and R123, which are a part of the electrostatic network believed to be important for stabilizing the tight binding substrates and inhibitors, also have alternate positions. Having alternate conformations in the crystal structure for two members of this important electrostatic network suggests that the network is even more destabilized with V36M. In addition, K136 both plays an important role in substrate recognition through its favorable electrostatic interactions with the acidic P6 position of the cleavage sites and makes favorable vdW contacts with telaprevir. However, no clear density was observed for the K136 side chain suggesting that

the distal V36M mutation also perturbed the likely stable interactions of K136 in α_1 helix with telaprevir.

Figure 5.2: Binding modes of inhibitors in crystal structures. The protease and inhibitor are represented as surface and sticks, respectively. The side chains of key residues are also shown as sticks: the drug resistance mutation sites R155 and V36 (pink), the catalytic triad (yellow), and key binding site residues (green).

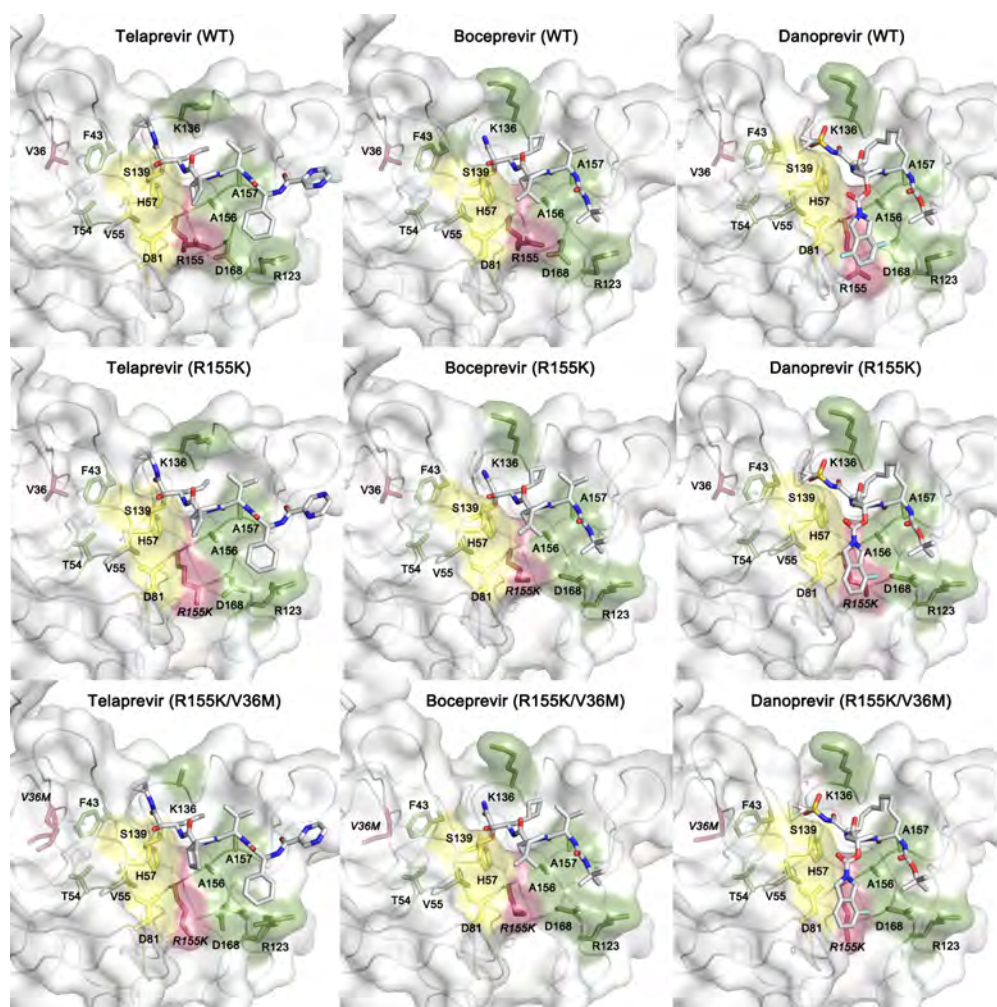
**Figure 5.2**

Figure 5.3: Conformational flexibility of danoprevir's P2 moiety (A) Crystal structures and (B) representative snapshots with equal intervals of 6 ns superimposed on the C2, N2, N5, C11, and C15 atoms of danoprevir.

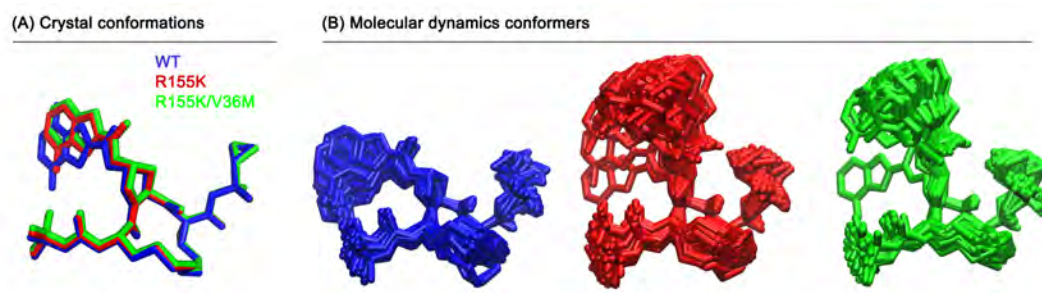


Figure 5.3

5.3.3 Intermolecular Hydrogen Bonds

The hydrogen bonding network between the protease and each inhibitor, observed in the crystal structures, has not been affected by the drug resistance mutations (Figure 5.4). The distance between the hydrogen bond donor and acceptor atoms vary 0.2 Å the most, but staying the same for the most part. However, the relative stability of these potential hydrogen bonds can be evaluated in conformational ensembles from the MD simulations (Table 5.3).

The inhibitors make hydrogen bonds mainly with the backbone hydrogen donor or acceptors in the binding site (Figure 5.4). Therefore, R155K does not necessarily lead to the loss of a hydrogen bond. However, because of the structural rearrangement in the binding site caused by V36M, the likelihood of a hydrogen bond with the backbone oxygen decreases $\approx 20\%$ for telaprevir and $\approx 10\%$ for boceprevir.

The most striking difference between the crystal and MD structures is that both telaprevir and danoprevir loses the hydrogen bond with the catalytic H57, which is observed $\approx 60\%$ of the time between boceprevir and WT and R155K proteases but drops to 34% in R155K/V36M complex. On the other hand, hydrogen bonds with R/K155 and A157 (on E2 β -strand) are preserved in the MD simulations suggesting that these bonds are important enthalpic contributors to the binding affinity.

Figure 5.4: Protease-inhibitor hydrogen bonds. (A) Telaprevir (B) boceprevir, (C) danoprevir. The hydrogen bonds are preserved in complexes of resistant variants, hence are shown only in wild-type structures. See Table 5.3 for the resistant variants.

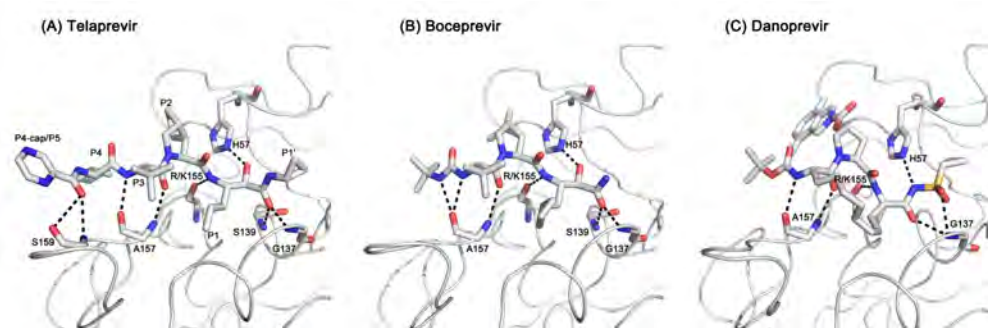


Figure 5.4

Table 5.3: Protease-inhibitor hydrogen bonds

		Telaprevir			Boceprevir			Danoprevir				
		WT	R155K	R155K V36M	WT	R155K	R155K V36M	WT	R155K	R155K V36M		
<i>Crystal structures - Distance between donor and acceptor (Å)</i>												
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	-	-	-	-	-	-	-	-
<i>MD simulations - %Time</i>												
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

5.3.4 Structural Changes in the Protease Backbone

Backbone changes in the crystal structures upon resistance mutations were assessed by computing the differences in the internal C_{α} - C_{α} distances between the resistant and wild-type variants bound to each inhibitor (Figure 5.5). Introducing V36M mutation to the R155K variant alters the backbone structure more than the R155K mutation by itself with respect to the wild-type complexes. From single to double mutant, per-residue average difference in the internal C_{α} - C_{α} distances went up from 0.06 Å to 0.08 Å for telaprevir, 0.07 Å to 0.09 Å for boceprevir, and 0.09 Å to 0.14 Å for danoprevir. Danoprevir, with almost a 300-fold reduction in inhibitory activity with R155K/V36M, alters the backbone of the protein more than the linear telaprevir and boceprevir. Telaprevir is also 220-fold less potent against R155K/V36M, however the structural changes in the backbone is less pronounced compared to danoprevir.

Figure 5.5: Double-difference maps. The differences in internal C_{α} - C_{α} distances were calculated between the R155K-WT, R155K/V36M-WT, and R155K/V36M-R155K structure pairs for telaprevir, boceprevir, and danoprevir complexes. The double-difference matrices are colored for each pair with the average double-difference per residue plotted below. The average double-difference values are also shown on the structure. The core of the protease is relatively rigid (colored blue on structure) while backbone structural changes occur around the binding site and in the loop regions.

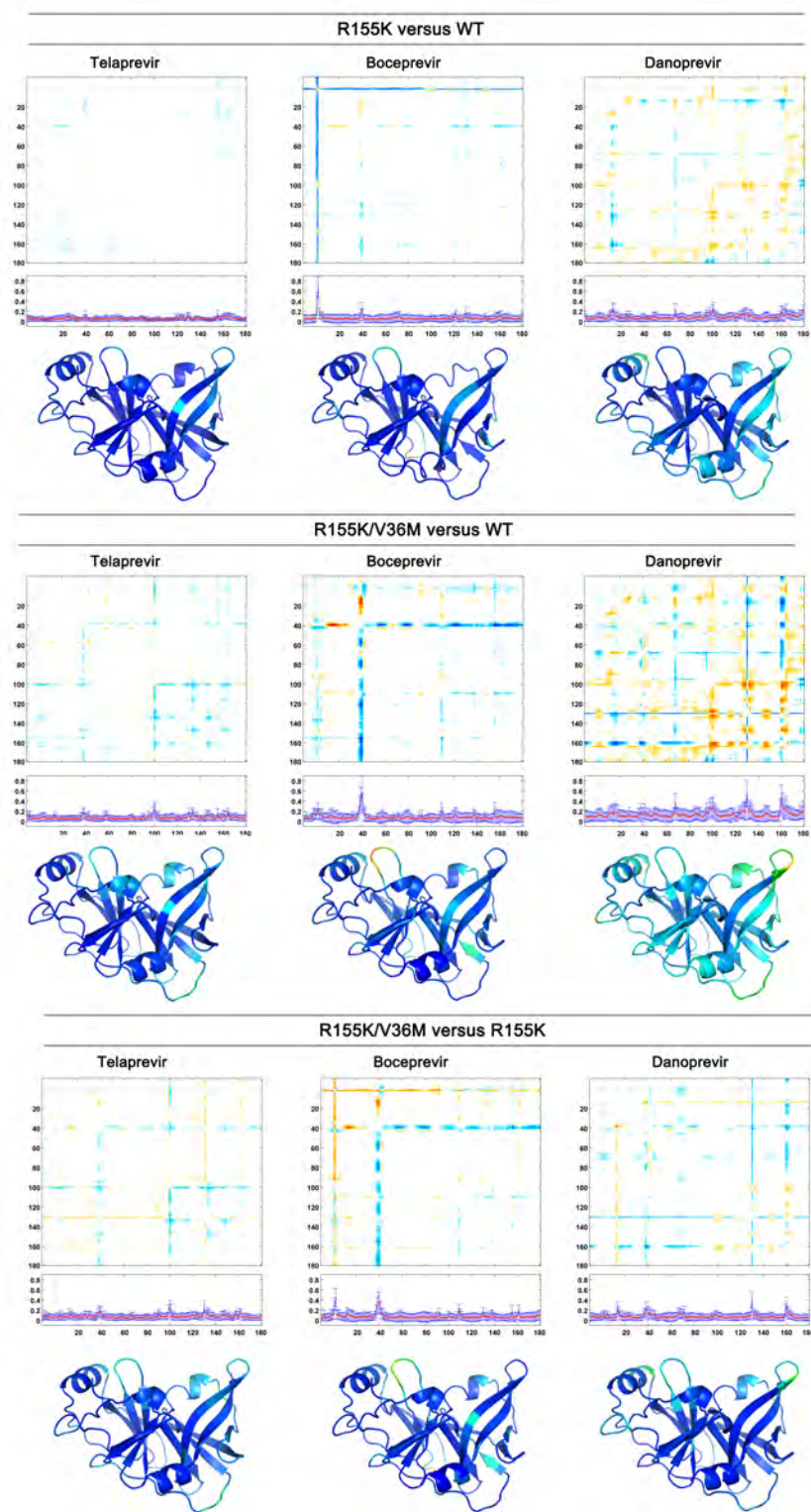


Figure 5.5

Not only the binding site residues were affected by V36M but also changes in distal regions were observed (Figure 5.4). More obvious binding site changes include the loop connecting the β -strands E2-F2 (155 is located on E2), and α_2 helix while subtle changes are observed on α_1 helix that locates the catalytic H57. On the other hand, the distal regions with clear backbone changes consist of the loop connecting A1 (residue 36 is located on A1) and B1 β -strands, N-terminal region of α_0 and C-terminal region of α_3 . Most of these regions are also flexible and exhibit differential dynamics upon inhibitor binding (Figure 5.6). The consistency between the crystallographic changes in the backbone structure and the degree of flexibility in the MD simulations suggests that the regions highlighted in the double-difference maps are not due to crystal contacts but the conformational preferences of these regions indeed vary upon mutations.

Figure 5.6: Changes in the atomic fluctuations in the wild-type NS3/4A.(A) The RMSFs for the protease in the apo form (gray) and in complex with telaprevir (blue), boceprevir (red), and danoprevir (green). (B) The RMSFs for the apo and danoprevir-bound protease are mapped onto the respective crystal structures. (C) The α_2 helix (130-137), which is conformationally stabilized upon inhibitor binding, is shown on the crystal structures. Consistent with the MD simulations, the apo crystal structure has two alternate conformations for Y134, which is located at the α_2 helix.

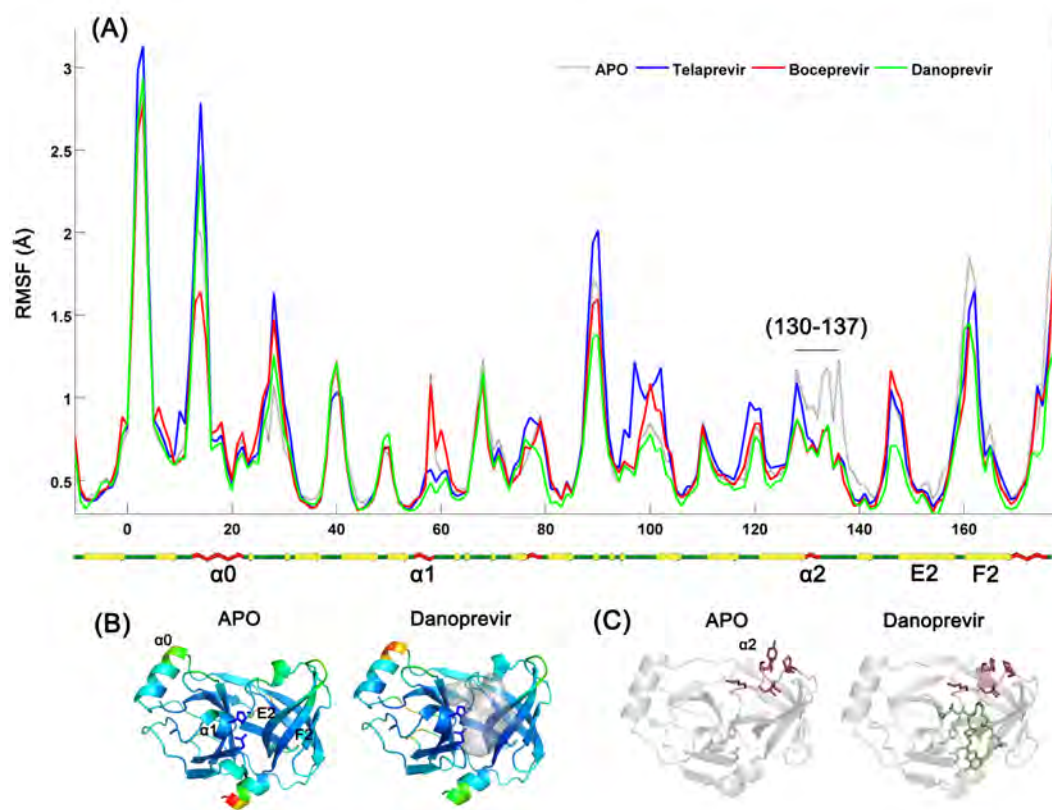


Figure 5.6

Characterizing the differential protein dynamics in the bound and unbound states of a protein can be useful to understand the structural and dynamic mechanisms of drug resistance. [128] For HCV NS3/4A, smaller local fluctuations are observed in inhibitor-bound complexes than the apo protease (Figure 5.6). In the apo protease, mostly the loop regions are undergoing changes. The core of the protein is highly stable, also supported by the stable secondary structures over the trajectories (Supporting Information). An 8 amino acid segment containing the residues 130-137 that are in close proximity to the binding site loses flexibility upon inhibitor binding (Figure 5.6), which is consistent with earlier simulation results reported by Zhu and Briggs. [220] In this segment, I132, L135, K136, and G137 and to an extent S138 make favorable vdW interactions with the inhibitor while the backbone amide of G137 hydrogen bonds to the inhibitors. All three inhibitors stabilize the α_2 helix upon binding.

The minimally fluctuating residues, the most rigid C_α 's, in the bound and unbound structures of the wild-type protease are not majorly affected by the resistance mutations (Figure 5.7). This suggests that the overall structure of the protein is vastly stable, which is also consistent with the minimal changes in the secondary structure throughout the MD simulations (Supporting Figures). Nevertheless, subtle changes upon mutations may be important for the alteration in molecular recognition. The most striking change in atomic fluctuations with V36M mutation is observed at telaprevir's A2, B2, and C2 strands, which become significantly more restrained compared to the wild-type, as evident in the atomic fluctuations (Figure 5.7). When bound to danoprevir, residues 70-80, which form E1 strand, are

relatively rigid in the wild-type protease and become mobile in both R155K and R155K/V36M variants. The other changes in atomic fluctuations are typically at loop regions. These statistically significant changes in atomic fluctuations upon mutations may also be significant in the structural mechanism of drug resistance.

Figure 5.7: Changes in the atomic fluctuations upon resistance mutations. The RMSFs are plotted for WT (blue), R155K (red), and R155K/V36M (green) proteases in the (A) unbound and (B-D) telaprevir-, boceprevir-, and danoprevir-bound states. In telaprevir-bound state, the A2 and B2 strands, significantly rigidified in R155K/V36M relative to WT and R155K, are shown in pink on the structure. The catalytic residues are colored yellow. Telaprevir is shown as sticks.

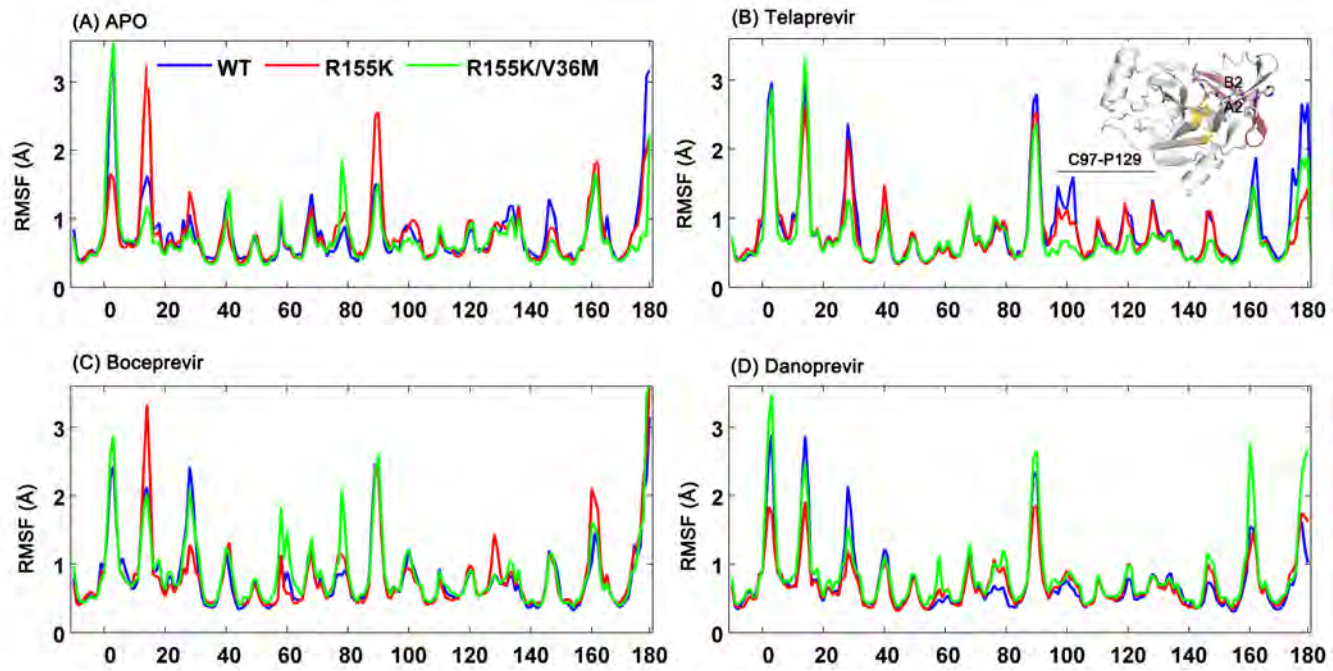


Figure 5.7

5.3.5 Intramolecular Salt Bridges

A key residue determining the electrostatic character of the binding surface is R155, which can make two salt bridges, one with D81 and the other with D168 in all inhibitor-bound structures. In the wild-type protease, R155-D168 salt bridge is more stable than the R155-D81 salt bridge, consistent with the distances observed in the crystal structures (6.4 Å for D81-R155 and 4.8 Å for R155-D168). While danoprevir binding further stabilize R155-D168, the ketoamide inhibitors appear to optimize the electrostatic network on the binding surface such that both these salt bridges exist more than $\approx 80\%$ of the time. However, R155K mutation favors D81-K155 over K155-D168 in telaprevir and danoprevir complexes but not in boceprevir. The higher stability of K155-D168 in boceprevir complex is probably interdependent on the higher stability of the H57-D81 salt bridge, since D81 is no longer available for a salt bridge with K155. This observation combined with the fact that boceprevir is less susceptible to R155K compared to telaprevir and danoprevir suggests that the geometry of the catalytic triad is a key factor stabilizing an inhibitor on the binding surface.

The disruption of the electrostatic network along the binding surface appears to be predominantly orchestrated by R155K mutation in telaprevir and boceprevir complexes. With danoprevir binding, a slight destabilization of D81-K155 is observed in the R155K/V36M protease (from $\approx 84\%$ to $\approx 78\%$).

Another contributor to the electrostatic network along the binding surface is R123, which is relatively solvent-exposed even in the bound state; therefore the

side chain conformation may be sensitive to crystal contacts. To test this possibility, all the potential salt bridges on the protein surface were evaluated (Table 5.4). Neither inhibitor binding nor mutations in the protease changed the salt bridges significantly suggesting that the changes observed on the electrostatic network involving H57-D81-R/K155-D168-R123 is due to the mutations in the protease but not alternate conformations of R123 favored in separate crystals.

Table 5.4: Salt bridges along the binding site of resistant NS3/4A variants

		APO			Telaprevir			Boceprevir			Danoprevir		
		WT	R155K	R155K V36M	WT	R155K	R155K V36M	WT	R155K	R155K V36M	WT	R155K	R155K V36M
<i>Crystal Structures (Distance between the side chain oxygen and nitrogen in Å)</i>													
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
<i>MD Simulations (% of the time the salt bridge criterion was satisfied)</i>													
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	91.2	90.8

^aIn only three structures, unambiguous electron density was observed for R92.

^bAll structures had relatively high B-factors for these residues.

5.3.6 vdW Interactions at the Binding Surface

The drug resistance mutations alter the local packing on the binding surface reflected in the difference in vdW contacts relative to the wild-type complexes (Figure 5.8). The most dramatic loss is at danoprevir's P2 moiety (Figure 5.9) with the residues R155K, A156 and A157 (Figure 5.8). This interaction loss is consistent with enhanced flexibility upon R155K (Figure 5.3). Danoprevir also loses significant interactions at P1' moiety with F43 on the B1 strand, which is located close to V36. The gain of interactions at P1 of danoprevir is not enough to compensate for the lost interactions at P2 and P1'.

Figure 5.8: vdW interactions at the NS3/4A binding surface. (A) Telaprevir, (B) boceprevir, (C) danoprevir. On the left, the vdW energies tracked over MD trajectories were plotted for WT (blue), R155K (green), and R155K/V36M (red) proteases. On the right, the difference of R155K (blue) and R155K/V36M (red) variants from the wild-type is shown.

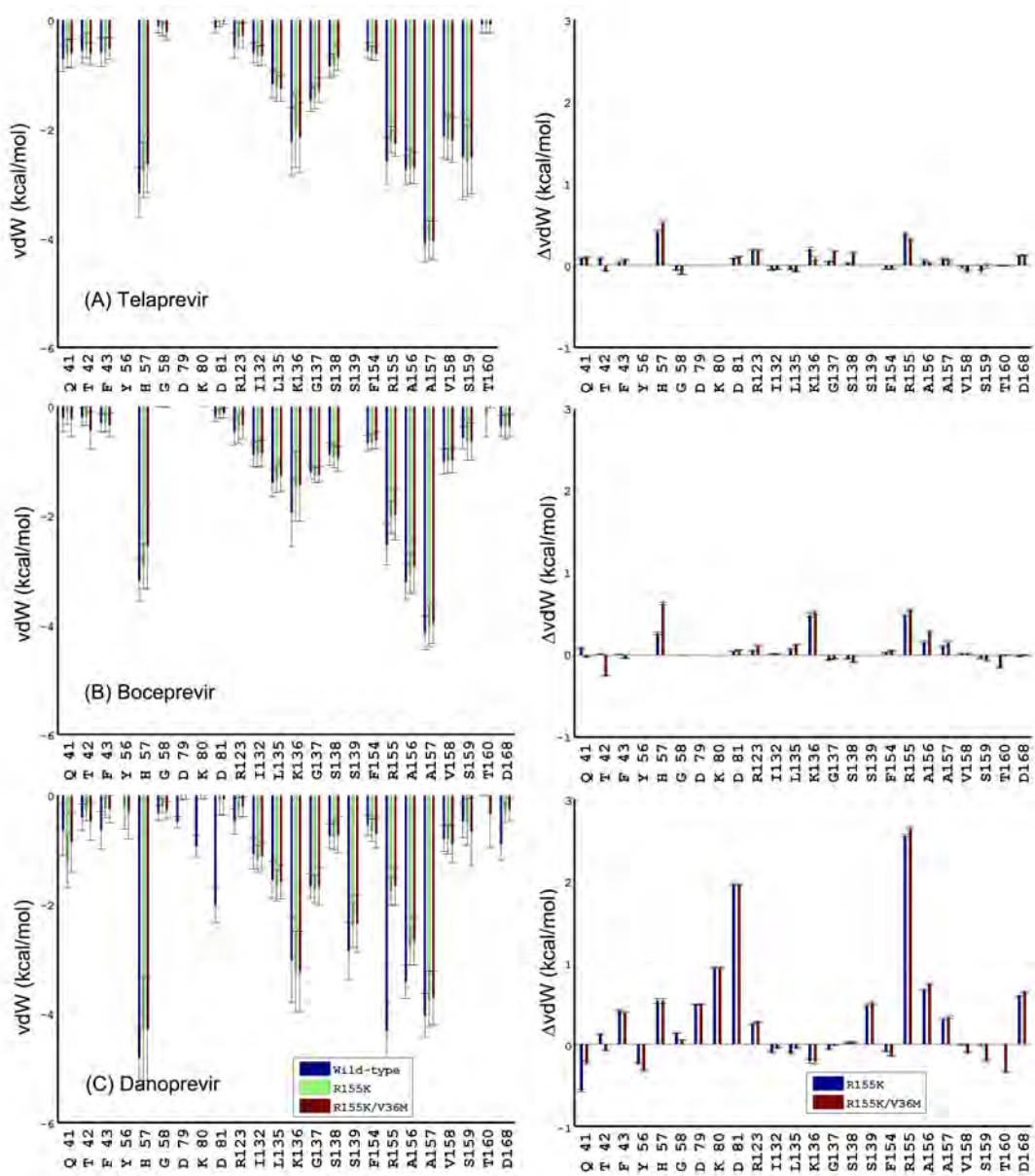
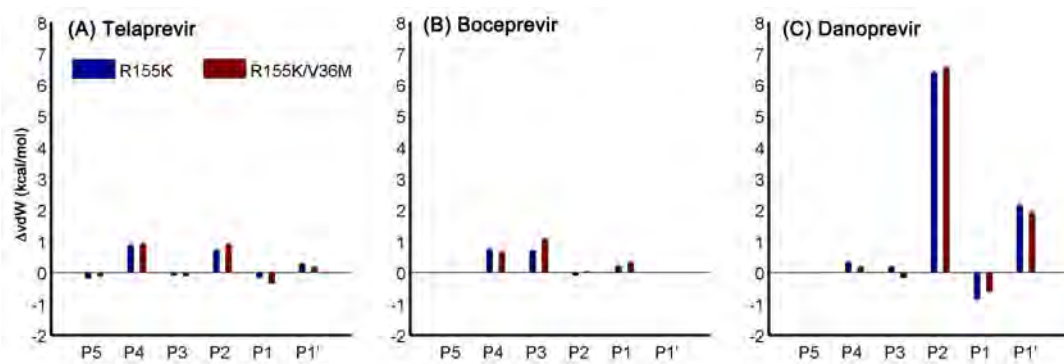


Figure 5.8

Figure 5.9: vdW interactions of the inhibitor moieties. (A) Telaprevir, (B) boceprevir, (C) danoprevir. The differences in inhibitor interactions with the R155K (blue) and R155K/V36M (red) proteases are shown with respect to the wild-type.

**Figure 5.9**

Upon R155K, telaprevir loses favorable vdW interactions mainly with the mutation site K155 and the catalytic H57 at its P2 and P4 positions. On the other hand, boceprevir, in addition to K155 and H57, loses interactions with K136, A156 and A157. A156 and A157 are located on the E2 β -strand, which makes up a significant portion of the binding surface. K136 is located on the α_2 helix. This 3-10 helix, spanning residues 130-137, appears to be important in binding events as it is highly flexible in the apo structure but is restrained significantly upon inhibitor binding (Figure 5.6). Ordering of this segment can only be achieved by tight, favorable interactions with the inhibitor. Therefore, losing these interactions contribute to the loss in the danoprevir affinity upon mutations.

V36M appears to further contribute to the loss of vdW contacts with H57 and K155. The differences in vdW energies between the R155K and R155K/V36M variants are small but significant. Danoprevir picks up new interactions with V158, S159, and T160 of the R155K/V36M protease compared to the WT and R155K variants. These are sites of more favorable interactions with the natural substrates than inhibitors (Figure 4.10). While these interactions are not enough to compensate for the dramatic loss of interactions at D81 and R/K155 for danoprevir binding, they may have a more pronounced affect on substrate recognition, shifting the equilibrium between substrate recognition and inhibitor binding towards substrate recognition. A comparison of the vdW energies in the crystal versus MD structures clearly show that the mutations significantly alter the dynamics of the protease and only the very obvious change in vdW energy of R155K can be observed in the crystal structures (Figure 5.8 and 5.10).

Taken all together, the changes in vdW contacts upon mutations is not limited to the site of mutation but a structural rearrangement is observed on the binding surfaces of the mutant proteases.

Figure 5.10: Crystallographic vdW interactions at the NS3/4A binding surface. (A) Telaprevir, (B) boceprevir, (C) danoprevir. The difference of R155K (dark gray) and R155K/V36M (light gray) variants from the wild-type is shown.

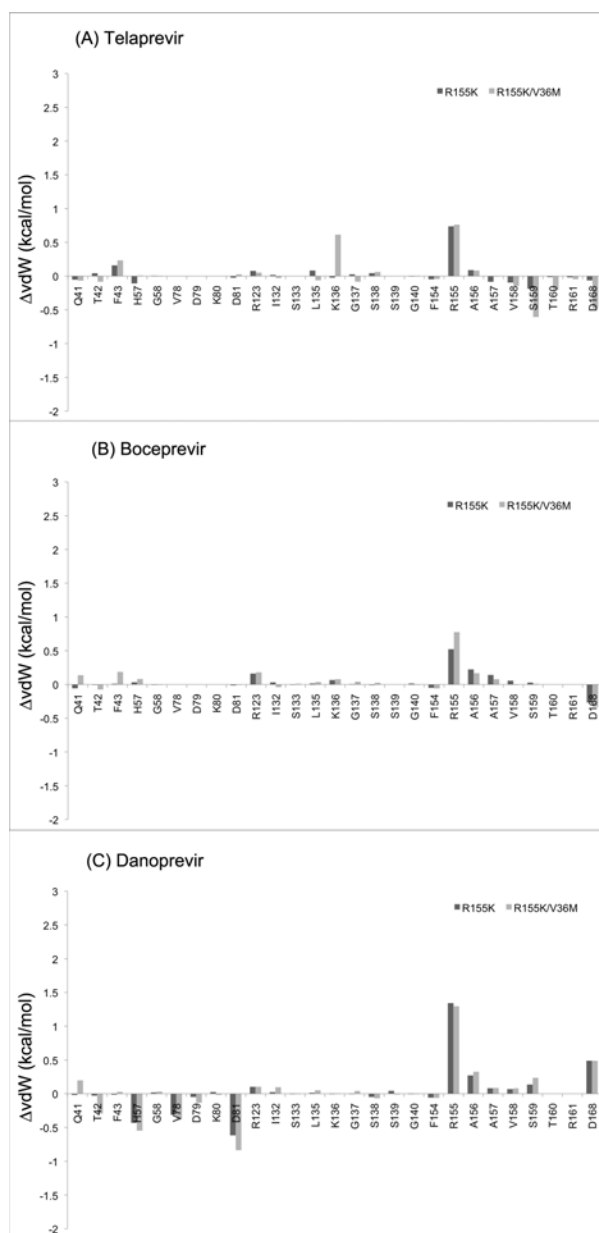


Figure 5.10

5.3.7 Inter-residue Distances in the MD Simulations

To better capture the structural rearrangement in the protease upon resistance mutations, first the inter-residue distances between V/M36 and F43 was measured (Figure 5.11). The A1 strand, locating V36, is bridged to the binding site via B1 strand. F43, which directly interacts with the substrates and inhibitors, is located on B1. In the apo structure, V36M mutation broadens the 36-43 distance distribution while in inhibitor-bound structures the two residues are relatively closer to each other compared to the wild-type and R155K complexes. The residues that are affected by the change in V/M36-F43 distance were determined by tracking their distances from F43 (Figure 5.12). The four residues, K136 in α_2 helix and R/K155, A156 and A157 on E2 β -strand get closer to F43 with V36M mutation. The fact that E2 and B1 β -strands and the α_2 helix are closer to each other suggests a slight shrinking in the binding site. The loops connecting A1-B1 and E2-F2 strands, which can be thought of as pseudo-flaps in analogy to HIV-1 protease, are also closer to each other in the apo and danoprevir-bound proteases while the width of the distribution changes in telaprevir-bound state (Figure 5.13). Because the inhibitors typically protrude beyond the dynamic substrate envelope (Figure 4.4) this shrinking is expected to impair inhibitor binding more than substrate recognition, hence may contribute to drug resistance.

Structural reorganization in the binding site involving the α_2 helix and B1 and E2 strands also impacts the geometry of the catalytic triad (Figure 5.14). In the apo protease, S139 gets pulled away from H57 and D81 with V36M in the pres-

ence of R155K. The distance between S139 and H57/D81 is also significantly larger in boceprevir and danoprevir complexes. In telaprevir and danoprevir complexes, R155K mutation causes H57 and D81 C_α atoms to get closer to each other while the side chains are not oriented properly for strong salt bridging (Table 5.4). On the other hand, the robust H57-D81 salt bridge in boceprevir complexes keep the residue backbone in place resulting in reasonably overlapping distance distributions for the three mutants.

Finally, the H57-K136 distance decreases in the unbound variants of R155K and R155K/V36M, with a bimodal distribution in R155K/V36M (Figure 5.15). Since K136 makes key interactions with the natural substrates, one would expect that a change in the relative orientation of K136 with respect to the catalytic H57 could alter the balance in the molecular recognition events in favor of substrate recognition versus inhibitor binding.

The V36M mutation may also interfere with the binding of the cofactor NS4A, which aids in the proper folding of NS3. Since the A1 strand, at which V36 is located, is in direct contact with the NS4A cofactor in structure (Figure 5.1), a mutation in A1 may as well affect the interactions with NS4A. In fact, while subtle shifts are observed in the inhibitor complexes, the distance between the residues V/M36 in A1 and R28 in NS4A are significantly larger in the apo protease (Figure 5.16). This shift supports the possibility of an interference of a methionine side chain in A1 with cofactor binding.

Figure 5.11: C_α - C_α distance between V/M36 and F43. The distance distribution is plotted for the WT (blue), R155K (red), and R155K/V36M (green) proteases in unbound and bound states. The residues with underlined labels, I25 and R28, are in NS4A. The NS3 residues V36 and Q34 are located in A1 β -strand while E13, one of the most flexible residues with high RMSF values, is located on α_0 helix.

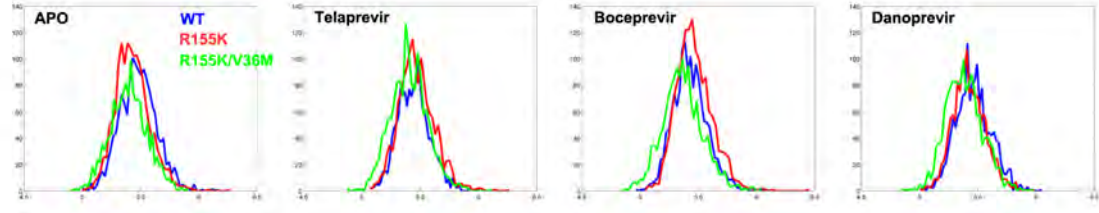


Figure 5.11

Figure 5.12: C_{α} - C_{α} distance between F43 and K136, R/K155, A156, A157.

The distance distribution is plotted for the WT (blue), R155K (red), and R155K/V36M (green) proteases in unbound and bound states.

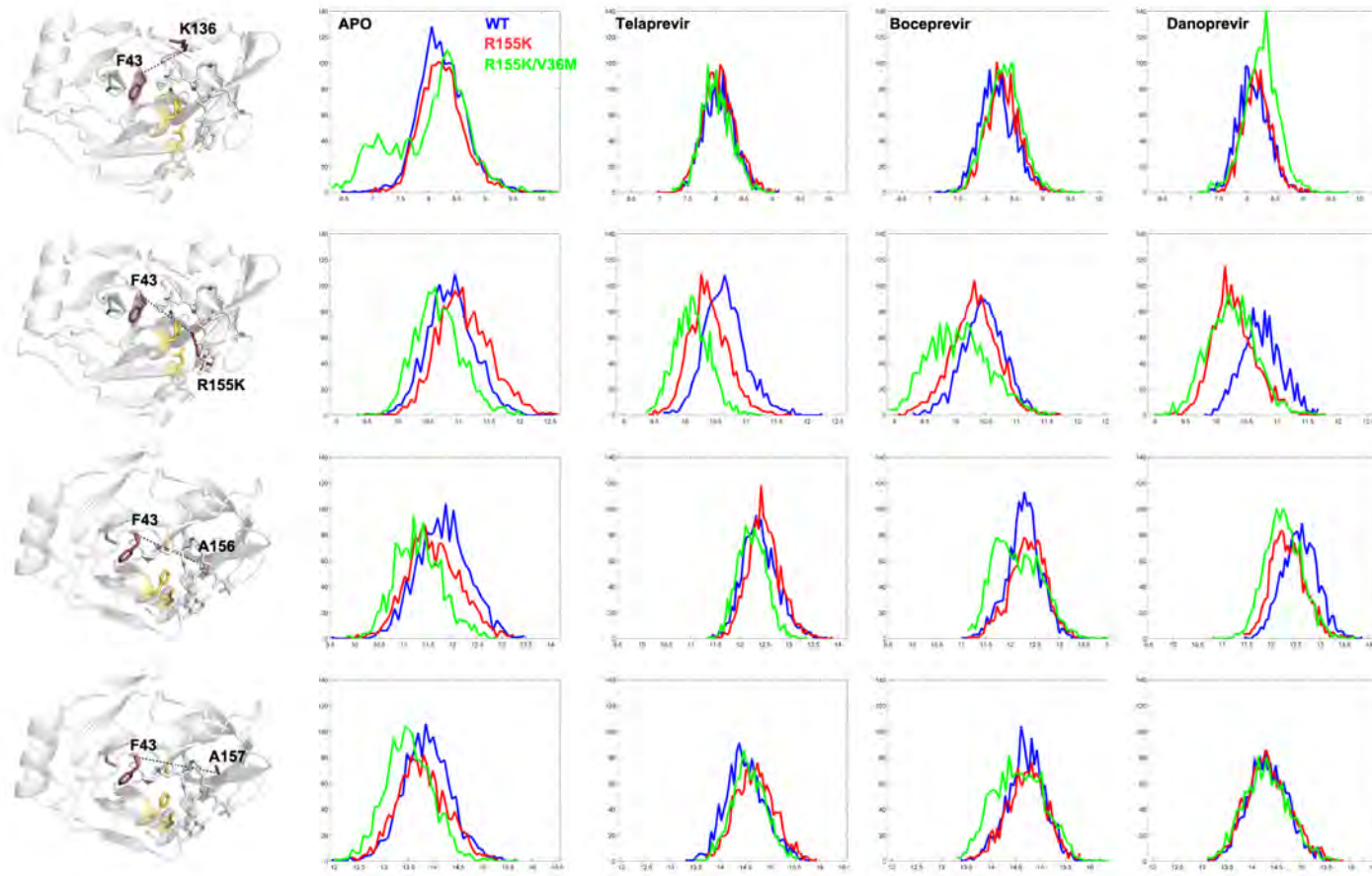


Figure 5.12

Figure 5.13: C_{α} - C_{α} distance between A39 and R161. The distance distribution is plotted for the WT (blue), R155K (red), and R155K/V36M (green) proteases in unbound and bound states. Residues A39 and R161 are located on the loops (“*pseudo-flaps*” analogous to HIV-1 protease) connecting A1-B1 and E2-F2, respectively.

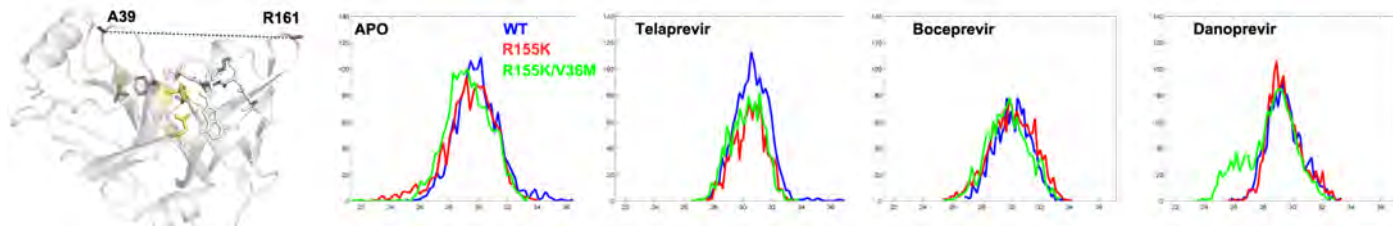


Figure 5.13

Figure 5.14: C_α - C_α distance within the catalytic triad. The distance distribution is plotted for the WT (blue), R155K (red), and R155K/V36M (green) proteases in unbound and bound states.

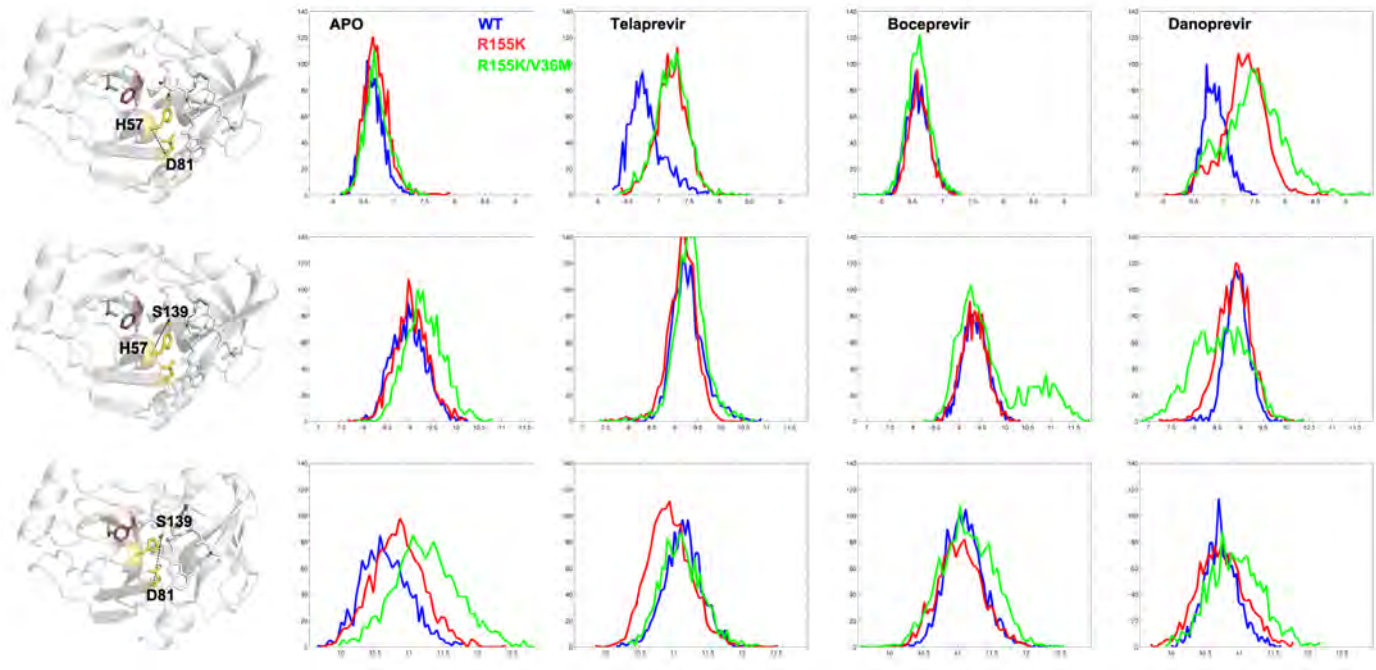


Figure 5.14

Figure 5.15: C_{α} - C_{α} distance between H57 and K136, A156, and A157. The distance distribution is plotted for the WT (blue), R155K (red), and R155K/V36M (green) proteases in unbound and bound states.

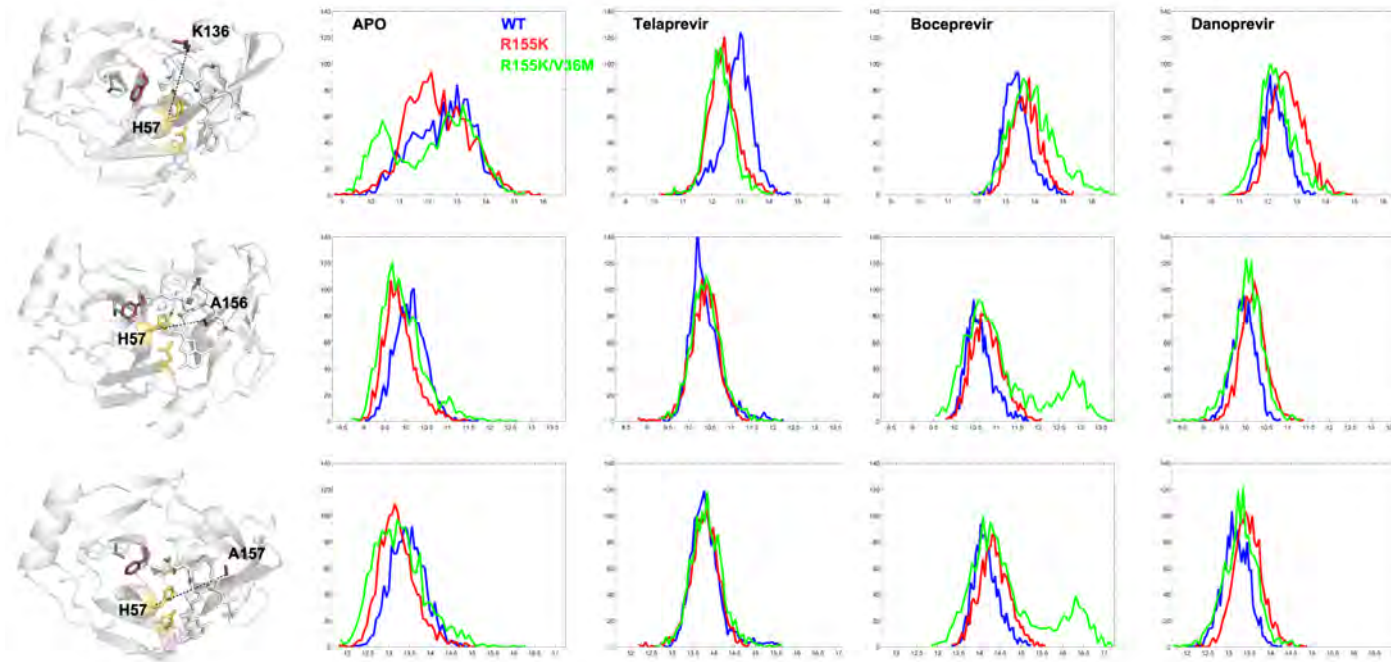


Figure 5.15

Figure 5.16: C_{α} - C_{α} distance between NS3 and the cofactor NS4A. The distance distribution is plotted for the WT (blue), R155K (red), and R155K/V36M (green) proteases in unbound and bound states.

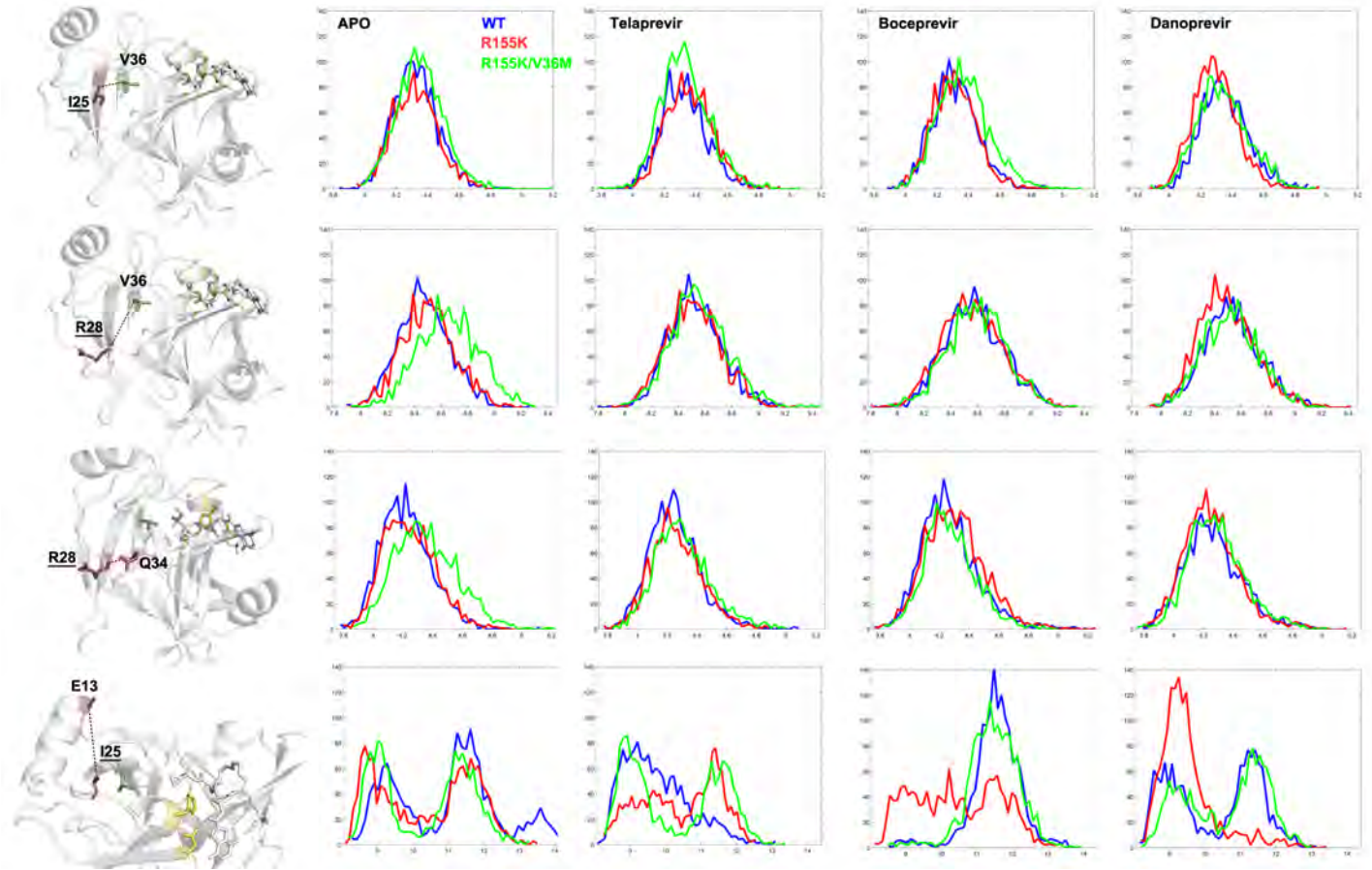


Figure 5.16

5.3.8 Correlation of Equilibrium Fluctuations from Gaussian

Network Model

Equilibrium fluctuations of NS3/4A were evaluated with the Gaussian network model (GNM) to decompose the overall dynamics of the protein into the more significant modes of motion. Overall conformational dynamics of a protein is mainly determined by the low frequency slow modes of motions. The cross-correlations of these motions in the first two slowest modes show that the domain motions in HIV-1 protease and HCV NS3/4A protease are strikingly similar despite the divergence in sequence and structure (Figure 5.17). The commonality of the domain motions suggests that the two proteases make use of very similar mechanistic motions to bind, process, and release their respective substrates independent of the substrate or protease sequence, the identity of the catalytic residues and the chemical mechanism of the actual hydrolysis reaction. In HIV-1 protease, the differential flap motions and their effect on the inter-residue distances between monomers has been reported for a highly resistant HIV-1 protease in comparison to the wild-type. [128] Similarly, in NS3/4A protease, while the active site mutation R155K directly alters the interactions at the binding surface, the distal V36M mutation may alter the orientation of the two β -barrel domains with respect to each other as reflected in the internal distances.

Figure 5.17: The correlation of equilibrium fluctuations in the slowest two modes from GNM. (A) HCV NS3/4A protease, (B) HIV-1 protease. In the slowest mode, the two monomers of HIV-1 protease fluctuate in a negatively correlated manner where in the second mode, the dimerization interface and the flap fluctuations are negatively correlated. The slow mode motions of NS3/4A protease is strikingly similar to HIV-1 protease.

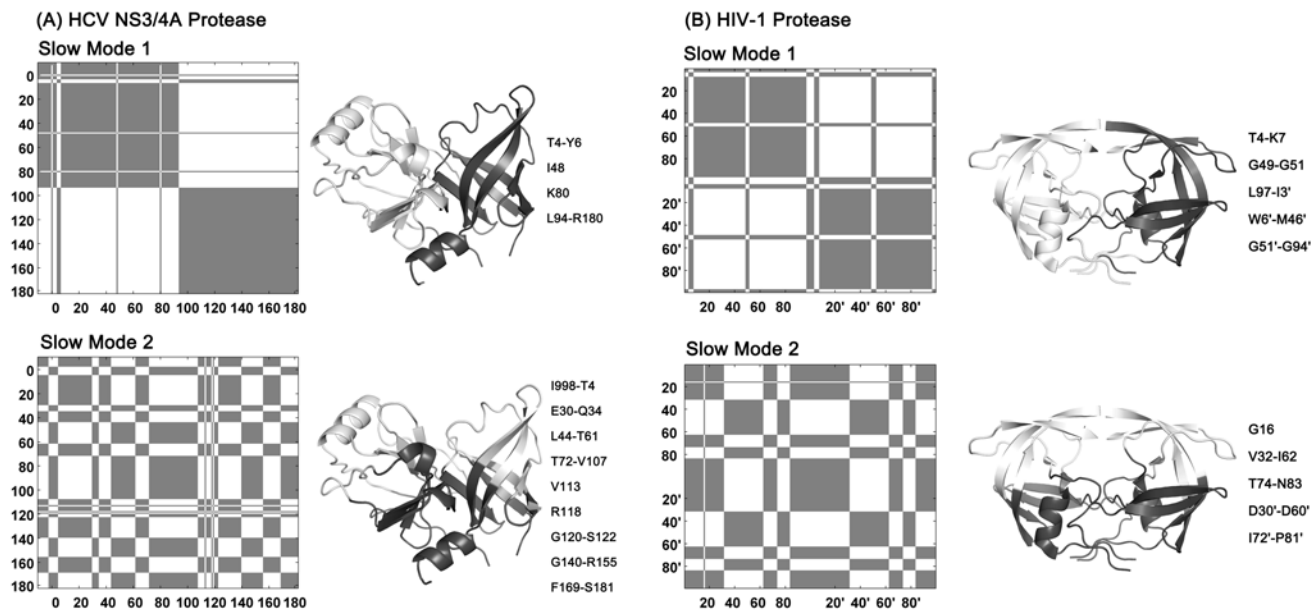


Figure 5.17

5.4 Conclusion

Molecular basis of drug resistance conferred by V36M in the R55K background has been investigated using a combination of drug susceptibility assays, x-ray crystallography, and molecular dynamics simulations. The replicon studies and binding assays show that the R155K/V36M is more resistant to telaprevir, boceprevir, and danoprevir than R155K. The impact of the distal V36M mutation on the crystal structure is subtle but MD simulations suggest that the conformational dynamics of the protease is altered via two separate mechanisms.

One possible mechanism is that the impact of V36M mutation may be propagated to the binding site through B1 β -strand. V36 is located on the A1 β -strand, right next to B1. β -strands are stable secondary structural elements that are able to pass a perturbation along to their spatial neighbors in a protein. The fact that F43, on B1 β -strand, is also a resistance mutation site supports the plausibility of this hypothesis. Studying the structural and dynamic effects of mutations at F43 in comparison to those of V36M variants would result in interesting and potentially useful insights into the structural adaptability of the protease and the mechanism of resistance.

The V36M mutation may also interfere with the binding of the cofactor NS4A, which aids in the proper folding of NS3. Substitution of a valine with a larger side chain, methionine, alters the local packing. Since the A1 strand containing V36 is in direct contact with NS4A in structure, a mutation in A1 may as well affect the interactions with NS4A. A hindrance in cofactor binding would also interfere

with substrate processing efficiency, which may be the reason for the reduced base-level replicative capacity of V36M and R155K/V36M variants of genotype 1b viruses in the absence of drugs (44% and 76% relative to wild-type). [84]

Because implications for both mechanisms are found in the presented data, the two mechanisms are probably not mutually exclusive, which makes it an even more interesting problem. The conventional understanding of allostery is a change of shape at one site of a protein surface due to the effects of ligand binding to another. However, mutations in regions distant from the binding site can have an impact on the ligand affinity. This impact clearly shows that allostery does not have to be brought about by ligand binding but proteins even in the unbound state can be thought of as a network of residues communicating with each other. The interdependence of the spatial neighbors in a protein can be characterized relatively easily because of the direct interactions among them; hence rationalizing the co-occurring mutations in these spatial neighbors is relatively less challenging. In contrast, elucidating the interdependence of residues that are located distal to each other is not straightforward. The results of this study show that V36M, in the context of R155K, significantly decrease the affinity to all three inhibitors. Nevertheless, this mutation does not drastically alter the structure trapped in a crystal but the conformational dynamics of the protein in solution is affected. In addition, investigating the conformational ensembles revealed considerable changes in the internal distances within the unbound bound state of NS3/4A variants. Because binding is an equilibrium event between the bound and unbound states, further characterization of the solution dynamics of both states by introducing mutations

in β -strands A1-B1 and NS4A will aid in elucidation of the complex mechanisms of allosteric communication in NS3/4A.

5.5 Materials and Methods

5.5.1 Mutagenesis and Gene Information

The HCV genotype 1a NS3/4A protease gene described in a Bristol-Meyers Squibb patent [221](1) was synthesized by Gen-Script and cloned into the pET28a expression vector (Novagen). Highly soluble NS3/4A protease domain contains 11 core amino acids of NS4A covalently linked at the N terminus. A similar protease construct exhibited catalytic activity comparable to that of the authentic full-length protein. [222] R155K and R155K/V36M protease variants were generated using the QuikChange Site-Directed Mutagenesis Kit from Stratagene and sequenced by Genewiz.

5.5.2 Expression and Purification

NS3/4A protease expression and purification were carried out as described previously. [64, 221] 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.

5.5.3 Crystallization

Danoprevir was prepared in-house using our convergent reaction sequence as described previously [9]; 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, 500 mM NaCl, 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.

5.5.4 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. 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. The diffraction intensities were indexed, integrated, and scaled using the program HKL2000. [223] All structure solutions were generated using simple isomorphous molecular replacement with PHASER. [224] The model of viral substrate N-terminal product 5A-5B (PDB ID: 3M5O) [9] 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 [225] were applied to improve the phases by building solvent molecules. [226] Crystallographic refinement was carried out within the CCP4 program suite or PHENIX with iterative rounds of TLS and restrained refinement until convergence was achieved. [227] The final structures were evaluated with MolProbity [228]. Five percent of the data was reserved for the R_{free} calculation to prevent model bias throughout the refinement process. [229] Manual model building and electron density viewing were carried out using the program COOT. [230]

5.5.5 Drug Susceptibility and Enzyme Inhibition

For enzyme inhibition experiments, 5 nM of the genotype 1a HCV NS3/4A protease domain was incubated with increasing boceprevir concentrations for 90 min in 50 mM Tris assay buffer (5% glycerol, 5 mM TCEP, 6 mM LDAO and 4% DMSO, pH 7.5). Proteolysis reactions were initiated by adding 100 nM HCV NS3/4A substrate [Ac-DE-Dap(QXL520)-EE-Abu-y-[COO]AS-C(5-FAMsp)-NH₂] (AnaSpec) and monitored using the EnVision plate reader (Perkin Elmer) at excitation and emission wavelengths of 485 nm and 530 nm, respectively. The initial cleavage velocities were determined from sections of the progress curves corresponding to less than 15% substrate cleavage. Apparent inhibition constants (K_i) were obtained by nonlinear regression fitting to the Morrison equation of initial velocity versus inhibitor concentration using Prism 5 (GraphPad Software). Data were generated in triplicate and processed independently to calculate the average inhibition constant and standard deviation.

5.5.6 Double-Difference Maps

The pair-wise atomic distances between each C_α of a given protease molecule and every other C_α in the same molecule were calculated. The differences of these C_α - C_α distances between each pair of protease molecules were then calculated and contoured as a map for visualization. These maps allow for effective structural comparisons without the biases associated with superimpositions. Double-differences in internal distances were then averaged for each protease residue and

plotted with the standard deviation and shown together with the corresponding double-difference map. The plots of per-residue averages of double-differences highlight the most variant regions between the two proteases that are being compared.

5.5.7 Hydrogen Bonds

A hydrogen bond was defined based on the criteria of donor-acceptor distance of less than 3.5 Å and hydrogen-donor-acceptor angle less than 30° and calculated with VMD for both molecular dynamics trajectories and crystal structures. [156] However, the exact coordinates for hydrogen atoms, even in structures of resolution as high as 1.5 Å, cannot be determined with confidence based on the electron density maps. Therefore, for proper evaluation of potential hydrogen bonds in a crystal structure, hydrogen atoms were added to the crystal structures such that the hydrogen bonding network was optimized using Maestro. [231] An energy minimization was also performed on the crystal structures by completely constraining the non-hydrogen atoms. These structures were then evaluated for the hydrogen bonds formed between the protease and the inhibitors.

5.5.8 Salt Bridges

Salt bridges were defined as an interaction between a side-chain oxygen atom of Asp or Glu within 4.0 Å of a nitrogen atom of Arg or Lys. The salt bridges in the molecular dynamics trajectories were tracked using VMD. [156]

5.5.9 van der Waals Interactions

The protease-inhibitor vdW contact energy was estimated by the simplified Lennard-Jones potential, the details of which was described in subsection 4.5.6. For consistency, the vdW parameters, ϵ and σ , were taken from the OPLS2005 forcefield as the OPLS2005 forcefield was also used in the MD simulations. For the pairs involving two separate atom types, the vdW parameters were geometrically averaged.

5.5.10 MD simulations

Structure preparation and the molecular dynamics simulations were performed following the energy minimization and molecular dynamics protocol described in subsection 4.5.3.

5.5.11 Molecular Modeling

Crystal structures for apo R155K and R155K/V36M variants were not available. These structures were modeled using two separate templates. First, the inhibitor coordinates were removed from the crystal structures of the danoprevir-bound R155K (PDB ID: 3SU0) and R155K/V36M variants and the resulting structure were relaxed using the energy minimization protocol described in subsection 4.5.3. Second, the mutations were introduced to the crystal structure of the wild-type apo protease and the resulting structures were again minimized with the same protocol. To ensure the equivalence of the independently obtained struc-

tural models, the energy-minimized structures were then compared with respect to backbone structural changes, side chain orientations of the mutated amino acids and the residues that are in direct contact with the sites of mutation.

5.5.12 Gaussian Network Model

Elastic network models, utilizing normal mode analysis, are physically plausible and mathematically tractable tools for characterizing molecular fluctuations near a given equilibrium state, hence for exploring protein dynamics in the native state. [232, 233] Among these, Gaussian network model (GNM) is the simplest to decompose a protein into structural domains, defined as portions of the protein that move independently. The accuracy of the method was verified for a large set of proteins and connections with graph theory were indicated. [234] HingeProt is a web implementation of GNM and its extension ANM (Anisotropic Network Model). [235] The wild-type apo x-ray structure of HCV NS3/4A protease was uploaded to HingeProt to calculate the cross-correlations of the low frequency motions based on GNM with a connectivity cutoff of 10 Å. For HIV-1 protease, the substrate CA-p2 was removed from the complex structure (PDB ID: 1F7A). Because the slowest modes contribute the most to the overall collective motions of proteins, the cross-correlations of residues in the first two slow modes were calculated to get an insight on the main domain motions of the enzymes.

Chapter VI

Discussion

6.1 Robust Drug Design: Hitting Multiple Targets at a Time

Pharmaceutical industry spends \$50 billion to produce only 20 new drugs annually. [236] Only $\approx 3\%$ of the drug discovery projects result in a marketed drug. [144] The reasons for drug failures are diverse and include unclear target biology, poor potency or selectivity, lack of efficacy, and drug resistance.

Drug resistance negatively impacts millions of patients and costs billions of dollars. As a result of resistance, new drugs quickly become less effective. [237] Arguably, the best way to address this problem of productivity in drug development is to focus on preclinical stages, which has a 35% success rate in delivering experimental drugs ready for clinical testing. [144] In the preclinical development, the goal of disrupting drug target's activity is necessary, but not sufficient for developing a robust drug that can avoid resistance.

To discover robust drugs, first a clear picture of drug target's native function should be obtained. Protein structure and dynamics are critical for function; therefore initial studies should be focused on these. This seemingly extra investment will eventually save time, money and lives by increasing the robustness of drugs.

Combating viral diseases, the target is not "a" virus but a viral population including many variants with various levels of fitness under different conditions. Thus, integrating evolution into drug discovery is required to come up with long-lived drugs.

In addition, at the molecular level, proteins exist as conformational ensembles but also receptor-drug complexes can be very dynamic. A mutation in the target enzyme can modulate inhibitor binding by disrupting direct interactions with the ligand but can also indirectly alter protein dynamics. Therefore, in quickly evolving diseases, the target is an ensemble in mutational and conformational space.

Structure-based drug design relies on static protein structures despite significant evidence for the need to include protein dynamics as a serious consideration. The traditional approach is to make the compounds bind to a static image of the wild-type enzyme as tight as possible and hope for the best for the potential mutants. However, mutational and conformational plasticity of the viral enzymes sets the goal to hitting multiple targets at a time for effective therapies. Simultaneous integration of evolution and conformational flexibility into drug discovery becomes feasible with ever-growing technological innovations in high-throughput assays, deep sequencing, and computational power.

Designing a robust drug is to hit multiple targets at a time, which is not an easy goal but certainly is worth the effort.

6.2 Drug Target's Function and Protein Dynamics

The HIV-1 and HCV NS3/4A proteases both recognize a diverse set of sequences as substrates. The recognition motif is not a consensus sequence but a consensus volume shared by the substrates. Deviations in the binding modes of existing inhibitors from the natural substrates create opportunities for drug re-

sistance mutations. According to the substrate envelope hypothesis, the inhibitors should be designed to mimic the consensus shape of natural substrates to avoid resistance, because a mutation that could potentially reduce inhibitor affinity would also interfere with substrate recognition, hence that mutant virus would not be populated under the selective pressure of drug therapy (Figure 6.1A).

Figure 6.1: Guiding drug design using the dynamic substrate envelope. (A)

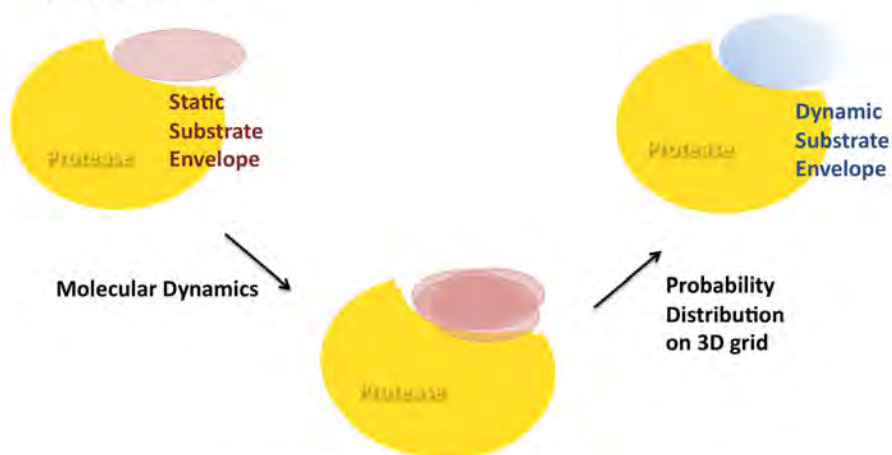
The most severe resistance mutations (red) occur at sites contacted by competitive inhibitors (orange) outside the substrate envelope (blue). (B) *Dynamic substrate envelope* can be defined as a probability distribution of the consensus substrate shape within the binding site by combining molecular dynamics simulations and three-dimensional grid based volume calculations. (C) Dynamic substrate envelope can be integrated into structure-based design of robust drugs by systematically optimizing two metrics: (1) V_{out} : The probabilistic volume of an inhibitor falling outside the dynamic substrate envelope, (2) $V_{remaining}$: The portion of the dynamic substrate envelope that is not fully occupied and, therefore can be better-utilized by an inhibitor.

(A) The substrate envelope hypothesis



(B) Integrating protein dynamics into the substrate envelope

Crystal structures



(C) Guiding drug design to avoid drug resistance

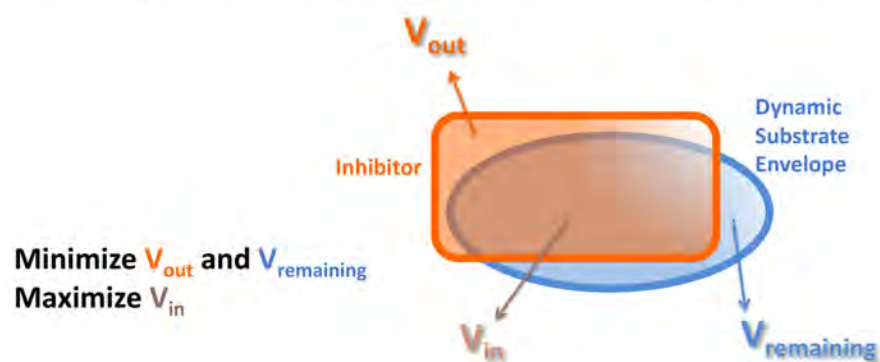


Figure 6.1

In my thesis, I reformulated the substrate envelope hypothesis by incorporating the conformational dynamics of protease-substrate complexes and defined the dynamic substrate envelope for HIV-1 and HCV NS3/4A proteases (Figure 6.1B).

HIV-1 protease has a mainly hydrophobic binding pocket covered by the flaps while HCV NS3/4A has a very shallow binding surface. As a result of this fundamental difference, the dynamic substrate envelope for HIV-1 protease reproduced the essentials of the static substrate envelope based on the crystal structures, whereas I observed significant deviations in the dynamic substrate envelope of HCV NS3/4A from its static substrate envelope. HIV-1 protease is a very flexible enzyme in the unbound state with flaps constantly sampling alternative conformations. Nevertheless, in the bound state with the flaps closed over the binding pocket, the HIV-1 substrates have less degrees of freedom compared to the highly exposed substrates of HCV NS3/4A on the flat binding surface. As a result, NS3/4A substrates are freer to move, while HIV-1 substrates are relatively restrained.

The development and clinical introduction of HIV-1 protease inhibitors is regarded as a prime example of structure-based rational drug design. [238] The conformational restriction on the bound ligands of HIV-1 protease turned these structure-based drug design efforts into a glorious success story, even though protein dynamics had not been explicitly considered in the design of these inhibitors. A good quality structure of a high affinity inhibitor turns out to be a good template for optimizing HIV-1 protease inhibitors for better pharmacological properties and flatter resistance profiles, since the active site is captured in its optimum

geometry for tight binding and does not undergo extreme conformational changes in solution. On the other hand, the dynamic substrate envelope of HCV NS3/4A protease has significant deviations from the static substrate envelope, mainly in P5 and P6 positions. In MD simulations, these two residues are very flexible; hence their consensus volume is relatively small as opposed to their ordered structure in crystals.

The local packing, quantified as van der Waals interaction energies, are extremely important for substrate recognition by the hydrophobic binding pocket of HIV-1 protease. NS3/4A compensates the lack of a well-defined binding pocket by stronger and longer ranged electrostatic interactions with P6 acid of substrates. This electrostatic interaction was identified in crystal structures between P6 and K165 only, however in the MD simulations, I observed a highly dynamic interaction between acidic P6 and a series of basic residues on the binding surface, R119, R123, R161, and K165. Substituting the P6 residue in substrates to a non-acidic side chain completely disrupts substrate binding. Similarly, the activity of the enzyme should drop drastically with the substitution of K165 to a non-basic residue if P6-K165 is the only electrostatic component of substrate recognition. However, based on my MD simulation data, the K165 mutants would have residual activity due to the presence of R119, R123, and R161 on the binding surface. The activity assays coupled with MD simulations, if performed on NS3/4A variants with single, double or triple mutations of these four basic residues, will provide insights to the importance of the electrostatic interactions in substrate recognition in solution.

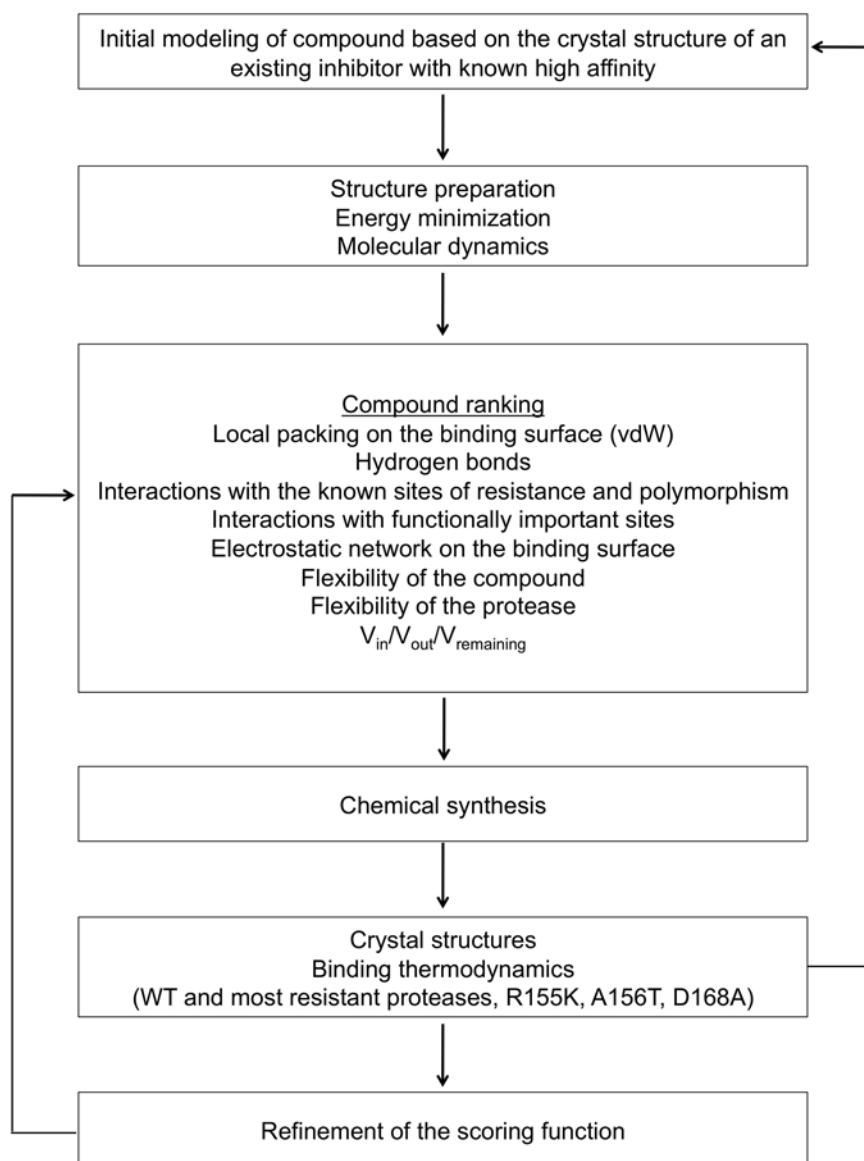
6.3 Dynamic Substrate Envelope: Integrating Dynamics into Drug Discovery while Avoiding Resistance

In this thesis, I introduce the dynamic substrate envelope as a significant improvement to the substrate envelope hypothesis to guide the design of robust NS3/4A inhibitors by systematically assessing the susceptibility to existing and potential resistance mutations.

The dynamic substrate envelope defines consensus mode of substrate binding more realistically as a probability distribution. This energetically smoother function can be added to the existing scoring functions to rank compounds based on their mimicry of the natural substrate binding (Figure 6.1C). In this process, new inhibitors should be designed to have an ether-linked quinoxaline or similar functional groups at P2 moiety that favor tight interactions with the catalytic triad rather than the resistance mutation sites, R155 and D168. Also compounds without the P2-P4 macrocycle should be designed to avoid D168A mutation. Structural models of the bound inhibitors can be generated using the crystal structure of a high-affinity inhibitor as the template. To achieve a fair ranking process, the conformational plasticity of the inhibitors should also be taken into account by simulating the bound-state dynamics of the inhibitors. The conformational ensemble of each inhibitor during the MD simulation can then be compared to the dynamic substrate envelope. The compounds with a smaller probable volume falling outside the dynamic substrate envelope (V_{out}) are ranked as less sus-

ceptible to resistance. The top-ranked inhibitors can be prioritized in the actual chemical synthesis. The top-ranked compounds may need further modification to optimize properties other than affinity and selectivity (i.e. solubility, metabolism, permeability, or efflux). In that case, $V_{remaining}$, the portion of the dynamic substrate envelope that is not fully used by that inhibitor, can guide these modifications to avoid possible contacts with resistance mutation sites. Other metrics to avoid resistance can also be easily computed from the MD trajectories such as the molecular interactions with the catalytic triad and functionally important residues versus sites of resistance mutations and polymorphisms. Compounds with more favorable interactions with the evolutionarily conserved residues than the resistance mutation sites should be ranked higher. In addition, the flexibility of the compounds can be evaluated from the trajectories since flexible P2 moieties are expected to adapt well to R155K mutation. As an indicator of tight binding, the stability of the electrostatic network involving the residues H51-D81-R155-D168-R123 can be assessed. The inhibitors that keep this network intact are expected to be tighter binders. The design scheme I describe (Figure 6.2) here is completely feasible as we are currently able to collect about 10 ns/day of simulated molecular dynamics data on explicitly solvated systems containing 25,000 atoms on average. In addition, the scheme I propose is an iterative process where the synthesized compounds are crystallized and tested against the wild-type and resistant variants. As more structural and binding data are available, then an ultimate scoring function may be formulated with tuning the weights assigned to each metric.

Figure 6.2: A structure-based drug design scheme that incorporates protein dynamics and potential resistance

**Figure 6.2**

The dynamic substrate envelope-based drug design process can potentially benefit from further improvements. Currently, no co-crystal structure is available for the full-length substrates of NS3/4A. Residual catalytic activity of inactive serine protease, with single, double or triple alanine mutations at the catalytic triad, still has residual catalytic activity. [239]The N-terminal cleavage products bind to the protease with higher affinities than C-terminal cleavage products. As a result, the crystal structures contain electron density only for N-terminal. To guide drug design, the dynamic substrate envelope defined based on the tight-binding N-terminal cleavage products is likely as useful as if it was defined based on the intact substrates. However, full-length substrate co-crystal may add to our understanding of substrate recognition by NS3/4A and allow more accurate predictions of substrate co-evolution. Generating molecular models is also an option in case good quality crystals cannot be obtained.

Another factor that may be important for the balance between substrate recognition and inhibitor binding is the role of water molecules in recognition. [190–192] Five water molecules are conserved in the substrate co-crystal structures of HIV-1 protease. [7] These water molecules mediate hydrogen bonding between the protease and its substrates in the crystal structures. While the NS3/4A substrates do not have conserved waters in the crystal structures, the inhibitors danoprevir and vaniprevir have eight crystallographic overlapping water molecules that are within 4 Å of both protease and inhibitor. Characterization of the thermodynamic stability of these waters may provide a more complete view of the differential binding characteristics between the substrates and inhibitors.

6.4 Predicting Substrate Co-evolution using the Dynamic Substrate Envelope

Protease-substrate co-evolution contributes to drug resistance in HIV-1. The HIV-1 protease substrates NC-p1 and p1-p6 has co-evolved with the V82A and D30N/N88D variants of protease. My molecular models followed by MD simulations show that the substrate envelope is preserved by the compensatory mutations in the cleavage sites, quantified as increased volume of the cleavage site staying inside the dynamic substrate envelope. However, the conformational interdependence of residues within the cleavage site sequences challenges direct prediction of substrate co-evolution from amino acid sequence or a single structure.

An initial report from the scientists at Vertex Pharmaceuticals, Inc. show that the four NS3/4A cleavage site sequences were well conserved in most of total 569 patients in response to a treatment regimen containing telaprevir but few changes in the cleavage site regions were observed in the 159 patients who failed telaprevir combination treatment, and no statistical evidence of directional selection after the acquisition of telaprevir resistance was detected. [240] Except for this report, the co-evolution of NS3/4A substrates with drug resistance has not been investigated in the HCV field. Based on the high flexibility and poor fit within the dynamic substrate envelope, I predict the 4A-4B cleavage site to be more sensitive to protease mutations, hence 4A-4B is more likely to co-evolve with resistant NS3/4A variants.

The dynamic substrate envelope can be used more systematically to predict potential NS3/4A protease-substrate co-evolution. Single mutations need to be introduced *in silico* at each position of the substrate. Each variant can be then subjected to energy minimization and short MD simulations to track V_{in} in the context of a drug resistance mutation in the protease. Similar to HIV-1, the NS3/4A substrate variants with improved V_{in} values relative to the mutant protease-wild-type substrate complex are more likely to compensate for the resistance mutation in the protease.

Being able to accurately predict the substrate co-evolution can be useful to reduce the likelihood of drug resistance in drug design. Some inhibitors may protrude beyond the envelope to make favorable contacts with a protease residue with the purpose of preserving affinity or to improve selectivity, solubility, permeability, etc. In those cases, accurately predicting how likely for one or more substrates to co-evolve with a potential resistance mutation in that particular protease residue is crucial. The availability of this probabilistic information can be used to decide at which location an inhibitor's protrusion beyond the envelope is more affordable in terms of avoiding the emergence of drug resistance via cleavage site mutations.

6.5 Concluding Remarks

Drug resistance will occur anytime rapid growth and evolution exists under the selective pressure of drug treatment but the growth is not completely inhibited by the drug, as it is widespread problem in invasive cancers and pathogenic microbes

such as bacteria, malaria, fungi, tuberculosis, and viruses. The mechanisms by which resistance can emerge include point mutations in the target protein. To overcome drug resistance, resistance should be predicted before it happens and drugs should be designed accordingly to avoid the accurately predicted resistance mutations. To achieve this goal, target identification is critical. The enzymes with multiple substrates that cannot easily tolerate mutations and maintain function are potentially good candidates. Moreover, all drug targets should be viewed as evolutionarily dynamic, not static. The inhibitors should be made as evolutionarily constrained as possible.

The crystal structures are extremely informative to provide insights into the most probable molecular interactions at the binding sites. But in reality, proteins are dynamic, they exist in conformational ensembles. Some proteins are inherently more flexible; others are less so. Some inherently flexible proteins may lose a great deal of flexibility upon inhibitor binding; others may not. Depending on the protein, ignoring protein dynamics may delay the successful discovery of perfect drugs that have high potency, good selectivity, and low toxicity and are also robust against the evolution of resistance.

Doing so, experimental techniques and computational methods should be used in concert, each according to its particular strengths. Dynamic substrate envelope is a useful tool to systematically incorporate the viral protein dynamics and evolution into structure-based rational drug design. Constant partnering of the dynamic substrate envelope with multiple disciplines is absolutely necessary to achieve the goal of discovering robust drugs, such as chemical synthesis, thermodynamics and

enzyme kinetics, crystallography and NMR, deep sequencing, virology, and many more.

I hope my work on the structure and dynamics of viral substrate recognition and drug resistance will serve as a helpful guide for the rational design of future generation drugs that remain active against diverse populations of drug targets.

Appendices

Appendix A

Assessing van der Waals Contacts

1.1 A Simplified Lennard-Jones Potential

How a particular ligand (substrate or inhibitor) interacts with a particular protease variant can be assessed in terms of their detailed molecular interactions; i.e. hydrogen bonds and van der Waals (vdW) interaction energies. vdW interactions at a binding surface represent the local packing and can be used to understand the degree of complementarity between a ligand and receptor. Assessing vdW contacts accurately is a subtle analysis since a simple distance criterion of 4.2 - 5.0 Å often does not distribute the energetic contributions accurately.

The vdW contacts in a macromolecular structure are typically evaluated using the *Lennard-Jones 6-12* potential function, a function of the distance between the

two atoms and the atom types:

$$V(r) = 4\epsilon \left[\left(\frac{\sigma}{r} \right)^{12} - \left(\frac{\sigma}{r} \right)^6 \right] \quad (1.1)$$

where ϵ and σ are the well depth and the collision diameter, respectively, which depend on the types of interacting atoms (Figure 1.1A). The positive sign in the first term of Equation 1.1 implies a repulsion, and the minus sign in the second term implies an attraction. The repulsive part of this potential, was originally chosen because it can be calculated rapidly by computers using the the square of r^{-6} . The virtue of this model is that it captures the universal features of a short-range attraction and even shorter-ranged repulsion and the two parameters ϵ and σ give enough flexibility for the model to predict experimental data accurately.

While this potential (Equation 1.1) works well in computing the interactions in a molecular dynamics simulation, the repulsive term is too restrictive in assessing particular interactions of an experimentally determined crystal structure, where slight changes in position can cause a favorable interaction to be considered unfavorable.

To overcome this problem and to quantitatively assess vdW contacts, I developed a "simplified van der Waals" calculation, where the attractive potential is retained, but the repulsive potential is removed (Figure 1.1B). Atoms that are "much" too close are flagged for the crystallographer to check. This approach allows the vdW potentials to be calculated and compared for each ligand (substrate or inhibitor) in complex with the wild-type or mutated protease. Also the

distribution of the vdW is shown for a ligand on a residue-by-residue basis. This the contacts can be quantitatively compared for a series of structures, such as a particular inhibitor with a series of protease variants, or a series of different inhibitors with the same protease. In addition, the role of a particular amino acid residue can be compared for a series of inhibitors. I have applied this technique to the analysis of various HIV-1 protease and HCV NS3/4A protease structures. [46, 90, 165, 187, 189, 241, 242]

Figure 1.1: The Original and Simplified Lennard-Jones 6-12 Potential Functions. (A) Lennard-Jones 6-12 potential function. (B) The Lennard-Jones 6-12 potential function simplified to consider only the attractive forces.

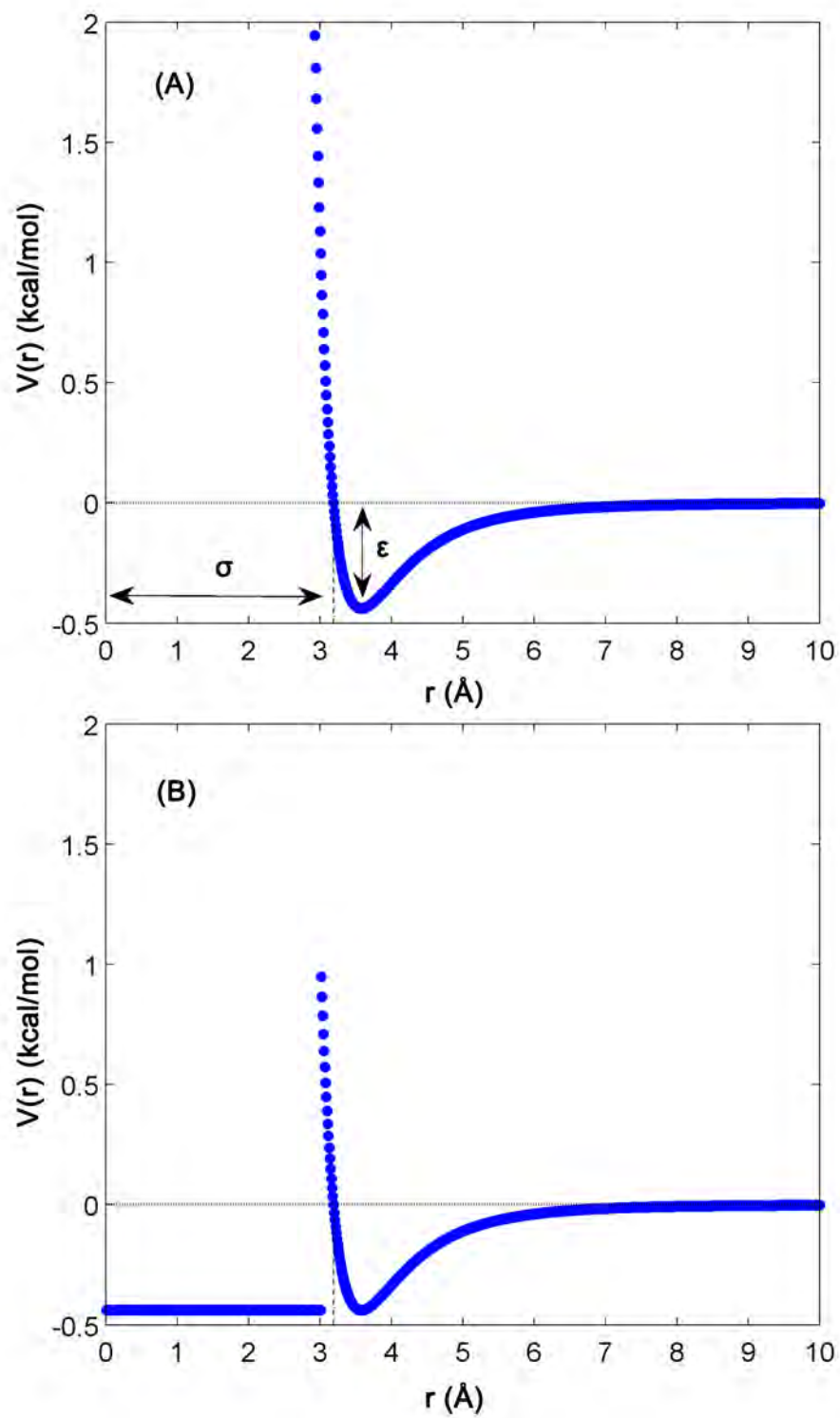


Figure 1.1

1.2 Forcefield

For macroscopic system like biological macromolecules, the total van der Waals force is often estimated as the sum over all interacting pairs. To get the van der Waals interaction energy of X-Y atom pair, ϵ and σ are obtained by arithmetically or geometrically averaging the parameters of the individual atoms, *i.e.*, The values of the ϵ and σ parameters were taken from the CNS forcefield for HIV-1 protease-substrate interactions (chapter II, and III):

$$\begin{aligned}\sigma_{XY} &= \frac{1}{2}(\sigma_{XX} + \sigma_{YY}) \\ \epsilon_{XY} &= \sqrt{\epsilon_{XX}\epsilon_{YY}}\end{aligned}\tag{1.2}$$

For the interactions of NS3/4A protease with its substrates and small-molecule inhibitors, the ϵ and σ parameters were taken from the OPLS2005 forcefield (chapter IV, and V):

$$\begin{aligned}\sigma_{XY} &= \frac{1}{2}(\sigma_{XX} + \sigma_{YY}) \\ \epsilon_{XY} &= \frac{1}{2}(\epsilon_{XX} + \epsilon_{YY})\end{aligned}\tag{1.3}$$

1.3 Limitations

The van der Waals contact potential of the atoms that have alternate conformations in a crystal structure was calculated based on the weighted averaged coordinates of those atoms. The alternate conformations were weighted by the occupancy factor of each conformer. To determine the contact potential of each alternate conformation separately, one should modify the input PDB file to keep the desired conformation and delete all others and alter the occupancy factors to one.

The sole purpose of this calculation is to get a quantitative picture of the favorable local packing at an interaction surface. The calculated contact potential values DO NOT intend to replicate the binding free energies or DO NOT NECESSARILY follow a similar trend with the binding free energies for a series of compounds bound to the same drug target. The change in free energy upon binding is dependent on a variety of components, only one of which is the van der Waals contact potential.

References

- [1] “Global report: Unaid report on the global aids epidemic 2012,” tech. rep., UNAIDS/JC2417E, UNAIDS.
- [2] A. F. Santos and M. A. Soares, “Hiv genetic diversity and drug resistance,” *Viruses*, vol. 2, no. 2, pp. 503–531, 2010.
- [3] J. Palella, F. J., K. M. Delaney, A. C. Moorman, M. O. Loveless, J. Fuhrer, G. A. Satten, D. J. Aschman, and S. D. Holmberg, “Declining morbidity and mortality among patients with advanced human immunodeficiency virus infection. hiv outpatient study investigators,” *N Engl J Med*, vol. 338, no. 13, pp. 853–60, 1998.
- [4] R. Hoggs, H. K.V., and B. Yip, “Improved survival among hiv-infected individuals following initiation of antiretroviral therapy,” *J Am Med Assoc*, vol. 279, pp. 450–454, 1998.
- [5] A. Fun, A. M. Wensing, J. Verheyen, and M. Nijhuis, “Human immunodeficiency virus gag and protease: partners in resistance,” *Retrovirology*, vol. 9, p. 63, 2012.
- [6] M. Prabu-Jeyabalan, E. Nalivaika, and C. A. Schiffer, “How does a symmetric dimer recognize an asymmetric substrate? a substrate complex of hiv-1 protease,” *J Mol Biol*, vol. 301, no. 5, pp. 1207–20, 2000.
- [7] M. Prabu-Jeyabalan, E. Nalivaika, and C. A. Schiffer, “Substrate shape determines specificity of recognition for hiv-1 protease: analysis of crystal structures of six substrate complexes,” *Structure*, vol. 10, no. 3, pp. 369–81, 2002.

-
- [8] N. M. King, M. Prabu-Jeyabalan, E. A. Nalivaika, and C. A. Schiffer, "Combating susceptibility to drug resistance: lessons from hiv-1 protease," *Chem Biol*, vol. 11, no. 10, pp. 1333–8, 2004.
- [9] K. P. Romano, A. Ali, W. E. Royer, and C. A. Schiffer, "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*, vol. 107, no. 49, pp. 20986–91, 2010.
- [10] N. E. Kohl, E. A. Emini, W. A. Schleif, L. J. Davis, J. C. Heimbach, R. A. Dixon, E. M. Scolnick, and I. S. Sigal, "Active human immunodeficiency virus protease is required for viral infectivity," *Proc Natl Acad Sci U S A*, vol. 85, no. 13, pp. 4686–90, 1988.
- [11] N. A. Roberts, J. A. Martin, D. Kinchington, A. V. Broadhurst, J. C. Craig, I. B. Duncan, S. A. Galpin, B. K. Handa, J. Kay, A. Krohn, and et al., "Rational design of peptide-based hiv proteinase inhibitors," *Science*, vol. 248, no. 4953, pp. 358–61, 1990.
- [12] B. D. Dorsey, R. B. Levin, S. L. McDaniel, J. P. Vacca, J. P. Guare, P. L. Darke, J. A. Zugay, E. A. Emini, W. A. Schleif, J. C. Quintero, and et al., "L-735,524: the design of a potent and orally bioavailable hiv protease inhibitor," *J Med Chem*, vol. 37, no. 21, pp. 3443–51, 1994.
- [13] D. J. Kempf, K. C. Marsh, J. F. Denissen, E. McDonald, S. Vasavanonda, C. A. Flentge, B. E. Green, L. Fino, C. H. Park, X. P. Kong, and et al., "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*, vol. 92, no. 7, pp. 2484–8, 1995.
- [14] S. Kaldor, V. Kalish, J. n. Davies, B. Shetty, J. Fritz, K. Appelt, J. Burgess, K. Campanale, N. Chirgadze, D. Clawson, B. Dressman, S. Hatch, D. Khalil, M. Kosa, P. Lubbehusen, M. Muesing, A. Patick, S. Reich, K. Su, and J. Tatlock, "Viracept (nelfinavir mesylate, ag1343): a potent, orally bioavailable inhibitor of hiv-1 protease," *J Med Chem*, vol. 40, no. 24, pp. 3979–3985, 1997.
- [15] E. E. Kim, C. T. Baker, M. D. Dwyer, M. A. Murcko, and et al., "Crystal structure of hiv-1 protease in complex with vx-478, a potent and orally

- bioavailable inhibitor of the enzyme,” *J Am Chem Soc*, vol. 117, pp. 1181–1182, 1995.
- [16] H. Sham, D. Kempf, A. Molla, K. Marsh, G. Kumar, C.-M. Chen, W. Kati, K. Stewart, R. Lal, A. Hsu, D. Betebenner, M. Korneyeva, S. Vasavanonda, E. McDonald, A. Saldivar, N. Wideburg, X. Chen, P. Niu, C. Park, V. Jayanti, B. Grabowski, G. Granneman, E. Sun, A. Japour, J. Leonard, J. Plattner, and D. Norbeck, “Abt-378, a highly potent inhibitor of human immunodeficiency virus protease,” *Antimicrob Agents Chemother*, vol. 42, no. 12, pp. 3218–3224, 1998.
- [17] B. S. Robinson, K. A. Riccardi, Y. F. Gong, Q. Guo, D. A. Stock, W. S. Blair, B. J. Terry, C. A. Deminie, F. Djang, R. J. Colonno, and P. F. Lin, “Bms-232632, a highly potent human immunodeficiency virus protease inhibitor that can be used in combination with other available antiretroviral agents,” *Antimicrob Agents Chemother*, vol. 44, no. 8, pp. 2093–9, 2000.
- [18] S. R. Turner, J. W. Strohbach, R. A. Tommasi, P. A. Aristoff, P. D. Johnson, H. I. Skulnick, L. A. Dolak, E. P. Seest, P. K. Tomich, M. J. Bohanon, M. M. Horng, J. C. Lynn, K. T. Chong, R. R. Hinshaw, K. D. Watenpaugh, M. N. Janakiraman, and S. Thaisrivongs, “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*, vol. 41, no. 18, pp. 3467–76, 1998.
- [19] S. De Meyer, H. Azijn, D. Surleraux, D. Jochmans, A. Tahri, R. Pauwels, P. Wigerinck, and M. P. de Bethune, “Tmc114, a novel human immunodeficiency virus type 1 protease inhibitor active against protease inhibitor-resistant viruses, including a broad range of clinical isolates,” *Antimicrob Agents Chemother*, vol. 49, no. 6, pp. 2314–21, 2005.
- [20] Y. Koh, H. Nakata, K. Maeda, H. Ogata, G. Bilcer, T. Devasamudram, J. F. Kincaid, P. Boross, Y. F. Wang, Y. Tie, P. Volarath, L. Gaddis, R. W. Harrison, I. T. Weber, A. K. Ghosh, and H. Mitsuya, “Novel bis-tetrahydrofuranylurethane-containing nonpeptidic protease inhibitor (pi) uic-94017 (tmc114) with potent activity against multi-pi-resistant human immunodeficiency virus in vitro,” *Antimicrob Agents Chemother*, vol. 47, no. 10, pp. 3123–9, 2003.

-
- [21] D. L. Surleraux, A. Tahri, W. G. Verschueren, G. M. Pille, H. A. de Kock, T. H. Jonckers, A. Peeters, S. De Meyer, H. Azijn, R. Pauwels, M. P. de Bethune, N. M. King, M. Prabu-Jeyabalan, C. A. Schiffer, and P. B. Wigerinck, "Discovery and selection of tmc114, a next generation hiv-1 protease inhibitor," *J Med Chem*, vol. 48, no. 6, pp. 1813–22, 2005.
- [22] J. M. Coffin, "Hiv population dynamics in vivo: implications for genetic variation, pathogenesis, and therapy," *Science*, vol. 267, no. 5197, pp. 483–9, 1995.
- [23] A. Ali, R. M. Bandaranayake, Y. Cai, N. M. King, M. Kolli, S. Mittal, J. F. Murzycki, M. N. Nalam, E. A. Nalivaika, A. Ozen, M. M. Prabu-Jeyabalan, K. Thayer, and C. A. Schiffer, "Molecular basis for drug resistance in hiv-1 protease," *Viruses*, vol. 2, no. 11, pp. 2509–35, 2010.
- [24] R. M. Gulick, J. W. Mellors, D. Havlir, J. J. Eron, A. Meibohm, J. H. Condra, F. T. Valentine, D. McMahon, C. Gonzalez, L. Jonas, E. A. Emini, J. A. Chodakewitz, R. Isaacs, and D. D. Richman, "3-year suppression of hiv viremia with indinavir, zidovudine, and lamivudine," *Ann Intern Med*, vol. 133, no. 1, pp. 35–9, 2000.
- [25] J. A. Bartlett, R. DeMasi, J. Quinn, C. Moxham, and F. Rousseau, "Overview of the effectiveness of triple combination therapy in antiretroviral-naïve hiv-1 infected adults," *AIDS*, vol. 15, no. 11, pp. 1369–77, 2001.
- [26] T. Wu, C. Schiffer, M. Gonzales, K. R., S. Chou, D. Israelski, A. Zolopa, J. Fessel, and R. Shafer, "Mutation patterns and structural correlates in hiv-1 protease following varying degrees of protease inhibitor treatment," *J Virol*, vol. 77, pp. 4836–4847, 2003.
- [27] S. Y. Rhee, M. J. Gonzales, R. Kantor, B. J. Betts, J. Ravela, and R. W. Shafer, "Human immunodeficiency virus reverse transcriptase and protease sequence database," *Nucleic Acids Res*, vol. 31, no. 1, pp. 298–303, 2003.
- [28] Y. M. Zhang, H. Imamichi, T. Imamichi, H. C. Lane, J. Falloon, M. B. Vasudevachari, and N. P. Salzman, "Drug resistance during indinavir therapy is caused by mutations in the protease gene and in its gag substrate cleavage sites," *J Virol*, vol. 71, no. 9, pp. 6662–70, 1997.

-
- [29] F. Bally, R. Martinez, S. Peters, P. Sudre, and A. Telenti, "Polymorphism of hiv type 1 gag p7/p1 and p1/p6 cleavage sites: Clinical significance and implications for resistance to protease," *AIDS Res Hum Retrovir*, vol. 16, no. 13, pp. 1209–1213, 2000.
- [30] F. Mammano, C. Petit, and F. Clavel, "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*, vol. 72, no. 9, pp. 7632–7, 1998.
- [31] M. F. Maguire, R. Guinea, P. Griffin, S. Macmanus, R. C. Elston, J. Wolfram, N. Richards, M. H. Hanlon, D. J. Porter, T. Wrin, N. Parkin, M. Tisdale, E. Furfine, C. Petropoulos, B. W. Snowden, and J. P. Kleim, "Changes in human immunodeficiency virus type 1 gag at positions I449 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*, vol. 76, no. 15, pp. 7398–406, 2002.
- [32] M. Kolli, E. Stawiski, C. Chappey, and C. A. Schiffer, "Human immunodeficiency virus type 1 protease-correlated cleavage site mutations enhance inhibitor resistance," *J Virol*, vol. 83, no. 21, pp. 11027–42, 2009.
- [33] J. Martinez-Picado, A. V. Savara, L. Shi, L. Sutton, and R. T. D'Aquila, "Fitness of human immunodeficiency virus type 1 protease inhibitor-selected single mutants," *Virology*, vol. 275, no. 2, pp. 318–22, 2000.
- [34] J. Martinez-Picado, A. V. Savara, L. Sutton, and R. T. D'Aquila, "Replicative fitness of protease inhibitor-resistant mutants of human immunodeficiency virus type 1," *J Virol*, vol. 73, no. 5, pp. 3744–52, 1999.
- [35] G. Croteau, L. Doyon, D. Thibeault, G. McKercher, L. Pilote, and D. Lamarre, "Impaired fitness of human immunodeficiency virus type 1 variants with high-level resistance to protease inhibitors," *J Virol*, vol. 71, no. 2, pp. 1089–96, 1997.
- [36] G. Bleiber, M. Munoz, A. Ciuffi, P. Meylan, and A. Telenti, "Individual contributions of mutant protease and reverse transcriptase to viral infectivity, replication, and protein maturation of antiretroviral drug-resistant human immunodeficiency virus type 1," *J Virol*, vol. 75, no. 7, pp. 3291–300, 2001.

-
- [37] L. Doyon, G. Croteau, D. Thibeault, F. Poulin, L. Pilote, and D. Lamarre, "Second locus involved in human immunodeficiency virus type 1 resistance to protease inhibitors," *J Virol*, vol. 70, no. 6, pp. 3763–9, 1996.
- [38] L. H. Robinson, R. E. Myers, B. W. Snowden, M. Tisdale, and E. D. Blair, "Hiv type 1 protease cleavage site mutations and viral fitness: implications for drug susceptibility phenotyping assays," *AIDS Res Hum Retroviruses*, vol. 16, pp. 1149–1156, 2000.
- [39] M. Kolli, S. Lastere, and C. A. Schiffer, "Co-evolution of nelfinavir-resistant hiv-1 protease and the p1-p6 substrate," *Virology*, vol. 347, no. 2, pp. 405–9, 2006.
- [40] M. Prabu-Jeyabalan, E. A. Nalivaika, N. M. King, and C. A. Schiffer, "Structural basis for coevolution of a human immunodeficiency virus type 1 nucleocapsid-p1 cleavage site with a v82a drug-resistant mutation in viral protease," *J Virol*, vol. 78, no. 22, pp. 12446–54, 2004.
- [41] S. Chellappan, V. Kairys, M. X. Fernandes, C. Schiffer, and M. K. Gilson, "Evaluation of the substrate envelope hypothesis for inhibitors of hiv-1 protease," *Proteins*, vol. 68, no. 2, pp. 561–7, 2007.
- [42] N. M. King, M. Prabu-Jeyabalan, E. A. Nalivaika, P. Wigerinck, M. P. de Bethune, and C. A. Schiffer, "Structural and thermodynamic basis for the binding of tmc114, a next-generation human immunodeficiency virus type 1 protease inhibitor," *Journal of Virology*, vol. 78, no. 21, pp. 12012–21, 2004.
- [43] E. Lefebvre and C. A. Schiffer, "Resilience to resistance of hiv-1 protease inhibitors: profile of darunavir," *AIDS Rev*, vol. 10, no. 3, pp. 131–42, 2008.
- [44] M. N. Nalam and C. A. Schiffer, "New approaches to hiv protease inhibitor drug design ii: testing the substrate envelope hypothesis to avoid drug resistance and discover robust inhibitors," *Curr Opin HIV AIDS*, vol. 3, no. 6, pp. 642–6, 2008.
- [45] M. D. Altman, A. Ali, G. S. Reddy, M. N. Nalam, S. G. Anjum, H. Cao, S. Chellappan, V. Kairys, M. X. Fernandes, M. K. Gilson, C. A. Schiffer, T. M. Rana, and B. Tidor, "Hiv-1 protease inhibitors from inverse design in the substrate envelope exhibit subnanomolar binding to drug-resistant variants," *J Am. Chem. Soc.*, vol. 130, no. 19, pp. 6099–113, 2008.

-
- [46] M. N. Nalam, A. Ali, M. D. Altman, G. S. Reddy, S. Chellappan, V. Kairys, A. Ozen, H. Cao, M. K. Gilson, B. Tidor, T. M. Rana, and C. A. Schiffer, "Evaluating the substrate-envelope hypothesis: structural analysis of novel hiv-1 protease inhibitors designed to be robust against drug resistance," *J Virol*, vol. 84, no. 10, pp. 5368–78, 2010.
- [47] A. Ali, G. S. Reddy, H. Cao, S. G. Anjum, M. N. Nalam, C. A. Schiffer, and T. M. Rana, "Discovery of hiv-1 protease inhibitors with picomolar affinities incorporating n-aryl-oxazolidinone-5-carboxamides as novel p2 ligands," *J Med Chem*, vol. 49, no. 25, pp. 7342–56, 2006.
- [48] S. Chellappan, G. S. Kiran Kumar Reddy, A. Ali, M. N. Nalam, S. G. Anjum, H. Cao, V. Kairys, M. X. Fernandes, M. D. Altman, B. Tidor, T. M. Rana, C. A. Schiffer, and M. K. Gilson, "Design of mutation-resistant hiv protease inhibitors with the substrate envelope hypothesis," *Chem Biol Drug Des*, vol. 69, no. 5, pp. 298–313, 2007.
- [49] V. Kairys, M. K. Gilson, V. Lather, C. A. Schiffer, and M. X. Fernandes, "Toward the design of mutation-resistant enzyme inhibitors: further evaluation of the substrate envelope hypothesis," *Chem Biol Drug Des*, vol. 74, no. 3, pp. 234–45, 2009.
- [50] S. Tuske, S. G. Sarafianos, J. Clark, A. D., J. Ding, L. K. Naeger, K. L. White, M. D. Miller, C. S. Gibbs, P. L. Boyer, P. Clark, G. Wang, B. L. Gaffney, R. A. Jones, D. M. Jerina, S. H. Hughes, and E. Arnold, "Structures of hiv-1 rt-dna complexes before and after incorporation of the anti-aids drug tenofovir," *Nat Struct Mol Biol*, vol. 11, no. 5, pp. 469–74, 2004.
- [51] S. Lesage, C. Condroyer, A. Lannuzel, E. Lohmann, A. Troiano, F. Tison, P. Damier, S. Thobois, A. M. Ouvrard-Hernandez, S. Rivaud-Pechoux, C. Brefel-Courbon, A. Destee, C. Tranchant, M. Romana, L. Leclere, A. Durr, and A. Brice, "Molecular analyses of the lrrk2 gene in european and north african autosomal dominant parkinson's disease," *Journal of Medical Genetics*, vol. 46, no. 7, pp. 458–64, 2009.
- [52] P. Simmonds, J. Bukh, C. Combet, G. Deleage, N. Enomoto, S. Feinstone, P. Halfon, G. Inchauspe, C. Kuiken, G. Maertens, M. Mizokami, D. G. Murphy, H. Okamoto, J. M. Pawlotsky, F. Penin, E. Sablon, I. T. Shin, L. J. Stuyver, H. J. Thiel, S. Viazov, A. J. Weiner, and A. Widell, "Consensus

- proposals for a unified system of nomenclature of hepatitis c virus genotypes," *Hepatology*, vol. 42, no. 4, pp. 962–73, 2005.
- [53] J. Bukh, R. H. Miller, and R. H. Purcell, "Biology and genetic heterogeneity of hepatitis c virus," *Clin Exp Rheumatol*, vol. 13 Suppl 13, pp. S3–7, 1995.
- [54] J. Bukh, R. H. Miller, and R. H. Purcell, "Genetic heterogeneity of hepatitis c virus: quasispecies and genotypes," *Semin Liver Dis*, vol. 15, no. 1, pp. 41–63, 1995.
- [55] Q. L. Choo, G. Kuo, A. J. Weiner, L. R. Overby, D. W. Bradley, and M. Houghton, "Isolation of a cDNA clone derived from a blood-borne non-a, non-b viral hepatitis genome," *Science*, vol. 244, no. 4902, pp. 359–62, 1989.
- [56] G. Kuo, Q. L. Choo, H. J. Alter, G. L. Gitnick, A. G. Redeker, R. H. Purcell, T. Miyamura, J. L. Dienstag, M. J. Alter, C. E. Stevens, and et al., "An assay for circulating antibodies to a major etiologic virus of human non-a, non-b hepatitis," *Science*, vol. 244, no. 4902, pp. 362–4, 1989.
- [57] M. J. Alter, D. Kruszon-Moran, O. V. Nainan, G. M. McQuillan, F. Gao, L. A. Moyer, R. A. Kaslow, and H. S. Margolis, "The prevalence of hepatitis c virus infection in the united states, 1988 through 1994," *N Engl J Med*, vol. 341, no. 8, pp. 556–62, 1999.
- [58] L. M. Blatt, M. G. Mutchnick, M. J. Tong, F. M. Klion, E. Lebovics, B. Freilich, N. Bach, C. Smith, J. Herrera, H. Tobias, A. Conrad, P. Schmid, and J. G. McHutchison, "Assessment of hepatitis c virus rna and genotype from 6807 patients with chronic hepatitis c in the united states," *J Viral Hepat*, vol. 7, no. 3, pp. 196–202, 2000.
- [59] R. K. Beran and A. M. Pyle, "Hepatitis c viral ns3-4a protease activity is enhanced by the ns3 helicase," *J Biol Chem*, vol. 283, no. 44, pp. 29929–37, 2008.
- [60] R. K. Beran, B. D. Lindenbach, and A. M. Pyle, "The ns4a protein of hepatitis c virus promotes rna-coupled atp hydrolysis by the ns3 helicase," *J Virol*, vol. 83, no. 7, pp. 3268–75, 2009.

-
- [61] D. N. Frick, R. S. Rypma, A. M. Lam, and B. Gu, "The nonstructural protein 3 protease/helicase requires an intact protease domain to unwind duplex rna efficiently," *J Biol Chem*, vol. 279, no. 2, pp. 1269–80, 2004.
- [62] R. K. Beran, V. Serebrov, and A. M. Pyle, "The serine protease domain of hepatitis c viral ns3 activates rna helicase activity by promoting the binding of rna substrate," *J Biol Chem*, vol. 282, no. 48, pp. 34913–20, 2007.
- [63] A. M. Lam, D. Keeney, P. Q. Eckert, and D. N. Frick, "Hepatitis c virus ns3 atpases/helicases from different genotypes exhibit variations in enzymatic properties," *J Virol*, vol. 77, no. 7, pp. 3950–61, 2003.
- [64] P. Gallinari, D. Brennan, C. Nardi, M. Brunetti, L. Tomei, C. Steinkuhler, and R. De Francesco, "Multiple enzymatic activities associated with recombinant ns3 protein of hepatitis c virus," *J Virol*, vol. 72, no. 8, pp. 6758–69, 1998.
- [65] R. Bartenschlager, L. Ahlborn-Laake, J. Mous, and H. Jacobsen, "Kinetic and structural analyses of hepatitis c virus polyprotein processing," *J Virol*, vol. 68, no. 8, pp. 5045–55, 1994.
- [66] Z. Chen, Y. Benureau, R. Rijnbrand, J. Yi, T. Wang, L. Warter, R. E. Lanford, S. A. Weinman, S. M. Lemon, A. Martin, and K. Li, "Gb virus b disrupts rig-i signaling by ns3/4a-mediated cleavage of the adaptor protein mavs," *J Virol*, vol. 81, no. 2, pp. 964–76, 2007.
- [67] M. H. Heim, "Innate immunity and hcv," *J Hepatol*, vol. 58, no. 3, pp. 564–74, 2013.
- [68] X. D. Li, L. Sun, R. B. Seth, G. Pineda, and Z. J. Chen, "Hepatitis c virus protease ns3/4a cleaves mitochondrial antiviral signaling protein off the mitochondria to evade innate immunity," *Proc Natl Acad Sci U S A*, vol. 102, no. 49, pp. 17717–22, 2005.
- [69] K. Li, E. Foy, J. C. Ferreon, M. Nakamura, A. C. Ferreon, M. Ikeda, S. C. Ray, J. Gale, M., and S. M. Lemon, "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*, vol. 102, no. 8, pp. 2992–7, 2005.
- [70] E. D. Brenndorfer, J. Karthe, L. Frelin, P. Cebula, A. Erhardt, J. Schulte am Esch, H. Hengel, R. Bartenschlager, M. Sallberg, D. Haussinger, and J. G.

- Bode, "Nonstructural 3/4a protease of hepatitis c virus activates epithelial growth factor-induced signal transduction by cleavage of the t-cell protein tyrosine phosphatase," *Hepatology*, vol. 49, no. 6, pp. 1810–20, 2009.
- [71] X. Kang, X. Chen, Y. He, D. Guo, L. Guo, J. Zhong, and H. B. Shu, "Ddb1 is a cellular substrate of ns3/4a protease and required for hepatitis c virus replication," *Virology*, vol. 435, no. 2, pp. 385–94, 2013.
- [72] C. Steinkuhler, G. Biasiol, M. Brunetti, A. Urbani, U. Koch, R. Cortese, A. Pessi, and R. De Francesco, "Product inhibition of the hepatitis c virus ns3 protease," *Biochemistry*, vol. 37, no. 25, pp. 8899–905, 1998.
- [73] M. Llinas-Brunet, M. Bailey, G. Fazal, S. Goulet, T. Halmos, S. Laplante, R. Maurice, M. Poirier, M. A. Poupert, D. Thibeault, D. Wernic, and D. Lamarre, "Peptide-based inhibitors of the hepatitis c virus serine protease," *Bioorg Med Chem Lett*, vol. 8, no. 13, pp. 1713–8, 1998.
- [74] R. De Francesco and G. Migliaccio, "Challenges and successes in developing new therapies for hepatitis c," *Nature*, vol. 436, no. 7053, pp. 953–60, 2005.
- [75] D. Lamarre, P. C. Anderson, M. Bailey, P. Beaulieu, G. Bolger, P. Bonneau, M. Bos, D. R. Cameron, M. Cartier, M. G. Cordingley, A. M. Faucher, N. Goudreau, S. H. Kawai, G. Kukolj, L. Lagace, S. R. LaPlante, H. Narjes, M. A. Poupert, J. Rancourt, R. E. Sentjens, R. St George, B. Simoneau, G. Steinmann, D. Thibeault, Y. S. Tsantrizos, S. M. Weldon, C. L. Yong, and M. Llinas-Brunet, "An ns3 protease inhibitor with antiviral effects in humans infected with hepatitis c virus," *Nature*, vol. 426, no. 6963, pp. 186–9, 2003.
- [76] H. Hinrichsen, Y. Benhamou, H. Wedemeyer, M. Reiser, R. E. Sentjens, J. L. Calleja, X. Forns, A. Erhardt, J. Cronlein, R. L. Chaves, C. L. Yong, G. Nehmiz, and G. G. Steinmann, "Short-term antiviral efficacy of biln 2061, a hepatitis c virus serine protease inhibitor, in hepatitis c genotype 1 patients," *Gastroenterology*, vol. 127, no. 5, pp. 1347–55, 2004.
- [77] T. Vanwolleghem, P. Meuleman, L. Libbrecht, T. Roskams, R. De Vos, and G. Leroux-Roels, "Ultra-rapid cardiotoxicity of the hepatitis c virus protease inhibitor biln 2061 in the urokinase-type plasminogen activator mouse," *Gastroenterology*, vol. 133, no. 4, pp. 1144–55, 2007.

-
- [78] R. B. Perni, S. J. Almquist, R. A. Byrn, G. Chandorkar, P. R. Chaturvedi, L. F. Courtney, C. J. Decker, K. Dinehart, C. A. Gates, S. L. Harbeson, A. Heiser, G. Kalkeri, E. Kolaczowski, K. Lin, Y. P. Luong, B. G. Rao, W. P. Taylor, J. A. Thomson, R. D. Tung, Y. Wei, A. D. Kwong, and C. Lin, "Preclinical profile of vx-950, a potent, selective, and orally bioavailable inhibitor of hepatitis c virus ns3-4a serine protease," *Antimicrob Agents Chemother*, vol. 50, no. 3, pp. 899–909, 2006.
- [79] A. D. Kwong, R. S. Kauffman, P. Hurter, and P. Mueller, "Discovery and development of telaprevir: an ns3-4a protease inhibitor for treating genotype 1 chronic hepatitis c virus," *Nat Biotechnol*, vol. 29, no. 11, pp. 993–1003, 2011.
- [80] B. A. Malcolm, R. Liu, F. Lahser, S. Agrawal, B. Belanger, N. Butkiewicz, R. Chase, F. Gheyas, A. Hart, D. Hesk, P. Ingravallo, C. Jiang, R. Kong, J. Lu, J. Pichardo, A. Prongay, A. Skelton, X. Tong, S. Venkatraman, E. Xia, V. Girijavallabhan, and F. G. Njoroge, "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*, vol. 50, no. 3, pp. 1013–20, 2006.
- [81] S. D. Seiwert, K. Kossen, L. Pan, J. Liu, and B. O. Buckman, "Discovery and development of the hcv ns3/4a protease inhibitor danoprevir (itm-191/rg7227)," in *Antiviral Drugs*, pp. 257–271, John Wiley & Sons, Inc., 2011.
- [82] S. Harper, J. A. McCauley, M. T. Rudd, M. Ferrara, M. DiFilippo, B. Crescenzi, U. Koch, A. Petrocchi, M. K. Holloway, J. W. Butcher, J. J. Romano, K. J. Bush, K. F. Gilbert, C. J. McIntyre, K. T. Nguyen, E. Nizi, S. S. Carroll, S. W. Ludmerer, C. Burlein, J. M. DiMuzio, D. J. Graham, C. M. McHale, M. W. Stahlhut, D. B. Olsen, E. Monteagudo, S. Cianetti, C. Giuliano, V. Pucci, N. Trainor, C. M. Fandozzi, M. Rowley, P. J. Coleman, J. P. Vacca, V. Summa, and N. J. Liverton, "Discovery of mk-5172, a macrocyclic hepatitis c virus ns3/4a protease inhibitor," *ACS Medicinal Chemistry Letters*, vol. 3, no. 4, pp. 332–336, 2012.
- [83] N. J. Liverton, S. S. Carroll, J. Dimuzio, C. Fandozzi, D. J. Graham, D. Hazuda, M. K. Holloway, S. W. Ludmerer, J. A. McCauley, C. J. McIntyre, D. B. Olsen, M. T. Rudd, M. Stahlhut, and J. P. Vacca, "Mk-7009, a

- potent and selective inhibitor of hepatitis c virus ns3/4a protease,” *Antimicrob Agents Chemother*, vol. 54, no. 1, pp. 305–11, 2010.
- [84] Y. He, M. S. King, D. J. Kempf, L. Lu, H. B. Lim, P. Krishnan, W. Kati, T. Middleton, and A. Molla, “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*, vol. 52, no. 3, pp. 1101–10, 2008.
- [85] T. L. Kieffer, C. Sarrazin, J. S. Miller, M. W. Welker, N. Forestier, H. W. Reesink, A. D. Kwong, and S. Zeuzem, “Telaprevir and pegylated interferon-alpha-2a inhibit wild-type and resistant genotype 1 hepatitis c virus replication in patients,” *Hepatology*, vol. 46, no. 3, pp. 631–9, 2007.
- [86] C. Lin, C. A. Gates, B. G. Rao, D. L. Brennan, J. R. Fulghum, Y. P. Luong, J. D. Frantz, K. Lin, S. Ma, Y. Y. Wei, R. B. Perni, and A. D. Kwong, “In vitro studies of cross-resistance mutations against two hepatitis c virus serine protease inhibitors, vx-950 and biln 2061,” *J Biol Chem*, vol. 280, no. 44, pp. 36784–91, 2005.
- [87] C. Sarrazin, R. Rouzier, F. Wagner, N. Forestier, D. Larrey, S. K. Gupta, M. Hussain, A. Shah, D. Cutler, J. Zhang, and S. Zeuzem, “Sch 503034, a novel hepatitis c virus protease inhibitor, plus pegylated interferon alpha-2b for genotype 1 nonresponders,” *Gastroenterology*, vol. 132, no. 4, pp. 1270–8, 2007.
- [88] X. Tong, S. Bogen, R. Chase, V. Girijavallabhan, Z. Guo, F. G. Njoroge, A. Prongay, A. Saksena, A. Skelton, E. Xia, and R. Ralston, “Characterization of resistance mutations against hcv ketoamide protease inhibitors,” *Antiviral Res*, vol. 77, no. 3, pp. 177–85, 2008.
- [89] X. Tong, R. Chase, A. Skelton, T. Chen, J. Wright-Minogue, and B. A. Malcolm, “Identification and analysis of fitness of resistance mutations against the hcv protease inhibitor sch 503034,” *Antiviral Res*, vol. 70, no. 2, pp. 28–38, 2006.
- [90] K. P. Romano, A. Ali, C. Aydin, D. Soumana, A. Ozen, L. M. Deveau, C. Silver, H. Cao, A. Newton, C. J. Petropoulos, W. Huang, and C. A. Schiffer, “The molecular basis of drug resistance against hepatitis c virus ns3/4a protease inhibitors,” *PLoS Pathog*, vol. 8, no. 7, p. e1002832, 2012.

-
- [91] E. Fischer, "Einfluss der configuration auf die wirkung der enzyme," *Berichte der deutschen chemischen Gesellschaft*, vol. 27, no. 3, pp. 2985–2993, 1894.
- [92] D. E. Koshland, "Application of a theory of enzyme specificity to protein synthesis," *Proc Natl Acad Sci U S A*, vol. 44, no. 2, pp. 98–104, 1958.
- [93] B. Ma, S. Kumar, C. J. Tsai, and R. Nussinov, "Folding funnels and binding mechanisms," *Protein Eng*, vol. 12, no. 9, pp. 713–20, 1999.
- [94] S. Kumar, B. Ma, C. J. Tsai, H. Wolfson, and R. Nussinov, "Folding funnels and conformational transitions via hinge-bending motions," *Cell Biochem Biophys*, vol. 31, no. 2, pp. 141–64, 1999.
- [95] C. J. Tsai, S. Kumar, B. Ma, and R. Nussinov, "Folding funnels, binding funnels, and protein function," *Protein Sci*, vol. 8, no. 6, pp. 1181–90, 1999.
- [96] B. Ma, M. Shatsky, H. J. Wolfson, and R. Nussinov, "Multiple diverse ligands binding at a single protein site: a matter of pre-existing populations," *Protein Sci*, vol. 11, no. 2, pp. 184–97, 2002.
- [97] M. R. Sawaya and J. Kraut, "Loop and subdomain movements in the mechanism of escherichia coli dihydrofolate reductase: crystallographic evidence," *Biochemistry*, vol. 36, no. 3, pp. 586–603, 1997.
- [98] F. M. McMillan, M. Cahoon, A. White, L. Hedstrom, G. A. Petsko, and D. Ringe, "Crystal structure at 2.4 a resolution of borrelia burgdorferi inosine 5'-monophosphate dehydrogenase: evidence of a substrate-induced hinged-lid motion by loop 6," *Biochemistry*, vol. 39, no. 15, pp. 4533–42, 2000.
- [99] F. Colonna-Cesari, D. Perahia, M. Karplus, H. Eklund, C. I. Braden, and O. Tapia, "Interdomain motion in liver alcohol dehydrogenase. structural and energetic analysis of the hinge bending mode," *J Biol Chem*, vol. 261, no. 32, pp. 15273–80, 1986.
- [100] S. Rozovsky, G. Jogl, L. Tong, and A. E. McDermott, "Solution-state nmr investigations of triosephosphate isomerase active site loop motion: ligand release in relation to active site loop dynamics," *J Mol Biol*, vol. 310, no. 1, pp. 271–80, 2001.

-
- [101] D. Freedberg, R. Ishima, J. Jacob, Y. Wang, I. Kustanovich, J. Louis, and D. Torchia, "Rapid structural fluctuations of the free hiv protease flaps in solution: relationship to crystal structures and comparison with predictions of dynamics calculations," *Protein Sci*, vol. 11, no. 2, pp. 221–232, 2002.
- [102] R. Ishima, D. Freedberg, Y. Wang, J. Louis, and D. Torchia, "Flap opening and dimer-interface flexibility in the free and inhibitor-bound hiv protease, and their implications for function," *Structure with Folding and Design*, vol. 7, no. 9, pp. 1047–1055, 1999.
- [103] E. Katoh, J. M. Louis, T. Yamazaki, A. M. Gronenborn, D. A. Torchia, and R. Ishima, "A solution nmr study of the binding kinetics and the internal dynamics of an hiv-1 protease-substrate complex," *Protein Sci*, vol. 12, no. 7, pp. 1376–85, 2003.
- [104] J. R. Collins, S. K. Burt, and J. W. Erickson, "Flap opening in hiv-1 protease simulated by 'activated' molecular dynamics," *Nat Struct Biol*, vol. 2, no. 4, pp. 334–8, 1995.
- [105] W. Scott and C. Schiffer, "Curling of flap tips in hiv-1 protease as a mechanism for substrate entry and tolerance of drug resistance," *Structure*, vol. 8, pp. 1259–1265, 2000.
- [106] R. P. Venkitakrishnan, E. Zaborowski, D. McElheny, S. J. Benkovic, H. J. Dyson, and P. E. Wright, "Conformational changes in the active site loops of dihydrofolate reductase during the catalytic cycle," *Biochemistry*, vol. 43, no. 51, pp. 16046–55, 2004.
- [107] N. Ozer, C. A. Schiffer, and T. Haliloglu, "Rationale for more diverse inhibitors in competition with substrates in hiv-1 protease," *Biophys J*, vol. 99, no. 5, pp. 1650–9, 2010.
- [108] D. A. Case and M. Karplus, "Stereochemistry of carbon monoxide binding to myoglobin and hemoglobin," *J Mol Biol*, vol. 123, no. 4, pp. 697–701, 1978.
- [109] C. R. Wagner, Z. Huang, S. F. Singleton, and S. J. Benkovic, "Molecular basis for nonadditive mutational effects in escherichia coli dihydrofolate reductase," *Biochemistry*, vol. 34, no. 48, pp. 15671–80, 1995.

-
- [110] A. Y. Istomin, M. M. Gromiha, O. K. Vorov, D. J. Jacobs, and D. R. Livesay, "New insight into long-range nonadditivity within protein double-mutant cycles," *Proteins*, vol. 70, no. 3, pp. 915–24, 2008.
- [111] A. S. Mildvan, D. J. Weber, and A. Kuliopulos, "Quantitative interpretations of double mutations of enzymes," *Arch Biochem Biophys*, vol. 294, no. 2, pp. 327–40, 1992.
- [112] G. Klebe, "On the validity of popular assumptions in computational drug design," *J Cheminform*, vol. 3, no. 1, p. O18, 2011.
- [113] B. G. Miller, M. J. Snider, S. A. Short, and R. Wolfenden, "Contribution of enzyme-phosphoribosyl contacts to catalysis by orotidine 5'-phosphate decarboxylase," *Biochemistry*, vol. 39, no. 28, pp. 8113–8, 2000.
- [114] B. Somogyi, Z. Lakos, A. Szarka, and M. Nyitrai, "Protein flexibility as revealed by fluorescence resonance energy transfer: an extension of the method for systems with multiple labels," *J Photochem Photobiol B*, vol. 59, no. 1-3, pp. 26–32, 2000.
- [115] W. L. Hubbell, D. S. Cafiso, and C. Altenbach, "Identifying conformational changes with site-directed spin labeling," *Nat Struct Biol*, vol. 7, no. 9, pp. 735–9, 2000.
- [116] J. Lipfert and S. Doniach, "Small-angle x-ray scattering from rna, proteins, and protein complexes," *Annu Rev Biophys Biomol Struct*, vol. 36, pp. 307–27, 2007.
- [117] P. Cozzini, G. E. Kellogg, F. Spyrakis, D. J. Abraham, G. Costantino, A. Emerson, F. Fanelli, H. Gohlke, L. A. Kuhn, G. M. Morris, M. Orozco, T. A. Pertinhez, M. Rizzi, and C. A. Sotriffer, "Target flexibility: an emerging consideration in drug discovery and design," *J Med Chem*, vol. 51, no. 20, pp. 6237–55, 2008.
- [118] J. S. Fraser, H. van den Bedem, A. J. Samelson, P. T. Lang, J. M. Holton, N. Echols, and T. Alber, "Accessing protein conformational ensembles using room-temperature x-ray crystallography," *Proc Natl Acad Sci U S A*, vol. 108, no. 39, pp. 16247–52, 2011.

-
- [119] A. Dhar, K. Girdhar, D. Singh, H. Gelman, S. Ebbinghaus, and M. Gruebele, "Protein stability and folding kinetics in the nucleus and endoplasmic reticulum of eucaryotic cells," *Biophys J*, vol. 101, no. 2, pp. 421–30, 2011.
- [120] H. Dong, S. Qin, and H. X. Zhou, "Effects of macromolecular crowding on protein conformational changes," *PLoS Comput Biol*, vol. 6, p. e1000833, 2010.
- [121] S. S. Cho, P. Weinkam, and P. G. Wolynes, "Origins of barriers and barrierless folding in bbl," *Proc Natl Acad Sci U S A*, vol. 105, no. 1, pp. 118–23, 2008.
- [122] J. L. Radkiewicz and C. L. Brooks, "Protein dynamics in enzymatic catalysis: exploration of dihydrofolate reductase," *Journal of the American Chemical Society*, vol. 122, no. 2, pp. 225–231, 2000.
- [123] A. C. Pan, D. W. Borhani, R. O. Dror, and D. E. Shaw, "Molecular determinants of drug-receptor binding kinetics," *Drug Discov Today*, 2013.
- [124] J. A. McCammon, B. R. Gelin, and M. Karplus, "Dynamics of folded proteins," *Nature*, vol. 267, no. 5612, pp. 585–90, 1977.
- [125] K. Lindorff-Larsen, S. Piana, R. O. Dror, and D. E. Shaw, "How fast-folding proteins fold," *Science*, vol. 334, no. 6055, pp. 517–20, 2011.
- [126] P. L. Freddolino, A. S. Arkhipov, S. B. Larson, A. McPherson, and K. Schulten, "Molecular dynamics simulations of the complete satellite tobacco mosaic virus," *Structure*, vol. 14, no. 3, pp. 437–49, 2006.
- [127] V. Y. Torbeev, H. Raghuraman, D. Hamelberg, M. Tonelli, W. M. Westler, E. Perozo, and S. B. Kent, "Protein conformational dynamics in the mechanism of hiv-1 protease catalysis," *Proc Natl Acad Sci U S A*, vol. 108, no. 52, pp. 20982–7, 2011.
- [128] Y. Cai, N. K. Yilmaz, W. Myint, R. Ishima, and C. A. Schiffer, "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*, vol. 8, no. 10, pp. 3452–3462, 2012.
- [129] I. M. devera, A. N. Smith, M. C. Dancel, X. Huang, B. M. Dunn, and G. E. Fanucci, "Elucidating a relationship between conformational sampling and drug resistance in hiv-1 protease," *Biochemistry*, 2013.

-
- [130] J. E. Foulkes-Murzycki, W. R. Scott, and C. A. Schiffer, "Hydrophobic sliding: a possible mechanism for drug resistance in human immunodeficiency virus type 1 protease," *Structure*, vol. 15, no. 2, pp. 225–33, 2007.
- [131] V. Mountain, "Astex, structural genomics, and syrrx. i can see clearly now: structural biology and drug discovery," *Chem Biol*, vol. 10, no. 2, pp. 95–8, 2003.
- [132] S. J. Teague, "Implications of protein flexibility for drug discovery," *Nat Rev Drug Discov*, vol. 2, no. 7, pp. 527–41, 2003.
- [133] A. Velazquez-Campoy, I. Luque, and E. Freire, "The application of thermodynamic methods in drug design," *Thermochimica Acta*, vol. 380, no. 2, pp. 217–227, 2001.
- [134] H. A. Carlson and J. A. McCammon, "Accommodating protein flexibility in computational drug design," *Mol Pharmacol*, vol. 57, no. 2, pp. 213–8, 2000.
- [135] F. Jiang and S. H. Kim, "'soft docking': matching of molecular surface cubes," *J Mol Biol*, vol. 219, no. 1, pp. 79–102, 1991.
- [136] J. A. McCammon, "Target flexibility in molecular recognition," *Biochim Biophys Acta*, vol. 1754, no. 1-2, pp. 221–4, 2005.
- [137] J. H. Lin, A. L. Perryman, J. R. Schames, and J. A. McCammon, "Computational drug design accommodating receptor flexibility: the relaxed complex scheme," *J Am Chem Soc*, vol. 124, no. 20, pp. 5632–3, 2002.
- [138] K. L. Meagher and H. A. Carlson, "Incorporating protein flexibility in structure-based drug discovery: using hiv-1 protease as a test case," *J Am Chem Soc*, vol. 126, no. 41, pp. 13276–81, 2004.
- [139] K. L. Meagher, M. G. Lerner, and H. A. Carlson, "Refining the multiple protein structure pharmacophore method: consistency across three independent hiv-1 protease models," *J Med Chem*, vol. 49, no. 12, pp. 3478–84, 2006.
- [140] A. C. Anderson, "Avoiding the rigid receptor: Side-chain rotamers," in *Computational and Structural Approaches to Drug Discovery: Ligand-Protein Interactions*, pp. 192–203, The Royal Society of Chemistry, 2008.

-
- [141] D. S. Goodsell, G. M. Morris, and A. J. Olson, "Automated docking of flexible ligands: applications of autodock," *J Mol Recognit*, vol. 9, no. 1, pp. 1–5, 1996.
- [142] M. P. Repasky, M. Shelley, and R. A. Friesner, "Flexible ligand docking with glide," *Curr Protoc Bioinformatics*, vol. Chapter 8, p. Unit 8 12, 2007.
- [143] D. J. Osguthorpe, W. Sherman, and A. T. Hagler, "Exploring protein flexibility: Incorporating structural ensembles from crystal structures and simulation into virtual screening protocols," *J Phys Chem B*, 2012.
- [144] D. W. Borhani and D. E. Shaw, "The future of molecular dynamics simulations in drug discovery," *J Comput Aided Mol Des*, vol. 26, no. 1, pp. 15–26, 2012.
- [145] N. M. King, L. Melnick, M. Prabu-Jeyabalan, E. A. Nalivaika, S. S. Yang, Y. Gao, X. Nie, C. Zepp, D. L. Heefner, and C. A. Schiffer, "Lack of synergy for inhibitors targeting a multi-drug-resistant hiv-1 protease," *Protein Sci*, vol. 11, no. 2, pp. 418–29, 2002.
- [146] J. Tozser, I. Blaha, T. D. Copeland, E. M. Wondrak, and S. Oroszlan, "Comparison of the hiv-1 and hiv-2 proteinases using oligopeptide substrates representing cleavage sites in gag and gag-pol polyproteins," *FEBS Lett*, vol. 281, no. 1-2, pp. 77–80, 1991.
- [147] S. Choudhury, L. Everitt, S. C. Pettit, and A. H. Kaplan, "Mutagenesis of the dimer interface residues of tethered and untethered hiv-1 protease result in differential activity and suggest multiple mechanisms of compensation," *Virology*, vol. 307, no. 2, pp. 204–12, 2003.
- [148] S. C. Pettit, J. C. Clemente, J. A. Jeung, B. M. Dunn, and A. H. Kaplan, "Ordered processing of the human immunodeficiency virus type 1 gagpol precursor is influenced by the context of the embedded viral protease," *J Virol*, vol. 79, no. 16, pp. 10601–7, 2005.
- [149] A. J. Olson and D. S. Goodsell, "Automated docking and the search for hiv protease inhibitors," *SAR QSAR Environ Res*, vol. 8, no. 3-4, pp. 273–85, 1998.
- [150] D. Case, T. Cheatham, T. Darden, C. Simmerling, J. Wang, R. Duke, R. Luo, K. Mertz, B. Wang, D. Pearlman, M. Crowley, S. Brozell, V. Tsui,

-
- H. Gohlke, J. Mongan, V. Hornak, G. Cui, P. Beroza, C. Schafmeister, J. Caldwell, W. Ross, and P. Kollman, "Amber 8," 2004.
- [151] D. A. Case, r. Cheatham, T. E., T. Darden, H. Gohlke, R. Luo, J. Merz, K. M., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods, "The amber biomolecular simulation programs," *J Comput Chem*, vol. 26, no. 16, pp. 1668–88, 2005.
- [152] W. L. Jorgensen, J. Chandrasekhar, J. Madura, R. Impey, and M. Klein, "Comparison of simple potential functions for simulating liquid water," *Journal of Chemical Physics*, vol. 79, pp. 926–935, 1983.
- [153] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, "Numerical-integration of cartesian equations of motion of a system with constraints - molecular-dynamics of n-alkanes," *Journal of Computational Physics*, vol. 23, no. 3, pp. 327–341, 1977.
- [154] H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola, and J. R. Haak, "Molecular-dynamics with coupling to an external bath," *Journal of Chemical Physics*, vol. 81, no. 8, pp. 3684–3690, 1984.
- [155] U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, "A smooth particle mesh ewald method," *Journal of Chemical Physics*, vol. 103, no. 19, pp. 8577–8593, 1995.
- [156] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," *J Mol Graph*, vol. 14, no. 1, pp. 33–8, 27–8, 1996.
- [157] M. A. Navia, P. M. Fitzgerald, B. M. McKeever, C. T. Leu, J. C. Heimbach, W. K. Herber, I. S. Sigal, P. L. Darke, and J. P. Springer, "Three-dimensional structure of aspartyl protease from human immunodeficiency virus hiv-1," *Nature*, vol. 337, no. 6208, pp. 615–20, 1989.
- [158] A. Wlodawer, M. Miller, M. Jaskolski, B. K. Sathyanarayana, E. Baldwin, I. T. Weber, L. M. Selk, L. Clawson, J. Schneider, and S. B. Kent, "Conserved folding in retroviral proteases: crystal structure of a synthetic hiv-1 protease," *Science*, vol. 245, no. 4918, pp. 616–21, 1989.
- [159] C. Flexner, "Hiv drug development: the next 25 years," *Nat Rev Drug Discov*, vol. 6, no. 12, pp. 959–66, 2007.

-
- [160] J. P. Ji and L. A. Loeb, "Fidelity of hiv-1 reverse transcriptase copying rna in vitro," *Biochemistry*, vol. 31, no. 4, pp. 954–8, 1992.
- [161] A. Feher, I. T. Weber, P. Bagossi, P. Boross, B. Mahalingam, J. M. Louis, T. D. Copeland, I. Y. Torshin, R. W. Harrison, and J. Tozser, "Effect of sequence polymorphism and drug resistance on two hiv-1 gag processing sites," *Eur J Biochem*, vol. 269, no. 16, pp. 4114–20, 2002.
- [162] S. La Seta Catamancio, M. P. De Pasquale, P. Citterio, S. Kurtagic, M. Galli, and S. Rusconi, "In vitro evolution of the human immunodeficiency virus type 1 gag-protease region and maintenance of reverse transcriptase resistance following prolonged drug exposure," *J Clin Microbiol*, vol. 39, no. 3, pp. 1124–9, 2001.
- [163] V. Pai and . -. Nahata, M.C. . Ann Pharmacother, "Nelfinavir mesylate: a protease inhibitor," *Ann Pharmacotherapy*, vol. 33, no. 3, pp. 325–39, 1999.
- [164] A. K. Patick, M. Duran, Y. Cao, D. Shugarts, M. R. Keller, E. Mazabel, M. Knowles, S. Chapman, D. R. Kuritzkes, and M. Markowitz, "Genotypic and phenotypic characterization of human immunodeficiency virus type 1 variants isolated from patients treated with the protease inhibitor nelfinavir," *Antimicrob Agents Chemother*, vol. 42, no. 10, pp. 2637–44, 1998.
- [165] A. Ozen, T. Haliloglu, and C. A. Schiffer, "Dynamics of preferential substrate recognition in hiv-1 protease: redefining the substrate envelope," *J Mol Biol*, vol. 410, no. 4, pp. 726–44, 2011.
- [166] "The pymol molecular graphics system, schrdinger, llc, new york, ny, 2011.."
- [167] H. C. Cote, Z. L. Brumme, and P. R. Harrigan, "Human immunodeficiency virus type 1 protease cleavage site mutations associated with protease inhibitor cross-resistance selected by indinavir, ritonavir, and/or saquinavir," *J Virol*, vol. 75, no. 2, pp. 589–94, 2001.
- [168] S. C. Pettit, N. Sheng, R. Tritch, S. Erickson-Viitanen, and R. Swanstrom, "The regulation of sequential processing of hiv-1 gag by the viral protease," *Adv Exp Med Biol*, vol. 436, pp. 15–25, 1998.

-
- [169] R. W. Shafer, "Rationale and uses of a public hiv drug-resistance database," *J Infect Dis*, vol. 194 Suppl 1, pp. S51–8, 2006.
- [170] "Stanford university hiv drug resistance database, <http://hivdb.stanford.edu/>."
- [171] "Hiv-1, human protein interaction database, <http://www.ncbi.nlm.nih.gov/projects/refseq/hivinteractions/>."
- [172] Y. Khudyakov, "Coevolution and hbv drug resistance," *Antivir Ther*, vol. 15, no. 3 Pt B, pp. 505–15, 2010.
- [173] W. Hu, "Correlated mutations in the four influenza proteins essential for viral rna synthesis, host adaptation, and virulence: Np, pa, pb1, and pb2," *Natural Science*, vol. 2, no. 10, pp. 1138–1147, 2010.
- [174] S. A. Qureshi, "Hepatitis c virus–biology, host evasion strategies, and promising new therapies on the horizon," *Medical Research Reviews*, vol. 27, no. 3, pp. 353–73, 2007.
- [175] J. Lara, G. Xia, M. Purdy, and Y. Khudyakov, "Coevolution of the hepatitis c virus polyprotein sites in patients on combined pegylated interferon and ribavirin therapy," *J Virol*, vol. 85, no. 7, pp. 3649–63, 2011.
- [176] J. M. Sayer, F. Liu, R. Ishima, I. T. Weber, and J. M. Louis, "Effect of the active site d25n mutation on the structure, stability, and ligand binding of the mature hiv-1 protease," *J Biol Chem*, vol. 283, no. 19, pp. 13459–70, 2008.
- [177] I. Miller and M. Miller, "Mathematical expectation," in *John E. Freund's Mathematical Statistics* (A. Heath and J. Grant, eds.), pp. 158–160, New Jersey, NJ: Prentice-Hall, 6th ed., 1999.
- [178] "Matlab, r2008b; the mathworks, inc.: Natick, ma, 2008.."
- [179] J. A. McCauley, C. J. McIntyre, M. T. Rudd, K. T. Nguyen, J. J. Romano, J. W. Butcher, K. F. Gilbert, K. J. Bush, M. K. Holloway, J. Swestock, B. L. Wan, S. S. Carroll, J. M. DiMuzio, D. J. Graham, S. W. Ludmerer, S. S. Mao, M. W. Stahlhut, C. M. Fandozzi, N. Trainor, D. B. Olsen, J. P. Vacca, and N. J. Liverton, "Discovery of vaniprevir (mk-7009), a macrocyclic hepatitis c virus ns3/4a protease inhibitor," *J Med Chem*, vol. 53, no. 6, pp. 2443–63, 2010.

-
- [180] B. Cabot, M. Martell, J. I. Esteban, M. Piron, T. Otero, R. Esteban, J. Guardia, and J. Gomez, "Longitudinal evaluation of the structure of replicating and circulating hepatitis c virus quasispecies in nonprogressive chronic hepatitis c patients," *J Virol*, vol. 75, no. 24, pp. 12005–13, 2001.
- [181] M. Martell, J. I. Esteban, J. Quer, J. Genesca, A. Weiner, R. Esteban, J. Guardia, and J. Gomez, "Hepatitis c virus (hcv) circulates as a population of different but closely related genomes: quasispecies nature of hcv genome distribution," *J Virol*, vol. 66, no. 5, pp. 3225–9, 1992.
- [182] M. D. Altman, E. A. Nalivaika, M. Prabu-Jeyabalan, C. A. Schiffer, and B. Tidor, "Computational design and experimental study of tighter binding peptides to an inactivated mutant of hiv-1 protease," *Proteins*, vol. 70, no. 3, pp. 678–94, 2008.
- [183] A. Ozen, T. Haliloglu, and C. A. Schiffer, "Hiv-1 protease and substrate coevolution validates the substrate envelope as the recognition pattern," *J Chem Theory Comput*, vol. 8, pp. 703–14, 2012.
- [184] C. Lin, J. A. Thomson, and C. M. Rice, "A central region in the hepatitis c virus ns4a protein allows formation of an active ns3-ns4a serine proteinase complex in vivo and in vitro," *J Virol*, vol. 69, no. 7, pp. 4373–80, 1995.
- [185] C. Lin, B. M. Pragai, A. Grakoui, J. Xu, and C. M. Rice, "Hepatitis c virus ns3 serine proteinase: trans-cleavage requirements and processing kinetics," *J Virol*, vol. 68, no. 12, pp. 8147–57, 1994.
- [186] N. Yao, P. Reichert, S. S. Taremi, W. W. Prosser, and P. C. Weber, "Molecular views of viral polyprotein processing revealed by the crystal structure of the hepatitis c virus bifunctional protease-helicase," *Structure*, vol. 7, no. 11, pp. 1353–63, 1999.
- [187] S. K. Lee, M. Potempa, M. Kolli, A. Ozen, C. A. Schiffer, and R. Swanstrom, "Context surrounding processing sites is crucial in determining cleavage rate of a subset of processing sites in hiv-1 gag and gag-pro-pol polyprotein precursors by viral protease," *J Biol Chem*, vol. 287, no. 16, pp. 13279–90, 2012.
- [188] V. Summa, S. W. Ludmerer, J. A. McCauley, C. Fandozzi, C. Burlein, G. Claudio, P. J. Coleman, J. M. Dimuzio, M. Ferrara, M. Di Filippo, A. T.

- Gates, D. J. Graham, S. Harper, D. J. Hazuda, C. McHale, E. Monteagudo, V. Pucci, M. Rowley, M. T. Rudd, A. Soriano, M. W. Stahlhut, J. P. Vacca, D. B. Olsen, N. J. Liverton, and S. S. Carroll, "Mk-5172, a selective inhibitor of hepatitis c virus ns3/4a protease with broad activity across genotypes and resistant variants," *Antimicrob Agents Chemother*, 2012.
- [189] K. P. Romano, J. M. Laine, L. M. Deveau, H. Cao, F. Massi, and C. A. Schiffer, "Molecular mechanisms of viral and host cell substrate recognition by hepatitis c virus ns3/4a protease," *J Virol*, vol. 85, no. 13, pp. 6106–16, 2011.
- [190] P. W. Snyder, J. Mecinovic, D. T. Moustakas, r. Thomas, S. W., M. Harder, E. T. Mack, M. R. Lockett, A. Heroux, W. Sherman, and G. M. Whitesides, "Mechanism of the hydrophobic effect in the biomolecular recognition of arylsulfonamides by carbonic anhydrase," *Proc Natl Acad Sci U S A*, vol. 108, no. 44, pp. 17889–94, 2011.
- [191] T. Beuming, Y. Che, R. Abel, B. Kim, V. Shanmugasundaram, and W. Sherman, "Thermodynamic analysis of water molecules at the surface of proteins and applications to binding site prediction and characterization," *Proteins*, vol. 80, no. 3, pp. 871–83, 2012.
- [192] D. D. Robinson, W. Sherman, and R. Farid, "Understanding kinase selectivity through energetic analysis of binding site waters," *ChemMedChem*, vol. 5, no. 4, pp. 618–27, 2010.
- [193] D. J. Huggins, W. Sherman, and B. Tidor, "Rational approaches to improving selectivity in drug design," *J Med Chem*, vol. 55, no. 4, pp. 1424–44, 2012.
- [194] A. M. Davis, S. J. Teague, and G. J. Kleywegt, "Application and limitations of x-ray crystallographic data in structure-based ligand and drug design," *Angew Chem Int Ed Engl*, vol. 42, no. 24, pp. 2718–36, 2003.
- [195] K. J. Bowers, E. Chow, H. Xu, R. O. Dror, M. P. Eastwood, B. A. Gregersen, J. L. Klepeis, I. Kolossvary, M. A. Moraes, F. D. Sacerdoti, J. K. Salmon, Y. Shan, and D. E. Shaw, "Scalable algorithms for molecular dynamics simulations on commodity clusters," *Proceedings of the ACM/IEEE Conference on Supercomputing (SC06)*, Tampa, Florida, 2006.

-
- [196] D. Shivakumar, J. Williams, Y. Wu, W. Damm, J. Shelley, and W. Sherman, "Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the opls force field," *J. Chem. Theory Comput.*, vol. 6, pp. 1509–1519, 2010.
- [197] T. Darden, Y. Darrin, and L. Pedersen, "Particle mesh ewald: An nlog(n) method for ewald sums in large systems," *Journal of Chemical Physics*, vol. 98, pp. 10089–93, 1993.
- [198] F. Poordad, J. McCone, J., B. R. Bacon, S. Bruno, M. P. Manns, M. S. Sulkowski, I. M. Jacobson, K. R. Reddy, Z. D. Goodman, N. Boparai, M. J. DiNubile, V. Sniukiene, C. A. Brass, J. K. Albrecht, and J. P. Bronowicki, "Boceprevir for untreated chronic hcv genotype 1 infection," *N Engl J Med*, vol. 364, no. 13, pp. 1195–206, 2011.
- [199] I. M. Jacobson, J. G. McHutchison, G. Dusheiko, A. M. Di Bisceglie, K. R. Reddy, N. H. Bzowej, P. Marcellin, A. J. Muir, P. Ferenci, R. Flisiak, J. George, M. Rizzetto, D. Shouval, R. Sola, R. A. Terg, E. M. Yoshida, N. Adda, L. Bengtsson, A. J. Sankoh, T. L. Kieffer, S. George, R. S. Kauffman, and S. Zeuzem, "Telaprevir for previously untreated chronic hepatitis c virus infection," *N Engl J Med*, vol. 364, no. 25, pp. 2405–16, 2011.
- [200] K. E. Sherman, S. L. Flamm, N. H. Afdhal, D. R. Nelson, M. S. Sulkowski, G. T. Everson, M. W. Fried, M. Adler, H. W. Reesink, M. Martin, A. J. Sankoh, N. Adda, R. S. Kauffman, S. George, C. I. Wright, and F. Poordad, "Response-guided telaprevir combination treatment for hepatitis c virus infection," *N Engl J Med*, vol. 365, no. 11, pp. 1014–24, 2011.
- [201] A. Chopra, P. L. Klein, T. Drinnan, and S. S. Lee, "How to optimize hcv therapy in genotype 1 patients: management of side-effects," *Liver Int*, vol. 33 Suppl 1, pp. 30–4, 2013.
- [202] P. Halfon and C. Sarrazin, "Future treatment of chronic hepatitis c with direct acting antivirals: is resistance important?," *Liver Int*, vol. 32 Suppl 1, pp. 79–87, 2012.
- [203] P. Halfon and S. Locarnini, "Hepatitis c virus resistance to protease inhibitors," *J Hepatol*, vol. 55, no. 1, pp. 192–206, 2011.
- [204] J. Vermehren and C. Sarrazin, "The role of resistance in hcv treatment," *Best Pract Res Clin Gastroenterol*, vol. 26, no. 4, pp. 487–503, 2012.

-
- [205] E. S. Svarovskaia, R. Martin, J. G. McHutchison, M. D. Miller, and H. Mo, "Abundant drug-resistant ns3 mutants detected by deep sequencing in hepatitis c virus-infected patients undergoing ns3 protease inhibitor monotherapy," *J Clin Microbiol*, vol. 50, no. 10, pp. 3267–74, 2012.
- [206] C. Welsch and S. Zeuzem, "Clinical relevance of hcv antiviral drug resistance," *Curr Opin Virol*, 2012.
- [207] M. P. Manns, M. Bourliere, Y. Benhamou, S. Pol, M. Bonacini, C. Trepo, D. Wright, T. Berg, J. L. Calleja, P. W. White, J. O. Stern, G. Steinmann, C. L. Yong, G. Kukulj, J. Scherer, and W. O. Boecker, "Potency, safety, and pharmacokinetics of the ns3/4a protease inhibitor bi201335 in patients with chronic hcv genotype-1 infection," *J Hepatol*, vol. 54, no. 6, pp. 1114–22, 2011.
- [208] V. Cento, C. Mirabelli, R. Salpini, S. Dimonte, A. Artese, G. Costa, F. Mercurio, V. Svicher, L. Parrotta, A. Bertoli, M. Ciotti, D. Di Paolo, C. Sarrechia, M. Andreoni, S. Alcaro, M. Angelico, C. F. Perno, and F. Ceccherini-Silberstein, "Hcv genotypes are differently prone to the development of resistance to linear and macrocyclic protease inhibitors," *PLoS One*, vol. 7, no. 7, p. e39652, 2012.
- [209] S. R. Lim, X. Qin, S. Susser, J. B. Nicholas, C. Lange, E. Herrmann, J. Hong, A. Arfsten, L. Hooi, W. Bradford, I. Najera, P. Smith, S. Zeuzem, K. Kossen, C. Sarrazin, and S. D. Seiwert, "Virologic escape during danoprevir (itm-191/rg7227) monotherapy is hepatitis c virus subtype dependent and associated with r155k substitution," *Antimicrob Agents Chemother*, vol. 56, no. 1, pp. 271–9, 2012.
- [210] M. Manns, H. Reesink, T. Berg, G. Dusheiko, R. Flisiak, P. Marcellin, C. Moreno, O. Lenz, P. Meyvisch, M. Peeters, V. Sekar, K. Simmen, and R. Verloes, "Rapid viral response of once-daily tmc435 plus pegylated interferon/ribavirin in hepatitis c genotype-1 patients: a randomized trial," *Antivir Ther*, vol. 16, no. 7, pp. 1021–33, 2011.
- [211] C. Sarrazin and S. Zeuzem, "Resistance to direct antiviral agents in patients with hepatitis c virus infection," *Gastroenterology*, vol. 138, no. 2, pp. 447–62, 2010.

-
- [212] T. L. Kieffer, A. D. Kwong, and G. R. Picchio, "Viral resistance to specifically targeted antiviral therapies for hepatitis c (stat-cs)," *J Antimicrob Chemother*, vol. 65, no. 2, pp. 202–12, 2010.
- [213] C. Sarrazin, T. L. Kieffer, D. Bartels, B. Hanzelka, U. Muh, M. Welker, D. Wincheringer, Y. Zhou, H. M. Chu, C. Lin, C. Weegink, H. Reesink, S. Zeuzem, and A. D. Kwong, "Dynamic hepatitis c virus genotypic and phenotypic changes in patients treated with the protease inhibitor telaprevir," *Gastroenterology*, vol. 132, no. 5, pp. 1767–77, 2007.
- [214] S. Susser, C. Welsch, Y. Wang, M. Zettler, F. S. Domingues, U. Karey, E. Hughes, R. Ralston, X. Tong, E. Herrmann, S. Zeuzem, and C. Sarrazin, "Characterization of resistance to the protease inhibitor boceprevir in hepatitis c virus-infected patients," *Hepatology*, vol. 50, no. 6, pp. 1709–18, 2009.
- [215] E. J. Gane, P. Pockros, S. Zeuzem, P. Marcellin, A. Shikhman, C. Bernaards, E. Yetzer, N. Shulman, X. Tong, I. Najera, A. Bertasso, J. Hammond, and S. Stancic, "Interferon-free treatment with a combination of mericitabine and danoprevir/r with or without ribavirin in treatment-naive hcv genotype 1-infected patients," in *63rd Annual Meeting of the American Association for the Study of Liver Diseases*, 2012.
- [216] G. Barbato, D. O. Cicero, M. C. Nardi, C. Steinkuhler, R. Cortese, R. De Francesco, and R. Bazzo, "The solution structure of the n-terminal proteinase domain of the hepatitis c virus (hcv) ns3 protein provides new insights into its activation and catalytic mechanism," *J Mol Biol*, vol. 289, no. 2, pp. 371–84, 1999.
- [217] D. Pan, W. Xue, W. Zhang, H. Liu, and X. Yao, "Understanding the drug resistance mechanism of hepatitis c virus ns3/4a to itm-191 due to r155k, a156v, d168a/e mutations: a computational study," *Biochim Biophys Acta*, vol. 1820, no. 10, pp. 1526–34, 2012.
- [218] C. Welsch, F. S. Domingues, S. Susser, I. Antes, C. Hartmann, G. Mayr, A. Schlicker, C. Sarrazin, M. Albrecht, S. Zeuzem, and T. Lengauer, "Molecular basis of telaprevir resistance due to v36 and t54 mutations in the ns3-4a protease of the hepatitis c virus," *Genome Biol*, vol. 9, no. 1, p. R16, 2008.

-
- [219] F. Bennett, Y. Huang, S. Hendrata, R. Lovey, S. L. Bogen, W. Pan, Z. Guo, A. Prongay, K. X. Chen, A. Arasappan, S. Venkatraman, F. Velazquez, L. Nair, M. Sannigrahi, X. Tong, J. Pichardo, K. C. Cheng, V. M. Girijavalabhan, A. K. Saksena, and F. G. Njoroge, "The introduction of p4 substituted 1-methylcyclohexyl groups into boceprevir: a change in direction in the search for a second generation hcv ns3 protease inhibitor," *Bioorg Med Chem Lett*, vol. 20, no. 8, pp. 2617–21, 2010.
- [220] H. Zhu and J. M. Briggs, "Mechanistic role of ns4a and substrate in the activation of hcv ns3 protease," *Proteins*, vol. 79, no. 8, pp. 2428–43, 2011.
- [221] M. S. Wittekind, S. Weinheimer, Y. Zhang, and V. Goldfarb, "Modified forms of hepatitis c ns3 protease for facilitating inhibitor screening and structural studies of protease:inhibitor complexes," 2002.
- [222] S. S. Taremi, B. Beyer, M. Maher, N. Yao, W. Prosise, P. C. Weber, and B. A. Malcolm, "Construction, expression, and characterization of a novel fully activated recombinant single-chain hepatitis c virus protease," *Protein Sci*, vol. 7, no. 10, pp. 2143–9, 1998.
- [223] Z. Otwinowski and W. Minor, "Processing of x-ray diffraction data collected in oscillation mode," *Methods Enzymol*, vol. 276, pp. 307–326, 1997.
- [224] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, "Phaser crystallographic software," *J Appl Crystallogr*, vol. 40, no. Pt 4, pp. 658–674, 2007.
- [225] P. D. Adams, P. V. Afonine, G. Bunkoczi, V. B. Chen, I. W. Davis, N. Echols, J. J. Headd, L. W. Hung, G. J. Kapral, R. W. Grosse-Kunstleve, A. J. McCoy, N. W. Moriarty, R. Oeffner, R. J. Read, D. C. Richardson, J. S. Richardson, T. C. Terwilliger, and P. H. Zwart, "Phenix: a comprehensive python-based system for macromolecular structure solution," *Acta Crystallogr D Biol Crystallogr*, vol. 66, no. Pt 2, pp. 213–21, 2010.
- [226] R. J. Morris, A. Perrakis, and V. S. Lamzin, "Arp/warp's model-building algorithms. i. the main chain," *Acta Crystallogr D Biol Crystallogr*, vol. D58, pp. 968–975, 2002.
- [227] N. . Collaborative-Computational-Project, "The ccp4 suite: programs for protein crystallography," *Acta Crystallogr D Biol Crystallogr*, vol. 50, pp. 760–763, 1994.

-
- [228] I. W. Davis, A. Leaver-Fay, V. B. Chen, J. N. Block, G. J. Kapral, X. Wang, L. W. Murray, r. Arendall, W. B., J. Snoeyink, J. S. Richardson, and D. C. Richardson, "Molprobity: all-atom contacts and structure validation for proteins and nucleic acids," *Nucleic Acids Res*, vol. 35, no. Web Server issue, pp. W375–83, 2007.
- [229] A. T. Brunger, "Free r value: a novel statistical quantity for assessing the accuracy of crystal structures," *Nature*, vol. 355, no. 6359, pp. 472–5, 1992.
- [230] P. Emsley and K. Cowtan, "Coot: model-building tools for molecular graphics," *Acta Crystallogr D Biol Crystallogr*, vol. 60, no. Pt 12 Pt 1, pp. 2126–32, 2004.
- [231] I. Maestro v9.2. Portland, OR: Schrodinger, 2011.
- [232] A. R. Atilgan, S. R. Durell, R. L. Jernigan, M. C. Demirel, O. Keskin, and I. Bahar, "Anisotropy of fluctuation dynamics of proteins with an elastic network model," *Biophys J*, vol. 80, no. 1, pp. 505–15, 2001.
- [233] I. Bahar, A. R. Atilgan, and B. Erman, "Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential," *Fold Des*, vol. 2, no. 3, pp. 173–81, 1997.
- [234] S. Kundu, D. C. Sorensen, and J. Phillips, G. N., "Automatic domain decomposition of proteins by a gaussian network model," *Proteins*, vol. 57, no. 4, pp. 725–33, 2004.
- [235] U. Emekli, D. Schneidman-Duhovny, H. J. Wolfson, R. Nussinov, and T. Haliloglu, "Hingeprot: automated prediction of hinges in protein structures," *Proteins*, vol. 70, no. 4, pp. 1219–27, 2008.
- [236] B. Munos, "Lessons from 60 years of pharmaceutical innovation," *Nat Rev Drug Discov*, vol. 8, no. 12, pp. 959–68, 2009.
- [237] A. C. Anderson, M. P. Pollastri, C. A. Schiffer, and N. P. Peet, "The challenge of developing robust drugs to overcome resistance," *Drug Discov Today*, vol. 16, no. 17-18, pp. 755–61, 2011.
- [238] A. Wlodawer and J. Vondrasek, "Inhibitors of hiv-1 protease: a major success of structure-assisted drug design," *Annu Rev Biophys Biomol Struct*, vol. 27, pp. 249–84, 1998.

- [239] P. Carter and J. A. Wells, "Dissecting the catalytic triad of a serine protease," *Nature*, vol. 332, no. 6164, pp. 564–8, 1988.
- [240] J. C. Sullivan, E. Z. Zhang, D. J. Bartels, A. Tigges, J. L. Dorrian, A. D. Kwong, and T. L. Kieffer, "Compensatory substitutions in the hcv ns3/4a protease cleavage sites are not observed in patients treated unsuccessfully with telaprevir combination treatment," *Viol J*, vol. 9, p. 147, 2012.
- [241] N. M. King, M. Prabu-Jeyabalan, R. M. Bandaranayake, M. N. Nalam, E. A. Nalivaika, A. Ozen, T. Haliloglu, N. K. Yilmaz, and C. A. Schiffer, "Extreme entropy-enthalpy compensation in a drug-resistant variant of hiv-1 protease," *ACS Chem Biol*, vol. 7, no. 9, pp. 1536–46, 2012.
- [242] S. Mittal, R. M. Bandaranayake, N. M. King, M. Prabu-Jeyabalan, M. N. Nalam, E. A. Nalivaika, N. K. Yilmaz, and C. A. Schiffer, "Structural and thermodynamic basis of amprenavir/darunavir and atazanavir resistance in hiv-1 protease with mutations at residue 50," *J Virol*, vol. 87, no. 8, pp. 4176–84, 2013.